



HAL
open science

Système d'aide à la gestion et planification de groupe en formation continue

Laurie Acensio

► **To cite this version:**

Laurie Acensio. Système d'aide à la gestion et planification de groupe en formation continue. Algorithme et structure de données [cs.DS]. Université de Lille, 2021. Français. NNT : 2021LILUI031 . tel-03503323v2

HAL Id: tel-03503323

<https://hal.science/tel-03503323v2>

Submitted on 16 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Lille

École Doctorale SPI-MADIS

Sciences Pour l'Ingénieur

Mathématiques, sciences du numérique et de leurs interactions

Laboratoire CRIStAL

THESE

Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ DE LILLE

Spécialité : Informatique et applications

**Systeme d'aide à la gestion et planification de groupe
en formation continue**

Présentée et soutenue publiquement par

Laurie Acensio

Le 15 juin 2021

JURY

Présidente : Laetitia Jourdan

Luigi LANCIERI	Professeur des Universités	Université de Lille	Directeur de thèse
Max CHEVALIER	Professeur des Universités	Université de Toulouse	Rapporteur
Thomas GUYET	HDR/Maître de conférences	AgroCampus-Ouest	Rapporteur
Bruno CREMILLEUX	Professeur des Universités	Université de Caen	Examineur
Frédéric HOOGSTOEL	Maître de conférences	Université de Lille Polytech	Examineur
Laetitia JOURDAN	Professeure des Universités	Université de Lille	Examineur
Michel BENSIMON	Directeur-Associé	Lexiane Formation	Invité

Résumé court

Dans le cadre des politiques des ressources humaines en entreprise, la formation continue devient un levier d'action stratégique pour l'évolution des carrières professionnelles. Parmi les modes de formation existants, la session de formation en présentiel est un mode d'action privilégié pour l'échange d'expériences professionnelles entre pairs. Dans ce contexte, notre projet de recherche consiste à proposer un système d'aide à la gestion de la formation continue visant à faciliter les tâches opérationnelles du gestionnaire de formation dont la recherche et l'analyse des données liée à l'activité de formation et la planification de sessions de formation. Les outils de gestion sont instrumentés par des tableaux de bords multiples associés à des techniques de visualisation pour faciliter l'interprétation des données.

Notre principale contribution scientifique se focalise sur l'aide à la planification semi-automatisée des sessions de formation. En pratique, cette tâche opérationnelle consiste à concilier des contraintes organisationnelles et des critères multiples (pédagogiques, sociaux et logistiques) constitué autour d'un groupe de formation en présentiel. Pour répondre à notre problématique, le prototype d'application vise à simplifier la composition des groupes de formation pour le gestionnaire de formation en proposant une liste de groupes de formation potentiels associés à une représentation graphique de type noeud-lien. Le système repose sur un principe de requête-réponse afin de s'adapter aux différents besoins de formation des commanditaires. Au niveau du traitement, la résolution de ce problème de décision de type combinatoire s'appuie sur une méthode méta-heuristique (et plus précisément un algorithme génétique, noté AG par la suite) qui vise à générer une liste de groupes de formation potentiels en privilégiant un temps de traitement relativement court. La principale difficulté est de paramétrer l'AG durant la phase d'initialisation notamment la fonction objective (couramment appelé *fitness*) qui détermine la qualité des résultats. Notre méthode s'appuie sur une analyse des données des questionnaires de satisfaction qui vise à définir les caractéristiques des groupes de formation qui conditionnent favorablement les interactions sociales entre les participants. Ainsi, les résultats obtenus permettent de justifier les paramétrages de l'AG qui oriente l'exploration et l'affectation des participants au sein des groupes de formation.

L'expérimentation distingue deux niveaux d'évaluation. La première série d'expérimentation vise à mesurer les paramètres et les résultats générés par l'AG à partir d'une collection de requêtes multicritères. La deuxième série d'expérimentation est centrée sur la performance du prototype d'application à partir de tests fonctionnels, techniques et utilisateurs selon une norme ISO.

Mots-clés : Groupe de formation pour adulte-Aide à la décision multicritère-Optimisation combinatoire- Algorithme Génétique-Visualisation des réseaux sociaux-Analyse des données d'apprentissage humain

Abstract

In the context of human resources management in companies, continuing training becomes a strategic action for the development of professional careers. Among the existing training methods, the training session is a privileged mode of action for the exchange of professional experiences between peers. In practice, training session planning consists of reconciling multiple criteria (educational, social and organizational) formed around a training group in face-to-face. In this context, our research project aims to propose an IT system for the management of continuing education consists of developing a system of support for the management of continuing training that will facilitate the operational tasks of the training manager : the planning of training sessions and the research and analysis of data related to the training activity. The management tools are consisted by multiple dashboards associated with visualization techniques to facilitate data interpretation.

Our main scientific contribution is focused on helping the semi-automated planning of training sessions. In practice, this operational task consists of reconciling organizational constraints and multiple criteria (educational, social and logistical) formed around a face-to-face training group. To respond to our problem, the application prototype aims to simplify the composition of training groups for the training manager by proposing a list of potential training groups associated with a graphical representation of the node-link type. The system is based on a request-response principle in order to adapt to the various training needs of sponsors. At the processing level, the resolution of this combinatorial type decision problem is based on a meta-heuristic method (and more precisely a genetic algorithm, denoted AG hereinafter) which aims to generate a list of training groups. potential by favoring a relatively short processing time. The main difficulty is to configure the GA during the initialization phase, in particular the objective function (commonly called *fitness*) which determines the quality of the results. Our method is based on an analysis of data from satisfaction questionnaires which aims to define the characteristics of training groups that favorably condition social interactions between participants. Thus, the results obtained make it possible to justify the settings of the GA which guides the exploration and assignment of participants within training groups.

The experiment distinguishes two levels of evaluation. The first series of experiments aims to measure the parameters and results generated by the GA from a collection of multicriteria queries. The second series of experiments is focused on the performance of the application prototype from functional, technical and user tests according to an ISO standard.

Keywords : Adult training group - Multi-criteria decision - Combinatorial optimization - Genetic Algorithm - Social visualization - Learning Analytics

*A mon fils Yobann que j'aime au-delà des mots,
A Joseph,
A Pouchkine.*

Remerciements

J'exprime ma reconnaissance à mes co-directeurs Luigi Lancieri et Frédéric Hoogstoel pour avoir encadré régulièrement cette thèse ainsi que toute l'équipe NOCE (Nouveaux Outils pour La Coopération et l'Education) pour leurs encouragements et conseils. J'adresse également mes remerciements aux membres du jury et rapporteurs qui ont assuré la continuité de l'évaluation de ce manuscrit malgré un contexte sanitaire particulier.

De même, je remercie les nombreux chercheurs rencontrés lors des conférences scientifiques, les partenaires industriels et les contributeurs anonymes sur internet qui ont permis l'enrichissement de ce travail à travers leurs points de vue diversifiés.

Je remercie vivement Michel et Joseph Bensimon ainsi que mes collègues qui ont contribué à la réalisation de ce projet tout en conciliant les contraintes de la vie quotidienne d'une petite entreprise.

Enfin, une pensée pour tous les étudiants rencontrés à Paris-Sorbonne et notamment les enseignants qui m'ont encouragé à poursuivre cette voie en me garantissant une expérience unique. Sans oublier, mes co-équipier(es) en natation synchronisée qui m'ont aidé à m'aérer l'esprit durant mes longues années d'études.

Table des matières

Résumé court	2
Remerciements	4
Liste des tableaux	8
Liste des figures	9
Liste des abréviations	11
Introduction	12

Partie 1. Etat de l'art

Chapitre 1. Système d'information en contexte organisationnel	18
1.1 Système d'information décisionnel	18
1.1.1 Notions	18
1.1.2 Technique d'aide à la décision multicritère	19
1.2 Système d'information de gestion	21
1.2.1 Système d'Information des Ressources Humaines	21
1.2.2 Evolutions	22
1.3 Système d'information dédié à l'apprentissage humain	23
1.3.1 Apprenants	23
1.3.2 Enseignants et formateurs	25
Chapitre 2. Analyse et visualisation des données en formation initiale et continue	27
2.1 Analyse et traitement des données	27
2.1.1 Analyse de l'apprentissage humain (ou <i>Learning Analytics</i>)	27
2.1.2 Analytique RH (ou <i>HR Analytics</i>)	28
2.2 Outils de gestion (ou management) des données	29
2.2.1 Indicateurs	29
2.2.2 Tableau de bord	31
2.3 Visualisation des données	32
2.3.1 Visualisation analytique	32
2.3.2 Visualisation descriptive	33
Chapitre 3. Analyse multidimensionnelle du groupe de formation pour adulte	35
3.1 Type de groupe de formation	35
3.1.1 Groupe de formation	35
3.1.2 Groupe professionnel	36
3.2 Acteurs du groupe de formation pour adulte	37
3.2.1 Apprenant-adulte	37
3.2.2 Formateur d'adulte	38
3.2.3 Organisation/Institution	38
3.3 Dimension sociale (ou relationnelle)	39
3.3.1 Relation professionnelle	39
3.3.2 Intensité des relations sociales	40
3.3.3 Symétrie/Asymétrie des relations sociales	41
3.4 Dimension cognitive	42
3.4.1 Interactions par les pairs	42
3.4.2 Expériences professionnelles	43
3.5 Dimension organisationnelle et environnementale	44

3.5.1	Type de regroupement	44
3.5.2	Mode environnemental (Présentiel <i>vs</i> Distanciel)	44
3.5.3	Taille du groupe	45

Chapitre 4. Algorithme d'aide à la composition du groupe de formation _____ **46**

4.1	Introduction	46
4.1.1	Méthodologie	46
4.2	Méthode méta-heuristique et heuristique	47
4.2.1	Principe de traitement	47
4.2.2	Algorithme génétique	48
4.3	Méthode de classification automatique	50
4.3.1	Classification basée sur les graphes (ou détection de communauté)	50
4.3.2	Classification non-supervisée	52

Chapitre 5. Conclusion _____ **54**

Partie 2 - LexRH : Méthodologie et développement d'un prototype d'aide à la gestion et à la planification de groupe en formation continue

Chapitre 6. Conception et développement d'un prototype d'aide à la gestion de la formation continue _____ **58**

6.1	Système d'aide à la gestion de la formation continue	58
6.1.1	Analyse de l'existant	58
6.1.2	Objectif(s)	59
6.1.3	Type d'utilisateurs	60
6.1.4	Type de données	61
6.2	Représentation et visualisation des données	62
6.2.1	Visualisation descriptive	63
6.2.2	Visualisation analytique	66
6.3	Ergonomie des interfaces-utilisateurs	67
6.3.1	Tableau de bord principal	67
6.3.2	Interface de recherche de données	68
6.3.3	Interface de visualisation interactive	69
6.3.4	Tableau de bord analytique	74
6.4	Implémentation technique	76
6.4.1	Type de prototype	76
6.4.2	Fonctionnement global	76
6.5	Conclusion	79

Chapitre 7. Aide à la planification par algorithme génétique _____ **81**

7.1	Types de planification en formation continue	81
7.1.1	Cas d'utilisations	81
7.1.2	Méthode de résolution	82
7.2	Identification des contraintes et des critères de décision	83
7.2.1	Type de contraintes	83
7.2.2	Catégorisation des critères de décision	84
7.3	Algorithme génétique	86
7.3.1	Fonction-objective	86
7.3.2	Principe et paramétrage du traitement de l'AG	87
7.3.3	Implémentation du code-source	92
7.4	Conclusion	93

Chapitre 8. Aide à la planification par analyse statistique	94
8.1 Analyse statistique de l'activité de formation	94
8.1.1 Finalités d'application	94
8.1.2 Objectif(s)	95
8.1.3 Méthodologie	96
8.2 Analyse et traitement statistique	97
8.2.1 Statistiques descriptives	97
8.2.2 Analyse globale	98
8.2.3 Traitement et tests statistiques	98
8.3 Analyse et traitement des données textuelles	103
8.4 Interprétation	107
8.5 Conclusion	108
Partie 3-Expérimentation	
<hr/>	
Chapitre 9. Evaluation de la performance du prototype d'application	111
9.1 Méthode et évaluation de l'AG	111
9.1.1 Jeu de données expérimentales	111
9.1.2 Collection des requêtes de test	113
9.1.3 Evaluation globale	114
9.1.4 Evaluation des paramètres de l'AG	117
9.1.5 Exemples de résultats	120
9.2 Tests fonctionnels et techniques	121
9.2.1 Cas d'utilisation	121
9.2.2 Mesures d'évaluation	122
9.2.3 Protocole de tests	122
9.2.4 Résultats	123
9.3 Tests-utilisateurs	125
9.3.1 Méthodologie	125
9.3.2 Résultats	127
9.4 Conclusion	128
Chapitre 10. Conclusion et Perspectives	129
10.1 Synthèse	129
10.2 Perspectives	130
Chapitre 11. Bibliographie	134

Liste des tableaux

1.1	Tableau comparatif : système d'information opérationnel, décisionnel et système de recommandation (Nègre E., 2015).....	19
2.1	Synthèse des catégories d'indicateurs des <i>Learning Analytics</i> (Holzer, A., 2016).....	30
3.1	Synthèse des travaux scientifiques utilisant l'AG dans le domaine de l'éducation (Kucak, D., 2019).....	50
4.1	Catégorisation et description des variables liées à l'activité de formation.....	61
4.2	Typologie des nœuds et des relations en situation de formation continue.....	63
4.3	Représentation visuelle (forme géométrique et coloration) des nœuds et relations	64
4.4	Représentation visuelle des statuts de présence des participants	73
5.1	Synthèse des caractéristiques des sessions de formation selon le mode organisationnel.....	82
5.2	Pseudo-code de l'AG	93
6.1	Indicateurs et description des variables associées.....	95
6.2	Récapitulatif des tests statistiques selon le type de variable	99
6.3	Tableau comparatif des analyses bivariées entre les variables du groupe de formation et le score de la satisfaction globale/score de la satisfaction liée à la collaboration selon la p-value associée.....	100
6.4	Tableau des moyennes estimées et la valeur p associée selon les modalités par taille.....	100
6.5	Type de polarité des réponses.....	104
6.6	Statistiques textuelles selon le type de polarité et exemple de commentaires.....	105
6.7	Extrait du mot-clé pivot « groupe » associé à son étiquetage et contexte droit/gauche.....	105
6.8	Répartition thématique en % selon le type de polarité positive/négative.....	106
7.1	Extrait du tableur de données artificielles (ou fictives).....	113
7.2	Statistique descriptive de la collection de requêtes selon le nombre de critères associés et valeur en % par type de jeu de données	114
7.3	Répartition et nombre de requêtes selon le taux de précision	115
7.4	Taux de précision moyenne par critères pour l'ensemble des requêtes.....	116
7.5	Score minimal, moyen, maximal et écart-type des listes de résultats	117
7.6	Valeurs des critères de la requête Q1.....	120
7.7	Liste des groupes restitués pour la requête Q1	121
7.8	Détails des scores bruts par itérations pour la requête Q1.....	121
7.9	Mesures d'efficacité et d'efficience de la norme ISO/IEC 25022	122
7.10	Extraits des résultats des tests selon des mesures d'efficacité et d'efficience	124
7.11	Nombre et caractéristiques des évaluateurs.....	126
7.12	Mesures de satisfaction de la norme ISO/IEC 25022	126

Liste des figures

1.1	Typologie simplifiée des méthodes d'agrégation (Sadok, W., 2008).....	21
2.1	Processus de visualisation de type analytique pour une aide à la décision interactive et évolutive (Keim D., 2010).....	32
3.1	Position de la communauté par comparaison à d'autres groupes sociaux (Dillenbourg, P., 2003)	35
3.2	Quadrant virtualité/intensité des liens (Bos-Ciussi, M. 2007)	40
4.1	Taxonomie des techniques de groupes de formation (Maqtary, N., 2019).....	47
4.2	Extrait de la taxonomie des algorithmes « bio-inspirés » (Binitha, S., 2012)	48
4.3	Schéma simplifié du processus d'un algorithme génétique (Dréo J., 2003).....	49
5.1	Exemples des représentations graphiques de type descriptives selon la tâche de recherche	65
5.2	Synthèse des représentations graphiques de type analytique.....	66
5.3	Maquette du tableau de bord principal.....	68
5.4	Exemple de saisie d'une valeur de pondération pour les critères « Secteur » et « Statut ».....	69
5.5	Interface d'une recherche simple d'un participant	70
5.6	Zoom d'une vue partielle du graphe <i>via</i> un filtrage dynamique.....	70
5.7	Représentation graphique de type « égocentré » d'une recherche simple d'un formateur.....	71
5.8	Illustration des étapes du partitionnement selon une méthode « divisive » pour la création d'un groupe de formation	72
5.9	Interface avec liste des résultats.....	73
5.10	Etape de la finalisation du groupe de formation.....	74
5.11	Maquette du tableau de bord analytique	75
5.12	Processus et fonctionnement global du prototype d'application	76
5.13	Schéma simplifié d'un tableur vers un modèle à base de graphe.....	77
5.14	Schéma simplifié de l'architecture technique.....	78
6.1	Arborescence des contraintes et des critères de décision	84
6.2	Interface de recherche de la fonctionnalité « Planification ».....	88
6.3	Principe du traitement global de l'AG.....	89
6.4	Exemple de codage binaire d'un groupe de formation.....	89
6.5	Principe de croisement à un point sur une chaîne de bits	91
7.1	Histogramme des comparatifs de la composition effective des groupes selon le genre des participants en variable X et la moyenne estimée du score de satisfaction de la collaboration en variable Y	101
8.1	Courbe des scores d'évaluations avec le nombre total de personnes trouvés (ordonnées) et requêtes associées en diagramme en barre (abscisses).....	118
8.2	Résultats de requêtes selon un mode « Inter-Organisationnel » avec courbe du temps d'exécution (ordonnées) et diagramme à barre du nombre d'itération de 10 à 50 et le score d'évaluation associé (abscisses).....	119
8.3	Exemple d'un enregistrement avec répartition des temps.....	123
8.4	Courbe du temps d'exécution (ordonnées) selon le nombre de critères des requêtes (abscisses) et (b) Courbe du temps d'exécution (ordonnées) selon le nombre de personnes (abscisses)	125

Liste des annexes

Annexe 1. Type de variables selon la catégorisation des données	156
Annexe 2. Représentation graphique de la modélisation conceptuelle	157
Annexe 3. Détails des modules techniques de développement.....	158
Annexe 4. Mode d’interactivité et détails des valeurs par critères de recherche	159
Annexe 5. Lexique comparée des AG.....	160
Annexe 6. Modèle d’évaluation de Kirpatrick.....	161
Annexe 7. Modèle du questionnaire de satisfaction.....	162
Annexe 8. Statistique descriptive des groupes de formation	163
Annexe 9. Statistique descriptive des participants	164
Annexe 10. Représentation graphique en boîte à moustache des profils des participants	165
Annexe 11. Représentation graphique en diagramme circulaire du score de satisfaction global et score de satisfaction lié à la collaboration.....	165
Annexe 12. Fréquence des formes actives par catégories grammaticales	166
Annexe 13. Lexique des mots-clés pivot pour les requêtes de cooccurrences.....	167
Annexe 14. Distribution du jeu de données mixte (Réel+Fictif).....	167
Annexe 15. Collection de requêtes et distribution des critères et des valeurs associées.....	168
Annexe 16. Identifiant unique, précision à dix et score d’évaluation par requête	170
Annexe 17. Taux de précision par critères et par requête	171
Annexe 18. Résultats d’évaluation des mesures d’efficacité et d’efficacité des tâches de simulation « Recherche » et « Planification »	173
Annexe 19. Liste des requêtes par nombre d’itération et temps d’exécution associées aux scores d’évaluations.....	176
Annexe 20. Questionnaire des tests-utilisateurs	178
Annexe 21. Résultats des tests-utilisateurs.....	179

Liste des abréviations

ACAO : Apprentissage Collaboratif Assisté par Ordinateur
AG : Algorithme Génétique
ARS : Analyse des Réseaux Sociaux
CSCL : Computer Supported Collaborative Learning
EDM : Educational Data Mining
EIAH : Environnement Informatique pour l'Apprentissage Humain
EAO : Enseignement Assisté par Ordinateur
EIAO : Enseignement Intelligent Assisté par Ordinateur
ERP : Enterprise Resource Planning
FTLV : Formation Tout au Long de la Vie
GPEC : Gestion Prévisionnelle de l'Emploi et des Compétences
GTA : Gestion des Temps et des Activités
GRH : Gestion des Ressources Humaines
IA : Intelligence Artificielle
ISO : International Organization for Standardization
LA : Learning Analytics
LMS : Learning Management System
MOOC : Massive Open Online Course
PGI : Progiciel de Gestion Intégré
PLA : Predictive Learning Analytics
RH : Ressources Humaines
RSE : Réseaux Sociaux d'Entreprise
SAD : Système d'Aide à la Décision
SIRH : Système d'Information des Ressources Humaines
SOLAR : Society for Learning Analytics Research
SPOC : Small Private Online Course
SNA : Social Networks Analysis

Introduction

- **Contexte**

Selon les politiques des ressources humaines des entreprises, la formation continue devient un levier d'action stratégique pour l'évolution des carrières professionnelles. Ainsi, les adultes sont encouragés à poursuivre des actions de formation afin de se perfectionner, adapter et consolider leurs compétences professionnelles.

En France, le système de la formation continue s'inscrit dans un environnement socio-économique complexe incluant des acteurs multiples dont les entreprises privés ou publics, les participants (salariés ou demandeurs d'emplois), les organismes de formation, les OPCA¹ et l'Etat. Parmi ces acteurs, l'organisme de formation est un intermédiaire pour les entreprises qui souhaitent déléguer leurs actions de formations. En pratique, les modalités de formation s'opèrent principalement par des sessions de formation en présentiel². D'une part, ce type de formation répond à une réglementation juridique relative au droit de la formation³ (gouvernance, financement, évaluation...). D'autre part, la formation en présentiel relève d'un besoin exprimé par les participants (ou les salariés) qui souhaitent interagir avec des personnes physiquement en bénéficiant d'un cadre spatio-temporel distancié de leurs lieux de travail.

Dans ce contexte, ce projet de recherche est soutenu par l'organisme de formation « Lexiane Formation » qui souhaite s'inscrire dans une démarche de qualité en optimisant la gestion courante de son activité de formation. Garant du dispositif de formation continue, ses missions peuvent être synthétisées de la manière suivante :

1. Un rôle pédagogique en proposant une offre de formation spécialisée

L'organisme de formation vise à proposer un dispositif de formation en adéquation avec les besoins des participants qui sont affiliés à des organisations (entreprises privés ou publics). Dans un cadre formel, son rôle pédagogique consiste à proposer une offre de formation⁴ et de mettre à disposition un formateur spécialisé afin de superviser le groupe de formation en présentiel. Contrairement à la formation initiale davantage axée sur la transmission de savoirs théoriques et formels par l'enseignant⁵, l'échange de pratiques professionnelles sont favorisées dans le cadre de la formation pour adultes. Ainsi, le groupe de formation en présentiel devient un vecteur de socialisation qui conditionne la qualité de la formation à travers les relations et interactions sociales qui s'instaurent entre ses membres. Principalement restreint, la composition du groupe de formation pour adulte n'est pas choisie par les membres du groupe mais planifié selon des contraintes et des critères multiples gérés par l'organisme de formation.

2. Un rôle de gestion pour l'organisation des sessions de formation

Au niveau organisationnel, la planification des sessions de formation consiste à composer un groupe de formation en présentiel. En complément des critères pédagogiques et sociaux

¹ Organisme Paritaire Collecteur Agréé

² Plus de 50 % des organismes privés de formation interrogés par l'observatoire de branche « Formation Professionnelle » ne réalisaient en 2016 aucune prestation en formations digitales. Seuls 4 % d'entre eux en tiraient plus de 50 % de leur chiffre d'affaires/SOURCE : IGAS-La transformation digitale de la formation professionnelle continue-Nicolas Amar, Anne Burstin-2017.

³ CPF (Compte Personnel de Formation) relatif à la loi n° 2014-288 du 5 mars 2014¹ de la formation professionnelle

⁴ Offre de formation spécialisée dans le secteur public

⁵ Apprentissage formel : dispensé dans un contexte organisé et structuré et qui débouche sur une certification officielle (diplôme...)

énumérées précédemment, cette tâche opérationnelle se heurte à des contraintes logistiques (ou spatio-temporelles) et économiques. D'une part, les participants sont dispersés géographiquement et doivent respecter un volume horaire plafonné lié à leurs obligations professionnelles et à une réglementation législative spécifique⁶. D'autre part, un nombre minimal de participants conditionne le projet de formation pour être viable économiquement. Par conséquent, la situation de formation s'inscrit dans un espace-temps en dehors du cadre de travail du participant et reste soumise à des contraintes économiques et juridiques.

Ainsi, le rôle de l'organisme de formation implique des tâches opérationnelles multiples et diverses (commercialisation de l'offre de formation, recrutement des formateurs, gestion administrative des contrats, planification et suivi des évaluations des sessions de formations). Dans ce contexte, notre proposition consiste à développer un système d'aide à la gestion de la formation continue en se focalisant particulièrement sur la planification semi-automatisée des sessions de formation.

- **Problématique**

La planification des sessions de formation demeure une mission fondamentale pour l'organisme de formation. Cette tâche opérationnelle consiste à conditionner favorablement l'action de formation en favorisant les interactions sociales entre les participants du groupe. Selon les théories socioconstructivistes⁷, de nombreux travaux scientifiques ont étudié le fonctionnement du groupe de formation dont principalement la dynamique et la collaboration entre les membres du groupe. Dans le cadre de la formation pour adulte, cette problématique est abordée différemment : (Bourgeois E., Nizet J. 1997) oriente cette question essentielle qui n'est plus de savoir si *globalement* les interactions entre pairs sont ou non de nature à favoriser l'apprentissage mais bien à *quelles conditions* se déroulent ces interactions sociales. En pratique, la composition des groupes de formation demeure une tâche manuelle et particulièrement complexe. En effet, le processus de décision se heurte à la difficulté de concilier des multiples critères de décision de différentes natures (sociale, pédagogique, logistique et économique) et interdépendants. Ainsi, le gestionnaire de formation est contraint d'explorer de nombreuses combinaisons de groupes de formation impliquant un traitement manuel excessivement long et dépendant de son expérience métier. Par conséquent, la principale problématique est de faciliter la planification des sessions de formation en proposant un outil informatique qui recherche et aide à la composition des groupes de formation potentiels dans un temps de traitement raisonnable.

En parallèle, l'analyse de l'activité de formation devient une mission de pilotage stratégique pour l'organisme de formation. Les données relatives à l'activité de formation sont particulièrement contextualisées liées à la spécificité de la formation continue. La problématique complémentaire est de proposer des outils adéquats permettant l'analyse et la représentation des données de type multidimensionnelle.

- **Contribution(s)**

A partir de l'état de l'art scientifique et des problématiques préalablement formulées, notre proposition consiste à développer un système d'aide à la gestion dédié essentiellement à la formation continue. Destiné au gestionnaire de formation, l'objectif principal est de faciliter les tâches opérationnelles suivantes :

-Aide à la recherche simple des données ;

⁶ Article L6321-2 du Code du travail

⁷ Courant de recherche initié par Lev Vygotsky et Jean Piaget

- Analyse des données liées à l'activité de formation ;
- Aide à la planification de sessions de formation.

Développées tout au long du manuscrit, ce projet de recherche opérationnelle se focalise particulièrement sur les contributions suivantes :

1. Planification semi-automatisée du groupe de formation

Dans le cadre de la formation continue, les commandes de formation ne sont pas génériques et les critères de décision peuvent varier selon le(s) contexte(s) et le(s) besoin(s) de formation exprimé(s). Au préalable, deux types de planification des sessions de formations sont à distinguer :

- Planification exploratoire : le gestionnaire de formation n'est soumis à aucune commande de formation ;

- Planification selon une commande : le gestionnaire de formation est soumis à une demande de formation explicite émises par une ou plusieurs organisations (ou entreprises)

Ces différents cas d'utilisations orientent notre démarche vers un prototype d'application basé sur un mode de requête/réponse. Ainsi, le gestionnaire de formation peut formuler des critères de recherche spécifiques et s'adapter aux différentes demandes de formation des commanditaires.

Au niveau du traitement, notre contribution consiste à proposer une méthode de résolution de type heuristique *via* un algorithme génétique dont les paramétrages sont justifiés par une analyse statistique.

Algorithme génétique. De type méta-heuristique, un algorithme génétique permet de simplifier la charge de travail du gestionnaire de formation en lui proposant une liste de groupes de formation potentiels. Cette méthode d'optimisation vise à résoudre des problèmes de planification de sessions de formation selon un mode « Inter-organisationnel »⁸ où la complexité combinatoire est plus élevée en comparaison avec le mode « Intra-organisationnel »⁹. En effet, cette modalité organisationnelle implique un nombre de participants potentiels plus large à explorer résultant de nombreuses configurations de groupes de formation admissibles. Au préalable, l'application d'un algorithme génétique implique la définition d'une fonction objective (couramment nommée *fitness*) et des choix de paramétrages lors de la phase d'initialisation (méthode de sélection, opérateurs, nombre d'individus, critère d'arrêt...) qui détermine la qualité des résultats. Dans notre cas d'application, la fonction objective est justifiée à partir d'une analyse statistique traitant des données liées à l'activité de formation. Au niveau de la restitution des résultats, la présentation d'une liste de groupes de formation potentiels est associée à une représentation graphique de type noeud-lien en privilégiant un temps de traitement raisonnable.

Analyse statistique. De type analytique, cette méthode issue de la théorie de la décision vise à identifier les valeurs (ou caractéristiques) d'un groupe de formation qui conditionnent favorablement les interactions sociales entre les participants. A travers la collecte, l'analyse et l'interprétation des données liée à l'activité de formation, cette étape vise plusieurs finalités d'applications. Au niveau du paramétrage de l'AG, les résultats obtenus permettent de définir la fonction objective de l'algorithme génétique. De plus, le type de planification exploratoire implique de connaître préalablement les valeurs du groupe de formation à spécifier lors de la formulation de la requête. Ainsi, la composition du groupe de formation est justifiée selon une approche rationnelle limitant ainsi une affectation des participants d'une manière subjective et arbitraire.

⁸ Participants issus d'organisation différentes

⁹ Participants issus d'une même organisation

2. Présentation et visualisation des données complexes

L'analyse des données liée à l'activité de formation est un enjeu stratégique pour l'organisme de formation qui souhaite s'inscrire dans une démarche de qualité. Au préalable, le traitement et l'analyse des données visent à identifier des phénomènes stabilisés et réduire l'incertitude des prises de décision dans un domaine d'activité spécifique. La présentation des données est une étape complémentaire au traitement qui influence et oriente les choix décisionnels. Au niveau de la restitution des résultats, notre contribution consiste à proposer des outils de gestion dont principalement des tableaux de bords associés à des techniques de visualisations pour faciliter l'interprétation des données.

Tableaux de bord multiples. Les tableaux de bord proposés visent à faciliter l'analyse et l'interprétation des données complexes. Parmi les besoins opérationnels, le gestionnaire de formation doit pouvoir rechercher des données et consulter l'activité de formation globale d'une manière instantanée. Cette démarche de conception implique de centraliser des multiples données hétérogènes (administratives, sociales, spatio-temporelles...) au sein d'une vue globale avec des techniques de visualisation adaptées. De même, le suivi et l'évaluation de l'activité de formation vise à répondre à une obligation juridique mais également stratégique dans une perspective de remédiation de l'activité de formation.

Notre contribution consiste à proposer des tableaux de bord multiples qui agrègent les indicateurs et les données associées. Le tableau de bord principal propose une vue globale de l'état actuel de l'activité de formation avec un accès aux différentes fonctionnalités du prototype d'application. Le tableau de bord secondaire de nature analytique centralise les indicateurs de performance, préalablement définis lors d'une analyse statistique, afin de faciliter l'analyse et le suivi des évaluations des sessions de formation.

Visualisation des données socio-professionnelles. Le contexte de la formation continue s'articule autour de plusieurs acteurs (participant, organisation et formateur) et implique des multiples relations professionnelles. Ainsi, notre démarche méthodologique s'oriente vers une représentation graphique de type nœud-lien qui se transpose intuitivement à la nature sociale du voisinage professionnel du participant. D'une part, ce type de visualisation implique de définir au préalable une représentation visuelle permettant de différencier les acteurs et la nature relationnelle entre eux. D'autre part, des techniques de visualisation interactive complémentaires sont appliquées afin de faciliter l'interprétation des données selon des différents niveaux de détails.

- Organisation du manuscrit

Le manuscrit est structuré autour de trois parties et se compose de neuf chapitres.

La **première partie** présente l'état de l'art scientifique qui met en évidence le caractère pluridisciplinaire de notre cadre de recherche. Cette partie est composée de quatre chapitres :

Le premier chapitre dresse un panorama des systèmes d'information dans un contexte organisationnel et plus précisément les SIRH (Système d'Information Ressources Humaines) dans le domaine des ressources humaines. Puis, les systèmes d'information dédiés à l'apprentissage humain sont distingués selon le type d'utilisateurs (apprenant/enseignant ou formateur).

Le deuxième chapitre se focalise sur l'analyse et la visualisation des données liée à l'apprentissage humain (ou de formation initiale ou continue). Le domaine de l'analyse des apprentissages (ou *Learning Analytics*) est présenté en comparaison avec le domaine de l'analytique

RH (ou *Ressources Humaines*). De même, les techniques de traitement ainsi que les outils de gestion associés dont les tableaux de bord et les techniques de visualisation sont communs.

Le troisième chapitre présente une analyse du groupe de formation pour adulte en détaillant les dimensions relationnelles, cognitives, sociales et environnementales.

Le quatrième chapitre présente les techniques existantes de planification du groupe de formation en se limitant aux principales méthodes informatiques appliquées (i.e. les méthodes d'optimisation et les méthodes de classification).

Enfin, le cinquième chapitre synthétise l'état de l'art dans la perspective de la deuxième partie dédiée à notre proposition de recherche développée.

La **seconde partie** présente le système d'aide à la gestion de la formation proposée incluant trois chapitres.

Le sixième chapitre présente la conception et l'implémentation technique du prototype d'application en détaillant les spécifications fonctionnelles et techniques. La représentation des données et les techniques de visualisation associées sont illustrées par des exemples d'utilisations.

Le septième chapitre présente l'aide à la planification *via* un algorithme génétique. De type heuristique, il détaille les différentes étapes méthodologiques : formulation des problèmes de décision, détermination des contraintes et des critères de décision, paramétrage et principe de traitement de l'AG.

Le huitième chapitre présente l'aide à la planification à partir d'une analyse statistique. A partir des questionnaires de satisfaction, l'analyse statistique détaille la collecte, les techniques de traitement des données utilisées et l'interprétation des résultats obtenus. De type analytique, cette méthode d'aide à la planification vise principalement à justifier les caractéristiques des groupes de formation qui conditionnent favorablement les interactions sociales entre les membres du groupe.

La **troisième partie** présente la démarche d'expérimentation axée sur l'évaluation de la performance du prototype d'application. A travers ce dernier chapitre, deux niveaux d'évaluation sont présentés :

- Evaluation des résultats de l'AG selon des métriques de performances prédéfinies ;
- Evaluation de la performance du prototype d'application à partir des tests fonctionnels, techniques et utilisateurs

Partie 1. Etat de l'art

L'état de l'art présente une synthèse de la littérature scientifique liée à notre projet de recherche pluridisciplinaire.

Le premier chapitre présente les systèmes d'information dans un contexte organisationnel en se limitant aux bases théoriques et aux méthodes d'aide à la décision les plus répandues dans la littérature scientifique. Puis, il se focalise sur les systèmes d'information de gestion et plus précisément les SIRH (Système d'Information Ressources Humaines) dans un domaine d'application spécifique. Puis, les systèmes d'information dédiés à l'apprentissage humain sont distingués selon le type d'utilisateurs (apprenant/ formateur et enseignant).

Le deuxième chapitre se focalise sur l'analyse et la visualisation des données d'apprentissage. Le domaine de l'analyse des apprentissages (ou *Learning Analytics*) est présenté en parallèle de l'analytique RH (ou *Ressources Humaines*) ainsi que les outils de gestion associés dont les tableaux de bord et les techniques de visualisation.

Le troisième chapitre présente une analyse multidimensionnelle du groupe de formation pour adulte en détaillant les dimensions relationnelles, cognitives, sociales et environnementales.

Le quatrième chapitre présente les techniques existantes de planification du groupe de formation en se limitant aux méthodes d'optimisation et aux méthodes de classification.

Le cinquième chapitre présente une synthèse de l'état de l'art pour aborder la deuxième partie du manuscrit dédiée au développement de notre prototype de recherche développée.

Chapitre 1. Système d'information en contexte organisationnel

Ce chapitre présente un bref historique des systèmes d'information dans un contexte organisationnel. Relatif à l'informatique décisionnelle, les techniques d'aide à la décision sont présentées tandis que la section suivante se focalise sur l'informatique de gestion et plus précisément les SIRH (Système d'Information Ressources Humaines) dans le domaine des ressources humaines.

Par la suite, les systèmes d'information dédiés à l'apprentissage humain sont distingués selon le type d'utilisateurs : les apprenants (EIAH¹⁰, LMS¹¹, MOOC¹², SPOC¹³...) et les enseignants ou formateurs (STI¹⁴, systèmes de recommandation...).

Compte tenu de l'abondance de la littérature scientifique, ce chapitre n'est pas exhaustif mais vise à mettre en évidence les spécificités des systèmes d'information.

1.1 Système d'information décisionnel

1.1.1 Notions

Initié dans le domaine de la recherche opérationnelle, la notion du système d'information est définie par (Reix R., 2004) comme étant un « ensemble organisé de ressources (matériel, logiciel, donnée, procédure...) permettant d'acquérir, de traiter, de stocker des informations (sous forme de données, textes, images, sons...) en interne des organisations ». Le traitement informatisé de l'information apparaît comme un avantage tout autant stratégique qu'économique dans le processus décisionnel d'une organisation. Cette notion du traitement de l'information a considérablement évolué avec l'avènement des techniques informatiques et l'accessibilité des données et des connaissances à travers le réseau internet. Plusieurs confusions et ambiguïtés terminologiques liées aux systèmes d'information sont constatées par (Nègre E., 2015) qui distingue les systèmes d'information selon plusieurs critères dont la finalité d'application, le type d'utilisateur et la nature des données utilisées selon le tableau suivant :

	Système d'information opérationnel	Système d'information décisionnel	Système de recommandation
Utilisateurs	Nombreux (employés)	Peu (décideurs)	Nombreux (employés et décideurs)
Données	Détaillées Volumineuses Internes (orientées application) Dynamique	Agrégées Volumineuses Internes et externes (orientées sujet) Statiques	Détaillées Internes et externes Dynamiques
Temps	Présent	Du passé/présent au futur	Du passé/présent au futur
Accès	Peu de données (courantes)	Beaucoup de données (historisées)	Beaucoup de données (comme très peu)
Mise à jour	Très souvent	Périodiquement	Très souvent

¹⁰ Environnement Informatique Apprentissage Humain

¹¹ Learning Management System

¹² Massive Open Online Course

¹³ Small Private Online Course

¹⁴ Système Tutoriel Intelligent

Objectif	Visualisation de l'état	Prévision, projection et décision	Aider l'utilisateur dans sa tâche
----------	-------------------------	-----------------------------------	-----------------------------------

Tableau 1.1 - Tableau comparatif : système d'information opérationnel, décisionnel et système de recommandation (Nègre E., 2015)

Relatif à l'informatique décisionnelle, selon (Gorry G., 1971) un système d'information décisionnel (nommé *Management Decision System*) vise à « aider les utilisateurs dans la résolution de leurs problèmes de décision non programmés et complexes ». Par la suite, de nombreuses terminologies se sont succédé dans la littérature scientifique : SIAD (Système d'Information d'Aide à la Décision), SAD (Système d'Aide à la Décision), SID (Système d'Informatique Décisionnel). Néanmoins, il demeure communément admis qu'un système d'information décisionnel vise à piloter et à modéliser des problèmes réels au sein des organisations dans différents domaines (économie, finance, gestion...)

Au niveau de la conception d'un système d'aide à la décision, les techniques et méthodes existantes peuvent être classées selon les problèmes d'optimisation à résoudre. (Power, D. J., 2008, 2014) propose une classification qui distingue les systèmes de décision basés sur des approches différentes : (i) modèles, (ii) données, (iii) communication, (iiii) connaissances, (v) environnement web. Cette classification a considérablement évolué avec l'émergence de l'intelligence artificielle. Plus récemment, (Atif, L. 2017) propose une classification des systèmes d'aide à la décision guidés par les données inspirées des types de problèmes de décision génériques proposés par (Longueville B., 2003) soit :

-*Descriptive DSS*¹⁵ (*Descriptive Analytics*) : ensemble d'artefacts numériques permettant d'aider le décideur à résoudre des problèmes associés à la caractérisation réelle de l'état courant de l'organisation ;

-*Explicative DSS* (*Explicative Analytics*) : ensemble d'artefacts numériques permettant d'aider l'utilisateur/décideur à résoudre des problèmes associés aux relations entre deux ou plusieurs éléments de données ou phénomènes ;

-*Diagnostic DSS* (*Diagnostic Analytics*) : ensemble d'artefacts numériques permettant d'aider le décideur à résoudre des problèmes associés à l'établissement d'une relation cause à effet ;

-*Predictive DSS* (*Predictive Analytics*) : ensemble d'artefacts numériques permettant d'aider le décideur à résoudre des problèmes associés à une projection basée sur des données historiques ;

-*Prescriptive DSS* (*Prescriptive Analytics*) : ensemble d'artefacts numériques permettant d'aider le décideur à résoudre des problèmes associés à la projection normative basée sur des données historiques.

Ces multiples classifications existantes démontrent l'exercice difficile de catégoriser les systèmes d'aide à la décision compte tenu de la diversité des utilisateurs, des besoins et des technologies utilisées. De même, les multiples terminologies reflètent en partie l'évolution qui caractérise ce domaine de recherche. Néanmoins, la volumétrie, le traitement, la diversité des sources des données orientent davantage la recherche scientifique vers les systèmes d'aide à la décision guidée par les données. Selon (Salles, M. 2015), l'aide à la décision de type prédictive et prescriptive demeure une voie de recherche prometteuse compte tenu des capacités d'auto-apprentissages des techniques algorithmiques.

1.1.2 Technique d'aide à la décision multicritère

Associé à l'informatique décisionnelle (ou *Business Intelligence*), les systèmes d'information décisionnels ont intégré les techniques d'aide à la décision de façon disparate selon le domaine d'application (ressources humaines, médical, marketing...).

¹⁵ Decision Support System

Historiquement, la théorie de la décision se fonde sur un « ensemble des descriptions des problèmes de décisions à partir desquelles des analyses cohérentes peuvent être menées (...) et aussi pouvoir justifier des solutions rationnelles » selon (Kast R., 1993). Selon une approche mathématique appliquée, le décideur peut avoir recours à une analyse statistique pour résoudre un problème de décision. Par définition, la théorie statistique de la décision est une méthode d'évaluation des décisions fondée sur une information statistiquement élaborée. Selon (Fericelli, A. 1978), cette théorie a pour « objet de mettre en évidence un ordre de préférence sur des actions ou sur des stratégies, lorsque l'environnement est incertain, éclairant par-là, même, le choix que le décideur doit faire entre ces actions ou ces stratégies ». Cette approche de pondération directe, nommé également « analyse inverse », implique que le décideur intervient pleinement dans la pondération des critères lors du formalisme du modèle de décision à partir de données expérimentales observées. Cette méthode implique d'avoir à disposition un échantillon de données observables et représentatif pour connaître préalablement l'importance des critères. Selon (Nadeau R., 1986), cette théorie a suscité un intérêt considérable dans la littérature scientifique mais demeure peu utilisée en pratique au sein des organisations du fait des obstacles techniques et organisationnels sur une grande échelle. De plus, des biais d'évaluation sont soulignés liées à une élicitation explicite des critères de décision. Néanmoins, elle semble constituer un point de référence pour les méthodes de décision développées par la suite dans le domaine de la recherche opérationnelle. En effet, le domaine de l'aide à la décision multicritère a engendré une multitude de méthodes diverses selon un principe de paramétrage des préférences du décideur dans le processus de décision. Le choix d'une méthode spécifique reste dépendant du problème de décision à résoudre (mono ou multicritère) et du contexte d'application.

A partir des années 1960, les méthodes d'aide à la décision multicritère se sont considérablement développées et visent à résoudre des problèmes réels dans différents domaines de recherche opérationnelle (économie, finance, commerce...). Historiquement, l'aide à la décision trouve ses racines au XIXème siècle avec les travaux en économie d'Edgeworth (Edgeworth F.Y, 1885) et de Pareto (Pareto V., 1896) qui proposent une vision classique des problèmes de décision selon l'optimisation d'un critère unique. Cette approche s'est progressivement étendue parmi plusieurs domaines scientifiques (science de gestion, science de management, science pour l'ingénieur...) mais se révèle inadéquate au regard de la complexité des problèmes de décisions réels. En France, les méthodes d'aide à la décision de type multicritère ont été initiées par (Roy B., 1960, 1971). Synthétisé par (David, A., 2011), la théorie de Bernard Roy marque une rupture avec les travaux en recherche opérationnelle classique en prenant en compte plusieurs critères afin de répondre à la complexité des problèmes de décision. Il propose une approche plus réaliste en favorisant la construction de solutions raisonnables plutôt que rationnelle au sens de Pareto. Par la suite, de nombreuses méthodes d'aide à la décision se sont développées reposant sur un principe d'agrégation. (Sadok, W., 2008) distingue trois méthodes d'agrégation synthétisée par la figure suivante :

- Méthodes basées sur une échelle quantitative commune ;
- Méthodes basées sur un surclassement ;
- Méthodes basées sur des règles de décision.

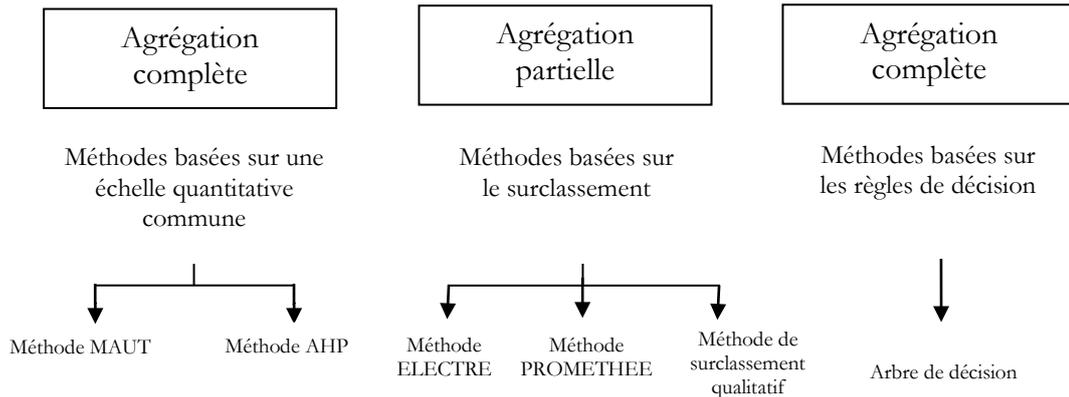


Figure 1.1 - Typologie simplifiée des méthodes d'agrégation (Sadok, W., 2008)

Selon (Le Gall A., 2009), les méthodes d'agrégation complètes incluent les méthodes basées sur une échelle quantitative commune (méthode MAUT¹⁶, méthode AHP¹⁷) et les règles de décision. Elles visent à synthétiser l'ensemble des indicateurs afin de faciliter la comparaison des décisions. Les méthodes d'agrégation partielle (méthode ELECTRE, méthode PROMETHEE, méthode de surclassement qualitatif) visent à refléter des situations où il n'y pas de classement précis des indicateurs à agréger. Les alternatives sont donc comparées deux à deux pour l'ensemble des indicateurs pour réaliser par la suite une synthèse de ces comparaisons. D'une manière globale, ces approches se basent principalement sur une interaction avec l'humain qui détermine l'importance de chaque critère en lui affectant une pondération directe.

Dans le domaine de la formation initiale, les méthodes d'aide à la décision multicritères sont appliquées comme par exemple pour la personnalisation interactive de parcours pédagogiques (Ho, H. 2015) ou pour optimiser l'analyse de l'apprentissage au sein de MOOC (Bouzayane, S., 2017), (Clivillé, V., 2018).

1.2 Système d'information de gestion

1.2.1 Système d'Information des Ressources Humaines

Les systèmes d'information de gestion ont pour objectif de soutenir les multiples activités de gestion au sein des organisations (contrôle des opérations, contrôle de gestion et de planification stratégique). Parmi les applications existantes, le SIRH (Système d'Information des Ressources Humaines) se définit comme un système d'information dédié essentiellement à la gestion des ressources humaines. A partir des années 1990, l'informatisation des outils de gestion s'est progressivement développée à travers les SIRH afin de centraliser et automatiser des tâches administratives courantes (gestion et contrôles des données de la paie, congés payés...) pour s'élargir par la suite à l'ensemble des domaines d'activités des ressources humaines (recrutement, évolution des carrières, formation, rémunération, qualité de vie du travail...).

Initialement, les SIRH ont recours à des PGI¹⁸ en les adaptant aux besoins des organisations. Par la suite, les méthodes informatiques ont permis de complexifier les SIRH selon une approche prévisionnelle en appliquant un processus de planification des activités. Cette évolution vise à appliquer un processus formel et systématique des activités de GRH en vue d'établir une adéquation entre les ressources humaines et la performance stratégique et organisationnelle de l'entreprise (Lemire, L., 2015). Les techniques algorithmiques se sont particulièrement focalisées

¹⁶ Multi Attribute Utility Theory

¹⁷ Analytic Hierarchy Process

¹⁸ Progiciel de Gestion Intégré

sur la planification automatique sous contraintes dans le domaine de la GTA¹⁹. Issue de la recherche opérationnelle, la planification automatique sous contraintes (ou ordonnancement) vise à ordonner la réalisation des différentes activités selon des contraintes temporelles ou de séquençement des ressources disponibles en quantité limitée. Les travaux scientifiques se sont particulièrement focalisés sur l'aide à la planification des activités pour automatiser et optimiser le traitement de la gestion du personnel (Franchini, L. 2000), (Draghici, C. 2005), (Redjem, R. 2013). Ainsi, la programmation par contrainte (ou PPC) est la principale technique utilisée pour la génération de planning en affectant le personnel durant une période fixe en prenant en compte les différentes contraintes de l'organisation (réglementation du travail, préférence individuelle...). D'autres travaux connexes utilisent cette technique pour optimiser le processus de décision interne comme par exemple : rationaliser la gestion prévisionnelle des emplois et des compétences (Dujarier M., 2010) ou l'aide à la constitution d'équipe et de groupe multifonctionnel en interne (Mansour, N., 2004), (Bennour, M., 2012).

Au niveau de la gestion de la formation, le rôle du SIRH est considéré comme une extension du LMS²⁰ (Cercle SIRH, 2017). Selon (Baudoin, E., 2019), ses fonctionnalités sont principalement axées pour l'aide à la gestion administrative et pédagogique des formations :

- Aide à l'élaboration du plan de formation ;
- Gestion du catalogue des formations ;
- Gestion administrative des formations en présentiel/distanciel ;
- Conception, diffusion et évaluation des contenus multimédias des formations en ligne ;
- Pilotage et déclaration administrative des formations.

Au niveau scientifique, les travaux de recherche axés sur la gestion et l'optimisation de l'activité de formation continue sont minoritaires et se focalisent principalement sur la gestion de la formation du personnel en interne (Bistorin, O., 2007). Plus récemment (Leon Blanco, M. 2019) propose des outils pour l'aide à la planification et la personnalisation des parcours de formation des apprenants en utilisant la programmation linéaire.

1.2.2 Evolutions

A partir des années 2010, les techniques algorithmiques liées à l'émergence de l'intelligence artificielle se sont sophistiquées dans le domaine des ressources humaines. Progressivement, le périmètre fonctionnel dédié à la gestion des SIRH s'est élargi à l'ensemble des activités des ressources humaines pour soutenir la prise de décision en intégrant la collecte, la modélisation, l'exploitation, la consolidation et la restitution des données de l'entreprise. De même, la volumétrie et la gestion des données se sont imposées pour permettre la compréhension des indicateurs d'activité et l'anticipation des actions pour piloter l'entreprise. Initialement, le SIRH centralise l'ensemble des données RH dans un système unique à travers des tableaux de bord. Ces outils visent à faciliter la gestion et l'exploitation des données dans un cadre réglementaire et juridique. Les fonctionnalités de simple contrôle (ou de « *reporting* ») utilisées pour le suivi des performances ont progressivement évolué vers l'aide à la prise de décision fondée sur des faits observables et quantifiables (Kun-fa, L., 2019). L'analyse de données liée aux ressources humaines (ou « *HR Analytics* ») vise à justifier des prises de décisions au-delà du simple outil de contrôle. Les travaux de recherche américains sont nombreux mais demeurent limités en France. Ce domaine de recherche s'est particulièrement observé d'une manière significative pour l'aide aux recrutements et la gestion des carrières (Jatobá, M., 2019). En effet, l'utilisation des algorithmes s'est étendue (raisonnement logique, méthode probabiliste, décisions en temps réel, théories des jeux) pour faciliter et optimiser le processus de recrutement (Russell, S., 2016), (Vedapradha, R., 2019). Ils sont particulièrement utilisés pour le traitement des données lors de la phase de sélection des candidats comme par exemple le tri des CV. De même, (Hemamou L., 2018) propose

¹⁹ Gestion des Temps et des Activités

²⁰ Learning Management System

de faciliter le recrutement par une analyse automatique de vidéos pour détecter les moments clés (ou significatifs) lors de l'entretien d'embauche. La modélisation repose sur un algorithme non linéaire pour émettre un avis au préalable de la décision humaine. (Bhalgat K., 2019) déclare que « cette évolution des pratiques dans le processus de recrutement va permettre d'optimiser l'acquisition de talents en supprimant les tâches répétitives et chronophages, minimiser les décisions humaines biaisées ainsi que les préjugés humains ». Néanmoins, ces nouvelles méthodes dans la gestion des ressources humaines impliquent de multiples défis dont l'adhésion et la prise en main technique des professionnels RH selon (Michaelides M., 2018). De plus, le caractère sensible de certaines données personnelles (âge, santé, sexe, orientation sexuelle...) induit des réflexions éthiques et juridiques (Black J., 2019).

1.3 Système d'information dédié à l'apprentissage humain

1.3.1 Apprenants

Les systèmes d'information dédiés aux apprentissages humains sont des objets sociotechniques complexes compte tenu du caractère pluridisciplinaire du domaine de recherche (science de l'éducation, psychologie cognitive, didactique, ergonomie...). La littérature scientifique s'est particulièrement focalisée sur les systèmes informatiques d'apprentissage humain destinés aux apprenants en formation initiale (élève, étudiant universitaire...). Historiquement, les plateformes utilisées pour l'apprentissage humain sont apparues avec l'émergence de l'ordinateur sous les abréviations EAO²¹ et EIAO²² initiant ainsi le principe de l'interactivité de l'apprenant avec le système informatique.

En France, ces systèmes informatiques ont été abordés sous le sigle EIAH²³ défini par (Tchounikine P., 2002, 2011) comme étant « un environnement intégrant des agents humains (apprenant ou enseignant) et artificiels (informatiques) et leur offrant des conditions d'interactions – localement ou à travers les réseaux informatiques – ou encore des conditions d'accès à des ressources formatives – humaines et/ou médiatisées – locales ou distribuées ». L'auteur souligne également une dissociation entre les systèmes « orientés pédagogie » qui se focalisent sur l'organisation et l'enseignement de contenus bien identifiés et les systèmes « orientés performance » qui se focalisent sur la réalisation d'activités par les apprenants. Parmi les systèmes « orientés pédagogie », l'avènement de la formation à distance (ou du e-learning) a permis l'émergence des systèmes d'apprentissage en ligne nommé LMS (ou *Learning Management System*). Ces systèmes informatiques visent la diffusion, la gestion, le stockage des contenus pédagogiques principalement destinés aux apprenants dans un contexte scolaire et universitaire. Ce type de support d'apprentissage unique présente l'inconvénient d'intégrer des contenus pédagogiques pré-fabriqués et ne pas pouvoir personnaliser le parcours pédagogique de l'élève (ou étudiant). Par ailleurs, les fonctionnalités axées sur la gestion de la formation sont peu mises en exergue et se limitent principalement à un tableau de bord instrumentés par divers outils (portfolio, wiki, reporting, réseautage social...) pour aider les enseignants à suivre la progression des apprenants.

En parallèle aux Etats-Unis, l'ACAO²⁴ plus connu sous l'abréviation anglo-saxonne CSCL²⁵, est officialisé lors d'une première conférence internationale organisé en 1995 à Bloomington (Koschmann T., 1995). Synthétisé par (Resta P., 2007), l'objectif principal de ce domaine de recherche est de favoriser la collaboration entre les apprenants en facilitant la création et le partage des connaissances à travers les interactions entre les pairs. Des outils de communication du web 2.0 (wikis, blogs, forums...) et des méthodes basées sur les techniques de l'analyse des

²¹ Enseignement Assisté par Ordinateur

²² Environnement Intelligent Assisté par Ordinateur

²³ Environnement Informatique d'Apprentissage Humain

²⁴ Apprentissage Collaboratif Assisté par Ordinateur

²⁵ Computer Supported Collaborative Learning

réseaux sociaux (ou SNA²⁶) sont proposés pour soutenir l'activité collaborative des apprenants afin de résoudre leurs problèmes d'apprentissage (Dalsgaard, C., 2006).

A partir des années 2010, les systèmes d'apprentissage en ligne se sont popularisés liés à l'accessibilité et à la massification des contenus pédagogique à travers les MOOC²⁷. (Bigrat, F., 2019) décrit le MOOC comme un environnement relativement figé et représenté par des dispositifs de transmission directe du savoir (enregistrement vidéo, documents pédagogiques et de dispositifs d'interactions et d'auto-gestion). Ces fonctionnalités diffèrent peu des LMS mis à part la possibilité d'exploiter des données plus volumineuses et diversifiées (profil des apprenants, question/réponse sur les forums...). L'analyse des données permet d'observer le comportement des apprenants dans une perspective de personnalisation des parcours pédagogiques. De plus, les MOOC permettent un accès élargi aux savoirs des apprenants au-delà du cadre scolaire et universitaire dans un contexte sociétal valorisant la FLTV²⁸.

Dans le cadre de la formation continue, les outils numériques utilisés sont similaires aux outils de la formation initiale. Il s'agit principalement des plateformes d'e-learning *via* les LMS, les MOOC et ses dérivés (SPOC²⁹, COOC³⁰...). Néanmoins, en comparaison avec la formation initiale, elles sont peu utilisées dans le cadre de la formation pour adulte où la formation en présentiel prédomine³¹. Les pratiques en formation continue peuvent également se traduire par un dispositif hybride (ou *Blended-Learning*) qui alterne des phases en présentiel et distanciel médiatisé *via* un système informatique d'e-learning. Particulièrement adaptée pour l'adulte, cette modalité de formation permet une souplesse d'organisation et d'auto-gestion pour l'apprenant combinant son temps de formation et ses obligations professionnelles comme le souligne (Poizat, G., 2017) dont « les formats, les espaces et les temporalités des MOOCs sont particulièrement adaptés à la formation pour adulte favorisant ainsi un apprentissage de type non-formel, en tout temps et en tout lieu ». A l'instar du MOOC, le SPOC est une formation à distance destinée exclusivement à un nombre restreint d'utilisateurs. Selon (Oliveri, N. 2017) (Filius, R. 2018), la valeur ajoutée du SPOC par rapport au MOOC est le nombre restreint de participants qui permet une rétroaction avec les pairs et ainsi un apprentissage plus approfondi. De même, (Lu H., 2018) constate une réduction du taux d'abandon en comparaison avec les MOOC. Il est également constaté que l'usage des plateformes d'e-learning peut être élargi à d'autres finalités d'applications. Au sein des entreprises, les MOOC permettent aux salariés d'acquérir des connaissances dans un domaine professionnel spécifique mais également de transmettre les valeurs de l'entreprise (Rafiq & al. 2019). Selon (Yao, J., 2019), les SPOC visent également à favoriser la mobilisation des salariés en interne d'une entreprise en complément de la recherche de perfectionnement des compétences professionnelles. Néanmoins, les SPOC sont destinés à des professionnels issus d'entreprises différentes, à l'inverse des COOC. Ces derniers sont exclusivement destinés aux salariés en interne d'une entreprise. Ce format numérique s'inscrit dans une démarche cognitive mais également sociale en favorisant la cohésion entre les salariés. (Acquatella, F. 2016) souligne également une dimension communicationnelle, sociale et coopérative des COOC permettant aux entreprises de créer une relation de confiance et de proximité avec les utilisateurs en facilitant la transmission d'informations.

²⁶ Social Networks Analyses

²⁷ Massive Open Online Courses

²⁸ Formation tout au Long de la Vie

²⁹ Small Private Online Course

³⁰ Corporate Online Open Course

³¹ 95 % des sessions de formation en présentiel (Rapport : IGAS-Inspection générale des Affaires Sociales, 2017)

1.3.2 Enseignants et formateurs

Les systèmes d'information dédiés à l'apprentissage humain sont principalement destinés aux apprenants. En comparaison, peu de travaux scientifiques se sont orientés vers les systèmes d'information destinés exclusivement aux enseignants. La notion de SAD (« Système d'Aide à la Décision ») est inexistante mais peut être associée au « Système Tutoriel Intelligent » (ou « *Intelligent Tutoring System* »). Selon (Garrot, É. 2007), le système d'aide au tuteur permet « la possibilité au tuteur en formation à distance de concevoir de façon dynamique les situations d'apprentissage de façon à coller au plus près aux besoins des apprenants ». Ainsi, le système d'aide au tuteur a pour but d'aider l'enseignant à être le relais entre des situations d'apprentissage génériques conçues par l'auteur (concepteur pédagogique) et des situations d'apprentissage réalisées par les apprenants. Ainsi, les STI visent initialement à faciliter les activités didactiques des enseignants. Définis par (Bourdeau J., 2011) comme des « environnements d'apprentissage informatisés qui visent à imiter le comportement d'un tuteur humain dans ses capacités d'expert pédagogue et d'expert du domaine », les STI sont destinés tout autant aux enseignants qu'aux apprenants selon des finalités d'application différentes. Pour les enseignants, les STI peuvent par exemple (i) aider à la planification de cours à distance pour optimiser les choix des activités d'apprentissage (Zaraté, P. 2005) ou (ii) diagnostiquer des profils des apprenants afin de définir leurs besoins d'apprentissage (Delozanne, É., 2008) (Chenevotot-Quentin, F 2012). Pour les apprenants, les STI visent principalement à fournir des rétroactions pertinentes lors de la résolution de problèmes ou d'exercices dans un domaine précis. Parmi les travaux existants, les usages d'application sont divers comme par exemple faciliter une évaluation automatisée et optimiser des corrections d'exercices (Prévit, D. 2008), (Grugeon-Allys, B. 2016). Dans le domaine de la formation professionnelle, les travaux scientifiques restent minoritaires. (Fougères, A. J. 2001) propose un STI basé sur une stratégie pédagogique de simulation, d'interaction et de coopération entre un agent pédagogique et un apprenant dans un domaine métier spécifique.

Plus récemment, les systèmes de recommandation appliquent une méthodologie de conception commune avec les STI en se basant sur une base de données (ou de connaissances) pour la modélisation de l'apprenant et du domaine. Initialement utilisés dans le commerce en ligne, les systèmes de recommandation se sont élargis progressivement dans le domaine de l'éducation. Ces systèmes d'information visent à pallier un problème de surcharge d'informations disponibles en proposant des items susceptibles d'intéresser les utilisateurs. Synthétisés par (Manouselis N., 2011) (Drachler, H., 2015), les axes de recherches sont multiples : recommandation de ressources pédagogiques et de scénarios d'apprentissage, personnalisation du modèle utilisateur... Parmi les axes de recherches existants, l'exploitation des données (ou traces d'activités) liées aux caractéristiques et aux activités des apprenants a été particulièrement étudiées dans le domaine de l'apprentissage (Dwivedi S., 2017, 2018). Les techniques de filtrage de données sont diverses : (i) le filtrage collaboratif, (ii) le filtrage sur le contenu et plus récemment (iii) le filtrage hybride c'est-à-dire combinant différentes sources hétérogènes pour pallier les limites de performances des approches classiques. Les techniques de filtrage restent principalement axées sur la suggestion d'items individuels³². Par la suite, les systèmes de recommandation se sont élargis vers des systèmes communément appelés « Système de Recommandation de Groupe » (ou *Group recommender System*). Ce type d'applications s'est particulièrement développé dans le domaine des loisirs et du divertissement (musique, films, tourisme...) et vise à suggérer des objets qui pourraient intéresser un groupe de personnes déterminé (Jameson, A. 2004, 2007), (Felfernig, A. 2018). Dans le domaine de la formation, les systèmes de recommandation de groupes sont essentiellement appliqués pour la recommandation de ressources pédagogiques spécifiques à un groupe d'apprenants (Dwivedi, P., 2015) (Yanhui, D., 2015) (Papamitsiou, Z., 2018). Le principal défi du traitement lié à la recommandation de

³² Terme usuel dans le domaine des systèmes de recommandation pour désigner un élément à recommander

groupe est de fusionner les préférences individuelles des apprenants tout en prenant en compte des données hétérogènes (styles d'apprentissages, niveaux de connaissances ...). Parmi les techniques existantes, la stratégie de l'agrégation des préférences est la plus répandue. Cette méthode consiste à traiter et optimiser les résultats en agrégeant des préférences individuelles permettant ainsi de déterminer une tendance collective (Márquez, J. 2015).

Chapitre 2. Analyse et visualisation des données en formation initiale et continue

Ce deuxième chapitre présente les méthodes de traitement des données selon les domaines scientifiques dont l'analyse de l'apprentissage (ou *Learning Analytics*) et l'analytique RH (ou *HR Analytics*). Les outils associés dont le tableau de bord et la visualisation des données principalement destinés à soutenir l'interprétation humaine sont détaillés.

2.1 Analyse et traitement des données

2.1.1 Analyse de l'apprentissage humain (ou *Learning Analytics*)

L'analyse de l'apprentissage humain³³ (ou *Learning Analytics*) est un domaine de recherche relativement récent et officialisé lors de la création de la communauté scientifique SOLAR³⁴ en 2011. Une définition communément admise est proposée par (Siemens G., 2010, 2011, 2012) synthétisée par (Elias, T. 2011) comme étant une discipline dont « la mesure, la collection, l'analyse et l'interprétation des traces des apprenants et de leurs contextes, pour comprendre et optimiser l'apprentissage et les environnements dans lesquels il se produit ». Selon (Bienkowski M., 2012), le domaine des Learning Analytics (noté LA par la suite) est un vaste domaine interdisciplinaire incluant plusieurs techniques issues du domaine de l'informatique décisionnelle (ou *Business Intelligence*), Web Analytics, exploration de données éducatives³⁵, systèmes de recommandation, apprentissage automatique, intelligence artificielle, recherche d'informations, statistiques et visualisation de l'information. Etroitement lié à l'EDM (ou *Educational Data Mining*), l'objectif commun avec les LA est de comprendre le processus d'apprentissage des étudiants sur la base de l'analyse de données éducatives à grande échelle. Selon (Charleer, S., 2014), la différence notable est la valorisation des données pour l'exploration du jugement humain à travers des outils de gestion tel que le tableau de bord et la visualisation des données.

Selon une méta-analyse des travaux scientifiques proposé par (Viberg, O. 2018) qui synthétise 252 publications de 2012 à 2017, la majorité des travaux de recherche applique une approche descriptive en utilisant des méthodes de collecte de données interprétatives et uniques (qualitative ou quantitative). Les méthodes mixtes c'est-à-dire combinant des méthodes qualitatives et quantitatives restent minoritaires. Plus récemment, des travaux de recherche appliquent une approche prédictive en mobilisant des données passées (ou historisées) pour prédire des comportements. Nommé PLA (ou *Predictive Learning Analytics*), cette branche du domaine vise à améliorer l'apprentissage notamment pour identifier les étudiants dont le risque d'abandon est élevé (Sclater, N., 2016), (Lu, O., 2018), (Herodotou C., 2019).

D'une manière plus restreinte, les techniques de LA sont également utilisées pour l'aide à la décision au service de la gouvernance des universités. Initialement nommée sous la notion «*Institutional Analytics*», cette branche du domaine fait « référence à l'analyse de données multiples et volumineuses pour améliorer des prises de décisions stratégiques au niveau institutionnel » (Daniel, B. 2016). Au sein des organisations universitaires, (Nieto, Y., 2019) identifie trois niveaux de décision :

- Niveau de décision de type stratégique relative à la gestion propre de la gouvernance de décision (par exemple les choix relatifs aux allocations de ressources financières) ;
- Niveau de décision de type tactique qui vise à mettre en œuvre les décisions stratégiques préalablement déterminés ;

³³ Traduction en français de l'expression Learning Analytics/Appelé également « Analytique de l'apprentissage »

³⁴ Society for Learning Analytics Research

³⁵ Appelé selon le terme anglophone EDM (Educational Data Mining)

-Niveau de décision de type opérationnelle relative aux tâches courantes destinées à l'ensemble des acteurs de l'organisation universitaire (personnel de direction, personnel enseignant, personnel administratif).

Pour ce dernier cas, les systèmes d'information de type ERP³⁶ sont principalement utilisés pour l'aide à la planification des ressources pédagogiques, logistiques et humaines. Initialement déployé au sein des organisations industrielles, les ERP sont des progiciels de gestion intégrés qui permettent de gérer l'ensemble des processus d'une entreprise (gestion des ressources humaines, gestion comptable et financière, l'aide à la décision...). Progressivement, les institutions universitaires ont déployé ce type de logiciel en adaptant leurs fonctionnalités selon les problématiques propres à la gouvernance des universités (Noaman, A.Y., 2015).

Au niveau du traitement, les données brutes sont analysées selon des techniques statistiques afin de mener une interprétation finale. (Baker, R., 2014) identifie cinq techniques statistiques :

- Regroupement des données par grappes (ou cluster) ;
- Exploration des relations selon une analyse de corrélation ;
- Analyse prédictive (ou analyse de régression) ;
- Méthode de visualisation.

Fondés sur la collecte des données, les méthodes liées aux LA restent dépendantes du volume et de la qualité des données recueillies. Ainsi, les critiques restent principalement axées sur les problèmes méthodologiques et le type des données collectées. Selon les considérations méthodologiques du domaine des SHS défini par (Paquay, 2006) repris par (Peraya D. 2019), trois critères sont à distinguer pour évaluer la qualité des données soit : (i) la pertinence, (ii) la validité et (iii) la fiabilité des données. (Ouakrat A., 2016) avance que « les hypothèses de recherche sont dépendantes de la nature des données disponibles dans les plateformes de type MOOC ou LMS pouvant impliquer un biais dans la pertinence des données ». De même, des limites concernant la validité des données sont révélés selon les auteurs qui déclarent que ces données « demeurent une traduction très partielle et limitée des pratiques. Elles décrivent des liens sociaux de façon restreintes et circonscrites et ne représentent qu'un pan de l'individu ». Selon (Boyer A., 2019), la notion d'utilité de la donnée consiste « (...) à déterminer ce que la donnée apporte relativement à ce qu'elle coûte, que ce soit en termes monétaires, éthiques, complexité de collecte ou niveau de performance ».

En parallèle, le traitement des données personnelles induit également des questionnements éthiques dont la protection de la vie privée des étudiants (confidentialité des données, consentement éclairé et le maintien de l'anonymat), la conservation et la sécurité des données personnelles (Gras, B., 2019), (Peraya, D., 2019).

2.1.2 Analytique RH (ou *HR Analytics*)

Selon (Coron C., 2019), l'omniprésence de la quantification est un phénomène sociétal qui s'est progressivement élargi à divers secteurs d'activité (associatif, gouvernemental et industriel). Dans le domaine des ressources humaines, la quantification des données devient une pratique centrale particulièrement au sein des grandes organisations. L'auteur distingue deux approches : une approche valorisant l'usage de la quantification comme gage d'objectivité (ou de scientificité) et une approche plus critique remettant en cause le mythe d'une mesure totalement neutre ou objective.

La première approche tend à considérer la quantification comme nécessaire pour appuyer et améliorer la prise de décision en RH selon une approche EBM (*Evidence-Based Management*). Son principe vise une élaboration de connaissances théoriques à partir d'une observation rigoureuse des faits afin d'améliorer l'efficacité et l'efficience des décisions managériales. Initialement, cette

³⁶ Enterprise Resource Planning

approche descriptive de l'analyse de données s'inscrit dans une démarche d'informatique décisionnelle pour aider les responsables de management à prendre des décisions sur le personnel (recrutement, formation, planification des carrières ...). Par la suite, la quantification des données en ressources humaines est définie sous le concept « *HR Analytics* » par (Mortensen, M., 2015), repris par (Angrave, D. 2016) comme étant « au croisement de l'ingénierie, de l'informatique, de la prise de décision et des méthodes quantitatives pour organiser, analyser et donner un sens aux quantités croissantes de données générées par les sociétés contemporaines ». L'émergence des algorithmes et la volumétrie des données disponibles ont progressivement modifié les usages de l'analytique RH. (Marler, J., 2017), élargit cette définition comme étant « une pratique RH permise par la technologie de l'information qui utilise des analyses descriptives, visuelles et statistiques des données liées aux processus RH, au capital humain, à la performance organisationnelle et aux repères économiques externes pour établir l'impact et permettre la prise de données en s'appuyant sur les données ». (Pertinant G., 2017) différencie la notion « *HR Analytics* » avec la notion du « *Big Data* » par la nature des données qui sont internes et structurés dans le cadre d'une analyse des ressources humaines tandis que les données du « *Big Data* » sont volumineuses, non-structurées et externes à l'organisation. Ainsi, (Bruno, I. 2015) explique que l'analyse statistique et la quantification des faits observables apportent des garanties de neutralité et d'objectivité pour justifier des prises de décision de gouvernance au sein des organisations. De même, les exigences méthodologiques concernant la qualité des données sont identiques au domaine des LA. (Dijkkamp J., 2019) note que la qualité des données comme source d'entrée devient une condition majeure pour produire un traitement et une analyse des résultats exploitable.

En parallèle, les approches de traitement de type prédictif ont pris un essor considérable, soutenues par les techniques émergentes d'intelligence artificielles. Selon (O'Neil C., 2016), les algorithmes de prédiction sont principalement utilisés dans le domaine du recrutement, la mobilité interne, la formation ou la prédiction des risques RH (absentéisme, démission...). La prédiction pour l'aide au recrutement est un domaine de recherche particulièrement étudié notamment durant la phase de sélection et de gestion des candidatures en réponse à des offres d'emploi. Ces méthodes de traitement peuvent néanmoins impliquer des biais de sélection potentiellement discriminatoire. (Besse P, 2020) préconise une mise en place d'audits et de documentation au préalable lors de la conception d'un système informatique pour assurer une transparence des choix des candidats. De même, l'utilisation et le traitement des données personnelles soulèvent des questions notamment sur le contrôle des salariés et la transparence des finalités d'utilisation des données.

Dans le domaine de la formation, (Cossette M., 2014) propose une approche globale de l'analytique des ressources humaines qui ne consiste pas seulement à quantifier un phénomène RH. Il s'agit d'interpréter les résultats obtenus et de les corrélérer avec différents indicateurs afin d'identifier des leviers d'actions. L'auteur souligne que l'analyse des données issues des questionnaires de satisfaction à l'issue des actions de formation sont pertinentes pour optimiser le coût de la formation mais également pour améliorer le transfert des apprentissages des salariés au sein de l'organisation.

2.2 Outils de gestion (ou management) des données

2.2.1 Indicateurs

Selon la définition proposée par (Boulangier P., 2004), « les indicateurs synthétisent ou simplifient des données (ou des variables) jugées pertinentes pour rendre compte d'un phénomène qui ne peut être décrit directement en raison de sa complexité ou pour des raisons de faisabilité. Ils reposent sur une relation connue ou supposée entre la ou les variable(s) et le phénomène évalué ». Un indicateur peut être de différentes natures listé par (Feschet, P., 2016) :

- Quantitatif (cardinal) ou qualitatif (nominal ou ordinal) ;

- Descriptif ou prescriptif ;
- Elémentaire (constitué de données brutes) ou agrégé (élaboré à partir de plusieurs indicateurs) ;
- Simple à obtenir (constitué de données brutes) ou complexe (calcul élaboré intégrant de nombreuses variables) ;
- Mesuré, calculé (calcul simple ou à partir d'un modèle) ou observé, déclaré (indicateur subjectif) ;

Dans le domaine des LA, les données brutes utilisées sont principalement les données personnelles de l'apprenant et les traces d'activités lors de sa navigation au sein d'un système informatique dédié à l'apprentissage humain. (Verbert, K., 2013) identifie les catégories de données qui peuvent être intégrées sous forme d'indicateurs au sein des tableaux de bord :

- Données liées à la consultation des ressources pédagogiques ;
- Données temporelles relatives au temps passé par l'apprenant ;
- Données relatives aux interactions sociales afin d'évaluer le degré d'engagement de l'apprenant ;
- Données relatives aux évaluations afin de suivre la progression des apprenants.

Plus récemment, (Holzer, A., 2016) identifie plus d'une centaine d'indicateurs parmi les travaux scientifiques existants. Il propose une catégorisation des indicateurs de la manière suivante :

Catégories	Questions	Exemples
Apprenant	Qui sont les apprenants ?	Education, âge, compétences, cours antérieurs, grade universitaire
Action	Que font-ils ?	Nombre de pages visités, nombre de fichiers téléchargés, temps consacré aux tâches, temps de connexion
Contenu	Quel est le contenu utilisé dans l'apprentissage ?	Nombre et sentiments des messages (forum par exemple), sujet traités
Résultat	Quels sont les résultat(s) ?	Note moyenne, répartition des notes dans un groupe
Contexte	Dans quel contexte se déroule l'apprentissage ?	Emplacement des apprenants autour de la table, emplacement dans une salle de classe, situation géographique
Social	Comment les apprenants interagissent avec autrui pendant l'apprentissage ?	Réseau exploitant les communications avec les autres membres du groupe, direction des interactions autour d'une table

Tableau 2.1 - Synthèse des catégories d'indicateurs des *Learning Analytics* (Holzer, A., 2016)

Dans le domaine des ressources humaines, le pilotage des activités de GRH à partir d'indicateurs est une pratique courante afin d'évaluer la performance des stratégies des organisations. Les indicateurs sont définis selon les besoins souhaités et les données disponibles de l'organisation. D'une manière synthétique, (Marler, J.R, 2017) proposent trois types d'indicateurs :

- Indicateurs d'efficacité qui se concentrent essentiellement sur les coûts et la productivité des activités de GRH ;
- Indicateurs d'efficacité des programmes GRH qui visent à développer les compétences, la motivation et l'attitude des salariés ;
- Indicateurs d'impact mesurant l'effet des processus et programmes de GRH sur la performance de l'organisation.

Au niveau de la qualité des indicateurs, (Voyer P, 2011) établit une liste de critères pour assurer une valeur optimale d'un indicateur soit : (i) la pertinence, (ii) la qualité et la précision de sa mesure, (iii) la faisabilité et (iiii) la convivialité de l'interprétation et d'utilisation.

Ainsi, les indicateurs initialement d'ordre économique se sont progressivement axés sur des indicateurs à caractère social et humain pour améliorer le processus RH. De même que les LA, la qualité des données en HR est un impératif méthodologique pour mener une analyse et une interprétation des données pertinentes.

2.2.2 Tableau de bord

La présentation des données sous forme de tableaux de bord vise à centraliser et agencer les indicateurs et les données préalablement définis.

Dans le domaine de l'analyse de l'apprentissage, le tableau de bord est défini selon diverses terminologies : tableau de bord d'analyse de l'apprentissage (ou *learning analytics dashboard*), tableau de bord de l'apprentissage (ou *learning dashboard*), tableau de bord éducatif (ou *educational dashboard*). (Schwendimann B., 2017) définit un tableau de bord comme « une synthèse unique agrégeant différents indicateurs sur les apprenants, les processus d'apprentissage et les contextes d'apprentissage au travers d'une ou de plusieurs visualisations ».

Les tableaux de bord sont principalement implémentés au sein des plateformes de type LMS³⁷ et MOOC destinés aux enseignants et aux apprenants (Holzer, A., 2016). D'une part, les tableaux de bord permettent aux étudiants de pratiquer une auto-observation de leurs activités pédagogiques et d'autre part, les enseignants peuvent assurer et personnaliser le suivi des étudiants (Bistodeau, A., 2019).

Au-delà du cadre universitaire, les tableaux de bord ont été largement utilisés comme outil de gestion et de management au sein des entreprises. Historiquement, le tableau de bord est défini comme une interface qui « représente une structure intermédiaire d'accès à l'information, de filtrage, de réorganisation et de présentation de cette information pertinente à la gestion » (Voyer P., 1994, réédité en 2011). Initialement, les tableaux de bord sont utilisés dans le domaine de la GRH comme des outils de « reporting » pour répondre à des contraintes légales (bilan social, rapport de situation comparée...). Progressivement, ils sont devenus des outils de gestion présentant de manière synthétique les activités et les résultats du service des ressources humaines. Dans ce cadre d'application, (Le Louarn J-Y., 2008) identifie quatre types de tableaux de bord :

- Tableau de bord opérationnel axé sur les processus RH (recrutement, rémunération, formation, évaluation) ;
- Tableau de bord sur les salariés (effectifs, attitudes, comportements) ;
- Tableau de bord stratégique (bilan des compétences) ;
- Tableau de bord de gestion prévisionnelle des dépenses.

Ainsi, les indicateurs visent à évaluer la réalisation d'objectifs fixés en amont et prendre des décisions nécessaires sur une période précise et dans un délai limité.

Initialement, les tableaux de bord RH centralisent principalement des données internes et administratives pour optimiser la fonction RH au sein d'une organisation. Avec l'avènement du « *Big-Data* », des données externes peuvent être collectées en complément des données internes pour optimiser l'analyse des processus RH (Garcia-Arroyo J., 2019). La notion de TBDM (Tableaux de Bord sur Données Massives) a été introduite par (Vayre, J. S. 2015) liée à l'expansion de la volumétrie des données.

³⁷ Moodle

2.3 Visualisation des données

La notion de « visualisation des données » (nommée également « *data-visualisation* ») correspond à une technique d'exploration et d'interprétation des données. Selon (Mizzaro, S. 1996), ce domaine de recherche nommé « Visualisation d'Information » vise à une représentation visuelle des données permettant de répondre à un besoin informationnel et de transiter d'un état de connaissance vers un autre. Ainsi, la visualisation de l'information vise principalement à amplifier la cognition, à construire du sens à partir de sources de données brutes, en s'appuyant sur les dimensions graphiques et interactives permises par l'informatique (Card M., 1999). Selon (Auber D., 2017), deux types de visualisation sont à distinguer :

- Visualisation analytique permettant à l'utilisateur d'explorer les données afin de rechercher des informations ou des phénomènes méconnus ;
- Visualisation descriptive permettant d'expliquer visuellement un phénomène déjà identifié.

2.3.1 Visualisation analytique

La visualisation analytique est une branche de la visualisation de l'information dont l'objectif est de permettre aux utilisateurs d'analyser des quantités croissantes de données. Selon (Van Wijk J., 2005), la finalité est de « produire des représentations graphiques capables de seconder l'utilisateur dans la formulation d'hypothèses sur le problème étudié, voire dans la découverte de nouvelles connaissances ».

Le choix du type de visualisation dépend de nombreux paramètres liés à la tâche et au contexte de l'utilisateur (Vuillemot R., 2010). Dans le cadre de l'aide à la décision, le processus de conception proposé par (Keim D., 2010) combine des méthodes d'analyse automatique en lien avec des interactions humaines. Comme illustré dans la figure suivante, cette méthodologie implique un processus selon les étapes suivantes :

- Pré-traitement et transformation des données d'entrée (« Data ») ;
- Choix de modélisation (« Models ») ;
- Extraction de connaissances (« Knowledge ») ;
- Visualisation pour l'interprétation des données (« Visualisation »).

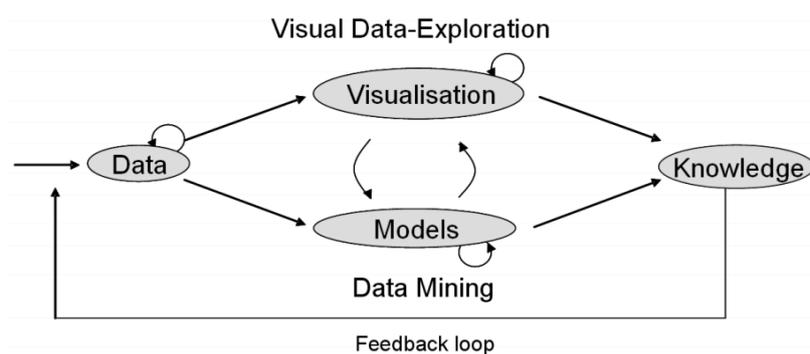


Figure 2.1 - Processus de visualisation de type analytique pour une aide à la décision interactive et évolutive (Keim D., 2010)

Selon l'auteur, la visualisation devient le support d'un processus analytique semi-automatisée, où les humains et la machine coopèrent pour une restitution des résultats plus efficaces.

Au niveau du graphisme, la difficulté majeure est la restitution de la volumétrie des données tout en permettant une lisibilité et une compréhension des données. Une visualisation unique n'est pas toujours suffisante pour restituer l'information utile d'une manière exhaustive. Des techniques d'interaction complémentaires (navigation, exploration, filtrage, sélection...) peuvent être utilisées pour faciliter l'interprétation des données à différents niveaux de détails. De plus, la complexité et la nature dynamique des données impliquent que les types de visualisation se sont enrichis tant au niveau du traitement que de la conception. Récemment, (Cuenca, E. 2018) propose une visualisation pour faciliter l'analyse et l'exploration de multiples séries temporelles à différents niveaux de granularités et une visualisation sous forme de graphes multicouches.

Dans le domaine de l'apprentissage humain, la visualisation des données connaît un regain d'intérêt dans la recherche scientifique avec l'émergence de l'analyse de l'apprentissage (ou *Learning Analytics*). Les techniques de visualisation sont utilisées pour faciliter l'aide à la décision et l'interprétation des données. Les outils de visualisation sont principalement déployés en complément d'un tableau de bord (Duval, E. 2011), (Leony, D., 2012), (Bodily, R., 2017, 2018). Néanmoins, (Schwendimann, B. 2016), (Vieira, C., 2018) précise que les techniques traditionnelles de visualisation statistique (diagramme à barres, diagramme de dispersion, graphique circulaire, graphique de réseaux) sont les plus couramment utilisées. Les travaux scientifiques s'appuyant sur des techniques de visualisation sophistiquées restent minoritaires comme (Poon, L. K. 2017) qui utilise un regroupement hiérarchique pour visualiser les traces collectées à partir d'un LMS.

2.3.2 Visualisation descriptive

Selon (Auber D., 2017), la visualisation descriptive permet de représenter et expliquer visuellement un phénomène déjà identifié.

Dans le domaine spécifique du CSCL³⁸, la visualisation descriptive a été particulièrement abordée en s'appuyant sur une approche structurale de l'analyse des réseaux sociaux (ou *SNA*)³⁹. Historiquement, l'analyse des réseaux sociaux comporte un ensemble de méthodes issues de la sociologie, de la psychologie et de l'anthropologie pouvant être mobilisé dans divers cadres théoriques. Initiée par le sociologue Georg Simmel, cette approche structurale permet de révéler des structures cachées et des phénomènes complexes à partir d'unités relationnelles élémentaires. Les structures relationnelles sont théorisées sous le concept de « dyade⁴⁰ », étendue à l'échelle des « triades », qui prend en compte trois acteurs (Simmel G., 1908). Par la suite, (Moreno J., 1934) formalise les relations sociales représentées sous forme de sociogramme initiant ainsi les bases de la sociométrie. La représentation graphique des réseaux relationnels entre personnes vise à faciliter « l'exploration des structures sociales à la lumière des attractions et des répulsions qui se sont manifestés au sein d'un groupe ».

Dans le cadre de l'apprentissage et de la formation en ligne, les outils de visualisation sont principalement déployés au sein d'environnement d'apprentissage à distance pour aider l'apprenant à interagir avec un environnement virtuel. Au niveau de l'ergonomie, les interfaces de visualisation sont principalement utilisées pour mettre en évidence le suivi de l'apprenant dans une perspective de métacognition (Cram, D., 2007), (Temperman, G. 2013) ou pour mettre en évidence les interactions sociales au sein des communautés d'apprentissage à partir des traces d'interactions en ligne (discussion des forums...) (Dimitracopoulou, A., 2006). Cet outil de visualisation est particulièrement utilisé au sein des LMS selon deux types de visualisation : une approche « socio-centrique » et une approche « égo-centrique » (Sie, R. L., 2012). La visualisation

³⁸ « Computer Supported Collaborative Learning »

³⁹ « Social Networks Analyses » en anglais

⁴⁰ Forme sociologique élémentaire soit la relation entre deux éléments

permet d'analyser le processus collaboratif selon une direction des relations qui peut être considérée comme « dirigée » (la personne A apprend de B mais pas l'inverse) ou « non-dirigé » (les personnes A et B apprennent mutuellement sans distinction). Plus récemment, des techniques de visualisations sous la forme de graphes multicouches permettent de modéliser les différentes relations notamment dans le contexte d'application lié aux réseaux sociaux (Cuenca, E. 2018). Néanmoins, il est à noter que les outils de visualisation reposent principalement sur une représentation de type nœud-lien. Ce type de visualisation implique des règles d'observations pour la restitution des données sous une forme intelligible. Historiquement, les travaux en sémiologie graphique du cartographe Jacques Bertin repris par (Metz, C. 1971) ont initié un ensemble de règles permettant le rendu visuel des différentes formes graphiques en privilégiant le sens de la vue et la perception pour le traitement et la représentation des données. L'utilisation des représentations de type nœud-lien implique également un volume restreint de données pour assurer une bonne lisibilité du graphe. Dans le même esprit, (Fruchterman et al., 1991) propose des règles ergonomiques pour l'utilisation des graphes dans une interface utilisateur : (i) disposition des sommets dans la fenêtre de représentation, (ii) minimisation des croisements des arêtes et le (iii) respect d'une certaine symétrie dans la disposition du graphe (répartition équitable).

Chapitre 3. Analyse multidimensionnelle du groupe de formation pour adulte

Le groupe de formation en présentiel demeure le mode d'action le plus couramment utilisé en formation continue. Au préalable, ce chapitre présente la notion du groupe de formation et les différents acteurs dans un contexte de formation continue. Par la suite, les multiples aspects du groupe de formation sont synthétisés selon une dimension sociale (ou relationnelle), cognitive, organisationnelle et environnementale.

3.1 Type de groupe de formation

3.1.1 Groupe de formation

En formation initiale, l'apprentissage humain s'est opéré traditionnellement à travers un groupe d'apprenants sous forme de classe. Dans un contexte de formation pour adulte, les mécanismes du groupe de formation diffèrent selon plusieurs facteurs.

En premier lieu, l'étude du groupe social s'appuie sur une vaste littérature scientifique pluridisciplinaire : sociologie, psychologie sociale et cognitive. (Fisher G., 2005) synthétise cette notion en proposant la définition suivante : « le groupe est une entité sociale identifiable et structurée, caractérisée par un nombre restreint de personnes liées entre elles et par des activités communes et interdépendantes et qui développent des interactions déterminées par des normes de conduites et de valeurs communes, dans la poursuite de leurs objectifs ». Il apparaît également que la notion du groupe peut être associée à des notions voisines dont la foule ou la communauté, qui sont assimilées à tort selon (Dillenbourg, P., 2003). Les auteurs opèrent une distinction des types de groupes selon le continuum suivant :

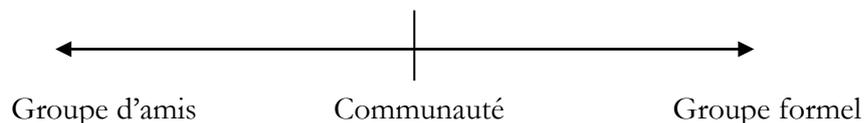


Figure 3.1 - Position de la communauté par comparaison à d'autres groupes sociaux (Dillenbourg, P., 2003)

Les trois formes d'organisation sociale se différencient selon l'initiateur du groupe : la création d'un groupe formel relève de l'initiative d'une personne (supérieur hiérarchique, expert, formateur...) tandis que l'adhésion à une communauté est volontaire selon les affinités professionnelles et/ou personnelles et se fédère autour d'un but partagé. Le groupe formel diffère également de la communauté selon les critères suivants : composition des membres, durée, taille et domaines d'expertises qui sont généralement préétablis. Contrairement à la communauté et au groupe formel, le groupe d'amis n'a pas de finalités précises et de règles explicites de participation.

La notion de « communauté » est multiple selon les disciplines et par conséquent ne fait pas l'objet d'un consensus scientifique autour de sa définition. En théorie de l'apprentissage social, la notion de « communauté de pratique » initié par (Wenger E., 2005, réédité en 2011), désigne « un groupe de personnes ayant en commun un domaine d'expertise ou une pratique professionnelle qui se rencontre pour échanger, partager et apprendre les uns des autres, face-à-face ou virtuellement ». Dans le cadre plus spécifique de l'apprentissage humain, (Riel, M., 2004) propose une définition de la « communauté d'apprentissage » qui peut être physique ou en ligne

comme étant « un groupe d'individus qui œuvrent ensemble dans un temps déterminé pour mener à bien une tâche, pour comprendre un nouveau phénomène ou pour réaliser une tâche de manière collaborative ». Les communautés d'apprentissage se sont particulièrement développées avec les plateformes d'apprentissage en ligne pour soutenir la collaboration entre les étudiants au sein des universités (Laferrière T., 2013). Avec l'avènement d'internet, les communautés d'apprentissage se sont décentralisées favorisant l'apparition de communautés non-formelles composés de personnes en dehors du cadre scolaire (professionnel, adulte...). Malgré les multiples évolutions technologiques et sociétales, (Cristol D., 2017) soutient que les communautés d'apprentissage restent « des groupes d'apprenants qui partagent formellement ou informellement en présence ou à distance l'intention d'apprendre ensemble ».

Dans le domaine des sciences de l'éducation, le groupe de formation, appelé également « groupe d'apprentissage », désigne un objet de recherche privilégié où s'articule l'apprentissage individuel et collectif. Historiquement, (Lewin K., 1959) a initié des expérimentations pour analyser les comportements individuels et collectifs au sein de petits groupes de formation selon des approches de recherche-action (ou dynamique de groupe). Par la suite, de nombreuses définitions du groupe de formation ont été proposées dans un cadre scolaire où « le groupe d'apprenants aussi appelé groupe secondaire est l'ensemble des apprenants qui sont plongés durant une même période dans une même situation d'apprentissage » selon (Vayer P., 1987).

D'une façon plus marginale, la notion du groupe a été également étudiée dans le contexte spécifique de l'apprentissage sportif. Selon la définition proposée par (Rey J.P, 2000), « un groupe n'est pas un simple rassemblement d'individus. C'est un espace social dans lequel les membres interagissent, communiquent, forment des sous-groupes en fonction des affects qu'ils ressentent les uns vis-à-vis des autres ». Ainsi, le rôle des émotions et des affects entre les membres du groupe devient des facteurs influençant la dynamique de groupe comme par exemple lors du choix du partenaire (Duprez, J. M., 2007). Cet angle de recherche est davantage mis en évidence dans un contexte d'éducation physique et sportif, à la différence des travaux de recherche dans le domaine des sciences de l'éducation qui se focalise principalement sur les processus cognitifs lié à l'apprentissage en groupe.

Dans un contexte de formation continue, le groupe de formation pour adulte est un objet variable selon la situation de formation (virtuel ou présentiel, durée courte ou longue, contenu générique ou sur-mesure...). Selon (Knowles, M. 1990), l'apprentissage en groupe est adapté pour un adulte car « il lui procure un sentiment d'approbation et de sécurité », ainsi qu'un « (...) sentiment d'appartenance lui apporte des satisfactions (...) ». Selon (Baudoin J.M, 2001), le groupe est posé comme le « lieu » véritable de la formation des adultes et considéré comme le levier principal permettant le changement individuel et collectif. Plus récemment, le groupe de formation devient « un support d'apprentissage dans le sens où les interactions sociales (confrontation des idées, questions posées au formateur...) ainsi que les tâches communes réalisées en groupe ou en sous-groupe, sont des activités qui permettent de mieux apprendre » selon (Solar C., 2011). Néanmoins, l'engagement de l'apprenant au sein du groupe reste secondaire : selon (Martin J.P, 1996), « le travail de groupe n'a pas pour but premier la réalisation d'une tâche mais le progrès individuel de chacun des membres ». Ainsi, le groupe apparaît comme un soutien à l'apprentissage individuel où la recherche du progrès individuel semble prévaloir sur la recherche de la performance du groupe.

3.1.2 Groupe professionnel

Initié par (Van Maanen J., 1984), la notion de « communauté de métier » se définit comme « un groupe de personnes qui se considère engagé dans un même travail, (...) qui partagent avec les uns et les autres un ensemble de valeurs et de normes au-delà des questions liées au travail (...) ». Plus récemment, (Gadéa, C., 2009) définit le groupe professionnel comme étant « un ensemble de travailleurs exerçant une activité ayant le même nom, et par conséquent

doté d'une visibilité sociale, bénéficiant d'une identification et d'une reconnaissance, occupant une place différenciée dans la division sociale du travail, et caractérisés par une légitimité symbolique ». Ainsi, le groupe professionnel est une unité élémentaire d'une communauté dont la profession (ou métier) devient le point commun entre les membres du groupe. Dans le prolongement de l'étude des représentations sociales (Durkheim E., 1898) (Moscovici S., 1961, 2003), le groupe professionnel est catégorisé selon une représentation professionnelle. Cette notion proposée par (Bataille M., 2000) et repris par (Piasser A., 2014) qui précise que « les représentations professionnelles (...) sont élaborées dans l'action et l'interaction professionnelle qui les contextualisent, par des acteurs dont les dynamiques identitaires, correspondent à des collectifs du champ professionnel ».

En interne d'une organisation, les dynamiques identitaires s'opèrent différemment au sein d'un groupe. Le groupe social (ou équipe de travail) a été principalement théorisé en psychologie du travail. Parmi les multiples définitions existantes, (Kozlowski S., 2008) souligne « qu'un groupe d'individus est qualifié d'équipe ou de groupe de travail si les individus se reconnaissent et se définissent en tant que groupe ou équipe remplissant des tâches dans des buts communs en interaction et interdépendance dans un contexte organisationnel plus large ». Par la suite, il élargit cette notion de groupe (ou équipe de travail) dans un contexte de formation. Selon lui, le groupe de formation « ne se limite pas à une simple agrégation de personnes véhiculant des connaissances mais reposant sur un principe de partage et de complémentarité de compétences entre les membres du groupe qui dispose d'un langage commun lié à une appartenance commune à une organisation (Kozlowski, S. 2013). De même, dans le cadre d'une formation, (Goguelin P., 1994) propose une distinction des groupes de travail selon une catégorisation socio-professionnelle du participant selon le type d'activité (production, administratif...) ou selon un niveau hiérarchique (cadre, exécutant...). Néanmoins, ces groupes (ou équipes de travail) restent formalisés dans un cadre bien défini lié à une appartenance commune à une organisation relevant d'une structure hiérarchique stable et identifiée.

3.2 Acteurs du groupe de formation pour adulte

3.2.1 Apprenant-adulte

En France, le domaine de recherche nommé « Education pour adulte » est rattaché aux sciences de l'éducation. Dédié exclusivement aux adultes en situation de formation, ce domaine a évolué en lien avec le développement de la formation continue. Théorisée sous le concept d'« andragogie » par Malcolm S. Knowles (1913-1997), l'auteur identifie des caractéristiques et spécificités cognitives propres aux apprenants-adultes :

- l'adulte a besoin de comprendre l'utilité des enseignements qui ont un impact immédiat dans sa vie personnelle et/ou professionnelle ;
- la maturité inhérente à la vie d'adulte l'amène à s'auto-diriger ;
- l'adulte a conscience d'être responsable de ses propres décisions et ses choix de vie ;
- les adultes sont davantage intéressés par le fait d'apprendre des choses ;
- la prise en compte de son expérience et ses erreurs constituent la base des activités d'apprentissage ;
- l'aptitude et la volonté d'apprendre sont renforcées si l'apprentissage est orienté vers des besoins de développement et des rôles sociaux.

Plus récemment, (Mucchielli R., 1991, réédité en 2016) complète cette liste en soulignant que l'adulte en formation bénéficie d'une expérience de vie professionnelle qu'il convient de prendre en considération en déclarant « les hommes et les femmes entrés dans la vie professionnelle, assumant des rôles sociaux actifs et des responsabilités familiales, ayant déjà une expérience directe avec l'existence ». Ainsi, la formation pour adulte n'est pas abordée

essentiellement sous un angle personnel mais particulièrement influencée par le contexte professionnel.

Au niveau sociétal, la formation en entreprise s'est particulièrement accélérée liée aux politiques publiques tant au niveau national qu'europpéen axées sur la FLTV⁴¹. Selon (Marchand, L. 1998), « les identités sociales, professionnelles et personnelles ne sont plus acquises une fois pour toutes ; elles ont besoin d'être parachevées, actualisées et reproduites dans une perspective d'apprentissage à vie ». Par la suite, plusieurs définitions se sont succédé dans la littérature scientifique : (Gohier C., 2000) la définit comme une « composante de l'identité globale de la personne et qui se développerait sur la base de l'identité personnelle par l'inscription de la personne dans des formes de vie sociale et plus particulièrement sur sa situation, sa perception et ses relations professionnelles ». Cette définition rejoint celle proposée par (Gentili, F. 2005) qui décrit l'identité professionnelle avant tout comme étant une « identité sociale ancrée dans une profession et constitue donc une socialisation secondaire ». D'autres définitions mettent également en évidence la notion de la profession (ou du métier) dont (Dubar, C. 2000) qui caractérise l'identité professionnelle selon une identité catégorielle (ou identité de métier) qui renvoie à « un collectif préexistant soit le métier ou la profession, vis-à-vis duquel l'individu développe un sentiment d'appartenance communautaire et une volonté de progresser ». Ainsi, l'identité professionnelle devient une notion centrale et en constante évolution pour l'adulte en situation de formation. Elle lui confère un statut particulier au sein du groupe de formation par son affiliation institutionnelle caractérisée par sa profession (ou métier) qui conditionne les relations et la socialisation avec les membres du groupe.

3.2.2 Formateur d'adulte

Au sein du groupe de formation pour adulte, le formateur anime et supervise le groupe de formation. (Champy, P., 2011) propose la définition suivante « le formateur d'adulte est celui qui assure surtout des prestations pédagogiques directes auprès des publics, ce qui exclut de la définition les activités de responsable de formation, de conseiller, de consultant, d'audit, d'ingénierie... ». (Brundseaux, M., 2009) différencie le rôle du formateur d'adulte avec celui de l'enseignant en lui attribuant un rôle supplémentaire. Outre son rôle de transmetteur de savoirs, l'auteur assigne aux formateurs d'adultes des « tâches de coordinateurs et de facilitateurs des processus d'apprentissage ». Il indique également que les formateurs sont également confrontés aux composantes d'un environnement social et culturel complexe. Les interactions avec les nombreux acteurs de la formation et les contraintes inhérentes aux cadres légaux de référence nécessitent la mobilisation de savoir-être qui transcendent l'expertise disciplinaire. (Betton, E. 2013) souligne la diversité des profils des formateurs qui peuvent être des formateurs permanents ou occasionnels, débutants ou expérimentés dans un domaine spécialisé. Ainsi, la supervision du groupe de formation est abordée d'une manière plus pragmatique et moins sous l'angle pédagogique à la différence des enseignants en formation initiale. Au niveau pédagogique, (Jobert, G., 2013) établit également une distinction entre l'enseignant « qui diffuse des savoirs décontextualisés » et le formateur pour adulte qui privilégie davantage « l'action et les interactions ».

3.2.3 Organisation/Institution

La formation continue constitue un levier stratégique pour les organisations. La formation du personnel peut être gérée en interne d'une organisation ou peut faire l'objet d'une externalisation. (Petit, A. 1993) indique que « la formation stratégique des ressources humaines est un ensemble planifié d'activités d'apprentissage variées, dont le but explicite est de développer

⁴¹ Formation Tout au Long de la Vie

les compétences propres des personnes et à l'atteinte de buts économiques et sociaux ». Selon la réglementation juridique française, l'entreprise reste le principal financeur de la formation. Par conséquent, le recours à la formation par l'entreprise est étroitement lié à sa politique interne (mode de gouvernance, objectifs managériaux, capacité de financement...). Selon (Guillemot, D., 2017), les objectifs assignés à la formation continue en entreprise peuvent être multiples : accroître l'efficacité, favoriser l'innovation, répondre aux obligations légales, accompagner le changement, motiver et fidéliser ses salariés, faciliter les mobilités, ou encore améliorer l'image de l'entreprise. Selon l'enquête menée par (Dubois J.M., 2016), la taille de l'entreprise influence particulièrement l'accès à la formation. Les entreprises de grande taille⁴² (plus de 500 salariés) ont davantage recours à la formation pour accompagner des changements structurels (organisationnels, technologiques...) tandis que les objectifs de formation pour les PME visent davantage à répondre à des exigences réglementaires.

3.3 Dimension sociale (ou relationnelle)

La dimension relationnelle se caractérise selon la nature, l'intensité, et le degré de symétrie ou d'asymétrie des relations entre les membres du groupe et le formateur.

3.3.1 Relation professionnelle

Initiée en sociologie, la notion du lien social désigne « l'ensemble des relations qui unissent des individus faisant partie d'un même groupe social et/ou qui établissent des règles sociales entre individus ou groupes sociaux différents ». Dans le domaine des sciences de l'éducation, la notion du lien social a été principalement théorisée par (Postic M., 2015) à travers la « relation pédagogique », appelée également « relation éducative ». Elle est définie comme « une relation sociale déterminée par des représentations préétablies ayant une finalité particulière » et s'inscrit essentiellement dans un contexte de formation initiale (scolaire ou universitaire).

Dans un contexte de formation pour adulte, la relation « maître-élève » est peu comparable avec la relation de l'adulte avec le formateur. Selon (Mucchielli R., 1998 réédité en 2016), le rôle du formateur n'est pas de transmettre un savoir académique mais induit une relation fonctionnelle avec l'apprenant-adulte pour faciliter son apprentissage. De même, les relations entre les membres du groupe induisent ainsi des mécanismes sociaux différents. Comme indiqué par (Mucchielli R., 1998 réédité en 2016), l'identité professionnelle de l'adulte en formation influence et conditionne les relations sociales entre les membres du groupe. Ainsi, la relation professionnelle se superpose avec la relation d'apprentissage au sein d'un groupe de formation pour adulte. D'une part, (Matthews T., 2011) définit la « relation professionnelle » comme étant « (...) la communication entre deux individus, généralement avec un minimum de formalité et de structure. Les objectifs communs pour les relations professionnelles comprennent le mentorat, la recherche de collaborateurs, la construction de réputation, et/ou la recherche de réponses ou de commentaires à propos d'une question donnée ». L'apprenant adulte étant affiliée à une organisation (ou entreprise), deux types de relations professionnelles sont à distinguer entre les membres d'un groupe de formation : les relations intra-organisationnelles où les individus travaillent au sein d'une même organisation (ou entreprise) et les relations inter-organisationnelles où les individus travaillent au sein d'organisations distinctes. Au niveau inter-organisationnel, la formation se structure avec des personnes issues d'organisations différentes. Ainsi, la mise en relation de personnes extérieures à l'entreprise vise à l'échange de pratiques professionnelles autour d'un groupe de formation. Au niveau intra-organisationnel, la formation est exclusivement centrée aux personnes d'une même organisation. Théorisée sous la notion « d'apprentissage organisationnelle », la formation s'inscrit dans une démarche de perfectionnement et transfert de

⁴² >500 salariés

compétence en situation de travail pour le développement de l'employabilité au sein d'une organisation.

3.3.2 Intensité des relations sociales

Initialement théorisée par (Granovetter, M. 1977), l'intensité du lien social distingue les liens « faibles » et les liens « forts ». Le lien « fort » correspond à une fréquence d'interactions tandis que le lien « faible » correspond à une fréquence d'interactions considérée comme minimale (ou considérée au-dessous de la moyenne estimée). Néanmoins, la force du lien « faible » permet de relier des liens « indirects » et servir de pont entre deux groupes. Par la suite, ce principe a été théorisé par (Burt R., 2001 réédité en 2017), sous le concept du « trou structural » désignant une absence de relations entre deux acteurs au sein d'une triade⁴³.

Dans un contexte pédagogique, la nature des relations sociales physiques s'est particulièrement transposée vers les réseaux sociaux numériques. Le lien technologique médiatisé par les outils de communications en ligne est considéré comme porteur d'un lien social. Ainsi, (Haythornthwaite, C. 2002) complète cette typologie fort/faible avec un lien dit « latent ». Il caractérise ce lien comme étant disponible techniquement mais qui n'a pas encore été activé par une interaction sociale. Cette typologie binaire du lien « fort » *vs* lien « faible » est reprise par (Bos-Ciussi, M. 2007) qui propose une approche selon quatre environnements d'apprentissage (distance *vs* local et lien fort *vs* faible) synthétisée dans la figure 4. D'une part, cette typologie vise à prendre en compte le degré de virtualité des lieux et d'autre part, l'intensité des liens sociaux schématisé de la manière suivante :

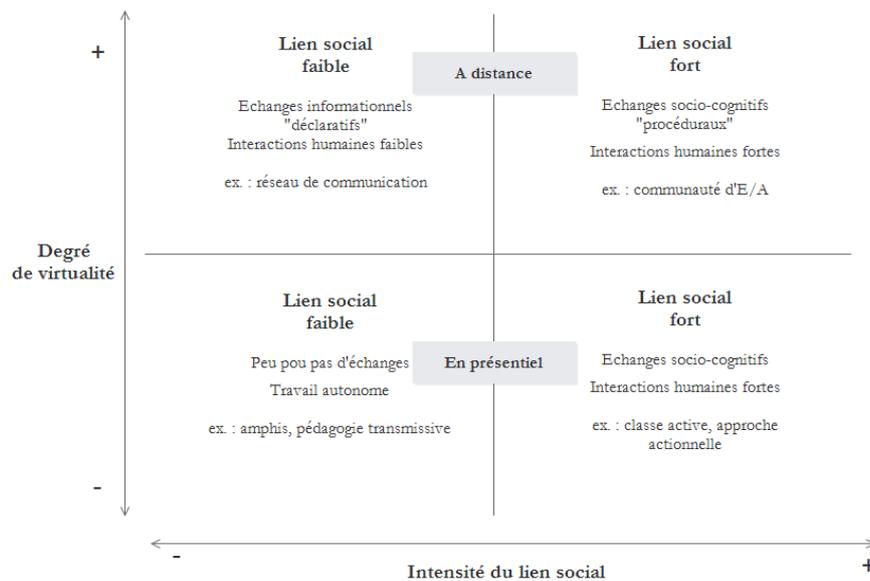


Figure 3.2 - Quadrant virtualité/intensité des liens (Bos-Ciussi, M. 2007)

Les relations affectives au sein du groupe déterminent également l'intensité des relations sociales. Bénéficiant d'une large littérature dans le domaine de la psychologie sociale, cette notion est synthétisée par (Gérard B., 2013) de la manière suivante : « les affinités reposent sur un attachement réciproque, personnalisé et sélectionné entre deux individus », la relation se fonde sur une reconnaissance des qualités personnelles des membres du groupe. Dans le domaine des sciences de l'éducation, l'étude de l'influence des relations socio-affectives est tardive du fait de

⁴³ Lien fort entre un individu A et un individu B et lorsque A et un autre individu C sont eux aussi liés par une forte relation, alors B et C sont au moins unis par un lien faible.

l'absence de modèle théorique et conceptuelle et que les méthodes de recherche traditionnelles (questionnaire, entretien qualitatif...) présentent des limites pour identifier les affinités entre les personnes selon (Dussarps C., 2014).

3.3.3 Symétrie/Asymétrie des relations sociales

Issu du domaine de la communication interpersonnelle, la notion de « relation symétrique » a été théorisée par Gregory Bateson⁴⁴ en 1936. Il définit la relation « symétrique » comme étant fondamentalement une relation d'égalité. À l'inverse, la relation complémentaire se définit par une position haute (ou position supérieur) où l'individu dirige et détient la responsabilité de l'interaction et une position basse (ou position inférieure) où l'individu s'ajuste à l'autre et répond à son initiative (Picard, D., 2013). Théorisé en psychologie sociale sous l'angle du conflit socio-cognitif, de nombreuses études issues de ce courant de recherche ont démontré l'influence de la représentation sociale sur la dynamique interactionnelle. En effet, les interactions sociales sont plus efficaces pour l'apprentissage lorsqu'elles s'inscrivent dans une relation symétrique entre les partenaires (De Paolis P., 1991), (Carugati F., 1980) et (Doise W., 1975, réédité en 2003).

Dans un contexte de formation scolaire, la relation pédagogique est abordée sous l'angle de l'autorité de l'enseignant. Selon (Postic M., 1982 réédité en 2015), la relation pédagogique reste nécessairement asymétrique entre l'enseignant et l'élève : elle reste une « intervention » d'une personne sur une autre et empêche un véritable dialogue, la finalité de l'échange étant définie par un des partenaires. Dans le cadre du travail en groupe, l'asymétrie de la relation sociale est liée aux caractéristiques des membres du groupe (âge, degré d'expertise en fonction des références socio-culturelles, niveau scolaire, statut professionnel...) qui peuvent varier d'un individu à l'autre. Selon (Huguet P., 1999), l'inégalité des statuts sociaux des membres au sein du groupe de formation induit une comparaison sociale. Initié par (Festinger L., 1954), la théorie de la comparaison sociale induit que les individus ont une tendance innée à évaluer ses opinions et ses aptitudes avec autrui. Il distingue trois types de comparaisons :

- « Comparaison latérale » correspond à se comparer avec des personnes qui sont semblables à soi ;
- « Comparaison sociale ascendante » correspond au fait de se comparer à meilleur que soi ;
- « Comparaison sociale descendante » correspond au fait de se comparer à un individu perçu comme inférieur.

Dans un cadre scolaire, (Butera, F., 2006) démontre que la comparaison sociale implique une focalisation qui entrave le niveau de l'attention, du raisonnement et de la motivation lors d'un apprentissage en groupe. Ce mode de comparaison basé sur une interdépendance négative implique une menace sur l'acquisition des compétences. Plus récemment, (Micari, M., 2011, 2014) confirme ce phénomène de comparaison sociale au sein des groupes d'élèves restreints en démontrant l'influence de l'intimidation et l'anxiété dans les interactions entre les élèves.

Dans un contexte professionnel, cette comparaison sociale est particulièrement amplifiée liée à la hiérarchisation des relations socio-professionnelles au sein des organisations. En science de gestion, (Fiol, M. 2006) souligne que dans toute équipe de management (regroupant un supérieur hiérarchique et ses collaborateurs directs), deux types de relations sont à distinguer : la relation hiérarchique, quand elle concerne les relations supérieurs/subordonnés et la relation latérale, lorsqu'elle caractérise les interrelations entre pairs au sein de l'équipe. Plus récemment, l'auteur identifie trois dimensions principales qui caractérisent les relations sociales au sein de l'organisation : la cohésion, la cohérence hiérarchique (ou verticale) et la cohérence latérale.

⁴⁴ Fondateurs de l'École de Palo Alto (Courant de pensée et de recherche ayant pris le nom de la ville de Palo Alto en Californie qui a initié les bases d'une « approche systémique et interactionniste des phénomènes humains ». (Picard, D. 2013)

Dans un contexte de formation continue, les relations socio-professionnelles entre les membres du groupe sont principalement caractérisées par l'identité et plus précisément le statut professionnel de l'apprenant-adulte. Selon la définition de (Raynal, F., 2018) « le statut se caractérise par l'influence ou la considération que les autres membres reconnaissent ou accordent à l'individu (...) basé sur des évaluations mutuelles que les personnes font de l'importance sociale des autres à l'intérieur d'un groupe ». Par nature, une asymétrie liée à la hiérarchisation des statuts et des relations professionnelles peut être inhérente selon la composition des groupes de formation. Selon (Bourgeois E., 2005), cette relation asymétrique d'ordre hiérarchique aura tendance à créer une résolution seulement sociale de l'interaction (et non cognitive), soit par complaisance (un point de vue est retenu au profit d'un autre), soit par juxtaposition (chacun conserve son point de vue, sans confrontation). Ainsi, le degré de l'asymétrie de la relation sociale et notamment hiérarchique est peu favorable à la collaboration et à l'interaction entre pairs au sein du groupe de formation pour adulte. La résolution du conflit socio-cognitif peut être régulée sur le plan relationnel en adoptant une attitude de complaisance (ou de « *statu-quo* ») ou par l'intervention d'un tiers qui dans un contexte de formation peut être représentée par le formateur.

Les relations socio-professionnelles entre les apprenants-adultes du groupe et le formateur est par nature asymétrique. Selon (Bourgeois, E., 1999), « le formateur a un statut supérieur, du fait de son expertise et sa position institutionnelle qui lui réserve un certain contrôle sur le processus de formation ». Malgré cette asymétrie relationnelle, l'auteur souligne « qu'une résolution socio-cognitive peut être obtenue à condition que le sujet en position de dominant ne propose pas immédiatement son point de vue ».

3.4 Dimension cognitive

La dimension cognitive se réfère aux méthodes favorisant les apprentissages humains au sein du groupe de formation.

3.4.1 Interactions par les pairs

Selon les théories socio-constructivisme de Lev Vygotsky (1896-1934)⁴⁵, l'apprentissage en groupe s'opère principalement à travers les interactions sociales entre membres du groupe. Issue d'une tradition éducative, la collaboration entre élèves a bénéficié d'une littérature significative et a été théorisée sous diverses terminologies : apprentissage social, apprentissage collectif, apprentissage coopératif... Parmi les définitions existantes, nous retenons la définition initiale proposée par (Smith BL, 1992) comme étant relative à « un groupe d'apprenants, impliquant un effort intellectuel conjoint, impliquant l'enseignant et les apprenants dans un processus collectif ». Une notion conjointe soit « l'apprentissage par les pairs » (ou *Peer Instruction*) est développée par (Mazur E., 1997) repris par (Crouch CH, 2001). Cette stratégie d'enseignement vise à encourager la discussion et l'entraide entre les élèves. Selon (Bouffard, G. 2015), la compréhension de notions difficiles (en mathématiques ou physique par exemple) est facilitée par l'argumentation et la force de persuasion des étudiants. Elle favorise également la confiance des étudiants dans leur sens critique. Néanmoins, cette méthode reste peu pratiquée en formation initiale et/ou universitaire, contrairement à la formation pour adulte.

Dans un contexte professionnel, la notion de « collaboration » diffère en s'inscrivant dans un cadre plus large que celle de l'apprentissage. Elle présente des variances terminologiques comme « collaboration », « partenariat », « coordination », « coopération », « concertation » « gestion horizontale » récapitulés par (Bentayeb, N. 2010).

⁴⁵ Lev Seminovich Vygotsky (1839-1934) : Penseur et l'un des pionniers du socio-constructivisme en défendant que l'être humain se caractérise par une sociabilité primaire. Par conséquent, le développement intellectuel de l'enfant est indissociable des interactions sociales (Pensée et langage, publiée en 1934 à Moscou réédité en français en 1997)

Dans le cadre de la formation pour adulte, les échanges et les interactions entre pairs sont centrales et reposent sur des mécanismes différents. La démarche de l'apprentissage en groupe s'avère particulièrement pertinente pour l'apprenant-adulte selon (Knowles, 1990) qui affirme que « la meilleure méthode d'enseignement pour adulte reste sans aucun doute aujourd'hui la discussion de groupe ». (Bourgeois, E. 2001) souligne également que la notion d'argumentation joue un rôle déterminant sur l'apprentissage individuel, indépendamment du fait que l'individu a plus ou moins pris une part active à l'argumentation. Ainsi, l'interaction sociale entre pairs est un élément fondamental pour l'apprentissage sans pour autant négliger le rôle du formateur. Plus récemment, (Le Boucher, C. 2015) aborde une problématique liée aux rythmes et temporalités multiples des acteurs de la formation (apprenant et formateur) pour prolonger les interactions entre pairs et l'appropriation des apprentissages. Ces nouveaux espaces-temps impliquent des modalités d'apprentissages adaptées pour favoriser la pérennisation des interactions entre pairs dans un contexte d'apprentissage formel et informel⁴⁶ (appelé également « co-formation »). Dans ce sens, (Rheingolt H., 2015) initie le concept de « pairagogie » (ou l'apprentissage par les pairs) qui souligne la nécessité de cette forme d'apprentissage en mettant en évidence les limites d'un seul formateur à transmettre des enseignements de plus en plus denses et complexes. Ainsi, ces limites peuvent être atténuées par les membres du groupe à travers les interactions sociales.

3.4.2 Expériences professionnelles

Initié par (Dewey J., 1916 réédité en 1973) en lien avec le mouvement de réforme sociale de l'éducation progressiste aux États-Unis, la notion d'expérience professionnelle fait l'objet de multiples définitions complexes à synthétiser. Dans le cadre de la formation des adultes, la prise en compte de l'expérience est un fondement théorique solide et constitue la base des activités d'apprentissage selon (McCarthy B. 1985) (Knowles M., 1990), (Mucchielli R., 1991). (Mayen, P. 2003) souligne également que l'expérience des adultes en formation est une composante dont il est difficile de faire abstraction et qui peut constituer un facteur de différenciation entre les membres du groupe.

Par la suite, la notion de « l'apprentissage expérientiel » (ou *Experiential Learning*), est définie comme une démarche de construction de savoirs dans une quête de sens. Cette notion a évolué selon les courants de recherche synthétisés par (Balleux A., 2000) où les multiples définitions existantes visent à différencier le savoir théorique du savoir pratique. Citons, par exemple, la définition proposée par (Coleman, 1976) pour qui « l'apprenant, au lieu de chercher à comprendre et à assimiler une information verbale ou écrite, doit pouvoir donner un sens à ce qu'il a vécu et construire des connaissances qui lui sont utiles ». Ce type d'apprentissage est particulièrement renforcé dans un contexte de formation pour adulte où la connaissance théorique est moins primordiale et doit être complétée par la pratique professionnelle. Ainsi, la situation de formation est davantage centrée sur la prise en compte de l'expérience et des acquis antérieurs des membres du groupe qui deviennent les objets des interactions sociales.

En parallèle, l'analyse des pratiques professionnelles est définie par (Fablet D., 2004) comme « des activités, (...) qui sont organisées [en France] dans un cadre institué de formation professionnelle initiale ou continue. Les praticiens se trouvent ainsi conviés à travailler, le plus souvent en groupe, et selon une certaine durée, à l'élucidation du sens de leurs pratiques et/ou à l'amélioration des leurs gestes professionnels (...) ». En pratique, cette méthode a été appliquée au sein de groupes composés de membres exerçant la même profession (médecins, enseignants ou travailleurs sociaux). Tout comme le principe de l'apprentissage expérientiel, cette notion postule que l'expérience est source de construction de savoirs contrairement aux savoirs académiques moins opérants pour résoudre des problèmes rencontrés dans le cadre d'une pratique professionnelle.

⁴⁶ Formel : activité d'apprentissage qui a lieu dans un contexte organisé, structuré et graduel
Informel : activité d'apprentissage peu structurée et autodirigé (Sauvé, L. 2014)

3.5 Dimension organisationnelle et environnementale

La dimension organisationnelle et environnementale est relative aux caractéristiques fonctionnelles du groupe de formation.

3.5.1 Type de regroupement

Le type de regroupement consiste à assigner des individus selon des caractéristiques préétablis soit d'une manière homogène ou hétérogène. Cette pratique organisationnelle est étroitement liée à la méthode de composition de groupe de formation dont (Depover, C., 2003) énumère plusieurs méthodes :

- Aléatoire qui s'opère sur la base de regroupements liés au hasard ;
- Autonome qui laisse le choix aux apprenants sur la base des préférences de chacun ;
- Opportuniste qui permet de prendre en considération certains critères (proximité géographique, disponibilité...);
- Raisonnée qui s'opère selon des critères préalablement définis.

Dans un cadre scolaire, la pratique de regroupement de type « homogène » consiste à assigner les élèves à des groupes sur la base de leurs habiletés cognitives ou de leurs résultats scolaires afin de faciliter l'enseignement et de favoriser l'apprentissage des élèves. A l'inverse, le type de regroupement « hétérogène » consiste à regrouper les élèves dans des groupes de manière à refléter toute l'étendue des habiletés d'une même cohorte. Ce type de regroupement consiste à pallier des inégalités sociales et à favoriser l'équité et la stimulation (Fortin, L., 2011).

Largement opposé dans un contexte de formation initiale (Dupriez, V., 2004), le type de regroupement hétérogène *versus* homogène est abordé différemment en formation pour adulte. En dehors du cadre scolaire, l'influence de la composition d'un groupe de formation en lien avec ses performances ont été initiées par (Levine J.M, 1998). Selon les auteurs, la composition d'un « groupe idéal » peut se baser selon les caractéristiques de ses membres mais également selon les interactions sociales. De même, selon (Decamps S., 2014), la collaboration en situation de formation peut être anticipée selon un mode de composition du groupe dit « raisonné » en se basant sur la similarité des caractéristiques des apprenants (âge, genre, niveau d'étude, ancienneté, profession ...). Historiquement, ce phénomène sociologique théorisé sous le concept d'« homophilie » par (Lazarfeld P., 1954) désigne l'effet d'attraction correspondant à la similitude des statuts sociaux. Ainsi, l'apprenant-adulte va avoir une tendance naturelle à interagir avec des pairs ayant des caractéristiques proches des siennes. Les caractéristiques sont multiples : similarité d'éducation (niveaux d'études), statut social (catégories professionnelles), âge (McPherson, L., 1987, 2001), style d'apprentissage (Alfonseca, E., 2006) et plus récemment le genre selon (Cen L., 2014). Cependant, bien que la similitude engendre une connexion de départ entre les individus, les liens sociaux évoluent selon la nature des relations et les attributs des individus (Lemieux V., 2004). En effet, les apprenants ajustent et régulent les interactions avec leurs pairs selon des critères subjectifs (rencontres antérieures, affinités sociales par exemple) qui peuvent influencer les logiques sociales à base de choix.

3.5.2 Mode environnemental (Présentiel *vs* Distanciel)

Le groupe de formation peut se pratiquer selon un mode en présentiel ou distanciel. Le groupe de formation en présentiel évolue dans un contexte spatio-temporel préalablement défini. Par conséquent, le cycle de vie du groupe de formation est dépendant de contraintes géographiques et temporelles.

Le groupes de formation à distance, (appelé également « groupe virtuel »), sont principalement des groupes constitués exclusivement *via* un système informatique d'apprentissage humain à distance. Ils sont principalement médiatisés à travers des outils de communication

asynchrones (outils de discussion *via* un forum par exemple). Dans un contexte professionnel, les contraintes organisationnelles et géographiques ont contribué à l'émergence des formations selon un mode d'organisation hybride (appelé également « *Blended-Learning* »). Une formation hybride se caractérise par des situations de formation dites « mixtes » (présentiel et distanciel) où des activités d'apprentissage s'effectuent en dehors de la présence physique du formateur et selon différents niveaux : formel et non-formel (ou informel) selon (Sauvé, L. 2014)⁴⁷. Néanmoins, malgré la multiplicité des technologies d'apprentissage à distance, la situation de formation en groupe en présentiel est indispensable pour réguler l'apprenant dans son parcours de formation (Ginns, P., 2007), (Kay, R., 2016).

3.5.3 Taille du groupe

De nombreux travaux scientifiques se sont intéressés à l'influence de la taille du groupe sur la collaboration entre les apprenants. Dans le domaine de la psychologie sociale, (Anzieu, D., 1971 réédité en 2013) propose un schéma de classification de la taille des groupes distingués selon les modalités suivantes :

- Restreint (3 à 5 individus) ;
- Moyen (moins de 6 à 13 individus) ;
- Etendu (14 à 24 individus) ;
- Large (25 à 50 individus) ;
- Assemblée (plus de 50 individus).

Particulièrement étudié par (Anzieu, D., 1971), le groupe dit « restreint » ne se caractérise pas par une simple agrégation de personnes mais par la nécessité d'une poursuite d'un objectif commun entre les membres du groupe. Selon (Mucchielli, R., 1992, réédité en 2019), le nombre d'individus, influence considérablement le fonctionnement et la cohésion du groupe en face-à-face. Il définit la taille optimale d'un groupe composé de 3 à 12 personnes avec un équilibre de 5 à 6 personnes pour faciliter les interactions entre les membres. De même, (Gallupe R., 1992) démontre que les interactions sociales diminuent quand le groupe devient large, et ceux-ci deviennent alors moins productifs.

Dans un contexte d'apprentissage humain, la taille du groupe de formation pour adulte reste peu étudiée et par conséquent une taille optimale du groupe reste indéterminée (Maskit, D. 1986). Dans le domaine de la formation initiale, selon une méta-analyse proposé par (Odo, C., 2019), l'ensemble des travaux ne mentionnent pas une taille de groupe dite « idéale » à fixer d'une manière définitive lors de la création de groupe. Ils s'accordent à justifier que les étudiants préfèrent collaborer au sein de petits groupes et deviennent moins productifs à la mesure que la taille du groupe augmente (Babar, M., 2007). De même, des travaux démontrent qu'un groupe restreint impacte favorablement l'apprentissage individuel selon une évaluation de satisfaction des étudiants (Kooloos, J. 2011) (Chou, P. 2018).

⁴⁷ Formel : activité d'apprentissage qui a lieu dans un contexte organisé, structuré et graduel
Informel : activité d'apprentissage peu structurée et autodirigé (Sauvé, L. 2014)

Chapitre 4. Algorithme d'aide à la composition du groupe de formation

Ce chapitre présente les techniques informatiques d'aide à la composition des groupes de formation les plus couramment utilisées à savoir les méthodes méta-heuristiques/heuristique et les méthodes de classification.

4.1 Introduction

4.1.1 Méthodologie

Comme indiqué lors du chapitre précédent, l'apprentissage collaboratif du groupe peut être influencé par les membres qui le compose. Ainsi, plusieurs stratégies de composition des groupes d'apprentissage peuvent être appliquées pour optimiser la collaboration entre les membres. Dans un cadre scolaire, l'assignation des élèves peuvent se faire manuellement par l'enseignant d'une manière aléatoire ou intentionnelle ou librement choisie par les élèves. Dans un cadre universitaire, la composition des groupes d'une manière libre et intentionnelle se renforce particulièrement où les étudiants ont tendance à créer des groupes de travail principalement selon les affinités personnelles. Bien que facile à mettre en œuvre, selon (Mesquita, C., 2018) ces méthodes de composition peuvent résulter un déséquilibre au niveau des compétences et un manque de diversité sociale au sein du groupe d'apprentissage. A partir de 2005, la composition des groupes soutenus par des outils informatiques pour pallier aux limites des assignations manuelles suscite un intérêt à travers divers travaux de recherche.

Au niveau de système informatique, (Putro, B. 2020) propose une méthodologie à partir d'une méta-analyse pour développer un agent intelligent (ou *framework*) selon les étapes suivantes :

(1) Comprendre les problèmes et les objectifs du développement de groupe

Cette première étape vise à définir le problème lié aux membres du groupe et de fixer des objectifs pour le développement d'un logiciel d'agent intelligent.

(2) Déterminer les critères (ou composants) pour les agents intelligents

Cette étape vise à sélectionner des critères intelligents (attributs des membres ou les attributs de groupe) et de choisir les techniques algorithmiques.

(3) Développement d'un logiciel d'agent intelligent

Cette étape consiste à modéliser les étapes de traitement d'un logiciel.

(4) Implémentation et test

Le résultat de cette étape est la migration des données, la mise en œuvre de l'algorithme et du test du logiciel.

(5) Évaluation des agents intelligents (ou du *framework*)

L'évaluation est le résultat de l'analyse des données de mise en œuvre et de test à la 4ème étape.

Au niveau des techniques algorithmiques, plusieurs méthodes informatiques visent à déterminer une stratégie de sélection selon des critères prédéfinis correspondant à la deuxième étape de la méthodologie proposée par (Putro, B. 2020). (Maqtary, N., 2019) propose une taxinomie qui synthétise l'ensemble des travaux publiés de 2005 à 2015 :

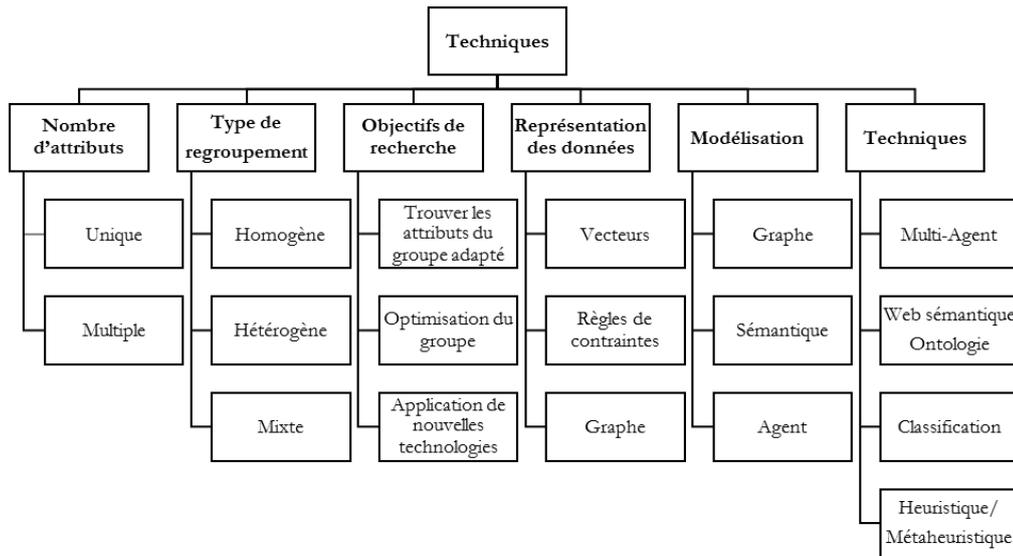


Figure 4.1 - Taxonomie des techniques de groupes de formation (Maqtary, N., 2019)

Les méthodes d'aide à la composition des groupes se distinguent principalement selon les attributs (ou les caractéristiques) des apprenants (âge, genre, niveau scolaire de connaissances, styles d'apprentissage...) ou selon les attributs du groupe de formation dans son ensemble (objectifs d'apprentissage, type de regroupement, la durée de vie du groupe et son adaptabilité). Parmi les techniques existantes, les techniques informatiques les plus couramment utilisés sont les méthodes méta-heuristiques et heuristiques et les méthodes de classification. Néanmoins, il est à noter que les travaux de recherche s'appliquent principalement dans un cadre scolaire et universitaire. De plus, la variété des contextes d'expérimentation rend difficile une analyse comparative et par conséquent d'affirmer qu'une technique informatique est plus pertinente à appliquer pour résoudre ce type de problème de décision.

4.2 Méthode méta-heuristique et heuristique

4.2.1 Principe de traitement

L'optimisation vise à trouver une solution optimale en prenant en compte un ensemble de contraintes et de variables inhérentes au problème de décision. Deux grandes familles de méthodes d'optimisation sont identifiées selon le mode de recherche de l'optimum :

- Méthodes déterministes (ou exactes) dont la résolution du problème donné vise à converger vers un optimum en parcourant un espace des solutions d'une manière identique ;

- Méthodes stochastiques (ou approchées) dont les méthodes sont basées sur une prospection aléatoire de l'espace des solutions à l'aide de règles de transition probabilistes. Parmi ces méthodes, deux sous-classes sont distinguées dont les méthodes méta-heuristiques/heuristiques.

Les méthodes méta-heuristiques sont définies par (Osman I., 1996) comme « un processus itératif qui subordonne et qui guide une heuristique, en combinant intelligemment plusieurs concepts pour explorer et exploiter tout l'espace de recherche. Une méta-heuristique peut donc regrouper plusieurs heuristiques. Des stratégies d'apprentissage sont utilisées pour structurer l'information afin de trouver efficacement des solutions optimales, ou presque-optimales ». Parmi cette catégorie, l'auteur inclut les méthodes basées généralement sur une amélioration itérative

d'une solution unique en construisant une trajectoire dans l'espace de recherche (méthode de recuit simulé, recherche tabou) et les méthodes à population de solutions. Le principe de traitement des méthodes à solution unique consiste à initier une seule solution et s'en éloigner progressivement tandis que les méthodes à population de solutions améliorent au fur et à mesure des itérations d'une population de solutions. Plus récemment, (Binitha, S., 2012) repris par (Yang, X., 2020) propose une taxonomie des méthodes méta-heuristiques pour l'optimisation inspirées de différents phénomènes du monde naturel qualifiés comme étant des « algorithmes bio-inspirés ». Les algorithmes sont classés en trois grandes catégories : les algorithmes évolutionnaires, les algorithmes à base d'intelligence distribuée et les algorithmes inspirés de l'écologie. Les algorithmes évolutionnaires sont considérés comme les premiers algorithmes bio-inspirés en intelligence artificielle pour l'apprentissage automatique et l'optimisation combinatoire. Ces algorithmes ont un principe de traitement stochastique selon lequel le processus d'optimisation se déroule de manière itérative avec une prise de décision souvent aléatoire, mais guidée par une fonction d'évaluation des individus appelée fonction objectif (ou *fitness*) (Back, T. 1996). Les principales méthodes sont illustrées selon la figure suivante :

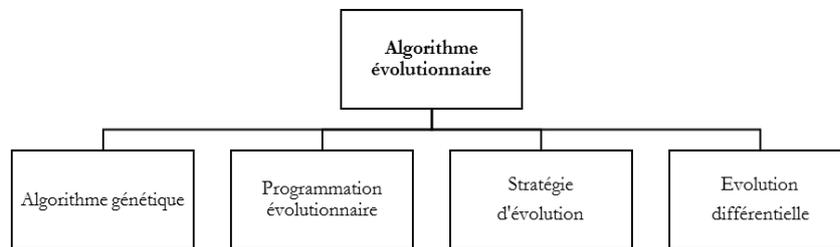


Figure 4.2 - Extrait de la taxonomie des algorithmes « bio-inspirés » (Binitha, S., 2012)

Parmi les différents algorithmes évolutionnaires, quatre catégories se distinguent :

-Algorithme génétique : proposé par (Holland J. 1975), l'AG est le plus utilisé pour la résolution de problèmes d'optimisation combinatoire (Zhou et al. 2011) ;

-Programmation évolutionnaire (ou génétique) : proposé par (Koza, J. R. 1994), cette méthode est considérée davantage comme un programme informatique en extension des algorithmes génétiques (Binitha, S., 2012) ;

-Stratégie d'évolution : cet algorithme est considéré comme l'une des premières versions des algorithmes évolutionnaires. Initialement appliqué dans le cas d'application d'un problème aérodynamique, les nouvelles versions sont proposées par (Beyer, H. G., 2002).

-Evolution différentielle : proposé par (Storn, R., 1997), le fonctionnement de cet algorithme est identique à celui de l'AG mais partitionne la population globale en plusieurs populations.

4.2.2 Algorithme génétique

Les algorithmes génétiques (noté AG par la suite) sont catégorisés comme étant des algorithmes d'optimisation de type méta-heuristique. Historiquement, les AG sont initiés par (Holland J., 1975) qui s'est inspiré du domaine de la théorie synthétique de l'évolution (ou TSE) de Charles Darwin en transposant les processus d'adaptation des systèmes naturels afin de concevoir des systèmes artificiels. A partir des années 1990, (Goldberg D., 1989) popularise cette technique et a particulièrement contribué à la diffusion de cette méthode dans le domaine scientifique. Le principe de traitement de l'AG vise à converger vers une solution dite « optimale » en procédant à une sélection des « meilleurs individus » contribuant à résoudre un problème donné comme schématisé selon la figure suivante par (Dréo J., 2003) :

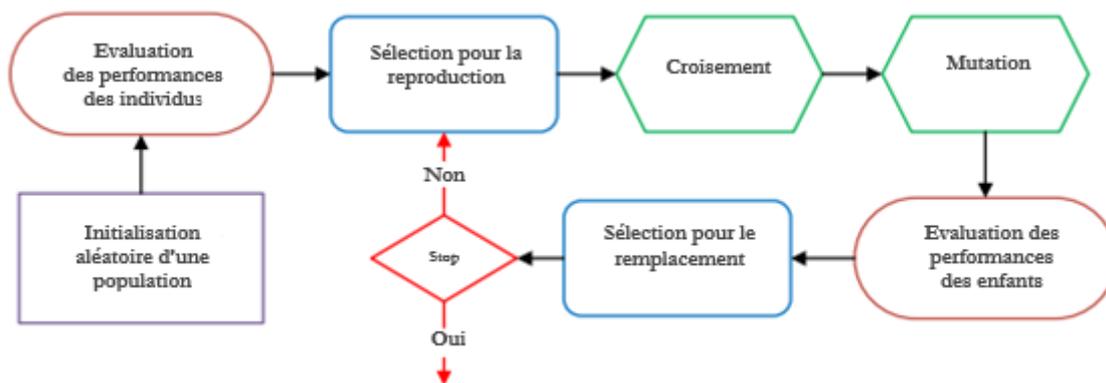


Figure 4.3 - Schéma simplifié du processus d'un algorithme génétique (Dréo J., 2003)

Le processus de l'AG démarre par une initialisation aléatoire d'une population d'individus représentée comme des « chromosomes » composés de « gènes » relatifs aux caractéristiques des individus. L'étape suivante consiste à évaluer les individus selon l'objectif de performance déterminé (couramment appelé *fitness*) et soumis à plusieurs opérateurs génétiques (sélection, croisement, mutation) pour obtenir une nouvelle population. Le but est de conserver les individus qui représentent les meilleures solutions. Le processus de traitement est répété jusqu'au critère d'arrêt qui détermine la fin du traitement.

Selon (Le Riche, R., 2007), les travaux liés aux AG se sont particulièrement orientés sur la recherche d'un optimiseur universel en se focalisant sur la représentation (ou codage) et l'importance des opérateurs de croisement et de mutation. Progressivement, les retours d'expériences ont démontré qu'un optimiseur universel reste théorique par manque de puissance de calcul. De plus, la variété et la complexité des problèmes posés impliquent qu'il devient nécessaire d'adapter a priori, les paramètres de l'algorithme (choix du codage, procédure d'initialisation et des opérateurs de variation) et la fonction à optimiser avec le problème de décision. Par la suite, les AG dit « modernes » sont abordés selon un angle adaptatif en intégrant plusieurs niveaux de complexité lors des paramétrages des opérateurs génétiques (multiples points de croisement, orientation de la mutation ...).

Dans le domaine de l'éducation (Kucak, D., 2019) recense environ 75 articles scientifiques de 2006 à 2019 qui utilisent les AG. Il catégorise l'ensemble des travaux selon les finalités suivantes :

- Identification de phénomènes spécifiques (taux de rétention des élèves et identification des étudiants en difficulté ou en décrochage scolaire) ;
- Analyse des données liées aux étudiants (prédiction des cours et des performances des élèves, identification de valeurs anormales liées aux fraudes...) ;
- Aide didactique pour les enseignants et aide logistique pour les personnels universitaires (création automatisée de tests d'évaluation, création de systèmes de recommandations, aide à la planification logistique des cours...).

L'ensemble des références est synthétisé dans le tableau suivant :

Finalités d'application	Détails des références
Identification de phénomènes spécifiques	(Sarafraz, Z., 2015), (Xing, W., 2015), (Yukselturk, E., 2014), (Zahedifard, M., 2015)
Analyse des données des étudiants	(Kalles, D. 2006), (Yemi, A. 2019)
Soutien aux personnels académiques	(Esteban, A.; 2018), (Ciguene, R.; 2019) (Dwivedi, P.; 2018), (Matias, J. B.; 2018) (Srisamutr, A.; 2018)

Tableau 3.1 - Synthèse des travaux scientifiques utilisant l'AG dans le domaine de l'éducation (Kucak, D., 2019)

D'une manière plus sporadique, les AG sont également utilisés pour semi-automatiser la composition des groupes de formation (Yannibelli, V. D., 2016), (Chniter, M., 2018). Au niveau des attributs du groupe, les techniques de l'AG permettent de paramétrer le type de regroupement (homogène *vs* hétérogène) souhaité. Dans un contexte de formation initiale, (Abnar, S., 2012) applique un traitement flexible où la nature du regroupement (homogène ou hétérogène) est paramétrable à partir d'une interface-utilisateur. De même, (Sukstrienwong, A. 2012) applique le principe de l'AG pour créer des groupes essentiellement de type « hétérogène » et ainsi assurer une forme d'équité dans l'acquisition des connaissances au sein du groupe. Ainsi, les résultats des tests effectués démontrent que les élèves peu performants ont peu de probabilités d'être réunis au sein d'un même groupe d'apprentissage. (Moreno, J., 2012) démontre que l'AG est particulièrement adapté pour traiter et sélectionner des individus ayant plusieurs caractéristiques (ou attributs) en comparaison avec des algorithmes déterministes. Néanmoins, il souligne que les choix stratégiques du paramétrage de l'AG restent dépendants du contexte d'expérimentation. Plus récemment, (Chen C.M, 2019) applique un paramétrage hybride de l'AG pour créer un groupe de nature « hétérogène » au niveau des caractéristiques des apprenants complété par des mesures de l'analyse des réseaux sociaux pour équilibrer la composition des groupes en milieu scolaire. De même, la méthode de composition peut influencer les résultats scolaires au sein d'un groupe d'élève selon (Zheng, Y., 2018), qui a mené une comparaison entre les méthodes de composition manuelle et automatique *via* un AG.

4.3 Méthode de classification automatique

Parmi les méthodes de classification automatique existantes, nous détaillons les méthodes de classification basée sur les graphes (ou détection de communauté) et les méthodes de classification non-supervisée.

4.3.1 Classification basée sur les graphes (ou détection de communauté)

La classification basée sur les graphes (ou détection de communauté) considère les clusters comme étant des ensembles de nœuds connectés reliés par des arêtes et modélisés sous forme de graphe. Issu du domaine des mathématiques discrètes, la théorie des graphes constitue le fondement de l'étude des réseaux complexes (ou issus du monde réel). Une littérature significative a abordé cette discipline selon des questions diverses (coloration des graphes, calcul optimal de chemin...) selon divers domaines d'applications (réseau informatique, réseau social, réseau biologique...).

Avec l'avènement de l'informatique, l'analyse des graphes s'est automatisée à travers des méthodes de détection de communauté dans les réseaux sociaux. Initialement, l'étude des réseaux sociaux constitue un champ d'application relativement ancien dont les nœuds sont considérés comme des individus (ou des entités sociales) relié par des relations de différentes natures. Néanmoins, l'application des méthodes de détection de communautés pour les réseaux sociaux est relativement récente. Le principe de partitionnement vise à détecter des sous-ensembles d'objets (appelés classes ou *clusters*) sans connaissance préalable de la structure de la communauté. La puissance de calcul des ordinateurs permet ainsi d'optimiser le temps de traitement notamment pour les grands graphes. Les techniques existantes sont nombreuses et demeurent un domaine de recherche actif en continuelle évolution. (Creusefond, J. 2017) synthétise les différentes stratégies de partitionnement :

- Méthodes de partitionnement de données où l'on cherche à grouper des objets génériques selon une fonction de similarité entre les objets ;

-Méthodes divisives qui se basent sur l'identification et la suppression successive des arêtes entre les clusters ;

-Méthodes dynamiques qui simulent un processus de parcours de graphe ;

-Méthodes spectrales qui se basent sur une matrice spécifique liée à la proximité entre les nœuds analysés. Cette technique vise à détecter au sein du graphe la présence de sous-groupes de sommets fortement liés entre eux et faiblement liés avec l'extérieur. Le processus calculatoire repose sur un principe de calcul de distance topologique entre les nœuds : deux nœuds proches sont considérés comme similaires et au contraire deux nœuds séparés par une large distance sont considérés comme différents. Cette méthode de classification est particulièrement utilisée pour l'aide à la composition des groupes d'apprentissages détaillée dans la section suivante.

La principale limite de cette méthode basée sur les graphes est que la nature des liens est dépendante de l'interprétation qu'on lui infère au préalable. Selon (Beauguitte, L. 2013), la valeur sémantique se limite selon trois types de graphes :

-Graphes binaires (ou booléens) où les liens traduisent l'existence d'une relation entre deux sommets ;

-Graphes valués (ou pondérés) où les liens sont associés à une fonction de valuation (ou de pondération) ;

-Graphes signés où les liens expriment l'attraction ou le rejet étiqueté par un signe positif (+) ou négatif (-).

La valeur d'un signe positif représente une relation favorable tandis qu'un signe négatif représente une relation défavorable. Elle peut également représenter une relation d'ordre de supériorité (+) et d'infériorité (-). Néanmoins, ces valeurs binaires selon le type de graphe peuvent s'avérer limités dans certains cas d'applications. En effet, les limites de ces approches de classification est l'absence d'informations qualitatives sur la nature des interactions et la prise en compte de la subjectivité du profil-utilisateur, la nature des liens existants et sa multi-appartenance à plusieurs communautés. Pour surmonter ces limitations (Degenne A., 1994 ; Mercklé P., 2004, Newman M., 2004, (Crossley, N., 2010) suggèrent une approche mixte en complétant l'approche quantitative avec des concepts sociologiques de l'analyse des réseaux sociaux. De plus, il est à noter que la plupart des techniques existantes n'est applicable qu'à partir d'un graphe mono-relationnel ⁴⁸. Ce constat a fait émerger les premiers modèles d'analyse multi-relationnelle de réseaux sociaux en introduisant des dépendances spécifiques pour correspondre davantage aux réseaux sociaux réels. Selon (Lazega E., 2014), la complexité des liens peut être analysé et traité en appliquant une « analyse des relations multiples entre toutes les paires d'acteurs qui procure une meilleure connaissance de la structure relationnelle du système social étudié et une meilleure mise à plat synthétique du système d'échange ». Ainsi, cette technique permet d'observer et d'expliquer des phénomènes sociaux plus réalistes notamment *via* les graphes bipartis qui permettent d'étudier et de visualiser les relations entre deux ensembles distincts de sommets (Beauguitte, L. 2013). Par la suite, les travaux ont évolué principalement vers l'étude de la dynamique des réseaux sociaux dont la prédiction des liens sociaux pour identifier des réseaux de collaborations. Dans un cadre universitaire, la recommandation de collaborations académiques est un objet de recherche privilégié compte tenu des relations inhérentes entre les auteurs affiliés à une institution académique (Backstrom, L., 2006) (Benchettara, N., 2011), (Bossard, A., 2016). Dans un cadre industriel, (Yahia, N. B., 2015) propose une approche hybride de détection de communautés avec une méthode méta-heuristique (optimisation par essaim particulière) afin de traiter simultanément les interactions sociales et la similarité des individus pour identifier un groupe de collaborateurs.

Dans le domaine de l'apprentissage humain, la détection d'une communauté est principalement utilisée pour identifier des réseaux d'élèves (ou étudiants) ainsi que pour la visualisation des graphes complexes. En comparaison avec les méthodes de classification non-supervisée détaillé dans la section suivante, cette technique reste minoritaire parmi la littérature

⁴⁸ Graphe ne comportant qu'un seul type de lien

scientifique. Par exemple, (Hmelo-Silver, C. E., 2014) applique cette méthode pour créer des groupes de discussions au sein des forums des MOOC. (Sadeghi, H., 2016) optimise le regroupement homogène d'élève en modélisant un groupe d'apprenants sous forme de graphe en attribuant un poids pour chaque lien indiquant le degré de similitude entre les individus. En comparaison avec une méthode de classification de type K-Means, le partitionnement en sous-groupe s'est opéré selon une durée d'exécution moins optimale mais a généré de meilleurs résultats au niveau de la similarité des profils au sein du groupe.

4.3.2 Classification non-supervisée

Parmi les travaux de recherche liés à l'éducation, (Pollalis Y., 2009) (Cruz W., 2014) souligne que la méthode de classification non-supervisée (ou *clustering*) soit K-Means (ou par k plus proches voisins) est particulièrement utilisée pour l'aide au regroupement des étudiants/élèves. Cette méthode consiste à partitionner de manière optimale l'ensemble des instances en un groupe fini de groupes (K). L'approche consiste à générer un partitionnement initial d'une manière aléatoire pour affiner par la suite l'affectation des classes en minimisant la distance entre deux instances de la même classe (ou intra-groupe) ou entre deux instances de classes différentes (ou inter-groupe)⁴⁹. Le principal inconvénient est la fixation préalablement du nombre de classes (K) d'une manière arbitraire qui déterminera la qualité des résultats. Ainsi, cette approche semble intuitive compte tenu que la classification consiste à partitionner les données selon un calcul de similarité en groupes distincts. De plus, la visualisation graphique des clusters permet une analyse et une interprétation intuitive des résultats. L'identification des groupes d'apprenants peut répondre à divers objectifs mais se base néanmoins selon les critères que l'on lui assigne préalablement. Parmi les travaux existants, l'approche classique consiste à semi-automatiser des regroupements en se basant sur les caractéristiques des étudiants/élèves comme par exemple : optimiser un regroupement d'élèves d'une manière homogène en se basant sur leurs caractéristiques personnelles (Christodouloupoulos, C. 2007) ou sur les critères de préférences d'apprentissage des élèves (Joseph, N., 2017). Dans ce cas d'application, (Webber, C., 2012) démontre que les préférences d'apprentissage ont généré des groupes où davantage d'interactions ont été observées entre les membres de groupes.

Selon (Hamalainen, W., 2015), une seconde approche consiste à utiliser les techniques de classification comme une étape de pré-traitement pour le paramétrage de modèles pour les STI⁵⁰ ou d'autres systèmes informatique adaptatif dans une perspective de prédiction. Par exemple, le partitionnement préalable des élèves vise à analyser et prévoir les interactions entre élèves dans le cadre d'un jeu éducatif (Davoodi, A., 2013) ou pour améliorer la précision de la prédiction du placement des élèves (Shukla, M., 2019).

(Hamalainen, W., 2015) souligne également que le jeu de données utilisé influence particulièrement le choix de la méthode de classification à appliquer. Le principal inconvénient est le nombre de variables restreint qui proviennent généralement d'un cours pendant un semestre impliquant un nombre d'étudiants limité sur une période temporelle courte. Il ajoute que la subjectivité des profils d'étudiants caractérisée par des valeurs dites « aberrantes » (correspondant aux étudiants dits « hors-normes ») ne peut pas être pris en compte, celles-ci étant systématiquement écartés. Au niveau de l'évaluation, les algorithmes de classification non-supervisée se mesure selon leur capacité à classer correctement les données selon le partitionnement souhaité et selon le temps d'exécution utilisée. D'une manière globale, le choix d'une méthode de classification influence les résultats selon la nature des données mise à disposition et la finalité de l'observation. Néanmoins, il reste difficile de définir la pertinence

⁴⁹ La distance entre groupes se réfère à quatre possibilités : la distance entre les centres de gravité des groupes, la distance minimum entre les deux groupes, la distance maximum entre les deux groupes et la distance moyenne entre toutes les paires d'individus

⁵⁰ Systèmes de Tutorat Intelligent

d'une méthode de traitement compte tenu de la spécificité des contextes d'expérimentation et par conséquent de comparer les résultats de partitionnement (Fortunato, S., 2016).

Chapitre 5. Conclusion

Ce chapitre résume l'état de l'art et les problématiques scientifiques associées.

Les deux premiers chapitres synthétisent les systèmes d'information existants dans un contexte organisationnel et scolaire en se focalisant sur l'analyse et la visualisation des données.

Les systèmes d'information demeurent un vaste objet de recherche scientifique et opérationnel (fonctionnel, technique, décisionnel...) en perpétuelle évolution. A l'intersection de l'informatique de gestion et décisionnel, le premier chapitre aborde les systèmes d'information au sein des organisations (ou entreprises). En informatique décisionnel, les techniques d'aide à la décision existantes se sont multipliées (méthode d'agrégation partielle ou complète) selon le type de problème à résoudre (nature des variables, mono ou multicritère ...). Ainsi, la modélisation d'un problème de décision s'appuie sur de nombreuses techniques pour optimiser des critères multiples et qui peuvent être antagonistes. Néanmoins, la résolution s'avère plus délicate pour explorer des solutions potentielles qui se heurte à la contrainte du temps de traitement.

En informatique de gestion, les SIRH sont essentiellement dédiés à la gestion des ressources humaines. Les techniques informatiques utilisées sont essentiellement la planification automatique sous-contraintes pour optimiser la gestion du personnel tandis que les travaux concernant la gestion de la formation continue sont minoritaires. Par la suite, les SIRH ont progressivement évolué vers des systèmes d'information décisionnel en intégrant diverses techniques informatiques (analyse de données stratégiques, aide au recrutement...)

Dans le domaine de l'apprentissage humain, les systèmes informatiques ont fait l'objet de nombreux courants de recherche et ont révélé des problématiques variées. Néanmoins, les travaux scientifiques sont principalement abordés sous l'angle de l'interactivité des systèmes informatiques (plus précisément les EIAH) avec le(s)apprenant(s) pour soutenir le processus d'apprentissage. Les outils d'aide aux enseignants sont abordés à travers les STI et visent principalement à faciliter les activités didactiques dans un domaine d'enseignement spécifique.

Dans un contexte de formation continue, les outils de la formation continue diffèrent peu des outils de la formation initiale (LMS, EIAH, MOOC/SPOC, systèmes de recommandation...) mais sont utilisés différemment liés à la spécificité de l'adulte-apprenant (contenus pédagogiques, nombres de participants, format temporel...). Néanmoins, la formation en présentiel reste un mode d'action favorisé permettant de mettre à disposition un espace-temps distinct de la situation de travail. Ainsi, les systèmes informatiques deviennent un support intermédiaire pour articuler les temps de formation (présentiel/distanciel) et favoriser la cohésion entre les salariés d'une organisation.

Après avoir présenté les domaines de recherche, le deuxième chapitre met en évidence les points communs de l'analyse de l'apprentissage humain (ou *Learning Analytic*) et le domaine de l'analyse des données liées aux missions des ressources humaines (ou *HR Analytics*). Liée à l'expansion de la volumétrie des données, ces domaines de recherche appliquent des techniques statistiques communes qui consistent à traiter l'ensemble des données provenant de sources multiples pour pouvoir mener une interprétation finale. L'objectif convergent est de mesurer des faits observables et quantifiables liées aux activités de formation pour assurer un suivi des performances et aider la prise de décision dans une perspective de remédiation (Kunfa, L., 2019). De même, ces domaines de recherche sont également confrontés à des problématiques communes liées au traitement des données (collecte des données, choix des méthodes, gestion et qualité des données...) et à l'évolution de la réglementation juridique. Contrairement au *LA*, le domaine de l'analytique des ressources (ou *HR Analytics*) n'est pas focalisé essentiellement sur l'analyse des données de la formation continue et s'applique aux diverses missions des ressources humaines (recrutement, gestion des carrières professionnelles...). Au niveau de la restitution des résultats, les outils associés dont les tableaux de bord et les types de visualisations sont également communs. Domaine de recherche à part entière, la visualisation des données propose des

méthodes et des types de représentations qui peuvent différer selon les domaines d'applications. Dans le domaine de la formation, la restitution des données utilise peu de techniques sophistiquées et se limite à une représentation graphique simple des données (diagramme à barre, histogramme, graphe...). En effet, les travaux appliquant la visualisation des données complexes notamment *via* les graphes multicouches sont minoritaires.

Les deux derniers chapitres se focalisent autour du groupe de formation pour adulte et les techniques informatiques associées.

Le troisième chapitre présente le groupe de formation pour adulte qui reste peu étudié dans la littérature scientifique. Le groupe de formation est un objet de recherche complexe compte tenu de ses dimensions multiples et interdépendants (relationnelle, cognitive, environnementale...). Dans le cadre de la formation continue, les mécanismes du groupe de formation diffèrent en comparaison avec la formation initiale mais restent un mode d'apprentissage privilégié pour l'apprenant-adulte. Au-delà du processus cognitif, le groupe de formation devient un vecteur social où les membres s'inscrivent principalement selon une approche relationnelle tout autant avec le formateur qu'avec autrui. Des phénomènes de comparaison sociale sont particulièrement amplifiés du fait, que l'identité et le statut professionnel deviennent des marqueurs socio-professionnels qui peuvent influencer la nature des interactions sociales. Ainsi, le degré de la symétrie et l'asymétrie de la relation sociale *via* la hiérarchisation des relations professionnelles est peu favorable à la collaboration et l'interaction entre pairs. De même, largement débattue en formation initiale, le type de regroupement du groupe selon l'homogénéité ou l'hétérogénéité des profils des membres du groupe est abordé différemment dans un groupe de formation pour adulte. La recherche de personnes similaires selon une approche catégorielle et hiérarchique est particulièrement favorisée dans un contexte où les interactions sociales sont principalement basées sur des échanges de pratiques professionnelles.

En prolongement du chapitre lié au groupe de formation pour adulte, le dernier chapitre aborde les méthodes informatiques pour semi-automatiser la composition des groupes de formation. Au niveau du traitement algorithmique, la génération des résultats peut être décentralisée des systèmes informatiques qui sont principalement des STI. Les techniques algorithmiques sont variées (Markov, ontologie, multi-agent...) dont les paramétrages se limitent principalement en fonction des caractéristiques des membres du groupe (genre, âge, niveau d'étude...) et du type de regroupement (homogène/hétérogène). Néanmoins, une légère tendance axée sur les méthodes d'optimisation et de classification supervisée est observée parmi la littérature scientifique. Parmi les méthodes d'optimisation, (Kucak, D., 2019) souligne l'intérêt grandissant des techniques d'AG pour la résolution de problèmes réels et multidimensionnels dans le domaine de l'éducation. Inspirés des concepts de la biologie évolutionniste, le principe de traitement de l'AG permet une exploration rapide et capable de générer des solutions admissibles. Les finalités d'application sont diverses mais restent principalement axées vers l'identification de phénomènes spécifiques liées aux profils des étudiants et le soutien des missions des enseignants et personnel dans un contexte universitaire. La principale difficulté est la collecte et le traitement simultané de plusieurs variables qui peuvent être de différentes natures (quantitatives/qualitatives).

Les méthodes de classification dont particulièrement l'algorithme *k*-moyennes (ou *k-means*) sont également étudiés compte tenu que le principe correspond à la finalité recherchée soit le partitionnement et regroupement d'individus à travers des classes (ou *clusters*) homogène. Ainsi, les méthodes de classification procèdent par itération à partir d'un point unique, à l'instar de l'AG dont le principe de traitement explore les solutions possibles à partir d'une population de points. La comparaison des résultats entre ces méthodes implique que certaines conditions expérimentales doivent être respectées (utilisation d'un jeu de données identique par exemple). Néanmoins, les travaux de la littérature scientifique ne démontrent pas qu'il existe une méthode

informatique comparativement meilleure qu'une autre ; le choix d'une méthode reste dépendant de la maîtrise de l'outil par l'utilisateur et du problème d'optimisation posé.

Partie 2 - LexRH : Méthodologie et développement d'un prototype d'aide à la gestion et à la planification de groupe en formation continue

Cette seconde partie présente le prototype d'aide à la gestion et à la planification de groupe en formation continue proposée et inclut les trois chapitres suivants :

Le sixième chapitre présente les spécifications fonctionnelles et techniques du prototype d'application. La méthodologie de la représentation des données et les techniques de visualisation associées sont détaillées et illustrées par des exemples d'utilisations.

Le septième chapitre présente le principe de traitement et l'implémentation de l'algorithme génétique (noté AG par la suite) qui consiste à planifier les groupes de formation potentiels. De type heuristique, ce chapitre détaille les différentes étapes méthodologiques : formulation des problèmes de décision, détermination des contraintes et des critères de décision, le principe de traitement et les paramétrages de l'AG.

Le huitième chapitre présente une analyse statistique des questionnaires de satisfaction. De type analytique, cette étape vise principalement à identifier les caractéristiques des groupes de formation qui conditionnent favorablement les interactions sociales entre les membres du groupe. Ainsi, cette étude observationnelle permet d'orienter la fonction objective (ou *fitness*) de l'AG et d'autre part, d'identifier les indicateurs au sein du tableau de bord dédié à l'évaluation de l'activité.

Chapitre 6. Conception et développement d'un prototype d'aide à la gestion de la formation continue

Ce chapitre décrit les spécifications fonctionnelles, ergonomiques et techniques du prototype d'aide à la gestion proposé selon les besoins et tâches opérationnelles identifiés.

6.1 Système d'aide à la gestion de la formation continue

6.1.1 Analyse de l'existant

Face à l'accroissement du volume des données en contexte industriel, les organisations (ou entreprises) ont intégré des systèmes informatiques de gestion connues sous diverses appellations : ERP⁵¹, EDD⁵², CRM⁵³... La finalité d'application vise à faciliter le stockage, l'accès et l'analyse des données selon des perspectives commerciales, organisationnelles et/ou collaboratives.

Dans le cadre de ce travail, l'outil de gestion existant repose sur un logiciel de relation client de type CRM qui vise à stocker l'ensemble des données relatives à plusieurs activités professionnelles⁵⁴. Ainsi, la gestion de l'activité de formation est essentiellement abordée selon une approche économique et commerciale et par conséquent, peu adapté pour les tâches quotidiennes de la formation continue. La couverture opérationnelle permet essentiellement de rechercher les données personnelles et les historiques de formation des participants. Au niveau technique, l'extension des fonctionnalités est rigide compte tenu de l'architecture de la base de données selon un modèle relationnel de type SGBD⁵⁵. Bien que couramment utilisé, ce type de modélisation est peu adapté pour le paramétrage de tâches opérationnelles spécifiques à l'activité de formation (planification des sessions, l'analyse des évaluations de formations...). La recherche et l'extraction des données sont formulées à travers une requête de type SQL selon des opérations d'algèbre relationnelle. En effet, le groupage d'objets complexes⁵⁶ peut être appliqué selon des opérations ensemblistes classiques (union, intersection, différence) pour traiter deux ensembles d'objets. Néanmoins, cette démarche nécessite de connaître préalablement le chemin d'accès aux données recherchées et implique une dépendance liée à la modélisation de la base de données. Dans un contexte d'utilisation élargie, la formalisation d'une requête de type SQL demeure une opération peu souple et difficile d'utilisation pour des utilisateurs non-spécialisés en base de données.

Dans un cadre de recherche scientifique, des travaux informatiques ont étudiés les techniques de fouilles de données (ou *data mining*) au sein de système de type progiciel pour faciliter la recherche et la prise de décision (recherche de groupes de clients spécifiques, prédiction de clients potentiels...) (Rygielski, C., 2002 repris par Liao, SH, 2012). Néanmoins, ces fonctionnalités restent peu implémentées au sein des principaux outils proposés sur le marché commercial des logiciels d'entreprises qui appliquent principalement une approche classique du système d'information de type entrepôt de données.

⁵¹ Entreprise Resource Planning ou Prologiciel de Gestion Intégré

⁵² EDD (base de données décisionnelles ou datawarehouse)

⁵³ Customer Relationship Management

⁵⁴ Activité de conseil et activité de formation continue

⁵⁵ Système de Gestion de Base de Données

⁵⁶ Objet composé lui-même d'autre objet

6.1.2 Objectif(s)

A partir de l'analyse de l'existant, notre proposition consiste à développer un système d'information de gestion de la formation continue qui peut être considéré comme une brique d'un Système d'Information des Ressources Humaines (SIRH).

L'objectif principal du prototype d'application vise à faciliter et à semi-automatiser les tâches opérationnelles essentiellement dédiées à l'activité de la formation continue. Il est destiné à un nombre restreint d'utilisateurs (employés/décideurs) n'ayant pas de connaissance préalable dans le domaine de l'informatique. Les principales fonctionnalités peuvent être synthétisés de la manière suivante :

- Centraliser et faciliter la recherche simple des données (participant, organisation, formateur, événement de formation) ;

La recherche de données demeure une tâche opérationnelle courante dans la gestion des données. Les outils de gestion associés sont les suivants :

Tableau de bord principal. Le tableau de bord principal propose une vue globale de l'état actuelle de l'activité de formation en temps réel. Le tableau de bord constitue le point d'entrée de l'arborescence du prototype d'application avec un accès aux différentes fonctionnalités :

- « Recherche » dédiée à la recherche simple de données ;
- « Planification » dédiée à la planification semi-automatisé des sessions de formation ;
- « Evaluation » dédiée à l'évaluation des sessions de formation.

Recherche simple. La fonctionnalité « Recherche » permet la recherche simple et la consultation des données liée à l'activité de formation. A partir d'une requête, cette fonctionnalité permet au gestionnaire de formation d'extraire des données dont les personnes physiques (participant et formateur), les personnes morales (organisation) et les événements de formation.

Dans un CRM traditionnel, les données sont organisées sous forme tabulaire où chaque enregistrement décrit les caractéristiques des différents acteurs (organisation, client, participant de formation...) au sein de la base de données. Ce type de modélisation ne facilite pas l'extraction des relations entre les entités, qui peuvent être porteuses d'information au sein d'un réseau social (par exemple X est_employé par Y). Pour pallier cette limite, le stockage et la recherche des données sont basés sur un modèle orienté « graphe »⁵⁷ qui permet de représenter les données sous la forme de graphes et d'extraire les relations entre les entités.

- Semi-automatiser la planification des sessions de formation *via* la composition des groupes de formation en présentiel ;

La planification des sessions de formation vise à automatiser une tâche opérationnelle dite « complexe ». La résolution consiste à prendre en considération des contraintes et de multiples critères spécifiques au contexte de formation. De plus, la méthode de planification diffère selon le type de planification et l'initiateur de l'action de formation. Pour répondre à cette problématique, l'outil de gestion proposé est le suivant :

Aide à la planification. La fonctionnalité « Planification » vise à faciliter la planification d'une session de formation qui se réalise essentiellement à travers un groupe de formation en présentiel. Dans un contexte de formation continue, la planification d'une session de formation est définie

⁵⁷ NoSql

préalablement selon deux cas d'utilisations : le mode « Inter ⁵⁸/Intra ⁵⁹ organisationnel ». Par conséquent, la composition des groupes de formation est soumise à une contrainte « stricte »⁶⁰ qui détermine préalablement l'affectation des participants au sein du groupe de formation. Par la suite, des multiples critères de différentes natures (social, cognitive et logistique) sont à prendre en considération mais ne sont pas considérés comme étant des contraintes.

- Contrôler et analyser des données liées à l'évaluation de la formation

L'évaluation de l'activité de formation est une obligation juridique imposé par la réglementation de la formation continue. Elle se révèle être également un enjeu stratégique pour améliorer la qualité du dispositif de formation. Les données et les résultats de l'évaluation de la satisfaction des participants sont centralisés à travers l'outil de gestion suivant :

Tableau de bord secondaire. Le tableau de bord secondaire est dédié à l'évaluation des sessions de formation. L'outil de gestion vise à stocker et traiter les résultats des questionnaires de satisfaction remis aux participants en fin de session de formation. Ainsi, l'agrégation et la visualisation des données visent à faciliter la consultation et l'analyse de l'activité de formation selon les besoins suivants :

- Analyse descriptive de l'activité de formation selon des indicateurs préalablement définis ;
- Identifier les facteurs qui favorisent la satisfaction et l'insatisfaction des participants

Les indicateurs du suivi de l'activité de formation sont déterminés à partir de l'analyse statistique détaillée dans le chapitre suivant ([Voir Section 8.4](#)). L'analyse des résultats vise à orienter les axes d'améliorations pour les prochaines sessions de formation.

6.1.3 Type d'utilisateurs

Le système d'aide à la gestion de la formation continue est principalement destiné à un type d'utilisateur spécifique principalement le gestionnaire (ou responsable) de formation. Ce métier vise à assurer la gestion et l'organisation de l'activité de formation en interne d'une entreprise (suivi des contrats, planification des sessions de formation, relation client...) ou peut être délégué à un tiers, généralement un organisme de formation. Selon la définition proposée par (Clarens V., 2008), les missions du « responsable de formation visent à maîtriser et coordonner des dispositifs complexes et multidimensionnelles (...) et exige ainsi de nombreuses compétences, notamment relationnelles pour se situer au cœur d'un réseau d'acteurs internes (direction, IRP, managers, salarié) et externes (OPCA, organisation de formation, services publics de l'emploi...) ». Ainsi, ce métier s'inscrit progressivement dans une démarche d'ingénierie de formation et doit particulièrement s'adapter à l'évolution de la réglementation juridique de la formation professionnelle.

Selon les appellations des métiers basées sur un référentiel officiel (ROME ⁶¹), les utilisateurs finaux sont les suivants :

- Responsable expert et généraliste du domaine de ressources humaines (rémunération, GPEC, recrutement, formation...);
- Responsable de la formation.

⁵⁸ Groupe composé des personnes travaillant au sein de différentes organisations

⁵⁹ Groupe composé des personnes travaillant au sein d'une même organisation

⁶⁰ Qualificatif d'une contrainte qui doit être impérativement respectée

⁶¹ Répertoire Opérationnel des Métiers et des Emplois

Compte tenu que ces utilisateurs finaux n'ont pas de connaissance préalable en informatique, l'ergonomie des interfaces fait l'objet d'une attention particulière pour faciliter la prise en main du système.

Au niveau de la conception, du développement et de la maintenance du prototype d'application, un nouveau métier est apparu liée à l'informatisation des activités des ressources humaines soit le « responsable SIRH ». Spécialisé dans les systèmes d'information dans le domaine des ressources humaines, le périmètre d'activité du responsable SIRH peut se limiter à l'activité de formation continue.

6.1.4 Type de données

La gouvernance et la gestion des données visent à garantir la qualité des données selon les finalités de traitement précédemment citées. Dans notre cas d'application, les données sont variées et principalement collectées à partir de sources internes et externes. Elles peuvent être réparties de la manière suivante :

- Données externes issues de bases de données professionnelles spécialisées ;
- Données internes liées à l'activité de l'organisation ;
- Données internes issues des questionnaires de satisfaction.

La centralisation des données est une étape préalable pour optimiser le traitement et l'analyse des données à venir. Ainsi, les données exploitées sont structurées par opposition aux données du « *Big Data* » qui sont externes et non-structurés.

Les données utilisées sont récapitulées dans le tableau suivant :

Catégories			Variables
Personne	Physique	Participant	Nom, Prénom, Genre, Statut, Profession, Organisation, Direction, Historique de formation, Type de participant (ou formation), Ancienneté
		Formateur	Nom, Prénom, Genre, Profession, Thématique de formation, Organisation
	Morale	Organisation	Nom, Secteur d'activité, Type d'établissement, Taille, Lieu géographique
Activité de formation	Session de formation		Mode organisationnel, Thématique, Formateur, Date de début, Date de fin, Lieu géographique, Nombre de participants, Montant de facturation
	Evaluation		Score de satisfaction global, Score de satisfaction (contenu pédagogique, animation et organisation matérielle/logistique)

Tableau 4.1 - Catégorisation et description des variables liées à l'activité de formation

Personnes physiques. Les données relatives aux personnes physiques sont personnelles et professionnelles. Elles se limitent aux variables « Nom », « Prénom », « Genre », « Profession⁶² » (i.e le métier), « Statut » et « Historique de formation ». En complément des données nominatives, des données complètent le profil des personnes physiques :

La variable « Historique de formation » consiste à connaître les actions de formation effectuées.

La variable « Statut » vise à déterminer la position hiérarchique et le niveau de responsabilité exercé par la personne au sein de l'organisation.

⁶² La définition du terme « Profession » est multiple selon les courants sociologiques. Dans le cadre de ce travail, la profession d'une personne est collecté en fonction de sa déclaration dans le questionnaire lors l'évaluation de la session de formation.

La variable « Type de participant » consiste à distinguer le type de formation effectué (Colloque ou Formation⁶³).

La variable « Profession » permet d'indiquer le domaine d'expertise de la personne.

Les données de la variable « Profession » présentent des limites. Elles peuvent être manquantes ou obsolètes compte tenu de la mobilité professionnelle des personnes. De plus, les données collectées à partir des questionnaires d'évaluation sont essentiellement déclaratives et peuvent impliquer des valeurs erronées liées à un biais de désirabilité sociale. En effet, ce concept initié par (Mc Crae R., 1983), repris par (Tournois J., 2000) décrit une tendance à déformer les descriptions de soi dans un sens favorable et à produire une image de soi positive. Elles présentent également une anomalie de type sémantique car les valeurs associées répondent à de multiples variations terminologiques selon les organisations.

Les données des variables « Age » et « Niveau d'éducation » sont également peu exploitables compte tenu de la proportion importante des valeurs manquantes. D'autre part, la signification de cette variable influence peu l'activité de formation à l'inverse de la variable « Ancienneté » où l'on peut déduire un niveau d'expertise d'une personne au sein de sa fonction (ou métier).

Les causes de ces anomalies sont variées (sources hétérogènes, saisie manuelle par différents utilisateurs, manque de protocole de saisie...). L'identification de ces anomalies est une étape indispensable pour définir par la suite une stratégie d'homogénéisation des données.

Personnes morales. Les données relatives aux personnes morales (ou organisations/entreprises) sont également nominatives et se limitent aux variables « Secteur d'activité », « Type d'établissement », « Forme juridique » et « Lieu géographique ». D'un point de vue juridique, ces données peuvent être librement compilées sans qu'il y ait une nécessité d'informer la personne morale (ou ses représentants) contrairement aux données des personnes physiques. Les données brutes sont collectées à partir d'annuaires professionnels assurant ainsi une qualité des données au niveau de la justesse des valeurs associées.

Activité de formation. Les données liées à l'activité de formation sont relatives à la description des sessions de formation et aux scores de satisfaction associés collectés à partir des questionnaires de satisfaction. Deux types de données sont à distinguer : la description des sessions de formation sont des données brutes tandis que les données liées à l'évaluation des sessions de formation sont des données complexes (ou calculés).

Ainsi, les données liées à l'activité de formation sont multidimensionnelles. Elles sont de différentes natures (personnelles, professionnelles, économiques ou financière...), collectées à partir de sources hétérogènes (sources externes et internes) et évolutives impliquant une veille et une mise à jour régulière.

La description des variables est détaillée en Annexe 1.

6.2 Représentation et visualisation des données

La visualisation des données vise à faciliter la compréhension et l'interprétation d'un ensemble de données préalablement collectées. Cette étape détaille la méthodologie utilisée pour analyser les données en vue d'un traitement et d'une représentation visuelle.

Deux types de visualisation sont à distinguer :

-Visualisation de type descriptive pour la recherche simple et la planification des sessions de formation ;

-Visualisation de type analytique pour l'analyse des données liées à l'évaluation des sessions de formation.

⁶³ Le colloque est une formation de type informel tandis la formation est formelle et contractualisée

6.2.1 Visualisation descriptive

La restitution des résultats liée aux fonctionnalités « Recherche » et « Planification » sont représentés sous forme de visualisation de type nœud-lien (appelé également diagramme *node-link*). Ce type de visualisation, particulièrement adapté aux réseaux sociaux (ou réseaux issus du monde réel), permet d'identifier les différents acteurs de la formation et les relations socio-professionnelles.

- Analyse des données

Au préalable, l'analyse des données vise à déterminer les éléments du graphe qui sont représentés par des nœuds⁶⁴ et des relations⁶⁵. Cette étape est déterminante car elle garantit une représentation/abstraction des données selon une situation réelle.

La nature des relations sociales au sein du groupe oriente notre démarche méthodologique selon une approche structurale de l'analyse des réseaux sociaux. Dans notre cas d'étude, les entités sociales peuvent être des personnes physiques (participant et formateur) ou des personnes morales (organisation) et sont représentées par des nœuds. L'ensemble des nœuds sont reliés par des relations socio-professionnelles représentées par des liens (ou arêtes). Cette approche descriptive vise à identifier la typologie des acteurs et de mettre en évidence les relations socio-professionnelles de différentes natures récapitulées dans le tableau suivant :

	Type	Description
Nœuds	Participant	Personne physique ayant effectué à minima une session de formation
	Formateur	Personne physique (expert de la thématique de la formation)
	Organisation	Personne morale
	Événement	Session de formation
Relations	Intra-organisation	Relation hiérarchique/fonctionnelle
	Inter-organisation	Relation de collaboration
	Emploi	Affiliation d'un participant (personne physique) avec une organisation (personne morale)
	Partenariat	Relation entre participant et formateur
	Animation	Supervision d'un événement par un formateur

Tableau 4.2 - Typologie des nœuds et des relations en situation de formation continue

Cette catégorisation des relations socio-professionnelles synthétise l'ensemble des relations existantes dans une situation de formation : la relation intra-organisation, la relation inter-organisation, la relation d'emploi et la relation de partenariat.

Relation intra-organisation. Les participants sont affiliés au sein d'une même organisation. Ainsi, les relations socio-professionnelles sont hiérarchiques et/ou fonctionnelles selon la structure interne de l'organisation. De nature permanente, ce type de relation peut être également considéré comme étant « asymétrique » (appelé également « vertical ») compte tenu de l'existence éventuelle d'une subordination entre les personnes.

Relation inter-organisation. Les participants sont affiliés au sein d'organisations différentes. Ainsi, les relations socio-professionnelles sont collaboratives du fait qu'elles sont occasionnelles et basées autour d'un intérêt professionnel commun. Ce type de relation professionnelle peut être également considéré comme étant « symétrique » (appelé également « horizontale ») compte tenu de l'absence de hiérarchisation formelle entre les personnes.

⁶⁴ Appelé « sommet » selon la terminologie de la théorie des graphes

⁶⁵ Appelé « arête » selon la terminologie de la théorie des graphes

Relation d'emploi. La relation professionnelle entre l'organisation (ou employeur) et le participant (ou employé) se déploie dans un cadre organisationnel et formalisée selon une contractualisation juridique.

Relation de partenariat. La relation pédagogique entre le formateur et le participant est considérée comme étant une relation de partenariat. La nature de la relation professionnelle est temporaire mais peut être considéré comme étant « asymétrique » compte tenu de l'autorité inhérente liée au statut du formateur.

Ainsi, la situation de formation s'articule autour de trois acteurs (participant, employeur et formateur). Dans le cadre organisé de la formation continue, des relations professionnelles spécifiques dont les relations bilatérales traditionnelles (relation employeur/employé et employé/employé) sont observées mais restent dépendantes d'un tiers extérieur (formateur). Cette différenciation des acteurs et des relations professionnelles est à prendre en considération compte tenu que l'événement de formation s'inscrit dans le prolongement de ce contexte professionnel.

Au niveau de la modélisation, cette étape permet de s'orienter vers un graphe simple⁶⁶ compte tenu de l'existence d'une seule relation entre les entités. Néanmoins, la modélisation de graphe plus sophistiquée (bipartis, multigraphe ou multi-niveaux) peut être envisagée pour représenter ce contexte d'énonciation. De même, un graphe statique a été privilégié pour simplifier l'analyse des données à un instant *T*. Les données multidimensionnelles et évolutives peuvent induire une modélisation sous forme de graphes dynamiques plus adaptée à ce contexte d'énonciation.

- Représentation visuelle

La représentation visuelle consiste à définir une légende visuelle (ou encodage visuel) sur les éléments du graphe pour distinguer les types de personnes, la nature et la densité des relations entre elles. La représentation visuelle appliquée est récapitulée dans le tableau suivant :

	Type	Forme géométrique	Coloration
Noeud	Participant	Type Ellipse	Bleu
	Formateur	Type Ellipse	Rouge
	Organisation	Type Diamant	Vert
	Evènement	Type Triangle	Gris
Relation	Intra-organisation	Lien solide	Bleu
	Inter-organisation	Lien en pointillé	Gris
	Partenariat	Lien solide	Noir
	Animation	Lien solide	Gris
	Emploi	Lien solide	Vert

Tableau 4.3 - Représentation visuelle (forme géométrique et coloration) des nœuds et relations

Les composantes graphiques visent à identifier les nœuds et relations d'une façon instantanée pour une interprétation humaine. L'intensité des couleurs⁶⁷, les formes géométriques, la densité des liens sont choisis pour minimiser l'effort cognitif de l'utilisateur. Selon (Bertin, J., 1977) repris par (Hurter, C. 2010), l'humain a la capacité de distinguer plus facilement certaines variables visuelles (position, taille, valeur, couleur, orientation, grain ou forme). De nombreuses

⁶⁶ Graphe simple : une seule arête peut exister entre deux nœuds

Multi-graphe : plusieurs arêtes peuvent exister reliant une même paire de nœuds

⁶⁷ Teinte basique soit rouge, vert, bleu et noir

combinaisons visuelles sont possibles mais ne favorisent pas pour autant une perception et une signification visuelle des données. Dans notre cas d'application, le formateur est caractérisé par la couleur rouge pour mettre en évidence son rôle central au sein du groupe. Ainsi, la perception d'un point rouge parmi un ensemble de points bleus est immédiate. De même, l'intensité des liens vise à caractériser le type de relations : le lien solide vise à mettre en évidence la relation intra-organisationnelle qui est par nature permanente tandis les liens en pointillés visent à mettre en évidence une relation inter-organisationnelle par nature occasionnelle entre les personnes.

A titre d'exemple, les figures suivantes présentent des exemples de représentations graphiques unitaires liées aux principales tâches de recherche simple des données :

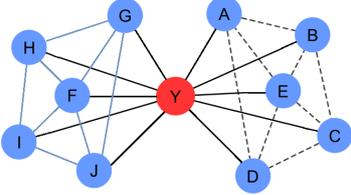
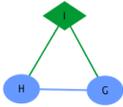
 <p>(a)</p>	<p style="text-align: center;">Formateur</p> <p>Représentation graphique de deux groupes (ou clusters) :</p> <ul style="list-style-type: none"> -Le formateur Y caractérisé par un nœud rouge ayant animé un groupe de formation organisé selon un mode inter-organisationnel incluant des participants caractérisés par des nœuds bleus (A, B, C, D, E) et un second groupe de formation organisé selon un mode intra-organisationnel (F, H, I, J, G) incluant des participants caractérisés par des nœuds bleus. <p>Les relations de type « Intra-organisationnel » sont représentées par une ligne complète en bleu tandis que les relations de type « Inter-organisationnel » sont représentées par une ligne pointillée en gris.</p> <p>Les relations de type « Partenariat » sont représentées par une ligne complète en gris.</p>
 <p>(b)</p>	<p style="text-align: center;">Evènement (ou session de formation)</p> <p>Représentation graphique ou « dyade »⁶⁸ incluant :</p> <ul style="list-style-type: none"> -Un formateur (G) qui anime un évènement de formation (H)
 <p>(c)</p>	<p style="text-align: center;">Organisation</p> <p>Représentation graphique ou « triade »⁶⁹ incluant :</p> <ul style="list-style-type: none"> -Les participants (H et G) caractérisés par des nœuds bleus Les relations de type « Intra-organisationnel » sont représentées par une ligne complète en bleu. -Une organisation spécifique (I) caractérisé par un nœud vert différencié sous une forme géométrique de type « Diamant » et un lien solide et vert qui caractérise la relation d'emploi avec les nœuds H et G

Figure 5.1 - Exemples des représentations graphiques de type descriptives selon la tâche de recherche avec (a) recherche d'un formateur, (b) recherche d'un évènement (ou session de formation) et (c) recherche d'une organisation

Ainsi, les attributs du graphe sont encodés essentiellement selon une représentation visuelle (ou légende visuelle) et ne font pas l'objet d'un calcul selon les propriétés topologiques du graphe. La visualisation repose sur un algorithme de dessin fondés sur les forces (*Force-based* ou *Force-directed algorithms*) inclut dans la librairie utilisée⁷⁰. Cette technique vise principalement à

⁶⁸ Sous-graphe composé de deux nœuds

⁶⁹ Sous graphe composé de trois nœuds

⁷⁰ Bibliothèque Python « Networkx ». Par défaut, algorithme de layout de force nommé « Spring layout ».

faciliter la visualisation des graphes pour agencer les nœuds du graphe d'une façon optimale dans l'espace (Fruchterman TM, 1991). Le résultat en sortie consiste à proposer une configuration acceptable selon un rendu visuel jugé « esthétique » tout en respectant le critère d'équilibre du système de forces du graphe : les nœuds les plus connectés sont placés au centre de la figure tandis que les nœuds les moins connectés sont placés en périphérie de la figure. De même, la position des nœuds vise à minimiser le chevauchement des liens pour assurer ainsi la lisibilité du graphe. Dans notre cas d'application, cette technique s'avère adaptée compte tenu que les graphes sont d'une taille réduite.

6.2.2 Visualisation analytique

L'évaluation de la formation vise à analyser les données liées à la satisfaction des participants. Les indicateurs de performances liés à l'activité de formation sont déterminés à partir d'une analyse statistique détaillée dans le chapitre 8 ([Voir section 8.1.1](#)). Les indicateurs et les données associées impliquent des représentations graphiques adaptées pour faciliter l'interprétation statistique des données complexes. Ils sont schématisés à travers les figures suivantes :

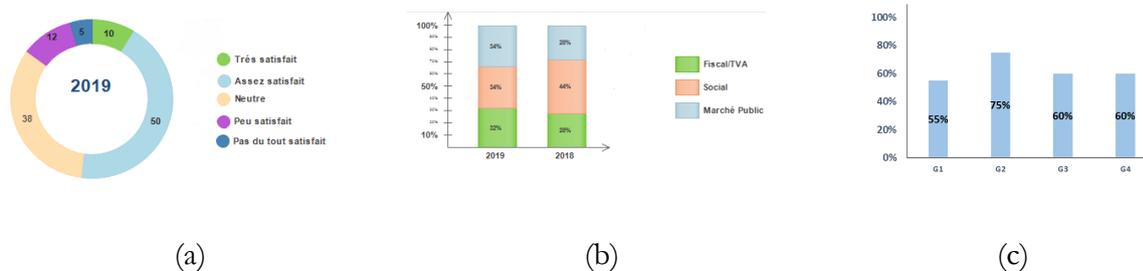


Figure 5.2 - Synthèse des représentations graphiques de type analytique avec (a) un camembert en anneau, (b) un histogramme à barre des sessions de formation et (c) un histogramme à barre des scores de satisfaction par groupes de formation

-Score global de satisfaction ;

La visualisation graphique de type « Camembert en anneau » (ou diagramme circulaire) met en évidence la proportion des scores de satisfaction sur un ensemble de participants selon un indicateur temporel (annuel). Les modalités d'échelles des réponses (« Très satisfait », « Assez satisfait » ...) se différencient par des couleurs alternatives pour mettre en évidence la répartition des valeurs.

-Nombre et thématique des sessions de formation

L'histogramme à barre vise à comparer le nombre total de sessions de formation sous forme de barres verticales en ordonnées (exprimés en %) et selon un axe temporel annuel en abscisse. Les longueurs des barres sont proportionnelles selon le nombre de sessions de formation effectués. Un code couleur supplémentaire met en évidence les thématiques de formation.

-Score de satisfaction globale par groupe de formation ;

L'histogramme à barre vise à afficher les scores de satisfaction globale par groupe de formation et permettre une comparaison rapide des valeurs associées.

6.3 Ergonomie des interfaces-utilisateurs

L'ergonomie des interfaces-utilisateurs vise à faciliter la consultation et l'analyse des données associées à des techniques de visualisation interactive. Compte tenu que les utilisateurs finaux n'ont pas de connaissance particulière dans le domaine de l'informatique, les interfaces-utilisateurs impliquent une ergonomie intuitive et facile d'utilisation. Les interfaces-utilisateurs proposées en adéquation avec les objectifs définis sont les suivantes :

- Tableau de bord principal ;
- Recherche de données ;
- Planification des sessions de formation ;
- Tableau de bord analytique.

L'ensemble des interfaces-utilisateurs utilise un environnement de développement commu⁷¹ (ou *framework*) pour assurer une uniformisation ergonomique standardisée (agencement des pages, bouton...). Pour l'ensemble des fonctionnalités, un dispositif de retour en arrière et un fil d'Ariane (ou *Breadcrumb*) en haut de la page indique à l'utilisateur la position de la page au sein de l'arborescence du prototype d'application.

6.3.1 Tableau de bord principal

Le tableau de bord principal est présenté en page d'accueil du prototype d'application. Cette interface-utilisateur propose une vue descriptive globale et synthétique de l'activité de formation à un instant t en temps réel, ce qui implique une mise à jour automatique des données. En complément, les fonctionnalités liées à la gestion de la formation continue (« Recherche », « Planification », « Evaluation ») sont accessibles à travers un panneau central en haut de page et un panneau latéral gauche afin de permettre une exploration détaillée des données vers des pages tierces.

Au niveau de l'ergonomie, la maquette de l'interface du tableau de bord est agencée selon des « vues multiples coordonnées » incluant des objets visuels. Selon la définition proposée par (Wang S., 2000) repris par (Vuillemot R., 2010), les « vues multiples coordonnées » consistent à assembler plusieurs vues distinctes agencées au sein d'une même interface selon des types de visualisations différentes. Ainsi, cette maquette de tableau de bord présente plusieurs indicateurs et les données de différentes natures associées selon quatre vues :

- Vue « Notifications » ;

Les données sous forme textuelle visent à lister les sessions de formations en cours selon une chronologie hebdomadaire et mensuelle.

- Vue « Carte géographique » ;

Les données spatiales (ou géospatiales) permettent de visualiser les lieux des sessions de formation en cours et ainsi d'identifier la distribution spatiale de l'activité de formation sur le territoire national.

- Vue « Calendrier » ;

Les données temporelles présentent la chronologie calendaire des sessions de formation.

- Vue « Réseau social » ;

Les données sociales proposent une visualisation exhaustive de l'ensemble des groupes de formation existants. La représentation graphique vise à faciliter la visualisation des structures relationnelles des données. Un filtrage de données peut être appliqué pour rechercher des groupes au sein du réseau social selon les critères associés.

⁷¹ Bootscrap



Figure 5.3 - Maquette du tableau de bord principal avec vue titrée (a) « Notification », (b) « Réseau social », (c) « Carte de France » et (d) « Calendrier »

Ainsi, l'agrégation et l'organisation des données hétérogènes permettent à l'utilisateur (ou gestionnaire de formation) une exploration simultanée de l'activité de formation en minimisant la charge cognitive. Des principes de conception visuelle complémentaires sont appliqués pour faciliter la consultation du tableau de bord dont le « défilement vertical » (ou *scroll*) réduit pour éviter une visualisation des données fragmentées ainsi qu'une ergonomie graphique minimaliste pour faciliter la lecture des données.

6.3.2 Interface de recherche de données

Les fonctionnalités « Recherche » et « Planification » s'appuient sur une interface de recherche pour faciliter la formulation de la requête. Les techniques de filtrage vont permettre à l'utilisateur de restreindre des résultats selon les valeurs sélectionnées. Deux techniques de filtrage sont prévues : le filtrage par liste de champ et le filtrage par pondération directe.

- Filtrage par liste de champs

La formulation de la requête *via* un filtrage par liste de champs a pour but de restreindre les résultats selon une liste de critères préalablement définie et détaillée dans le chapitre suivant (Voir section 7.2).

La fonctionnalité « Recherche » s'appuie sur une interface de recherche basée essentiellement sur une liste de champs associée à un menu déroulant pour sélectionner le(s) valeur(s) choisie(s). De même, l'interface de recherche pour la fonctionnalité « Planification » distingue les différentes dimensions du groupe de formation (Groupe, Formation et Logistique) agencées sous forme de formes de colonnes.

- Filtrage par pondération directe

Une fonction complémentaire est disponible lors de la formulation de la requête pour la planification d'une session de formation. L'utilisateur peut soumettre une requête en utilisant une pondération directe⁷² pour certains critères. Ce mode de filtrage permet une recherche exploratoire parmi l'ensemble des solutions potentielles.

⁷² « Secteur », « Typologie du profil », « Statut » et « Historique de formation », « Région » et « Département »

Par défaut, aucune valeur de pondération n'est prédéfinie pour l'ensemble des critères et l'absence de valeur équivaut à un choix indifférent. L'utilisateur peut émettre un pré-ordre partiel⁷³ sur plusieurs critères en attribuant à un critère C_1 un poids plus important que le critère C_2 selon une échelle de mesure prédéfinie de 1 à 5 (la valeur 1 correspond à une valeur prioritaire par rapport à un critère selon une valeur 2) dont la figure suivante présente un exemple suivant :

Figure 5.4 - Exemple de saisie d'une valeur de pondération pour les critères « Secteur » et « Statut »

Cette méthode de classement consiste à prioriser les critères de recherche selon les préférences de l'utilisateur. Par exemple, la valeur d'une pondération associée à la variable « Région » vise à organiser un groupe composé de participants proches géographiquement sans pour autant déterminer d'une manière exacte le lieu de formation. En comparaison avec un filtrage par liste de champs, l'obligation de satisfaire le(s) critère(s) n'est pas impératif mais indicatif.

6.3.3 Interface de visualisation interactive

Les interfaces de visualisation interactive visent à présenter les résultats sous forme de représentation graphique selon une identification visuelle préalablement définie ([Voir section 6.2.1](#)).

La visualisation des résultats est générée selon une vue distincte associée aux critères de recherche en panneau latérale pour ne pas surcharger visuellement la lecture des résultats. L'interface de visualisation est « zoomable »⁷⁴ permettant ainsi d'obtenir une visualisation détaillée au sein du graphe. Pour chaque nœud, un dispositif de pointage détaille une information *via* une étiquette selon le « focus » de l'utilisateur. Cette technique de navigation appelé « *Focus+Contexte* » initiée par (Leung Y., 1994) vise à faciliter l'identification et l'interprétation d'une information sans déformation géométrique au sein d'un graphe.

Ce type de visualisation implique que les résultats associés doivent être principalement des graphes restreints afin de favoriser la lisibilité des données. En effet, la minimisation du nombre de croisements, l'homogénéité des arêtes et la minimisation de la taille des dessins sont des critères visuels pour faciliter l'interprétation du graphe.

- Présentation d'une recherche simple

Deux types de visualisation diffèrent selon la tâche de recherche : le résultat d'une recherche simple (participant, organisation et événement) est restitué sous forme d'un réseau complet tandis que le résultat d'une recherche d'un formateur est restitué sous forme d'un réseau égo-centré.

Réseau complet. Le réseau complet délimite un ensemble de personnes et les relations associées d'une manière exhaustive selon un critère déterminé. La représentation graphique de type nœud-

⁷³ Type structure préordinaire

⁷⁴ Zommable User Interface (ZUI)

lien implique que chaque nœud est directement lié à tous les nœuds du graphe. La conception de ce type de visualisation implique d'avoir à disposition les données d'une manière exhaustive.

Cette approche consiste à sélectionner une entité indépendamment les unes des autres et présente l'avantage de mettre en évidence le voisinage socio-professionnel de la personne(collaborateur(s) et organisation). La figure suivante présente un exemple de résultat d'une recherche simple d'un participant caractérisé par un nœud au centre du graphe :

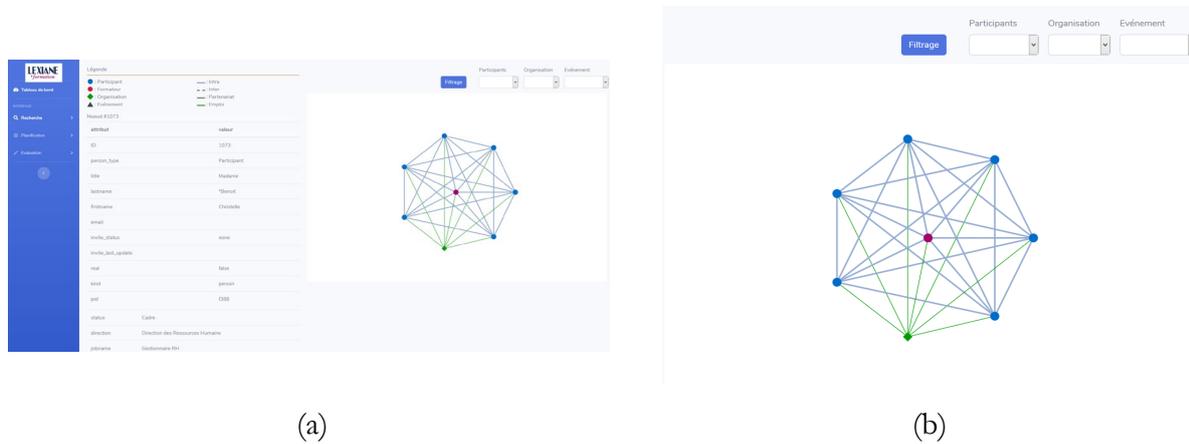


Figure 5.5 - Interface d'une recherche simple d'un participant avec (a) Affichage du résultat avec légende visuelle et vue des critères de recherche et (b) Vue de visualisation interactive avec filtrage des données

Une approche « top-down »⁷⁵ est appliquée pour cette tâche de recherche c'est-à-dire que le résultat génère initialement une visualisation du graphe entier. Par la suite, l'utilisateur peut interagir *via* un filtrage de données en haut et à droite de l'interface. Cette technique d'exploration interactive complémentaire permet d'obtenir une vue partielle du graphe selon les critères sélectionnés afin de permettre à l'utilisateur de recentrer son exploration visuelle. Comme présenté dans la figure suivante, un filtrage selon la valeur « Intra-organisation » permet essentiellement de visualiser les relations internes du nœud sélectionné en supprimant le nœud « Organisation » caractérisé sous forme de diamant.



Figure 5.6 - Zoom d'une vue partielle du graphe *via* un filtrage dynamique avec (a) Graphe complet et non-filtré, (b) Filtrage dynamique (c) Graphe filtré

⁷⁵ Niveau d'abstraction *via* trois types de filtrage de données : approche « top-down » (visualisation commence avec le graphe entier), une approche « bottom-up » (visualisation à partir de la sélection d'un nœud) et une approche hybride « middle-out » combinant les approches citées précédemment (Von Landesberger, T., 2011).

Réseau égo-centré. La visualisation d'un réseau « égo-centré » (*egocentric network* ou *ego network*) se focalise à l'échelle d'un seul individu (appelé « égo »). Ce type de visualisation est particulièrement adapté pour le formateur, qui est un acteur central au sein du groupe de formation. La visualisation selon une représentation de type nœud-lien permet de consulter l'activité de formation et le capital social du formateur théorisé par (Burt R., 2017).

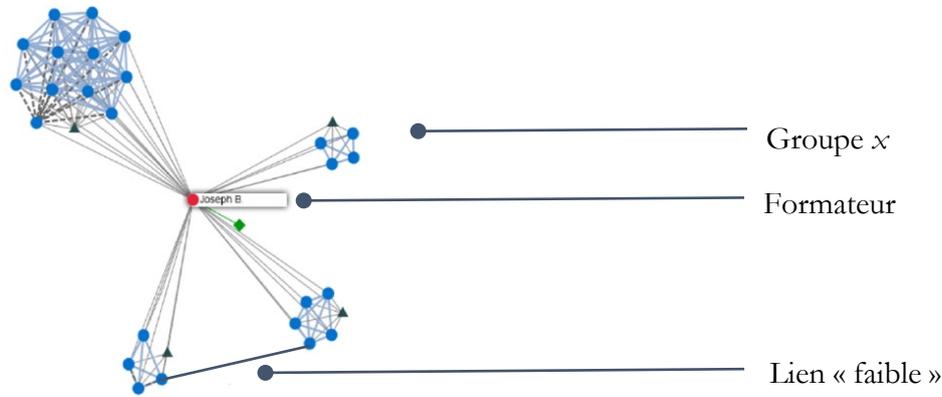


Figure 5.7 - Représentation graphique de type « égo-centré » d'une recherche simple d'un formateur

En premier lieu, ce type de visualisation permet d'identifier aisément l'activité de formation d'un formateur d'une manière instantanée. En prolongement de la représentation visuelle, ce graphe peut faire l'objet d'un partitionnement pour créer des groupes de formation autour d'une thématique de formation. Dans le cadre de notre contexte d'application, le formateur est qualifié comme étant un expert dans un domaine spécialisé et par conséquent porteur d'une thématique de formation d'une manière constante. Dans le cadre théorique de l'ARS, la position centrale du formateur peut être considéré au sein du graphe comme étant un nœud intermédiaire⁷⁶ entre les différents participants des groupes de formation qui n'ont pas de liens directs sans l'entremise de celui-ci. Ainsi, les nœuds non-adjacents ayant des liens relationnels dit de « faible » intensité théorisé par (Granovetter M., 1973) sont facilement détectable afin d'émettre des déductions sur les intérêts de la formation. Par exemple selon la représentation graphique ci-dessus, la visualisation égo-centrée du réseau du formateur permet de déduire que l'ensemble des nœuds au sein d'un groupe partage des intérêts de formation communs avec les nœuds voisins malgré l'absence de lien « direct » entre les participants. Cette approche méthodologique permet d'utiliser les propriétés du graphe existant afin de créer des nouveaux groupes de formation. Parmi les méthodes de partitionnement existantes, la méthode « divisive » développée par (Girvan M., 2002) consiste à appliquer un découpage itératif du graphe en supprimant le lien ayant la plus forte centralité, puis en exploitant les liens « faibles » comme étant des ponts entre les nœuds pour créer un nouveau groupe de formation. A titre d'exemple, cette approche méthodologique d'optimisation combinatoire basée sur les graphes est décrite selon le processus suivant qui propose un nouveau groupe de formation selon un partitionnement manuel et arbitraire :

⁷⁶ Centralité d'intermédiation

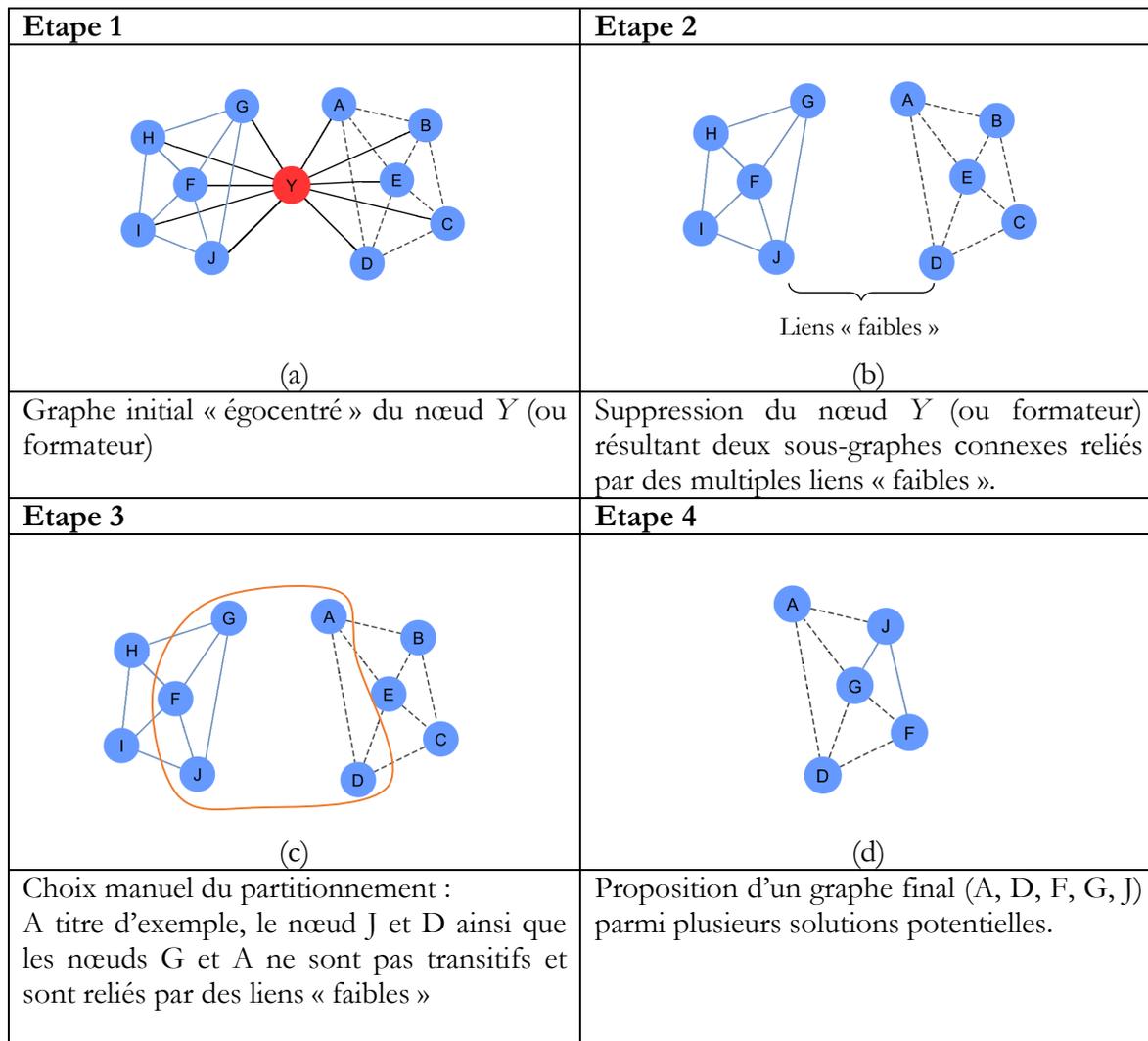


Figure 5.8 - Illustration des étapes du partitionnement selon une méthode « divisive » pour la création d'un groupe de formation

Ainsi, les liens « faibles » permettent d'exploiter des zones fermées et éviter une redondance d'affectation de participants lors de la composition des nouveaux groupes de formation.

- Présentation des groupes de formation

Les groupes de formation générés selon la requête formulée sont présentés selon les interfaces-utilisateurs suivantes :

- Liste de résultats ;
- Finalisation du groupe de formation.

Liste de résultats. L'interface-utilisateur présente les groupes de formation potentiels sous forme d'une liste ordonnée associée à une représentation graphique de type nœud-lien. Le nombre est limité à dix et classé selon un score de pertinence par ordre décroissant. Le type de visualisation est un réseau complet, ce qui implique que les nœuds sont directement liés entre eux. En comparaison avec la présentation des résultats d'une recherche simple, le nœud « Organisation » n'est pas associé au nœud « Personne ». Cette information est disponible *via* une étiquette associée au nœud selon une technique de visualisation « *Focus+Context* » tandis que la nature de la

relation entre les nœuds (inter/intra-organisationnelle) est caractérisée par une représentation visuelle préalablement déterminée (Voir 6.2.1). Ainsi, la représentation graphique est simplifiée pour éviter une redondance et une surcharge visuelle lors de l'exploration du graphe comme présenté dans la figure suivante :

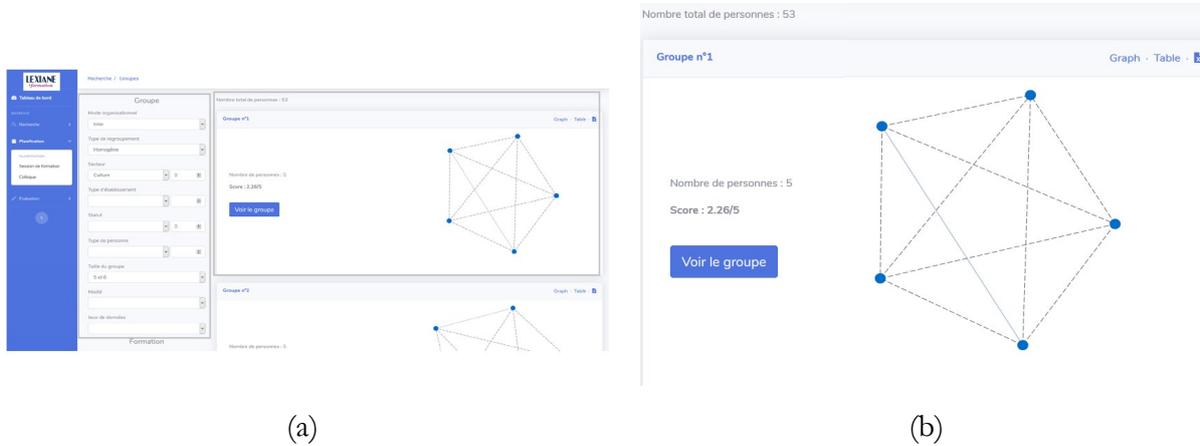


Figure 5.9 - Interface avec liste des résultats incluant (a) Liste de résultats avec détails des critères de recherche *via* un menu panneau latéral gauche et (b) Zoom du résultat n°1 associé à un score d'évaluation sous forme de représentation graphique par défaut.

En complément, la liste des personnes est également consultable sous forme d'un tableau. Il est à noter que des participants peuvent être communs parmi la liste des groupes de formation potentiels (exemple : le groupe n°1 inclut 5 personnes dont 3 sont communs au groupe n°2).

Finalisation du groupe de formation. Après la sélection d'un groupe de formation parmi la liste de résultats, l'étape suivante consiste à proposer aux participants le projet de formation. Compte tenu qu'un nombre minimal de participants conditionne la finalisation du projet de formation, des informations additionnelles liée au statut de présence du participant sont intégrées au niveau du nœud du graphe. La représentation visuelle des nœuds distingue les différents statuts de présence selon un code graphique universel récapitulé dans le tableau suivant :

	Coloration	Statut de présence	Détails
Noeud	Vert	Oui	Confirmation de présence au projet de formation
	Orange	En attente	Attente de confirmation de présence au projet de formation
	Rouge	Non	Refus de présence au projet de formation

Tableau 4.4 - Représentation visuelle des statuts de présence des participants

Ainsi, l'utilisateur peut visualiser l'état d'avancement du projet de formation selon la disponibilité des participants. A titre d'exemple, la figure suivante présente un exemple d'une vue globale de l'état d'avancement du groupe de formation : deux participants ayant confirmé leurs présences, deux participants en attente de confirmation et un refus d'un participant⁷⁷ :

⁷⁷ Par confidentialité, les données indiquées sont fictives

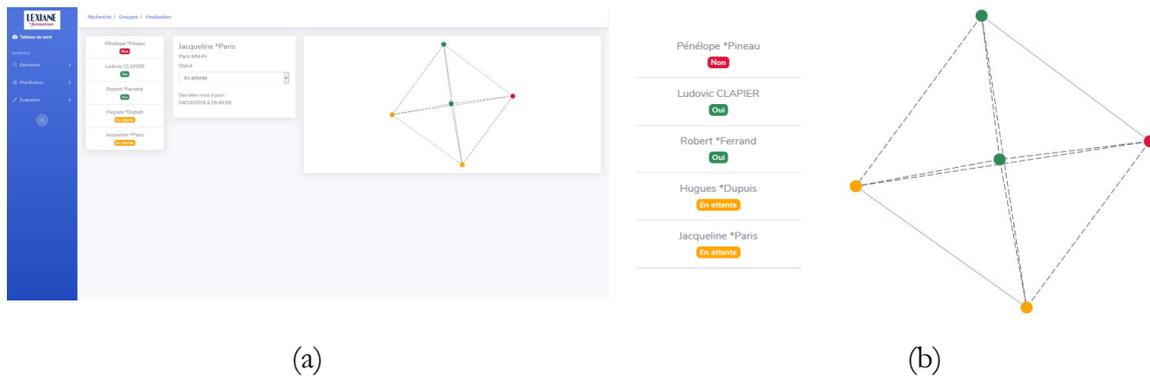


Figure 5.10 - Etape de la finalisation du groupe de formation avec (a) Interface de l'état de l'avancement d'un groupe et (b) Zoom de la représentation graphique des statuts de présence des personnes avec légende visuelle selon un code graphique universel

A ce stade du développement du prototype, les résultats sont enregistrés et mis à jour manuellement par l'utilisateur. Les réponses positives des personnes sur l'ensemble des dix groupes de formation générées peuvent être fusionnées afin de respecter le critère de la taille du groupe.

6.3.4 Tableau de bord analytique

Le tableau de bord analytique vise à permettre à l'utilisateur de quantifier et d'analyser des données liées à l'activité de formation. Il peut être qualifié comme étant un tableau de bord secondaire compte tenu de son emplacement au sein de l'arborescence au sein du prototype d'application. L'ensemble des données est synthétisé selon deux niveaux d'analyse : l'analyse globale de l'activité de formation et l'évaluation de l'activité du formateur.

- Analyse globale de l'activité de formation

Ce niveau d'analyse vise à quantifier l'activité de formation à partir des données des questionnaires de satisfaction remis aux participants en fin de session de formation. Les différentes versions du questionnaire utilisés sont centralisées afin d'assurer la traçabilité des modèles d'évaluation. Les données sont représentées sous forme de visualisation de type analytique pour faciliter l'interprétation humaine ([Voir section 6.2.2](#)). Ainsi, les résultats obtenus visent à soutenir la prise de décision stratégique et de mener des réajustements nécessaires pour l'amélioration des sessions de formation à venir.

- Analyse de l'activité du formateur

Dans un contexte de formation continue, le formateur devient le garant de la qualité de la session de formation. Ainsi, ce niveau d'analyse synthétise l'activité de formation du formateur à partir des indicateurs suivants :

- Profil professionnel du formateur

Le profil du formateur est associé à son identité professionnelle incluant ses données personnelles (nom, prénom, genre) et son domaine d'expertise et pédagogique (fonction professionnelle).

-Score de satisfaction global par groupe de formation

Cette vue détaille le score de satisfaction global associé à chaque session de formation. Le pourcentage de scores de satisfaction agrégés des groupes de formation se présente sous forme de tableau. En complément, une vue détaille les scores de satisfaction émis individuellement par les participants. En complément, une représentation graphique sous forme d'un histogramme à barre vise à comparer visuellement les scores de satisfaction par groupe de formation. Contrairement aux indicateurs élémentaires du tableau de bord principal, les indicateurs de l'activité de formation font l'objet d'un calcul pour obtenir des indicateurs agrégés. Ils sont déterminés à partir de l'analyse statistique détaillée dans le chapitre suivant. Les indicateurs sont de type « statique » c'est-à-dire que les résultats bénéficient d'un score défini et invariable.

Au niveau de l'ergonomie, la maquette de l'interface du tableau de bord analytique propose une vue globale et synthétique de l'ensemble des indicateurs. Selon le même principe du tableau de bord principal, l'emplacement des composants visuels et les données associées sont agencées selon des « vues multiples coordonnées » comme présenté selon la figure suivante :

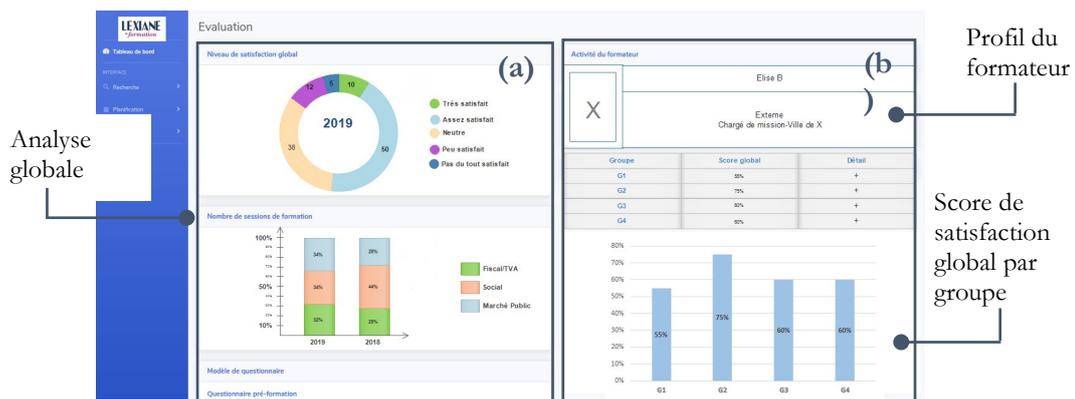


Figure 5.11 - Maquette du tableau de bord analytique avec (a) Vues des indicateurs globaux des sessions de formation, (b) Vue du profil et de l'activité du formateur

Au niveau technique, cette maquette unifie les différentes vues et indique le positionnement des composants visuels pour une implémentation technique à venir.

La maquette suivante présente le niveau de détail d'un groupe de formation qui applique la représentation visuelle citée précédemment.

6.4 Implémentation technique

6.4.1 Type de prototype

Le développement d'un système d'information est une étape complexe compte tenu des contraintes humaines, financières et juridiques. Ces exigences impliquent d'appliquer une méthode de conception du système d'information selon une démarche itérative et incrémentale. Cette approche consiste à développer et tester progressivement les différentes versions successives du prototype d'application qui serviront de point de référence pour optimiser l'étape de production lors du passage à l'échelle de l'application. Dans le cadre de notre projet, cette étape vise principalement à bénéficier d'un premier retour d'expérience sur l'implémentation fonctionnelle et technique de l'environnement de développement (choix des langages de programmation, évaluation et estimation des performances matérielles ...).

6.4.2 Fonctionnement global

Cette section décrit les différentes étapes du processus fonctionnel du prototype d'application schématisé selon la figure suivante :

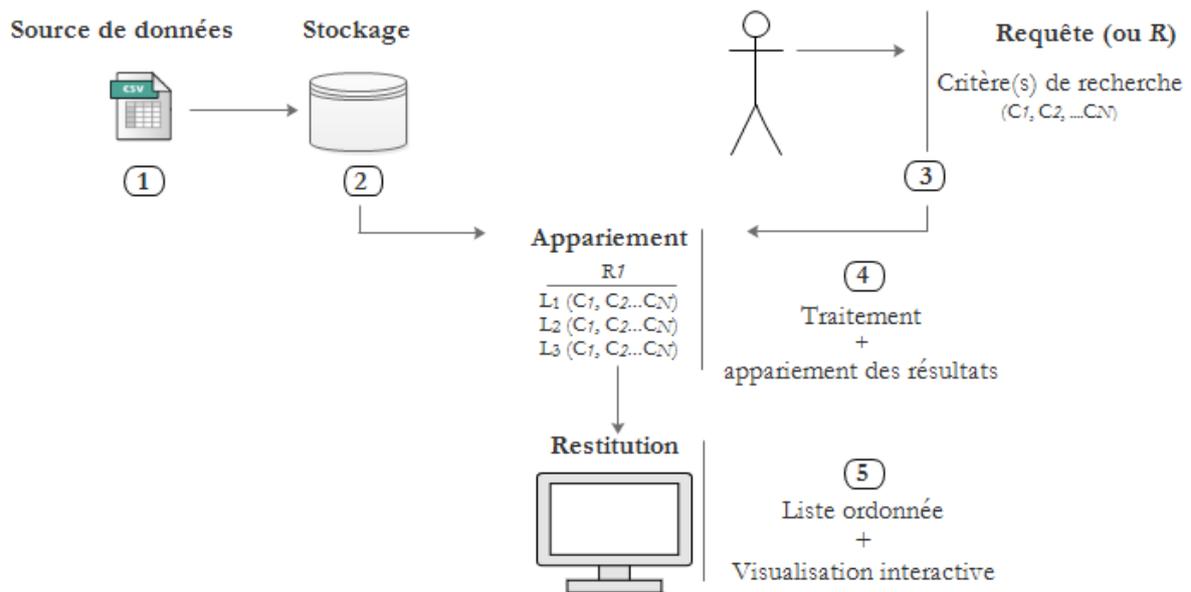


Figure 5.12 - Processus et fonctionnement global du prototype d'application

-Etape 1 : Source de données

Les sources de données en entrée sont multiples et de différentes natures (quantitative et qualitative) ([Voir section 6.1.4](#)). La migration du jeu de données brutes initialement sous forme tabulaire⁷⁸ consiste à s'assurer de la transformation des données vers une modélisation à base de graphe.

-Etape 2 : Modélisation et stockage à base de graphe

Les données sont stockées dans une base de données orientée graphe⁷⁹. Ce système de stockage représente les entités comme des nœuds (ou des sommets) tandis que les relations (ou

⁷⁸ Fichier de type excel

⁷⁹ NoSql

arêtes) sont labellisées pour caractériser les relations entre les entités. Ainsi, l'ensemble des données interconnectées facilite l'extraction des relations entre les nœuds en évitant ainsi la complexité des jointures des modèles relationnels.

L'exemple suivant présente un cas simplifié d'un extrait d'un tableur avec les valeurs associées vers un modèle à base de graphe. Le nœud « Personne » est relié par la relation « employé_par » avec le nœud « Organisation ».



Figure 5.13 - Schéma simplifié d'un tableur vers un modèle à base de graphe

La principale contrainte est la mise à jour des données notamment liées à la mobilité professionnelle des personnes. Cette étape implique une veille régulière et une mise à jour manuelle des variables (ou caractéristiques) des personnes pour s'assurer de l'exactitude des valeurs.

-Etape 3 : Requête de recherche

Le traitement de la recherche repose essentiellement sur un processus d'interaction de l'utilisateur avec le système à travers une interface de recherche, contrairement à un système de recommandation qui anticipe les besoins de l'utilisateur. Ainsi, la formulation de la requête (noté *R1*) est spécifiée explicitement par l'utilisateur selon des critères de type mono ou multicritère (noté *C1, C2, ... Cn*) en étape 3 dans la figure précédente.

-Etape 4 : Traitement et appariement des résultats

En fonction de la requête spécifiée, la fonction d'appariement s'appuie sur un principe de correspondance entre les critères de recherche avec les valeurs stockées au sein de la base de données. En amont, le traitement de la requête diffère selon la nature de l'élément (ou item) de type individuel ou groupal. La recherche simple résulte d'éléments (ou items) individuels⁸⁰ tandis que l'aide à la planification résulte d'éléments (ou items) groupaux :

Item individuel. La tâche liée à la recherche simple consiste à restituer une liste d'items individuels (personne, organisation, formateur, événement de formation) selon la requête formulée par l'utilisateur. Le processus de traitement repose sur une mise en correspondance puis une extraction des valeurs à partir de l'indexation de la base de données.

Item groupal. La tâche liée à l'aide à la planification consiste à restituer une liste d'items de groupes de formation potentiels selon la requête formulée par l'utilisateur. A l'inverse de la recherche d'items individuels, le traitement et la génération d'items groupaux impliquent un traitement algorithmique spécifique et plus précisément un algorithme génétique détaillé dans le chapitre suivant.

-Etape 5 : Classement et présentation des résultats

La dernière étape consiste à restituer les résultats sous forme de liste ordonnée (noté *L1, L2, L3...*). Concernant la tâche de recherche simple, le système génère une liste partielle et

⁸⁰ Terme usuel dans le domaine des systèmes de recommandation pour indiquer d'un élément à recommander

ordonnée d'items individuels en fonction de l'exploration de l'indexation de la base de données. Cette approche traditionnelle diffère du classement des items groupaux générés pour l'aide à la planification qui sont triés selon un score d'évaluation. Par défaut, la présentation des résultats est sous forme d'une représentation graphique de type nœud-lien. Un panneau latéral à gauche permet de lister les critères spécifiés lors de la formulation de la requête par l'utilisateur, qui peut modifier sa recherche selon ses besoins.

6.4.3 Architecture technique

L'architecture technique décrit l'implémentation et les outils associés pour permettre le fonctionnement global du prototype d'application.

- Architecture globale

L'architecture technique repose sur un *framework* web en interactions avec diverses composantes associées synthétisé selon la figure suivante :

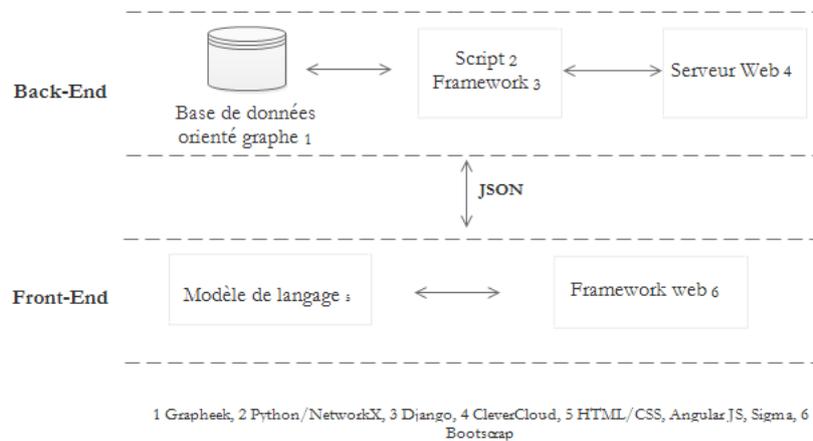


Figure 5.14 - Schéma simplifié de l'architecture technique

Les deux couches du système d'information (couramment nommé « *Back-End* » et « *Front-End* ») interagissent selon les choix de l'utilisateur :

- Fournisseur de service Web (ou *Back-End*)

La partie arrière-plan (couramment nommée « *Back-End* ») est l'ensemble du code source exécuté côté serveur (modélisation conceptuelle, préparation des données...). Le cadre de développement web, les modules techniques et les bibliothèques associées sont choisis selon les prédispositions du langage, leurs utilisations globales et selon la communauté d'utilisateurs actifs.

Au niveau applicatif, la modélisation conceptuelle des données du système a pour but de représenter graphiquement les entités et les liens qui existent entre chacune d'elles. La restriction du modèle conceptuel implémente un ensemble minimal d'éléments nécessaires tout en assurant une couverture de l'ensemble des besoins. La complétude et la simplicité sont les recommandations pour assurer par la suite la performance du système d'application. La modélisation conceptuelle est présentée selon une représentation graphique en Annexe 2.

-Client web (ou *Front-End*)

Le développement web frontal (couramment nommée « *Front-End* ») est l'ensemble du code exécuté par le navigateur pour la génération des interfaces utilisateurs. Le stockage et la synchronisation des données sont assurés par l'intermédiation d'une application web⁸¹ selon le format JSON⁸². L'avantage d'un service web d'échange permet une simplicité d'utilisation pour l'installation et la mise à jour des données accessibles *via* le réseau internet.

La gestion des codes-sources est assurée à l'aide d'un logiciel de gestion de versions décentralisé⁸³.

La description des modules techniques de développement est détaillée en Annexe 3.

- Hébergement

L'hébergement du prototype d'application s'appuie sur une architecture orientée « service » c'est-à-dire selon une configuration virtualisée à travers des conteneurs logiciels⁸⁴. Cette approche permet de déployer un prototype d'application avec un équipement minimum requis en évitant une installation d'une couche matérielle lourde basées sur des machines virtuelles.

- Cycle de développement

Le développement du prototype d'application est partiel compte tenu des contraintes techniques. A ce stade du cycle de production du prototype d'application, la finalisation technique n'est pas l'objet de ce projet de recherche. L'automatisation des fonctionnalités vise à être implémenté selon une bibliothèque de fonctions accessibles *via* des API⁸⁵ lors du passage à l'échelle du prototype d'application (implémentation des composants du tableau de bord principal, automatisation du statut de présence des participants *via* une connexion d'une messagerie⁸⁶). De même, la taille de la base de données est limitée dans le volume des données pour générer des données sous forme de visualisation. Au préalable, l'intégration de l'ensemble des composants technique nécessite une évaluation supplémentaire au niveau des tests de performance pour assurer la phase de test de montée en charge (temps de chargement et de réponse, gestion de la mémoire...).

6.5 Conclusion

Cette proposition de prototype d'aide à la gestion de la formation continue vise à rechercher et à explorer les données à travers des tableaux de bord multiples dédiées à l'activité de formation. D'une part, le tableau de bord principal permet une consultation et une visualisation descriptive d'une manière instantanée de l'activité de formation et un accès aux fonctionnalités associées au système d'application. D'autre part, le tableau de bord analytique dédié à l'évaluation des sessions de formation vise davantage une quantification et une analyse plus approfondie des données pour l'aide à l'interprétation et à la prise de décision stratégique. L'aide à la planification des sessions de formation est une également une fonctionnalité centrale au sein du prototype d'application qui implique un traitement et un déploiement spécifique détaillé dans le chapitre suivant.

⁸¹ Application Programming Interface/Google Drive

⁸² Javascript Object Notation

⁸³ GitHub

⁸⁴ Docker

⁸⁵ Widget calendrier et géolocalisation

⁸⁶ Protocole XMPP (Extensible Messaging and Presence Protocol)

L'exploration des données à travers des techniques de visualisation vise à faciliter l'interprétation de l'utilisateur (ou gestionnaire de formation). Dans le cas spécifique de la formation pour adulte, le statut socio-professionnel du participant est un marqueur central du fait qu'il peut influencer la perception et les interactions sociales entre les membres du groupe. Ainsi, cette information est mise en évidence à travers une représentation graphique nœud-lien qui présente les avantages suivants :

- Différencier le statut et les multiples relations des participants selon une norme professionnelle (relation inter et intra-organisationnelle) ;
- Identifier le voisinage du réseau socio-professionnel du participant.

Pour aller plus loin, dans le cadre théorique de l'ARS la représentation graphique permet également d'émettre des déductions sur les intérêts de formation des participants. Cette méthode permet de déduire des « liens faibles » en utilisant la position centrale (ou intermédiaire) du formateur et procéder ainsi à un partitionnement pour créer des groupes de formation.

Au niveau technique, le développement du prototype d'application est réalisé selon une méthode incrémentale c'est-à-dire implémenté par étapes successives. L'environnement et les composantes techniques s'avèrent nécessaire à décrire compte tenu qu'elles déterminent la performance du prototype d'application ainsi que le traitement et l'exécution des résultats à venir. A ce stade, cette version de l'application utilise une infrastructure technique minimale basée sur des conteneurs logiciels dont la souplesse et la facilité du déploiement favorise une première validation du prototype.

Chapitre 7. Aide à la planification par algorithme génétique

Ce chapitre présente une méthode d'aide à la planification en appliquant un algorithme génétique (noté AG par la suite). De type méta-heuristique, cette méthode vise à traiter et générer le(s) groupe(s) de formation potentiel(s) selon les étapes suivantes :

- Description des types de planification en formation continue ;
- Identification des contraintes et des critères de décision spécifiques ;
- Paramétrage et implémentation de l'AG.

7.1 Types de planification en formation continue

7.1.1 Cas d'utilisations

Dans le domaine de la formation continue, la planification des sessions de formation s'opère différemment selon les contraintes suivantes :

- Mode organisationnel qui détermine l'affectation des participants en fonction de son affiliation professionnelle ;
- Initiateur de l'action de formation qui fixe les contraintes et les critères de décision.

- Mode organisationnel

Les sessions de formation peuvent être planifiées selon un mode « Inter-organisationnel » ou « Intra-organisationnel ». Ainsi, la personne physique n'est pas affectée à titre personnel au sein d'un groupe de formation mais en fonction de son affiliation professionnelle.

Ces modes organisationnels répondent à des besoins de formation différents synthétisés selon le tableau suivant :

Mode organisationnel		
Catégorie	Inter-organisation	Intra-organisation
Pédagogique	Contenu pédagogique	
	Rédaction du contenu pédagogique par le formateur en fonction de son domaine d'activité et des objectifs pédagogiques prédéfinis	Rédaction et adaptation du contenu pédagogique selon les besoins de formation spécifiques de l'organisation. Personnalisation du contenu pédagogique liée à une intermédiation avec une organisation unique
Social	Composition	
	Participants issus d'organisations différentes	Participants issus d'une même organisation
	Taille	
	Taille fixée par l'organisme de formation en fonction de la demande	Taille fixée par l'entreprise commanditaire
	Collaboration/Interaction	
Interactions sociales visant à l'échange d'expériences de bonnes	Interactions sociales visant à répondre à des problématiques spéci-	

	pratiques professionnelles liées à un domaine d'expertise spécifique	liées à l'organisation Renforcer le sentiment d'appartenance en consolidant les relations professionnelles entre collaborateurs internes
	Statut professionnel	
	Statut professionnel pouvant freiner ou accentuer la collaboration (comparaison sociale, relation non hiérarchique)	Statut professionnel hiérarchique et lien de subordination peuvent favoriser une réserve et freiner la collaboration
Logistique	Dates	
	Fixé par l'organisme de formation en fonction des disponibilités des participants et formateurs	Fixé par l'entreprise (ou organisation) en fonction des contraintes internes
	Géographique	
	Lieu géographique fixé par l'organisme de formation en fonction des déplacements géographiques possibles des participants	Lieu géographique fixé librement par le commanditaire

Tableau 5.1 - Synthèse des caractéristiques des sessions de formation selon le mode organisationnel

- Initiateur de l'action de formation

Deux types de planification peuvent être distingués selon l'initiateur de l'action de formation :

Organisation (ou entreprise). Ce type de planification à la demande vise à planifier des sessions de formation répondant à des objectifs précis et explicites. Le besoin de formation est formulé par un ou plusieurs commanditaires (organisations/entreprises externes), intéressés par une thématique de formation et soumis à des contraintes spécifiques (budget, durée, zone géographique ...). Ainsi, l'organisme de formation est sollicité en fonction de son offre de formation pour organiser une session de formation (formateur, gestion administrative, logistique...). L'objectif principal consiste à générer des groupes de formation respectant les contraintes et les besoins de formation des commanditaires et d'assurer un nombre minimal de participants.

Organisme de formation. Contrairement à la planification selon la demande d'un ou plusieurs commanditaire(s), la planification de type exploratoire est principalement à l'initiative de l'organisme de formation. Ainsi, la composition du groupe de formation est gérée librement par le gestionnaire de formation qui n'est soumis à aucune contrainte précise liée à une commande de formation. La principale difficulté pour le gestionnaire de formation est qu'il n'a pas à disposition les caractéristiques d'un groupe de formation dit « optimal » qui influencent favorablement les interactions sociales entre les membres du groupe.

7.1.2 Méthode de résolution

Dans notre cas d'étude, la recherche de solutions potentielles s'appuie sur un problème de décision multicritère. En effet, la nature multidimensionnelle du groupe de formation nous incite

à appliquer une méthode de type méta-heuristique et plus précisément, une méthode visant la conversion d'un problème d'optimisation multi-objectif (ou multi-critère) en un problème d'optimisation mono-objectif (ou mono-critère). Selon les types de planification, un principe de traitement commun est appliqué : l'exploration, le traitement et le classement des solutions potentielles en fonction de la requête formulée. Ce type de méthode privilégie davantage la génération de solutions potentielles jugées « satisfaisantes » en écartant une solution unique et exacte. Ainsi, un compromis entre plusieurs critères de décision est privilégié afin d'assurer un temps de traitement raisonnable. Selon les méthodes d'aide à la décision listé par (Horn, J. 1997), ce type de méthode est relative à la méthode à préférence *a priori* c'est-à-dire que les préférences du décideur sont spécifiées avant le lancement du processus de décision multicritère. Ce type de méthode implique de s'appuyer sur des valeurs absolues pour simuler le problème de décision dans l'espace de recherche. Au préalable, cette méthode d'optimisation implique les étapes suivantes détaillées dans les sections suivantes :

-Identification des contraintes et des critères de décision

Cette étape consiste à identifier les contraintes et de lister les critères de décisions à prendre en considération pour planifier une session de formation autour d'un groupe en présentiel.

-Paramétrage de la fonction objective (ou *fitness*) de l'AG

La fonction objective (ou *fitness*) de l'AG vise à maximiser le nombre de solutions admissibles lors du processus d'exploration et de sélection du traitement en écartant la condition d'une solution unique et optimale. Cette étape peut impliquer une incertitude ou une absence d'informations pour définir les valeurs à optimiser lors de la phase d'initialisation. Pour résoudre cette difficulté, la fonction objective de l'AG est partiellement justifiée à partir d'une analyse statistique. Notre approche méthodologique consiste à définir les caractéristiques d'un groupe de formation dit « satisfaisant » à partir des données collectées des questionnaires de satisfaction. Cette étape fait l'objet d'un chapitre dédié pour décrire les méthodes de traitement utilisées et expliciter les résultats obtenus ([Voir chapitre 8](#)).

7.2 Identification des contraintes et des critères de décision

Cette étape consiste à identifier et à prioriser les contraintes et les critères de décision du problème de décision posé.

7.2.1 Type de contraintes

Plusieurs contraintes et critères de décision sont à prendre en considération et peuvent être qualifiés comme étant « strictes » (ou obligatoires) ou « souples » (ou optionnelles). Ainsi, les types de contraintes sont distingués de la manière suivante :

-Contrainte(s) stricte(s)

Les contraintes strictes (ou obligatoires) doivent impérativement respecter un critère de recherche spécifique. Comme indiqué dans le chapitre précédent, l'affiliation professionnelle du participant induit un critère indissociable avec l'organisation (ou entreprise). Par conséquent, le critère lié au « Mode organisationnel » est considéré comme une contrainte stricte. Cette contrainte « stricte » est par nature binaire et correspond au niveau 1 de l'arborescence présenté dans la figure suivante.

Au niveau de l'interface-utilisateur, la nature de cette contrainte implique une valeur obligatoire et fixe. Ainsi, le choix d'un mode organisationnel est impératif selon une liste de champs⁸⁷ et ne peut pas faire l'objet de l'attribution d'une valeur de pondération.

-Contrainte(s) souple(s)

A l'inverse de la contrainte stricte, les contraintes souples (ou optionnelles) sont les critères de décision facultatifs. Le gestionnaire de formation peut spécifier d'une manière interactive la valeur des paramètres de recherche qui n'ont pas nécessairement la même importance. Le choix des valeurs peut être mené *via* une sélection des champs selon une liste préétablie ou en attribuant une valeur de pondération.

7.2.2 Catégorisation des critères de décision

Cette étape consiste à lister et catégoriser d'une manière organisée l'ensemble des critères de décision. Cette arborescence schématisée selon la figure suivante met en évidence les contraintes et les critères de décision :

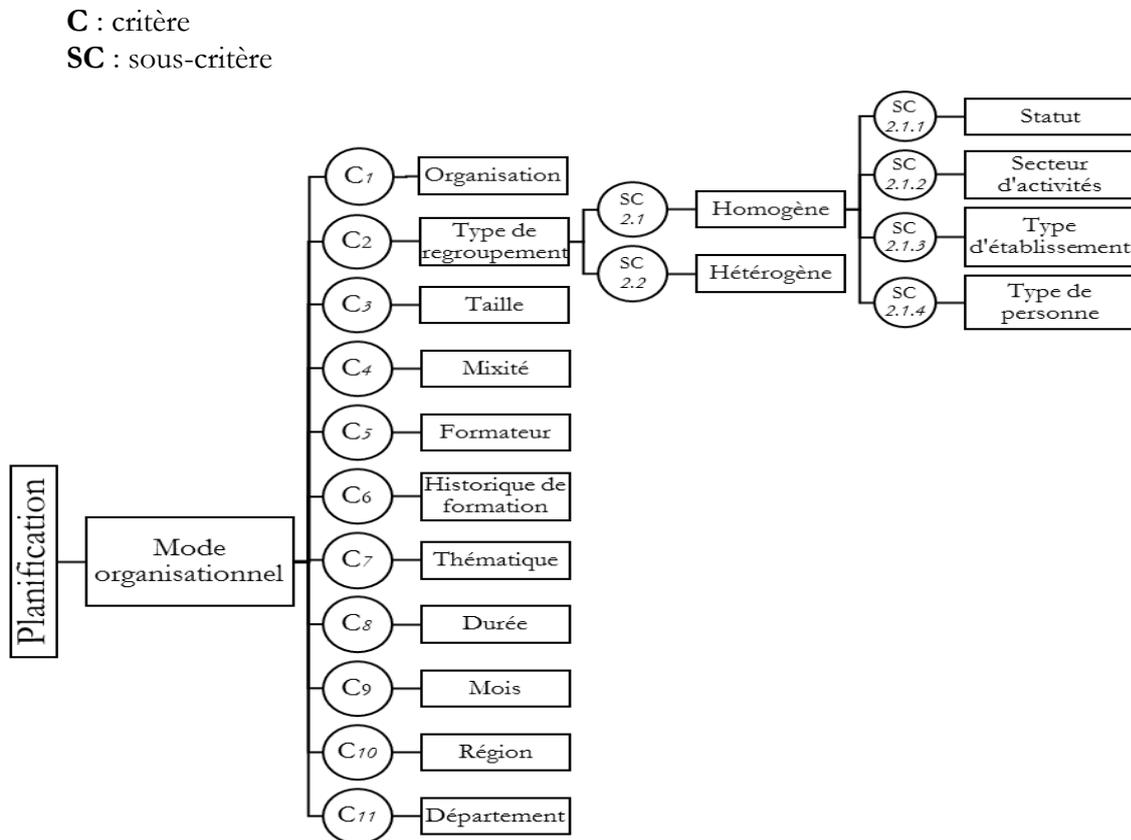


Figure 6.1 - Arborescence des contraintes et des critères de décision

Le niveau 0 correspond à l'objectif du problème de décision à résoudre : la planification des sessions de formation qui se situe au niveau le plus haut de la hiérarchie.

⁸⁷ Valeurs des champs (inter ou intra-organisationnel)

Le niveau 1 correspond à la contrainte stricte du problème de décision : le mode organisationnel dont les valeurs peuvent être « Inter-organisationnel » ou « Intra-organisationnel ».

Le niveau 2, 3 et 4 correspond aux critères et sous-critères de décision. Ils sont au nombre total de 18 et peuvent être catégorisés selon les dimensions suivantes :

Dimension sociale. Cette dimension inclut les critères de décision liés au groupe de formation :

-C₁ : Organisation

Ce critère de décision est obligatoire selon une liste de champs préétablie lors d'une planification d'une session selon un mode « Intra-organisationnel »

-C₂ : Type de regroupement

Ce critère de décision est obligatoire et les valeurs associées peuvent être « Homogène » ou « Hétérogène ». Le type de regroupement selon la valeur « Homogène » repose sur des sous-critères liés aux caractéristiques des participants (type de participant, statut, secteur d'activité et type d'établissement).

-C₃ : Taille

-C₄ : Mixité

Dimension cognitive (ou pédagogique). Cette dimension inclut les critères de décision de nature pédagogique :

-C₅ : Formateur

-C₆ : Historique de formation

-C₇ : Thématique de la formation

Dimension logistique. Cette dimension inclut les critères de décision de nature spatio-temporelle :

-C₈ : Durée

-C₉ : Mois

-C₁₀ : Région

-C₁₁ : Département

Ces critères de décision de nature logistique ont une importance spécifique pour la planification d'un groupe de formation en présentiel. En effet, la disponibilité et la proximité géographique entre les participants facilitent le regroupement physique.

Dans le cas du mode « Intra-organisationnel », le lieu géographique du déroulement de la session de formation est par défaut la domiciliation géographique du siège social de l'entreprise.

Dans le cas du mode « Inter-organisationnel », le lieu du déroulement de la session de formation résulte d'un compromis avec l'ensemble des participants selon un périmètre géographique délimité.

Ainsi, l'identification des contraintes et des critères de décision met en évidence que le mode organisationnel affecte préalablement la composition des groupes de formation et implique des critères de décisions multiples. Selon les propriétés définies par (Roy B., 1997), la cohérence des critères est respectée par son caractère exhaustif et non-redondant. En effet, l'ensemble des critères assure tous les aspects concourants à l'évaluation des solutions potentielles et par conséquent aucun des critères n'est superflu. La difficulté suivante est de combiner et prioriser des critères de nature catégorielle⁸⁸.

Le mode d'interactivité et détails des valeurs des contraintes et des critères de recherche sont détaillées en Annexe 4.

⁸⁸ Critères dont les informations appartenant à une catégorie dont la distance ne peut être établie entre deux valeurs, contrairement aux informations d'ordre numériques et ordinales.

7.3 Algorithme génétique

7.3.1 Fonction-objective

La résolution des problèmes de recherche de nature multicritère et combinatoire oriente notre démarche à appliquer un algorithme génétique. Le principe de traitement vise à explorer et générer une liste de groupes de formation potentiels en prenant en compte les contraintes et les critères de décision préalablement identifiés dans un temps de résolution relativement court.

La fonction objective (couramment appelé *fitness*)⁸⁹ de l'AG consiste à guider les objectifs d'évaluation de l'algorithme selon la stratégie d'exploration souhaitée. Lors du processus de traitement de l'AG, ce paramétrage est calculé lors de l'étape « Evaluation » ([Voir section 7.3.2](#)). Parmi les alternatives existantes, notre démarche consiste à définir une fonction objective selon des caractéristiques du groupe de formation qui vont conditionner les interactions sociales entre les participants. Les valeurs des paramètres sont justifiées partiellement à partir d'une analyse statistique détaillée dans le chapitre suivant ([Voir section 8.4](#)). Ainsi, la fonction objective de l'algorithme génétique vise à optimiser le traitement de résolution selon les valeurs suivantes :

- Type de regroupement « Homogène »

Le groupe de formation est qualifié comme étant homogène selon les caractéristiques suivantes :

-Groupe catégoriel (ou sectoriel)

Le groupe catégoriel (ou sectoriel) consiste à regrouper des participants appartenant à une catégorie socio-professionnelle commune selon les sous-critères « Secteur d'activité » et « Type d'établissement ». Le critère « Profession » n'a pas pu être exploité compte des anomalies relevées lors de la collecte des données ([Voir section 6.1.4](#)).

Au niveau théorique, ce choix est justifié par le fait, que les échanges et l'analyse de pratiques professionnelles sont facilités entre des participants confrontés à des problématiques métiers communs. Par déduction, les participants auront ainsi des intérêts de formation convergents. De plus, l'analyse statistique résulte un score de satisfaction positif lors des groupes de formation selon un type de regroupement « homogène » ([Voir section 8.2.3](#)). Il est à préciser que l'ensemble des participants sont employés dans le secteur public et se différencient selon un domaine d'activité spécifique⁹⁰. Ainsi, l'offre de formation proposée par l'organisme de formation est spécialisée selon les spécificités juridiques des domaines d'activités. Les contenus pédagogiques sont adaptés selon les législations qui peuvent différer selon la nature des activités exercées par l'établissement (par exemple, le droit social dans le secteur hospitalier est soumis à des dispositions différentes aux règles statutaires du secteur universitaire).

-Groupe hiérarchique

Le groupe hiérarchique consiste à réunir des participants ayant le même niveau hiérarchique principalement caractérisé par le critère « Statut ». Il est catégorisé selon les valeurs suivantes :

- « Cadre+ » : personnes catégorisées comme « Cadre administratif » supervisant une direction ou une équipe > 5 personnes ;
- « Cadre » : personnes catégorisées comme « Cadre administratif » ;

⁸⁹ Également nommé « fonction adaptation » ou « fonction objective »

⁹⁰ Secteur d'activité : Centre Hospitalier, Université, Ministère...

- « Non-Cadre » : personnes catégorisées comme des employés administratifs.

En complément des critères liés à la catégorie socio-professionnelle des personnes, des critères supplémentaires sont pris en considération notamment « Typologie des participants » et « Historique de formation ».

- Groupe paritaire et de taille restreinte

Le groupe paritaire consiste à favoriser la mixité des groupes selon un nombre de participants restreints (minimum cinq personnes). Ces caractéristiques du groupe de formation favorisent un score de satisfaction global positif selon les résultats obtenus lors de l'analyse statistique ([Voir section 8.2.3](#)).

Le groupe de taille restreint favorise un nombre d'interactions adapté entre les participants. Pour un groupe de 5 à 6 personnes, le nombre potentiels d'interactions d'individu à individu s'élèvent à 10 à 15 relations interindividuelles⁹¹. Par conséquent, l'augmentation du nombre d'individus au sein du groupe augmente les potentiels relations interindividuelles et provoqués des sous-groupes défavorisant les interactions sociales.

-Temps de traitement raisonnable

Le principe de traitement de l'AG implique un grand nombre de générations (ou itérations) pour obtenir des résultats pouvant nécessiter un temps de traitement long indépendamment des ressources matérielles utilisées. La typologie du problème de planification induit que le nombre de combinaisons peut augmenter en fonction du nombre de caractéristiques des individus ce qui implique un temps de calcul dépendant du nombre d'entrée⁹². Bien que cette notion soit subjective, le temps de traitement de l'algorithme est une problématique à prendre en considération avec un certain seuil à ne pas dépasser pour assurer la performance du prototype d'application. Dans un contexte industriel, la réactivité de la prise de décision est une exigence qui implique un délai raisonnable du temps de réponse pour obtenir le(s) résultat(s). Dans notre cas d'application, le temps de réponse n'est pas « critique », comme par exemple, dans le cas d'application des urgences médicales où une réponse quasi-instantanée est impérative (Chennaoui, A. 2015). Néanmoins, le temps d'exécution impliquant des résultats approximatifs est privilégié au détriment de l'exactitude des résultats inhérente aux méthodes déterministes.

7.3.2 Principe et paramétrage du traitement de l'AG

Cette étape consiste à détailler le principe et le paramétrage de l'AG selon les contraintes et les critères de recherche préalablement définis. Deux étapes principales sont à distinguer : la requête-utilisateur puis le traitement de l'AG.

-Etape 1 : Requête-utilisateur

La planification d'une session de formation s'appuie sur une forte interaction avec l'utilisateur qui détermine le choix et l'importance des critères de décision lors de la formulation de la requête. Lors de cette étape, le gestionnaire de formation spécifie obligatoirement le mode organisationnel et les critères du groupe de formation souhaités (social, cognitive et spatio-temporel) d'une manière facultative. Comme présenté dans la figure suivante, l'interface de

⁹¹ Selon la formule de calcul $n(n-1)/2$

⁹² Selon la théorie de la complexité en temps, différentes classes sont catégorisées (classe P, classe NP, classe NP-complet). Concernant les problèmes combinatoires, (Garey M. 1979) définit une classe spécifique soit NP-difficile du fait qu'aucun algorithme polynomial n'a été démontré.

recherche présente l'ensemble des critères de décision agencés sous forme de colonnes pour faciliter la formulation de la requête :

The image shows a web application interface for searching training groups. On the left is a blue sidebar with the 'LEXIANE' logo and a navigation menu with options like 'Recherche', 'Planification', 'Suivi', and 'Evaluation'. The main area is divided into three vertical panels, each with a title and several filter fields:

- (a) Groupe:** Contains filters for 'Mode organisationnel' (with a dropdown menu), 'Titre', 'Type de regroupement', 'Taille du groupe', 'Matière', and 'Niveau de données' (with a dropdown menu).
- (b) Formation:** Contains a filter for 'Thématique de formation' with a dropdown menu and a search icon.
- (c) Logistique:** Contains filters for 'Durée', 'Mois', 'Région' (with a dropdown menu), 'Département' (with a dropdown menu), and 'Statut de présence' (with a dropdown menu). A blue 'Valider' button is located at the bottom of this panel.

Figure 6.2 - Interface de recherche de la fonctionnalité « Planification » avec critères de décision liés au (a) Groupe (b) Formation et (c) Logistique

Au niveau du traitement de l'algorithme, la contrainte stricte relative au mode organisationnel consiste à rejeter impérativement les solutions potentielles qui ne respectent pas cette condition. Ainsi, cette première étape de filtrage séquentiel permet de réduire au préalable l'ensemble des solutions potentielles. Le cas d'utilisation selon une modalité « Intra-organisationnel » limite considérablement la complexité combinatoire compte tenu que la recherche de participants potentiels se restreint au sein d'une même organisation. A l'inverse, le critère selon une modalité « Inter-organisationnel », l'espace de recherche s'étend plus largement selon un ensemble de participants issus de plusieurs organisations complexifiant le nombre de solutions potentielles.

-Etape 2 : traitement de l'AG

Inspirés de la théorie darwinienne⁹³, le fonctionnement de l'AG s'appuie sur la sélection et l'adaptation des individus ainsi que sur la recombinaison génétique de Grégor Mendel⁹⁴. Tranposé par (Holland, J. 1975) au sein des systèmes artificiels, ce principe de traitement emploie une terminologie du modèle évolutionniste détaillée en Annexe 5.

Dans notre cas d'application, après la validation de la requête-utilisateur, les principes de traitement de l'AG sont appliqués selon les étapes suivantes comme illustrées dans la figure :

- (1) Initialisation de la population
- (2) Evaluation
- (3) Opérateurs génétiques (sélection, croisement, mutation)
- (4) Renouvellement d'une génération
- (5) Critère d'arrêt
- (6) Génération d'une liste des résultats

⁹³ Charles Darwin (1809-1882) : en 1859, il publie l'ouvrage « De l'origine des espèces » considérée comme un ouvrage fondateur en biologie évolutive. Il y présente les fondements de la théorie du principe d'évolution selon la sélection naturelle et l'adaptation selon l'environnement

⁹⁴ Gregor Mendel (1822-1884) : botaniste autrichien, il théorise le principe de l'hérédité en démontrant la transmission des caractères innés sous le nom des « lois de Mendel » (Loi d'uniformité des hybrides de première génération, Loi de disjonction, Loi de disjonction indépendante)

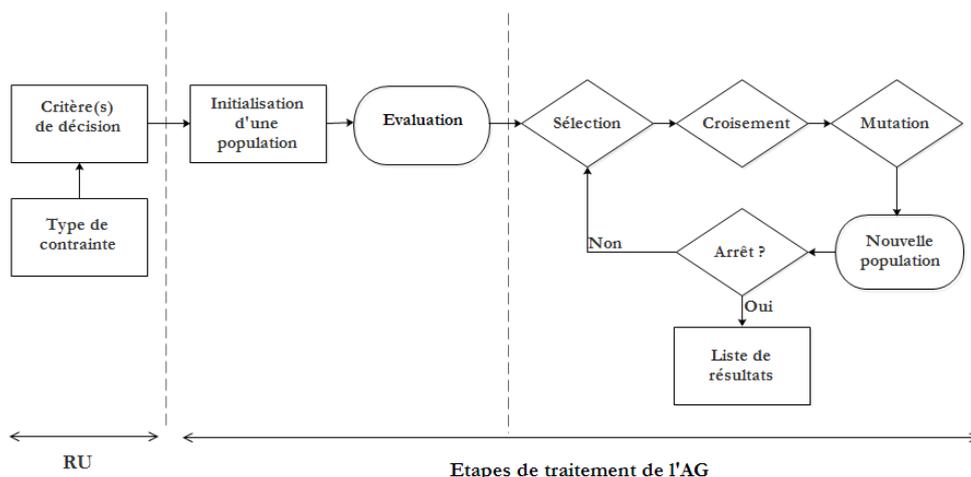


Figure 6.3 - Principe du traitement global de l'AG

- Initialisation de la population

La population initiale est composée d'un ensemble de groupes de formation qui devient l'unité élémentaire de la population. Au préalable, un codage des données est nécessaire pour procéder à une génération aléatoire de la population initiale.

Codage des données. Cette étape consiste à représenter les groupes de formation selon un codage binaire. Ainsi, chaque vecteur de taille fixe est représenté sous la forme d'une chaîne de bits (0,1) correspondant à un ensemble de participants potentiels associé à des caractéristiques. La valeur associée indique la présence (1) ou l'absence (0) de l'individu dans le groupe dont les caractéristiques (genre, type de participant, statut, secteur d'activité, type d'établissement et historique de formation) sont rattachées à un index comme illustré dans la figure suivante :

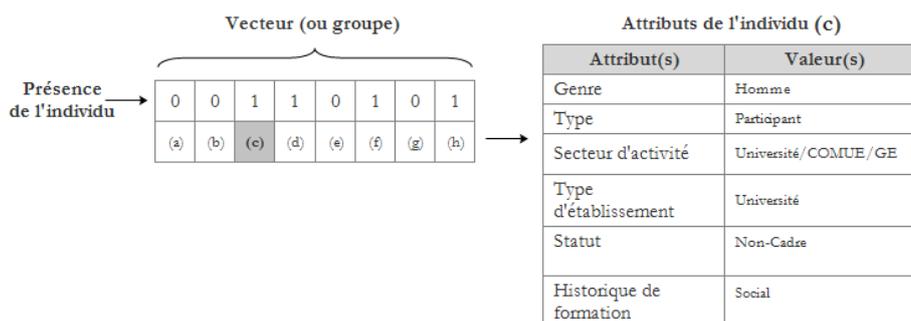


Figure 6.4 - Exemple de codage binaire d'un groupe de formation (ou chromosome) incluant des individus (ou gènes) rattachés à un ensemble d'attributs (ou allèles)

Les caractéristiques des individus sont codées également à l'aide d'attributs binaires : la valeur 1 indique que l'individu possède la caractéristique et la valeur 0 indique qu'il ne la possède pas. Au préalable, ce codage va permettre d'appliquer les opérateurs génétiques.

Génération aléatoire de la population initiale. Au début du processus du traitement, l'AG génère une population composée de groupes de formation incluant des participants affectés

d'une manière aléatoire. Cette première génération vise à servir de point de départ pour générer d'autres générations selon les principes de traitement de l'AG. Par nature, la génération aléatoire de la population initiale est à priori non adaptée aux objectifs de performance préalablement définis c'est-à-dire des groupes de formations homogène de taille restreinte et paritaire. La diversité de la population initiale vise à permettre l'exploration de l'espace de recherche et à converger vers des solutions dites « satisfaisantes » en utilisant les opérateurs génétiques à venir (Sélection, Croisement, Mutation).

- Evaluation

Cette étape de traitement de l'AG consiste à attribuer un score d'évaluation (ou une valeur numérique) pour chaque groupe selon la fonction objective (ou *fitness*) déterminé préalablement. Dans ce cas d'étude, il s'agit de maximiser l'homogénéité (ou minimiser l'hétérogénéité) des groupes de formation en fonction des caractéristiques des individus.

Cette étape d'évaluation vise à sélectionner les individus dont les caractéristiques vont permettre au processus de recherche de définir un groupe selon un type de regroupement homogène. Le score final du groupe de formation repose sur un calcul de la variance de la moyenne pondérée (ou *weight sum method*) de l'ensemble des caractéristiques des individus. Les caractéristiques des individus sélectionnées sont les suivantes⁹⁵ :

- Typologie de l'individu ⁹⁶
- Secteur d'activité
- Type d'établissement
- Statut
- Historique de formation

La méthode de calcul du score final se base sur les principes de l'équation suivante soit :

$$\text{Score}(g) = \sum_{i=1}^N P_i * C_i \quad (1.1)$$

Ainsi, le score d'un groupe (g) correspond pour un ensemble de n caractéristiques. Un coefficient de poids est affecté pour chaque caractéristique soit le poids affecté noté P_i à une caractéristique spécifique noté C_i .

Cette méthode consiste à agréger l'ensemble des critères en un critère unique et de convertir ainsi un problème d'optimisation multi-objectif à un problème mono-objectif. Ainsi, l'AG se base sur cette valeur numérique pour mener une comparaison avec d'autres groupes de formation potentiels et déterminer ainsi si un groupe sera sélectionné ou pas selon les opérateurs génétiques à venir. Au niveau de l'implémentation, il convient de s'assurer que les attributs (ou caractéristiques) des individus soient ordonnés identiquement au sein de l'index.

- Opérateurs génétiques

Les opérations génétiques (ou artificielles) visent l'exploration et le renouvellement de la population de groupes de formation à chaque génération. Ainsi, le principe de traitement repose sur un principe itératif qui consiste à générer successivement des solutions approximatives qui se rapprochent graduellement de la solution recherchée selon les étapes suivantes :

⁹⁵ La caractéristique liée au « Genre » (femme ou homme) n'est pas inclût pour éviter de générer des groupes exclusivement composé d'hommes ou de femmes.

⁹⁶ La typologie de l'individu se distingue selon les valeurs suivantes en fonction du cadre de formation : « Participant » (apprentissage formel) et « Colloque » (apprentissage non-formel)

Opération de sélection. Cette opération consiste à choisir les « meilleurs » participants parmi les groupes de formation potentiels selon le score d'évaluation associé. Couramment utilisé, cette technique de sélection nommée « élitiste » consiste à conserver systématiquement le « meilleur » groupe en écartant les groupes qui ne répondent pas aux objectifs visés. Les solutions potentielles considérées comme étant les plus pertinentes sont celles qui maximisent l'homogénéité des caractéristiques des individus au sein du groupe de formation. L'avantage de cette méthode est que les groupes de formation potentiels répondant aux critères de la fonction objective sont systématiquement conservés. Le taux de sélection est fixé à 20%.

Opération de croisement. Cette opération consiste à choisir au hasard un point de croisement au sein de chaque vecteur représenté par une chaîne de bits. Il s'agit d'intervertir et de fusionner des individus à partir de deux parents nommés P_1 et P_2 vers les deux descendants nommés D_1 et D_2 comme illustré selon la figure suivante :

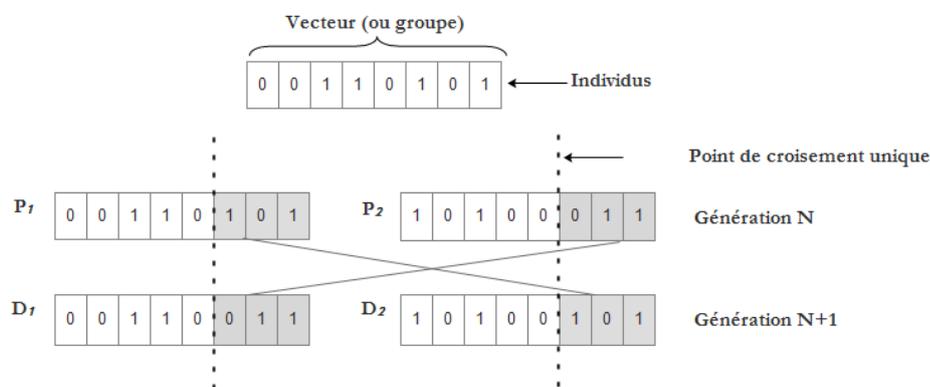


Figure 6.5 - Principe de croisement à un point sur une chaîne de bits d'une génération avec P (Parent) et D (Descendance)

Le principe de comparaison implique que le « meilleur groupe » de la nouvelle génération se croise avec le « meilleur groupe » de la génération précédente. Ainsi, l'objectif de cette opération est de conforter les solutions préalablement générées par l'opération de sélection en considérant qu'en combinant essentiellement des solutions « satisfaisantes », la seconde génération est supposée résulter des solutions encore plus « satisfaisantes ». Ainsi, ce remplacement générationnel de type élitiste vise à préserver et à converger vers des solutions adaptées mais au détriment de la diversité des individus.

Opération de mutation. Cette opération consiste à intégrer une valeur aléatoire nécessaire pour diversifier l'exploration des solutions potentielles. En effet, la méthode préalable de sélection implique un biais compte tenu que la population se restreint à l'exploration d'un seul sous-ensemble des solutions possibles. L'opérateur de mutation permet de pallier ce biais et vise à modifier une partie de la population en réintroduisant à minima des groupes qui ont été éliminés lors du renouvellement de génération. Le groupe « muté » ne sera pas forcément meilleur mais il apportera des possibilités d'exploration supplémentaires. Le taux de mutation est généralement minimal. Dans notre cas d'étude, il est fixé à 1%. Néanmoins, un fort taux de mutation peut être également envisagé pour intégrer des participants avec des profils plus diversifiés au sein des groupes de formation potentiels sélectionnés.

- Critère d'arrêt

Le critère d'arrêt se définit selon un nombre d'itérations préalablement fixé. Le nombre de génération est considéré comme suffisant lorsque le processus de traitement ne s'améliore plus. Si le critère d'arrêt n'est pas atteint, les opérations de sélection, de croisement et de mutation se réitèrent pour générer une nouvelle génération. Dans le cas de figure où le critère d'arrêt est atteint, le processus de l'AG interrompt le traitement de recherche et génère les résultats obtenus. Dans notre cas d'application, le nombre d'itérations est fixe et limité à 10. D'après une série de tests, un nombre d'itérations supérieurs n'améliorent pas significativement le score d'évaluation et augmente le temps de traitement.

- Liste de résultats

Les résultats proposent une liste de 10 groupes de formation potentiels classés par ordre décroissant selon le score d'évaluation attribué sur une échelle de 0 à 5. Cette conversion vise à simplifier l'interprétation pour l'utilisateur qui peut mesurer la pertinence du groupe de formation proposée. De même, le classement par ordre décroissant est le plus fréquemment utilisé et répond ainsi à une consultation intuitive de l'utilisateur. Le principe est de classer le groupe de formation le plus pertinent en première position et le moins pertinent en fin de liste. La restitution de l'ensemble des résultats est représentée par défaut sous forme de représentation graphique de type nœud-lien ([Voir section 6.3.3](#)).

7.3.3 Implémentation du code-source

L'implémentation de l'AG est codée sous un langage de programmation spécifique⁹⁷. Cette étape implique de choisir des paramètres statiques qui fixe le processus de traitement : taille de la population initiale et le nombre d'itération (ou critère d'arrêt). Dans notre cas d'application, des paramètres supplémentaires sont intégrés manuellement par défaut :

-Taille restreint

La modalité d'une taille de groupe (5 et 6 participants) correspond au nombre moyen des groupes de formation observé dans le cadre de l'activité de formation et justifiée selon l'analyse statistique ([Voir section 8.2.3](#)).

-Groupe paritaire

La parité implique un nombre équivalent au niveau de la variable « Genre » des individus soit respectés au sein du groupe. D'un part, ce paramétrage permet d'éviter de générer des solutions potentielles composées exclusivement d'hommes ou de femmes. D'autre part, la parité au sein du groupe de formation est une modalité qui influence favorablement le niveau de satisfaction des participants en formation continue ([Voir section 8.2.3](#)).

Le processus de traitement est le suivant :

⁹⁷ Python

Etapes	Actions
	Entrée : $R = \{c_1, \dots, c_h, \dots, c_Z\}$: requête
	Sortie : $G = \{g_1, \dots, g_i, \dots, g_n\}$: liste des « meilleurs » groupes
	$n_iterations = 10$ //Fixation du critère d'arrêt
	$n_G = 1000$ //Taille de la population initiale
	parité H=F //Valeurs par défaut
	taille 5-6 //Valeurs par défaut
1	Début
2	Initialisation Pop
3	Evalue (Pop) $G \{g_1, \dots, g_i, \dots, g_n\}$ //Evaluation pour chaque groupe
4	Tant que (critère d'arrêt non-atteint) faire : //Génération n à la génération $n+1$.
5	Sélection
6	Mutation
7	Croisement
8	Fin
9	Return Génération d'une liste limitée à 10

Tableau 5.2 - Pseudo-code de l'AG

Détails de la notation :

Pop : population

G : collection des groupes est notée $G = \{g_1, \dots, g_i, \dots, g_N\}$ où N représente le nombre de groupe dans la population.

R : requête est notée R incluant un ensemble fini de critères notée C $\{c_1, \dots, c_h, \dots, c_Z\}$ où Z représente le nombre de critères.

7.4 Conclusion

La résolution des problèmes de décision de nature multicritère et combinatoire oriente notre démarche à implémenter un algorithme génétique pour optimiser et générer les solutions potentielles. Cette méthode présente l'avantage de convertir un problème de décision multicritères liés aux multiples caractéristiques (ou attributs) des individus en un problème de décision monocritère à travers un score d'évaluation.

Au niveau de l'implémentation, les principes élémentaires de l'AG sont appliqués : un codage binaire, un seul point de croisement, une sélection selon un principe « élitiste » sont les paramètres les plus couramment utilisée dans le domaine de l'AG. Des paramètres supplémentaires sont fixés par défaut dont la taille et la parité du groupe permettant ainsi d'appliquer une approche adaptative selon le contexte d'énonciation.

De nombreuses perspectives peuvent améliorer le principe de traitement notamment au niveau des opérateurs génétiques (codage réel, fixation de plusieurs points de croisement, orientation de la mutation, optimisation de la fonction d'évaluation...). Néanmoins, l'évaluation du traitement de l'AG reste dépendante de la distribution de la population initiale qui doit être suffisamment large afin de permettre une exploration et une convergence vers des solutions potentielles.

Au niveau de l'administration du système, le processus de traitement de l'AG est modifiable essentiellement à travers le code-source, ce qui complexifie les tests de paramétrage. La conception et le développement d'une interface graphique peuvent être envisagés pour faciliter le paramétrage de l'algorithme génétique afin d'observer les variations des résultats.

Chapitre 8. Aide à la planification par analyse statistique

Ce chapitre présente le traitement et l'analyse des données à partir des questionnaires de satisfaction liée à l'activité de formation en détaillant la méthode de collecte de données, les tests statistiques, les variables utilisées et les résultats obtenus. D'une part, cette approche de type analytique permet principalement de justifier les paramétrages pour l'aide à la planification des sessions de formation. D'autre part, cette approche vise à définir les indicateurs et les données multidimensionnelles à synthétiser au sein du tableau de bord analytique du prototype d'application pour soutenir l'aide à la décision stratégique.

8.1 Analyse statistique de l'activité de formation

8.1.1 Finalités d'application

Issue de la théorie statistique de la décision, l'analyse des données vise à orienter et justifier des prises de décision fondée sur une approche rationnelle. Cette approche implique d'avoir à disposition des données d'observations exploitables pour parvenir à une généralisation des résultats. Malgré les biais d'évaluation, le principal avantage de cette méthode de décision est de prendre en compte la complexité des situations de formation à partir d'observations réelles.

Dans notre cas d'application, l'analyse statistique vise à répondre à plusieurs finalités d'applications qui peuvent être distinguées de la manière suivante : l'aide à la planification et l'aide à la décision stratégique.

- Aide à la planification

-Paramétrage de l'AG

Lors de l'implémentation d'un AG, la phase de paramétrage s'avère être une étape déterminante pour optimiser le processus de traitement lors de l'exploration de la population d'individus. L'analyse statistique permet d'orienter la fonction objective (ou *fitness*) de l'AG en identifiant les caractéristiques et les valeurs associées du groupe de formation à optimiser. Dans notre cas d'application, il s'agit de définir les caractéristiques du groupe de formation qui vont conditionner les interactions sociales entre les participants. En pratique, l'objectif de cette méthode n'est pas de définir un groupe de formation dit « optimal » d'une manière absolue mais de générer un ensemble de groupes de formation potentiels en évitant un choix de paramétrage arbitraire ([Voir section 7.3.1](#)).

-Aide à la planification de type exploratoire

Comme indiqué lors du chapitre précédent, la planification de type exploratoire des sessions de formation est à l'initiative de l'organisme de formation. Ainsi, la composition du groupe est gérée librement par le gestionnaire de formation qui n'est pas soumis à de(s) contrainte(s) prédéfini(s) par une commande de formation. Parmi les méthodes existantes, la nature du problème de décision nous oriente vers une méthodologie exploitant des résultats d'une analyse statistique. Cette méthode inductive consiste à justifier le choix des valeurs des critères de planification à partir de données expérimentales observées. Ainsi, l'imprécision et la subjectivité du processus décisionnel humain sont limitées.

- Aide à la décision stratégique

-Indicateurs de performance

Le tableau de bord analytique vise à présenter une analyse synthétique de l'activité de formation et à mesurer des objectifs de performances en vue d'une remédiation. Il s'appuie sur des indicateurs clés de performance (ou *Key Performance Indicator*) dont la combinaison offre une vision synthétique et complémentaire des données étudiées. Selon (Feschet, P., 2016), un bon indicateur doit être utile et clair en couvrant les besoins des utilisateurs. Ainsi, cette analyse statistique vise à identifier les indicateurs de performance les plus pertinents à intégrer au sein du tableau de bord analytique ([Voir section 6.3.4](#)).

8.1.2 Objectif(s)

Dans un contexte de formation continue, la notion de « satisfaction » n'est pas communément définie mais reste étroitement liée à l'évaluation de la formation. Il s'agit d'apprécier la valeur d'une formation sur la base de standards ou de critères spécifiques (...) dans une démarche de qualité (McCain 2005). Selon (Hadjji 2012), le participant est un acteur central de l'évaluation : la qualité d'une formation résulte d'une adéquation entre ce que le participant est légitimement en droit d'attendre et la réalité évaluée. Par conséquent, la perception de la qualité de la formation est généralement définie en termes d'écart entre les attentes et les objectifs pédagogiques du participant et l'évaluation finale de la satisfaction.

Dans ce cas d'application, l'analyse statistique des données issues des questionnaires de satisfaction vise à identifier les conditions pédagogiques, sociales et organisationnelles du groupe de formation qui influencent favorablement la satisfaction du participant. Ainsi, les objectifs peuvent être formulés de la manière suivante :

- Evaluer le score de satisfaction globale et le score de satisfaction de collaboration
- Identifier les caractéristiques du groupe de formation qui impactent favorablement le score de satisfaction globale et le score de satisfaction liée à la collaboration
- Identifier les facteurs d'insatisfaction et d'indifférence

Les indicateurs et les types de variables associés sont récapitulés dans le tableau suivant :

Indicateurs	Variables
Participant	-Identité personnelle (âge, genre) -Identité professionnelle (organisation, secteur d'activité, fonction, ancienneté, statut)
Groupe	-Mode organisationnel (inter/intra-organisation) -Taille -Durée des sessions de formation -Type de regroupement (homogène/hétérogène) -Environnement logistique (support matériel, salle...)
Pédagogique	-Formateur -Objectif(s) de formation -Thématique de formation -Contenu du support pédagogique

Tableau 6.1 - Indicateurs et description des variables associées

8.1.3 Méthodologie

- Mode de collecte des données d'observation

Questionnaire. La collecte des données d'observation repose sur un questionnaire basé sur le modèle d'évaluation créé par Donald Kirkpatrick (Kirkpatrick D., 1960, réédité en 2009), pionnier dans la recherche liée à l'évaluation de la formation. Largement utilisé par les professionnels, les différentes versions du questionnaire d'évaluation se focalisent sur quatre niveaux (réaction, apprentissage, comportement, résultat) détaillé en Annexe 6. Le caractère hiérarchique et causal du modèle émet l'hypothèse que le premier niveau (réaction) réalisé en fin de l'action de formation qui vise à mesurer la satisfaction générale conditionne la suite du processus de formation.

La collecte des données s'étend sur 5 années d'activités de formation⁹⁸. Cette période relativement longue permet de bénéficier ainsi de données d'observation stables et significatives. Le traitement implique une méta-analyse des données pour mener une synthèse objective. Cette démarche de mise en commun des résultats est réalisable compte tenu que l'échelle d'évaluation et la structuration thématique des modèles du questionnaire sont identiques. Les modifications du modèle d'évaluation principalement selon le taux de remplissage des réponses et l'évolution des réglementations juridiques sont minoritaires⁹⁹.

Le questionnaire regroupe 20 questions dont 14 questions fermées et 6 questions ouvertes classées selon les sections suivantes :

- Identité professionnelle
- Satisfaction globale
- Organisation de la formation
- Satisfaction de la collaboration (ou de l'animation)
- Contenu pédagogique
- Transfert

Les questions ouvertes sont liées à l'identité personnelle et professionnelle du participant (Nom, Age, Ancienneté, Fonction et Organisation) et une seule question est dédiée à une évaluation complémentaire de la session de formation sous forme de commentaires libres.

Le modèle du questionnaire, les variables et les différents formats de réponse sont présentés en Annexe 7.

Echelle d'évaluation. L'échelle de mesure utilisée est de type Likert qui vise à mesurer le degré d'adhésion à une affirmation selon les cinq items suivants : l'item 0 correspond à « Pas du tout satisfait », l'item 1 correspond à « Peu satisfait », l'item 2 correspond à « Neutre », l'item 3 correspond à « Plutôt satisfait » et l'item 4 correspond à « Très satisfait ».

L'échelle de Likert est de type ordinale et numérique dans le sens où l'ordre des modalités proposés est transitive. Le choix d'une modalité impair permet au participant de choisir une réponse médiane. Cette modalité centrale évite de contraindre le participant de se positionner selon une mesure « ipsative » c'est-à-dire « à choix forcé ». Comme précisé dans l'Annexe 7, les échelles de Likert sont associées aux thématiques suivantes : « Satisfaction globale », « Satisfaction de la collaboration », « Contenu pédagogique » et « Organisation ».

⁹⁸Axe temporel : 2014 à 2018

⁹⁹ Absence de réponses pour la question relative au niveau « Transfert »/« Loi Informatiques et Libertés » devenu « RGPD »

- Mode d'administration

L'évaluation de la session de formation est une obligation juridique pour l'organisme de formation qui vise à vérifier l'atteinte d'objectifs pédagogiques préalablement définis¹⁰⁰. Par conséquent, un créneau temporel est consacré en fin de session de formation pour administrer un questionnaire en face-en face d'une manière non-anonymisé et en format papier. Ainsi, l'évaluation de la satisfaction du participant se base selon une appréciation immédiate de l'action de formation (couramment appelé « évaluation à chaud ») et se limite sur une expérience unique. Il est à noter que la divulgation de l'identité des répondants peut constituer un biais d'évaluation. En effet, ce mode de réponse peut induire des réponses surévaluées et empêcher les participants d'exposer ouvertement leurs critiques ou déceptions.

- Echantillonnage

Le nombre de questionnaire à disposition correspond à 387 observations au total. Compte tenu du nombre réduit d'observation, une méthode de recensement s'impose pour l'analyse des données, ce qui implique un traitement exhaustif des observations à l'inverse d'une méthode d'échantillonnage de tirage aléatoire. L'avantage de cette méthode est de contourner les difficultés liées à la représentativité de l'échantillon (couramment appelé « biais de sélection »).

8.2 Analyse et traitement statistique

8.2.1 Statistiques descriptives

Cette étude statistique comporte 387 observations répartis sur 54 groupes de formations sur cinq années consécutives¹⁰¹. La saisie des données s'est effectuée manuellement sur un tableur Excel selon des instructions de formatage (sujets par lignes, nom de variables unique...) pour pouvoir les exporter vers un logiciel de statistique¹⁰².

La distribution des participants est la suivante : la majorité des participants sont des femmes (environ 62%) et un tiers d'hommes (37%). La moyenne d'âge est d'environ 40 ans pour une ancienneté en moyenne comprise entre 6 et 7 ans (dans 38,7% des cas). Cet indicateur lié à l'ancienneté peut suggérer que le besoin de formation est davantage lié à un besoin de perfectionnement et non d'acquisition de compétences.

Une forte majorité des participants occupe une fonction de « Cadre » (81,7%). Parmi ces participants, environ 30% ont le statut « Cadre+ » qui équivaut à un poste ayant une responsabilité de supervision et de gestion d'équipe composé de 5 à 20 personnes. Il est à noter que les participants ayant une position « Cadre+ » sont majoritairement des hommes (58,3%).

Près de la moitié des participants ont assisté à une session de formation à leur initiative (50,4%) tandis que pour environ un quart d'entre eux, leurs présences relèvent d'une exigence hiérarchique¹⁰³ (25,6%). Cette démarche volontariste est en adéquation avec les objectifs pédagogiques de formation déclarés : (i) acquérir ou développer de nouvelles compétences (52,2%) et (ii) s'adapter à une évolution réglementaire (44,7%)¹⁰⁴. Il est à noter que la situation de formation évaluée est la première expérience de formation pour environ 91% des participants impliquant qu'ils ne peuvent pas la comparer avec une expérience de formation antérieure.

¹⁰⁰ En application de l'article L.6362-5 du Code de Travail, les organismes de formation et les formateurs indépendants sont dans l'obligation de justifier la réalisation des actions de formation qu'ils dispensent.

¹⁰¹ 2014 à 2018

¹⁰² Xlstats

¹⁰³ 24% ne l'ont pas précisé

¹⁰⁴ Possibilité de réponses multiples

Les sessions de formation sont organisées majoritairement selon un mode inter-organisationnel (63%) avec des groupes de formation principalement de taille restreinte comptant moins de 7 participants (61%) et mixte (environ 80%)¹⁰⁵. La durée des sessions de formation est principalement d'une journée (50%)¹⁰⁶. Compte tenu que la majorité des participants sont des femmes, les groupes dont le nombre de femmes est supérieurs aux hommes sont majoritaires (46,3%).

L'ensemble des statistiques descriptives est détaillé en Annexes 8, 9,10 et 11.

8.2.2 Analyse globale

L'analyse statistique résulte un niveau de satisfaction global nettement positif pour l'ensemble des sessions de formation. En effet, la majorité des participants se déclare « Plutôt satisfait » (51,9%) voire « Très satisfait » (29,2%)¹⁰⁷. La moyenne attribuée est de 3,1 (basé sur une échelle allant de 0 à 4). Néanmoins, ce score de satisfaction globale est à nuancer compte tenu du nombre de réponses selon l'item « Neutre » qui s'élève à environ 19%.

Le niveau de la satisfaction lié à la collaboration révèle des résultats également positifs, avec des scores quasi-équivalents soit une majorité de réponse selon l'item « Plutôt satisfait » (57%) voire « Très satisfait » (30%). Une différence notable est qu'environ 10% se déclare « Peu satisfait », en comparaison avec le score de satisfaction global où cette modalité de réponse est inexistante.

Les différents items croisés (qualité du formateur, collaboration/animation, contenu du support pédagogique, composition du groupe, conditions matérielles) démontrent logiquement que la satisfaction sur ces items augmente le score de satisfaction globale. Ce constat nous permet de supposer une cohérence dans les déclarations des participants.

La répartition du score de la satisfaction globale et du score de la satisfaction liée à la collaboration est détaillée en Annexe 11.

8.2.3 Traitement et tests statistiques

Les tests statistiques diffèrent selon la nature des questions qui peuvent être fermées ou ouvertes. Une analyse bivariée est appliquée pour les questions fermées consistant à comparer le score de satisfaction global et le score de satisfaction lié à la collaboration (variables dépendantes) avec les caractéristiques des participants et du groupe de formation (variables indépendantes).

Tests statistiques. La contrainte méthodologique est liée à la nature multidimensionnelle des données. Ainsi, le choix des tests statistiques dépend des variables à analyser récapitulés dans le tableau suivant :

	Variables		Tests statistiques
	Noms	Description	
Participant	Genre	Catégorielle/nominale	Test de student ¹⁰⁸
	Age	Discrète ¹⁰⁹	Régression linéaire
	Statut/Position	Catégorielle/nominale	Analyse de la variance (ANOVA)
	Ancienneté	Discrète	Régression linéaire
Groupe	Type de	Catégorielle/nominale	Test de student

¹⁰⁵ 20% des groupes sont non-mixtes (16.7% sont composé exclusivement de femme et 3.7% sont composés exclusivement d'hommes)

¹⁰⁶ 33% des sessions de formations ont une durée de 2 jours et 17% ont une durée de 3 jours

¹⁰⁷ Selon un intervalle de confiance avec risque d'erreur de 5%

¹⁰⁸ Test t de Student sur deux échantillons indépendants ^s

¹⁰⁹ Variable de type quantitative qui ne prend que des valeurs isolées

	regroupement		
	Mode d'organisation	Catégorielle/nominale	Test de student
	Taille	Continue ¹¹⁰	Test de student
	Mixité	Catégorielle/nominale	Test de student
	Durée	Discrète	Régression linéaire
Pédagogique	Objectif(s) de formation	Catégorielle/nominale	Test de student ⁴
	Thématique de formation	Catégorielle/nominale	Analyse de la variance (ANOVA)
	Support pédagogique	Catégorielle/nominale	Régression linéaire
	Animation du formateur	Continue	Régression linéaire
	Genre du formateur	Catégorielle/nominale	Test de student ⁴

Tableau 6.2 - Récapitulatif des tests statistiques selon le type de variable

Dans le cadre de cette analyse, les facteurs intrinsèques et extrinsèques sont distingués pour expliquer leurs influences sur le score de satisfaction des participants. Le style d'animation du formateur et le contenu pédagogique sont considérés comme des facteurs extrinsèques de la session de la formation tandis que les caractéristiques du groupe de formation (taille, mixité, type de regroupement, durée et mode organisationnel) sont considérées comme des facteurs intrinsèques de la session de la formation.

- Facteurs extrinsèques

Animation du formateur. Le rôle du formateur à travers son style d'animation est considéré comme un facteur extrinsèque au groupe de formation. La relation observée entre la satisfaction globale avec l'évaluation du formateur est particulièrement marquée, ce qui signifie que la qualité de l'animation du formateur est une condition essentielle pour obtenir une satisfaction positive du participant ($p < 0,0001$)¹¹¹. De même, le score de satisfaction axé sur la collaboration est particulièrement corrélé avec le formateur. Cette tendance est particulièrement confirmée à travers l'analyse textuelle des réponses ouvertes. En effet, plus de 73% des réponses sont axées sur les compétences du formateur (exemple : « expertise du formateur », « pédagogie du formateur », « formateur professionnel »...).

Contenu pédagogique. Le contenu du support pédagogique influence également très significativement la satisfaction globale et la satisfaction de la collaboration ($p < 0,0001$). Dans le cadre de la formation continue, le formateur est l'auteur principal du contenu pédagogique de la formation. Ainsi, il paraît cohérent que ces deux variables (animation du formateur et contenu pédagogique) sont intrinsèquement liées et fortement corrélées avec les scores de satisfaction. Néanmoins, l'analyse textuelle des réponses ouvertes fait apparaître des appréciations positives et négatives équilibrées. D'une part, les motifs d'insatisfaction sont axés sur le contenu pédagogique (« formation un peu technique », « formation trop pointue ») et sur les aspects matériels des supports pédagogiques (« pas de support papier remis en fin de session »). D'autre part, les réponses positives sont également axées sur le contenu pédagogique dont notamment une appréciation positive des cas pratiques et des exercices (« cas pratiques et exercices adapté aux problématiques », « caractère synthétique de la présentation »).

¹¹⁰ Variable de type quantitative dont les valeurs sont comprises en deux nombres

¹¹¹ $R^2=0,72$, p-valeur du test de Fisher inférieure à 0,0001

- Facteurs intrinsèques

Les variables du groupe de formation considérés comme étant des facteurs intrinsèques sont les suivants :

- Taille
- Mixité
- Type de regroupement
- Durée
- Mode organisationnel.

A partir des résultats obtenus, la significativité statistique (ou p-value) est synthétisée selon le tableau suivant :

		Variable explicative	
		Satisfaction globale	Satisfaction de la collaboration
		Signification (p-value)	
Variable discriminante	Taille	<0,0001	0,23
	Mixité	<0,0001	0,15
	Type de regroupement	<0,0001	0,12
	Durée	0,02	0,38
	Mode organisationnel	0,001	0,54

Tableau 6.3 - Tableau comparatif des analyses bivariées entre les variables du groupe de formation et le score de la satisfaction globale/score de la satisfaction liée à la collaboration selon la p-value associée

Taille du groupe. La taille des groupes est répartie selon les modalités suivantes : 2-4, 5-6, 7-9, 10-14, 15+. Le nombre moyen sur l'ensemble des groupes de formation est de 7 participants. Cette modalité peut être considérés comme un groupe de formation « restreint » selon la définition de (Anzieu, D., 1971). Les résultats révèlent que la taille du groupe a un lien significatif avec le score de la satisfaction globale selon la valeur de p associée (<0,0001) tandis que la significativité avec le score de la satisfaction de la collaboration n'est pas exploitable (p=0,23). Bien que l'écart soit faible, les participants semblent préférer les groupes comportant 2-4 ou 10-14 participants. Néanmoins, les moyennes estimées pour chaque modalité ne permettent pas d'affirmer la prépondérance d'une taille optimale d'un groupe détaillé dans le tableau suivant :

Modalité	Moyenne estimée	Erreur standard	Pr > t
2 à 4	3,30	0,08	0,00
5 à 6	3,10	0,07	0,88
7 à 9	2,81	0,09	0,00
10 à 14	3,20	0,07	0,07
15 et plus	3,03	0,08	0,40

Tableau 6.4 - Tableau des moyennes estimées et la valeur p associée selon les modalités par taille

Il peut être également constater que la valeur de p fait apparaître un lien relativement faible pour les groupes de 15 et +. Une analyse complémentaire permet d'observer que ces groupes de formation composés sont essentiellement des sessions de formations organisés selon un mode « Intra-organisationnel ». A contrario, les groupes restreints sont plutôt organisés selon le mode « Inter-organisationnel ». Diverses contraintes d'ordre économique et logistique peuvent expliquer cet écart lié à la difficulté de réunir physiquement des participants dispersés géographiquement selon une période donnée.

Mixité du groupe. Au niveau individuel, il apparaît que le genre du participant influence le score de la satisfaction globale : les femmes attribuent un score de satisfaction global légèrement plus élevé (3,2 en moyenne) que les hommes (2,9 en moyenne). Ce résultat n'est pas dépendant d'une autre variable spécifique (âge, ancienneté, initiative de la formation). De la même manière, les femmes attribuent un score de satisfaction liée la collaboration plus positif (3,1 en moyenne) que les hommes (2,9 en moyenne) malgré un écart faible. Un autre constat peut être observé au niveau de la mixité des groupes : le score de la satisfaction liée à la collaboration diminue lorsque que les groupes sont composés essentiellement d'hommes. La figure suivante met en évidence cette différence d'évaluation : le score de satisfaction augmente si le groupe est mixte ou exclusivement composé de femmes et diminue quand le groupe est composé exclusivement d'hommes :

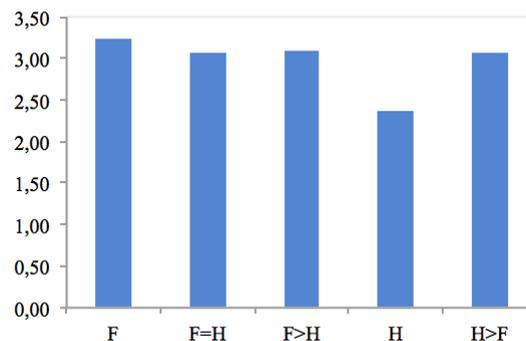


Figure 7.1 - Histogramme des comparatifs de la composition effective des groupes selon le genre des participants en variable X (F : groupe exclusivement composé de participants de genre féminin, F=H : groupe paritaire, F>H : groupe dont le nombre de participants de genre féminin est supérieur aux participants de genre masculin, H : groupe composé exclusivement de participants de genre masculin, H>F : groupe dont le nombre de participants de genre masculin est supérieur aux participants de genre féminin) et la moyenne estimée du score de satisfaction de la collaboration en variable Y :

Concernant le genre du formateur, les tests de différence de moyenne démontrent que le genre du formateur (homme ou femme) n'a pas d'impact sur l'évaluation de la satisfaction globale et la satisfaction liée à la collaboration.

Type de regroupement. Dans le contexte de cette étude, deux types de regroupement sont à distinguer : le type de regroupement dit « Homogène » ou « Hétérogène ». Le caractère « Homogène » du groupe de formation se limite au profil du participant et de son affiliation professionnelle. Les variables pris en considération pour caractériser l'homogénéité du groupe sont les suivantes :

- Secteur d'activité¹¹²

¹¹² Centre Hospitalier/Collectivités territoriales/Etat/Médico-social/Privé/Université/Autre

- Statut professionnel

Le statut socio-professionnel du participant indique son niveau hiérarchique et implique une déduction sur son niveau de responsabilité et de management au sein de son organisation.

D'autres variables telles que « Ancienneté », « Profession », « Type d'établissement » induisent un niveau de détails qui ne peut pas être exploiter compte tenu du nombre de valeurs manquantes. De plus, la variable « Profession » présente plusieurs limites : la diversité terminologique des noms de professions des participants ne permet pas d'utiliser cette variable qui diffère selon la gouvernance des organisations. De plus, la nature déclarative de cette variable peut impliquer un biais de désirabilité sociale¹¹³.

A contrario, le type de regroupement est qualifié comme « Hétérogène » quand les participants ne travaillent pas dans le même secteur d'activité. Il est à noter que par nature, les groupes de formation organisés selon un mode « Intra-organisationnel » ne sont pas considérés par défaut comme étant « Homogène » malgré le fait que les participants appartiennent à la même organisation et par conséquent au même secteur d'activité. En effet, la composition du groupe peut être considérée comme « Hétérogène » dès lors que le statut professionnel entre les participants diffère au sein du groupe (par exemple, un groupe de formation incluant des participants ayant un statut de « Cadre » et « Non-Cadre »).

Parmi les 54 groupes de formation analysés, il est à noter que la distribution des groupes selon le type de regroupement est peu équilibrée. En effet, il apparaît que le nombre de groupe de formation selon un regroupement de type « Homogène » est nettement supérieur (40 groupes) au groupe de formation selon un regroupement de type « Hétérogène » (14 groupes). Néanmoins, il peut être constaté que le type de regroupement « Homogène » influence le score de satisfaction global ainsi que le score de satisfaction lié à la collaboration. En effet, ces groupes de formation sont évalués plus favorablement au niveau du score de la satisfaction globale (moyenne de 3,2) tandis que le score de satisfaction globale diminue pour les groupes de composition de type « Hétérogène » (moyenne de 2,7). Cet écart se confirme également en analysant le score de la satisfaction lié à la collaboration avec une moyenne de 3,1 pour le type de regroupement « Homogène » et une moyenne de 2,9 pour le type de regroupement « Hétérogène ».

Durée. L'action de formation continue implique un format temporel court. La durée des sessions de formation s'échelonne sur un, deux ou trois jours principalement de manière consécutive. Par conséquent, le cycle de vie du groupe de formation est dépendant de la durée de la session de formation. En pratique, il peut être considéré qu'une durée de deux jours à minima est nécessaire pour créer une expérience de groupe. Il est à noter que la durée des sessions de formation est fortement corrélée selon le mode organisationnel : les sessions de formations organisées selon un mode « Intra-organisationnel » sont majoritairement limitées à une seule journée (66%) tandis que les sessions de formations organisées selon un mode « Inter-organisationnel » s'échelonnent selon une durée de deux à trois jours (environ 60%). Bien que modéré, le score de satisfaction globale des participants augmente en fonction de la durée de la formation. En effet, il peut être observé que le nombre de jours supplémentaires augmente le score de satisfaction (p-value 0,02). Par ailleurs, les sessions de formation ayant une durée minimale (un jour) est un format temporel qui génère des résultats négatifs. Cette tendance est particulièrement observée lors de l'analyse et le traitement des données textuelles issues des questions ouvertes. En effet, la durée des sessions de formation est un point d'insatisfaction particulièrement fort caractérisé par environ 74% de réponses négatives (exemple : « durée trop courte » « durée insuffisante »).

Mode organisationnel. Les moyennes des scores de satisfaction font apparaître que les sessions de formations selon le mode « Intra » ou « Inter-organisationnel » sont quasi-équivalentes. Néanmoins, il convient de nuancer ce résultat en fonction de l'initiateur de l'action de formation.

¹¹³ Tendence à déformer les descriptions de soi dans un sens favorable (Mc Crae et Costa, 1983) et à produire une image de soi positive (Tournois et al., 2000)

En effet, l'hypothèse avancée est que l'action de formation selon une initiative personnelle et non contraint par l'organisation est un indicateur de motivation et d'engagement du participant dans le processus de formation. Un écart est particulièrement observé selon le mode organisationnel. Dans le cas des sessions de formation organisés selon un mode « Inter-organisationnel », la majorité des participants (74,04%) s'engage dans une action de formation selon leurs propres initiatives tandis qu'une minorité est contrainte par une exigence hiérarchique (8,17%). À l'inverse, l'action de formation est majoritairement à l'initiative de la direction (45,81%) dans le cas d'une session de formation organisée selon un mode « Intra-organisationnel ». Nous pouvons avancer l'hypothèse que le score de satisfaction exprimée de manière positive peut impliquer un biais d'appréciation compte tenu de la relation hiérarchique sous-jacente.

De plus, il est à noter que le profil du participant diffère particulièrement selon le mode organisationnel au niveau du statut professionnel : les sessions de formation organisées selon un mode « Intra-organisationnel » sont essentiellement composés de participants ayant le statut de « Non-Cadre » (78%). Cette tendance est significativement différente pour les sessions de formation organisées selon un mode inter-organisationnel, qui sont essentiellement composés de participant ayant le statut « Cadre » (62%).

8.3 Analyse et traitement des données textuelles

L'analyse des données textuelles se focalise sur le traitement d'une seule question relative aux points positif(s) et points négatif(s) perçus lors de la session de formation. Le principal avantage est d'avoir à disposition des avis qualitatifs et par conséquent à forte valeur ajoutée. Cette analyse des données textuelles vise à identifier principalement les motifs de satisfaction et d'insatisfaction afin d'émettre des axes d'amélioration pour les sessions de formation à venir.

- Méthodologie

Parmi les 387 observations, 205 réponses ont été recensées correspondant ainsi à un taux de réponse d'environ 53%. Cette question ouverte¹¹⁴ invite le participant à émettre un avis, un commentaire ou une suggestion d'une manière personnalisée et argumentée. La méthode de traitement se déroule selon les quatre étapes suivantes :

- (1) Constitution du corpus
- (2) Analyse lexicale et partitionnement
- (3) Recherche de cooccurrences
- (4) Catégorisation et analyse thématique

Constitution du corpus. Un prétraitement des données textuelles est requis pour une analyse automatique du corpus. Cette étape consiste à un nettoyage et une normalisation des données¹¹⁵ pour permettre un traitement informatisé *via* des outils d'aide à l'analyse lexicale¹¹⁶. L'ensemble des réponses constitue un corpus unique¹¹⁷ qui représente 2701 occurrences pour 595 formes différentes (dont 321 hapax¹¹⁸, soit 53 % des formes et 11 % des occurrences). Par la suite, chaque commentaire est balisé comme une unité de codage c'est-à-dire comme « un segment de contenu à considérer comme unité de base en vue de la catégorisation et du comptage fréquentiel » selon la définition proposée par (Bardin D., 2001).

¹¹⁴Q13. Quels sont les points forts et les points faibles de la formation ?

¹¹⁵ Codage en UTF-8 , balisage, nettoyage (erreurs de saisie, suppression de majuscules, caractères spéciaux...) et lemmatisation

¹¹⁶ Iramuteq et Plateforme open-source TXM (Textométrie)

¹¹⁷ Format texte (ou .txt)

¹¹⁸ Mot qui n'a qu'une seule occurrence dans un corpus donné.

Analyse lexicale et partitionnement. Une analyse lexicale permet un comptage des termes les plus utilisés pour les lister à des fins d'interprétation. Cette étape consiste à extraire des formes actives, les formes supplémentaires et les hapax après une étape de lemmatisation. Par la suite, le corpus linguistique initial est partitionné en trois sous-corpus selon le type de polarité des réponses. Ainsi, les réponses sont assignées manuellement selon le type de polarité détaillé dans le tableau suivant :

Type de polarité	Description
Positive	Commentaire ayant exclusivement une tonalité positive
Mixte	Commentaire ayant une tonalité positive et négative
Négative	Commentaire ayant exclusivement une tonalité négative

Tableau 6.5 - Type de polarité des réponses

Cette méthode s'avère adaptée pour traiter les commentaires d'une manière exhaustive. En effet, des commentaires peuvent impliquer un chevauchement des thématiques dans les contenus des réponses (comme par exemple « + : petit groupe (5 personnes), formateur expert »). Ainsi, une itération par une analyse thématique n'est pas adaptée car cette méthode assignerait des commentaires selon une seule thématique au détriment d'une autre thématique. Compte du faible volume de données et du format des réponses, l'assignation est effectuée selon une lecture humaine et une interprétation sémantique. Le traitement manuel est facilité compte tenu qu'elles sont généralement précédé d'un signe tel que « + » ou « - ». Certaines réponses ne peuvent pas être assignées du fait qu'elles s'assimilent davantage à une suggestion (« Formation complémentaire nécessaire ») ou la formulation s'avère être ambiguë ou incomplète (comme par exemple « Fractionner la formation ? »).

Les statistiques textuelles de l'ensemble des unités linguistiques (occurrences, formes et hapax) catégorisé par type de polarité sont synthétisés dans le tableau suivant :

		Statistiques textuelles		
		Nombre d'occurrences	Nombre de formes	Nombre d'hapax ¹¹⁹
Type de polarité	Positive	948 (N=2701)*	370 (N=595)	187 (19.73% occurrences 60.32% formes)
		Exemple		
		« Journée supplémentaire et dédiée pour les cas pratiques » « Formation complète-Articulation équilibré entre la théorie et la pratique »		
	Mixte	1174 (N=2701)	370 (N=595)	226 (19.25% occurrences 61.08% formes)
		Exemple		
		« Bonne introduction en théorie mais durée insuffisante pour comprendre les modalités de calcul »		

¹¹⁹ Forme dont une seule occurrence est relevé dans un corpus

	Négative	Nombre d'occurrences	Nombre de formes	Nombre d'hapax
		579 (N=2701)	205 (N=595)	117 (20.21% occurrences 57.07% formes)
		Exemple		
		« Formation non consécutive/Salle trop petite » « Partie technique non adapté à notre secteur d'activité » « Vague interprétation du texte juridique »		

* : Nombre total de réponses

Tableau 6.6 - Statistiques textuelles selon le type de polarité et exemple de commentaires.

Recherche de cooccurrences. Dans notre cas d'application, l'analyse lexicale se focalise principalement sur les formes actives¹²⁰ de la catégorie grammaticale « Nom » pour les utiliser comme des mots-clés pivot pour les requêtes de cooccurrences. Ce type de requête permet d'extraire le terme sélectionné avec son contexte droit et son contexte gauche (e.g « formation » « médicale »). Il est à noter que l'extraction brute du corpus¹²¹ implique que les formes fléchies et la synonymie ne sont pas pris en compte lors de l'exploitation du corpus. Par exemple, au niveau morphologique, le mot « formateur » n'est pas distingué du genre féminin (« formatrice ») et des synonymes peuvent être utilisés (« animateur » ou « intervenant »). Par conséquent, un lexique de mots-clés relatifs à la thématique est recensé pour s'assurer de l'exhaustivité des unités linguistiques du corpus. L'ensemble des mots-clés pivot associé pour les requêtes de cooccurrences sont les plus fréquemment cités (>5) et détaillé en Annexe 13. Ainsi, cette analyse de cooccurrences permet d'affiner le partitionnement préalable notamment pour les réponses assignées selon une polarité « mixte ». De cette manière, l'analyse thématique des réponses permet de distinguer les contenus textuels selon une polarité positive et négative. Ainsi, la réponse contextualisée permet une interprétation sémantique comme indiquée dans le tableau suivant :

Mots-clés	Etiquette	Contexte gauche	Pivot	Contexte droit
groupe	positif	+ : petit	groupe	(5 personnes)-Formateur expert
	positif	qualité de l'intervenant,	groupe	homogène , support complet
	positif	+ : taille du	groupe	, durée adapté
	positif	temps d'échange avec le formateur et le	groupe	, cas pratique adapté en situation réelle
	positif	+ : échange avec le formateur et le	groupe	, formation généraliste, pas uniquement de la théorie
	positif	Formateur spécialisé, petit	groupe	adapté pour la prise de parole

Tableau 6.7 - Extrait du mot-clé pivot « groupe » associé à son étiquetage et contexte droit/gauche

L'analyse des voisinages des mots est variable : il peut être limité aux unités qui suivent ou qui précèdent immédiatement le mot-clé pivot (exemple « petit groupe », « groupe homogène »)

¹²⁰ Une forme active participe à la construction de classes lexicales en éliminant des formes des catégories grammaticales (préposition, auxiliaire...)

¹²¹ Non-lemmatisé (opération qui consiste à transformer les termes sous forme canonique)

ou élargi à plusieurs unités (« taille du groupe ») ou à la phrase « échange avec le formateur et le groupe »).

En complément, une analyse de cooccurrences similaires est appliquée concernant les adjectifs qualificatifs (>3) qui indiquent l'intensité du jugement d'évaluation d'une manière plus précise que la polarité binaire (Positive/Négative).

Catégorisation et analyse thématique. La diversité des réponses ouvertes implique de procéder à une analyse thématique afin de classer les termes lexicaux. Selon la définition proposée par (Deschenaux F., 2007), l'analyse thématique peut s'exercer de deux manières différentes : « (...) d'une manière inductive, en partant du corpus pour générer des thèmes ou de manière déductive en ayant, préalablement à l'analyse, identifié les thèmes à repérer ». Dans le cadre de ce travail, notre démarche est déductive du fait que les thématiques sont identifiées au préalable selon les catégories suivantes : contenu pédagogique, formateur, groupe de formation et logistique. À partir des 205 commentaires analysés, le tableau suivant présente la répartition des thématiques selon le type de polarité :

Thématique	Répartition en %		Total
	Polarité		
	Positive	Négative	
Contenu pédagogique	44 (54,32%)	37 (49,33%)	81 (100%)
Formateur	39 (92,86%)	3 (7,14%)	42 (100%)
Groupe	22 (68,75%)	10 (31,25%)	32 (100%)
Logistique	17 (24,29%)	53 (75,71%)	70 (100%)

Tableau 6.8 - Répartition thématique en valeurs absolues et % selon le type de polarité positive/négative

- Résultats

A partir des résultats obtenus, plusieurs constats peuvent être observés. En complément de la polarité positive/négative, l'analyse de cooccurrences de la catégorie grammaticale « Adjectif » met en évidence plusieurs phénomènes linguistiques :

-Le formateur est un facteur déterminant lors de l'évaluation de la session de formation. Parmi les réponses existantes, le formateur est apprécié majoritairement selon une polarité positive (92,86%) tandis que les réponses selon une polarité négative sont peu nombreuses (7,14%). Ses aptitudes professionnelles et plus précisément ses compétences dans un domaine spécialisé sont particulièrement citées (exemple : « Expertise du formateur », « Expérience du formateur » ou « Formateur professionnel »). Le style d'animation du formateur est également apprécié à travers des réponses comme par exemple « Disponibilité du formateur », « Ecoute de la formatrice », « Pédagogie du formateur ». Il est à noter que le terme « échange » est principalement associé avec le terme « formateur ».

-La durée des sessions de formation est un motif d'insatisfaction particulièrement significatif. Les participants associent le format temporel de la session de formation avec des adjectifs négatifs comme « trop court » (10) et « insuffisant » (6). Des formulations proches comme par exemple « durée de formation pas adapté », « durée un peu limitée dans le temps » sont également citées. Le format court des sessions de formation est également jugé inadéquat avec le contenu pédagogique (« durée trop limitée pour formation de type spécialisée ») trop dense (« pas le temps pour les cas pratiques ») et peu adapté pour les échanges (« peu de temps pour les questions personnalisées »).

- L'analyse lexicale liée à la thématique « Contenu pédagogique » relève que le mot « pratique » ne peut pas être considéré comme étant un adjectif mais doit être interprété selon le groupe nominal « cas pratique ». Ce type d'exercice pédagogique lié à l'évaluation du contenu pédagogique est particulièrement apprécié parmi l'ensemble des commentaires (78%) avec quelques commentaires négatifs liés au niveau pédagogique des participants (« cas pratique trop technique » ou inversement « cas pratique trop généraliste »).

8.4 Interprétation

En fonction des finalités d'application préalablement formulées, les résultats obtenus permettent de justifier des choix rationnels pour l'aide à la planification des sessions de formations et l'aide à la décision stratégique.

- Aide à la planification

Les résultats obtenus à partir de cette analyse statistique permettent de définir les caractéristiques (ou valeurs) d'un groupe de formation qui influencent favorablement la satisfaction des participants. Des observations stables peuvent être identifiées et justifier les paramétrages de fonction-objective (ou *fitness*) lors de la phase d'initialisation de l'AG. Ils peuvent être synthétisés de la manière suivante :

-Type de regroupement : Homogène

Les interactions et échanges entre pairs ayant des caractéristiques similaires sont favorisés pour échanger sur des problématiques métiers communs. L'homogénéité des profils des participants dépend des attributs qu'on lui confère. Dans notre cas d'application, l'homogénéité du groupe est principalement lié aux attributs professionnels du participant ([Voir section 7.1.3](#)).

-Taille : 5-6 personnes

Les scores de satisfaction au niveau de la taille des groupes sont équilibrés et aucune(s) modalités ne se démarque nettement pour affirmer la prépondérance d'une taille optimale. Cette modalité de taille de 5-6 personnes résulte d'un choix médian selon les données à disposition.

-Mixité : H=F

La composition paritaire des groupes de formation influence favorablement la satisfaction des participants. D'un point de vue juridique et éthique, il ne peut pas être envisagé de composer délibérément des groupes de formation exclusivement d'hommes ou de femmes.

Dans le cas d'application d'une planification de type exploratoire, des critères complémentaires peuvent être pris en compte pour optimiser la session de formation. Au niveau du groupe, les caractéristiques (ou valeurs) de nature sociale citées précédemment sont communes. Néanmoins, les résultats obtenus par l'analyse statistique apportent des orientations complémentaires notamment au niveau des conditions organisationnelles et logistiques :

-Mode organisationnel : Inter-organisationnel

A l'instar du mode « Intra-organisationnel », l'action de formation occasionne des rencontres professionnelles informelles et principalement selon une initiative personnelle. Ainsi, les échanges de pratiques d'expériences professionnelles sont favorisés entre des personnes externes qui ne sont pas contraint par une obligation hiérarchique.

-Lieu géographique : Département

La session de formation en présentiel organisée selon un mode « Inter-organisationnel » s'effectue à l'extérieur des organisations auxquels les participants sont rattachés. Par conséquent, la proximité géographique selon un périmètre défini (région ou département) conditionne leurs présences physiques.

-Durée : 2 jours minimum

La durée de formation est élément déterminant pour améliorer la satisfaction des participants. L'analyse lexicale a mis en évidence que le format court de l'action de formation est un facteur d'insatisfaction particulièrement élevée et peu adapté avec un contenu pédagogique fortement spécialisé.

- Aide à la décision stratégique

L'évaluation de la session de formation est une étape primordiale pour mener des actions de formation de qualité. Les résultats obtenus permettent de définir les indicateurs à mettre en évidence au sein du tableau de bord analytique :

-Analyse globale de l'activité de formation

Ce niveau d'analyse présente une analyse synthétique de l'activité de formation à travers les indicateurs suivants : (i) Score global de satisfaction, (ii) Répartition thématique des sessions de formation.

-Analyse de l'activité du formateur

Les résultats ont démontré l'importance du rôle du formateur durant la session de formation. Le suivi de son activité est un indicateur stratégique pour le gestionnaire de formation. Au sein du tableau de bord analytique, une vue est dédiée à l'activité du formateur à travers les indicateurs suivants ([Voir section 6.3.4](#)) : (i) Profil professionnel du formateur et (ii) Score global de satisfaction par groupe de formation.

8.5 Conclusion

L'analyse statistique de l'activité de formation s'inscrit dans une démarche expérimentale globale. D'une part, les résultats obtenus permettent de soutenir les choix rationnels pour l'aide à la planification. D'autre part, l'aide à la décision stratégique devient un enjeu essentiel pour le pilotage de l'activité de formation et qui nécessite au préalable l'identification des indicateurs.

L'analyse des questions fermées fait l'objet d'un traitement statistique simple et multivarié tandis que l'analyse des réponses ouvertes rédigées en langage naturel nécessite un traitement lexical. Dans notre cas d'étude, la méthode manuelle est possible compte tenu du faible volume des données mais n'est pas recommandé pour analyser un corpus volumineux. Néanmoins, de nombreuses perspectives peuvent être envisagées. D'une part, une analyse lexicale peut être approfondie notamment les catégories grammaticales dont les adverbes qui peuvent marquer l'intensité des adjectifs qualificatifs des commentaires négatifs ou positifs. D'autre part, cette approche manuelle permet d'étiqueter des patrons (ou *patterns*) de catégories morfo-syntaxiques qui peuvent être stockées au sein d'un dictionnaire (ou une ontologie). Ainsi, cette méthode permet un traitement semi-automatisé des extractions des unités linguistiques pour faciliter l'exploitation et l'analyse du corpus qui devient volumineux. D'autres perspectives peuvent être

envisagées à travers l'amélioration du questionnaire de satisfaction. L'intégration des questions supplémentaires qui élicitent les préférences spatio-temporelles des participants (disponibilité, durée de formation, jours de formation consécutifs, zone géographique) vise à analyser des nouvelles variables et par conséquent obtenir des informations complémentaires.

Partie 3-Expérimentation

L'évaluation d'un système d'information est une tâche multidimensionnelle complexe compte tenu des différents niveaux d'évaluation (ergonomie, architecture technique, sécurité...). Dans un contexte industriel, l'évaluation de la qualité d'un système d'information est principalement abordée selon des normes de qualité définie par l'OSI¹²². Les standards normatifs sont divers et référencés selon les objectifs d'évaluation visés¹²³.

Cette troisième partie présente la démarche d'expérimentation axée sur l'évaluation de la performance du prototype d'application développé. A travers ce dernier chapitre, deux niveaux d'évaluation sont présentés :

- Evaluation des résultats de l'AG selon des métriques de performances prédéfinies ;
- Evaluation du fonctionnement du prototype d'application à partir des tests fonctionnels, techniques et utilisateurs.

L'ensemble des scripts, les scénarios de tests et les résultats détaillés des expérimentations sont consultables sur un site internet dédié au projet de R&D¹²⁴.

¹²² Organisation Internationale de Normalisation

¹²³ ISO 9126 (Génie logiciel : qualité des produits) ISO 25000 (Ingénierie des systèmes et du logiciel : exigences de qualité des systèmes et du logiciel et évaluation), ISO 9241 (Ergonomie de l'interaction homme-système)

¹²⁴ www.lexianerd.fr

Chapitre 9. Evaluation de la performance du prototype d'application

Ce dernier chapitre présente une évaluation de la performance du prototype d'application selon deux niveaux d'expérimentation.

Le premier niveau d'évaluation vise à mesurer les paramètres et les résultats obtenus par l'AG décrit lors du chapitre 7 en détaillant le jeu de données expérimentales, les mesures de performance utilisées et les résultats obtenus. Le second niveau d'évaluation est centré sur le fonctionnement et la performance du prototype d'application à partir de tests d'utilisateurs, fonctionnels et techniques.

9.1 Méthode et évaluation de l'AG

L'évaluation des résultats est une étape primordiale pour mesurer la capacité de traitement de l'AG selon les paramétrages définis détaillés dans la partie précédente ([Voir section 7.3](#)). Notre démarche expérimentale s'inspire des méthodes et des mesures d'évaluation des systèmes de recherche d'information¹²⁵. Ce choix est principalement justifié du fait que le système repose sur un principe de requête/réponse. Au préalable, ce cadre d'évaluation implique d'avoir à disposition un jeu de données, la constitution d'une collection des requêtes et la définition des critères d'affectations pertinents. Ainsi, cette première série d'expérimentation vise à évaluer d'une manière globale les résultats obtenus à partir des requêtes multicritères puis d'évaluer les paramètres de l'AG.

9.1.1 Jeu de données expérimentales

L'évaluation d'une méthode de résolution d'algorithme dépend sensiblement du jeu de données utilisé. Par conséquent, il convient d'assurer au préalable la qualité du jeu de données pour garantir une évaluation optimale des résultats.

Au niveau de la volumétrie, l'état de l'art scientifique relève la difficulté d'établir un consensus pour déterminer un seuil standard compte tenu des contextes d'expérimentation différents et par conséquent peu comparables. Néanmoins, les bonnes pratiques conviennent de favoriser un jeu de données volumineux pour améliorer la significativité des résultats.

Dans ce cas d'application, la principale limite du jeu de données est le nombre d'observations à disposition (387 observations). Le volume de données peut être considéré comme étant faible car il restreint l'exploration dans l'espace de recherche nécessaire pour un AG. De plus, il peut également impliquer un problème de démarrage à froid¹²⁶ lors de la simulation des requêtes de test. Pour pallier ce biais d'évaluation, notre méthode consiste à optimiser le jeu de données à disposition en le complétant avec des données artificielles (ou fictives). Ainsi, cette approche nous permet de maximiser l'espace de recherche inhérent à l'utilisation d'un AG et d'évaluer la performance du prototype en utilisant un nombre volumineux de données.

- Données réelles

Les données « réelles » correspondent à un ensemble d'individus existants¹²⁷. Les valeurs associées ont été collectées manuellement à partir de plusieurs sources de données et stockées au

¹²⁵ Approche de type laboratoire (*laboratory-based model*) initiée par Cleverdon (Cleverdon, 1967)

¹²⁶ Incapacité du système à fournir des résultats compte tenu du faible volume de données

¹²⁷ Compte tenu que les données ne sont pas anonymisées, la déclaration est soumise à une réglementation juridique liée au RGPD selon des finalités d'utilisations scientifiques.

sein de l'application existante de type CRM¹²⁸ ([Voir section 6.1.4](#)). Néanmoins, ce jeu de donnée induit des limites inhérentes à une saisie manuelle des valeurs qui partiellement peut ne pas correspondre au degré de précision attendu pour une évaluation optimale des résultats. Par conséquent, l'utilisation de ce jeu de données nécessite au préalable une étape corrective qui consiste à nettoyer les données brutes (doublons, valeurs aberrantes, valeurs incomplètes, valeurs obsolètes liée à la mobilité professionnelle ...).

Parmi les profils existants, seuls les individus qui ont préalablement effectué une formation de type formel (session de formation) ou non-formel (colloque) sont inclus dans le jeu de données expérimentales¹²⁹. Notre hypothèse est de considérer que les participants ayant un historique de formation sont dans une démarche volontariste et ont davantage de probabilités de vouloir renouveler une expérience de formation. Ainsi, ce jeu de donnée comporte 335 individus ayant des caractéristiques équilibrées selon certaines variables (« Genre », « Age », « Ancienneté » ...). Néanmoins, il est à noter un déséquilibre au niveau des profils des individus qui majoritairement appartiennent à une catégorie socio-professionnelle spécifique. Comme indiqué dans le chapitre précédent ([Voir section 8.2.1](#)), certains profils d'individus sont sous-représentés notamment les hommes ayant un statut Non-Cadre et exerçant dans certains secteurs d'activités (secteur culturel, syndicats)¹³⁰. Ainsi, cette limite oriente notre démarche à compléter le jeu de données réel avec des données artificielles (ou fictives) afin de corriger la volumétrie mais également la représentativité des données à disposition.

- Données artificielles

La génération de données artificielles (ou fictives) vise à compléter le jeu de données initial pour évaluer et analyser les résultats de l'AG. Cette démarche implique de créer des données dites « réalistes » incluant des valeurs d'attributs proches des données réelles. La principale difficulté est de générer des données qualitatives (ou catégorielles) dont notamment la variable « Profession ». Cette contrainte implique qu'une simple génération de données fictives d'une manière exclusivement aléatoire n'est pas adaptée. Par conséquent, la méthode de génération d'un jeu de données artificielle s'appuie sur une approche semi-automatisée. Une première génération de données selon une approche aléatoire est menée à partir d'un script pour certaines variables (« Genre », « Nom », « Prénom », « Ville », « Statut », « Type », « Département » et « Région »). Par la suite, cette première liste est complétée en affectant manuellement des valeurs pour les variables catégorielles (« Profession » et « Direction ») qui sont extraites à partir d'un référentiel-métier spécialisé existant¹³¹. L'intervention et l'appréciation humaine est requise pour vérifier et éviter des incohérences dans le profil du participant (exemple : la profession « Assistant de formation » doit être associée obligatoirement à une direction RH et un statut hiérarchique équivalent). Ainsi, ce jeu de donnée comporte 625 individus ayant des caractéristiques et selon une proportion similaire au jeu de données réel.

Au final, cette approche méthodologique résulte un jeu de données « mixte » c'est-à-dire incluant un ensemble d'individus dont les valeurs peuvent être fictives et réelles distingués par un symbole associé au sein de la base de données. Le nombre total d'individus s'élève à 960 dont 335 individus « réels » et 625 individus « fictifs ».

L'Annexe 14 détaille la distribution du jeu de données expérimentales.

¹²⁸ Customer Ressource Manager/Base de données de type relationnel

¹²⁹ Typologie des profils : prospect, client, formation, colloque

¹³⁰ Plus de 80% des participants sont des « Cadre » dont parmi 30% travaille dans le secteur hospitalier, 14% pour le secteur « Université », 9% pour le secteur Conseil Régional, 6% pour les collectivités territoriales

¹³¹ Répertoire des métiers de la fonction publique territoriale de la CNFPT (Secteur public)

- Importation

L'importation vise à implémenter les données expérimentales au sein du prototype d'application connectées à travers une API¹³². Les données d'entrées sont importées sous format tabulaire dont le tableau suivant présente un extrait d'une manière simplifiée :

pid	type	genre	nom	prenom	type_etablissement	secteur_activite	statut
f8	Participant	F	*Poulain	Alexandra	Office Public d'Habitat	Habitat Social	Cadre
f9	Colloque	H	*Masson	Alphonse	Grande Ecole	Université/ COMUE/ GE	Cadre
f10	Participant	H	*Nicolas	Claude	Ville	Collectivité territoriale	Cadre
f11	Colloque	H	*Barbier	Olivier	Ville	Collectivité territoriale	Non-Cadre
f12	Colloque	H	*Blandiard	Emmanuel	Ville	Collectivité territoriale	Cadre
f13	Participant	H	*Leveque	Jean	EPCI	Collectivité territoriale	Cadre+
....

Tableau 7.1 - Extrait du tableur de données artificielles (ou fictives)

Selon la requête effectuée, le prototype d'application interroge l'API afin d'extraire les sorties (ou résultats) souhaités. La principale limite de cette approche est qu'une première exécution de l'importation des données est nécessaire afin de détecter des erreurs en situation réelle (saisie des variables, format des données, doublons...). L'interface utilisateur de l'API permet d'effectuer des mises à jour de façon simplifiée à travers le fichier de données et synchronise automatiquement l'ensemble des modifications effectuées côté client. Ainsi, des corrections nécessaires ont été effectuées en amont pour assurer une standardisation de l'ensemble des données pour mener une évaluation des résultats.

9.1.2 Collection des requêtes de test

Cette étape vise à définir un ensemble de requêtes afin de simuler des tâches de planification. Comme indiqué lors du chapitre 7 ([Voir section 7.1.1](#)), les besoins de formations ne sont pas génériques impliquant que les tâches de planification diffèrent selon les demandes des commanditaires. Par conséquent, les requêtes sélectionnées ne sont pas exhaustives et visent à tester des cas d'utilisations réalistes. Selon (Buckley, C., 2017), la collection des requêtes fixes doit compter entre 25 et 50 requêtes pour garantir une évaluation optimale. Dans le cadre de notre évaluation, la collection de test comporte un nombre total de 45 requêtes multicritères. La majorité des requêtes (41 requêtes) sont déterminées selon les objectifs de performance préalablement définis ([Voir section 7.3](#)) selon les critères suivants :

-Mode « Inter-organisationnel »

Le choix du critère selon le mode « Inter-organisationnel » est justifié pour plusieurs motifs. D'une part, il permet d'évaluer la performance de l'AG à traiter la complexité combinatoire de participants qui sont issus d'organisations différentes et par conséquent l'exploration d'un espace de recherche large. D'autre part, la planification des sessions de formation selon ce mode organisationnel est la demande la plus fréquente des commanditaires auprès de l'organisme de formation.

-Type de regroupement « Homogène »

¹³² Google Drive API

Le critère selon le type de regroupement « Homogène » est particulièrement testé afin d'évaluer le score d'évaluation attribué pour chaque groupe de formation selon les sous-critères associés (« Secteur d'activité », « Type d'établissement », « Statut », « Type de personne », « Historique de formation ») (Voir section 7.3.2).

En complément, des critères sociaux et logistiques peuvent être spécifiés dans les requêtes (« Taille », « Mixité », « Département » et « Région »).

Le tableau suivant présente une analyse descriptive de la collection des requêtes répartie selon le jeu de données et le nombre de critères :

Jeu de données « Réel »						
	Nombre de critères					Nombre total
	3	4	5	6	7	
Nombre de requêtes	1	4	6	2	1	14
%	7,1%	28,5%	42,8%	14,3%	7,1%	100%

Jeu de données « Mixte »						
	Nombre de critères					Nombre total
	3	4	5	6	7	
Nombre de requêtes	1	4	12	14	0	31
%	3,2%	12,9%	38,7%	45,2%	0%	100%

Tableau 7.2 - Statistique descriptive de la collection de requêtes selon le nombre de critères associés et valeur en % par type de jeu de données

Les requêtes sont spécifiées manuellement et exécutées une par une à partir de la collection préalablement établie. Dans un premier temps, ces requêtes sont testées selon le jeu de données « Réel » puis avec le jeu de données « Mixte ». Cette distinction de tests selon la nature du jeu de données présente plusieurs intérêts. D'une part, les requêtes testées avec le jeu de données « Réel » permet de bénéficier d'une évaluation des résultats basée sur une « vérité terrain » et d'assurer par la suite des expérimentations à partir d'observations réelles. D'autre part, les requêtes testées avec le jeu de données « Mixte » permettent d'observer l'évolution du score d'évaluation et le fonctionnement de l'AG en cas d'ajout de données (ou d'individus) supplémentaires. Ainsi, chaque requête renvoie une liste ordonnée de groupes de formation potentiels qui fait l'objet par la suite d'une évaluation selon des mesures de performances.

La collection des requêtes et la distribution des critères et des valeurs associées sont détaillées en Annexe 15.

9.1.3 Evaluation globale

Les résultats obtenus à partir de la collection de requêtes multicritères préalablement déterminées sont évalués à partir d'une valeur unique soit le taux de précision moyenne (ou *Mean Average Precision*). Cette mesure d'évaluation consiste à mesurer la capacité du système à proposer des individus correctement affectés en adéquation avec les critères de la requête spécifiés par l'utilisateur. D'une part, la valeur moyenne des taux de précision est effectuée par requêtes puis par critères. L'automatisation de l'évaluation pour l'ensemble des résultats est effectuée partiellement à partir d'un script afin de minimiser le taux d'erreur.

- Evaluation par requêtes

Selon la formule 2.1, le principe de calcul du taux de précision moyenne repose sur une agrégation de l'ensemble des taux de précision des critères (noté $\{c1, \dots, cn\}$) pour une requête spécifique (noté k) :

$$Prec_{moy}(R) = \sum_{k=1}^{|Q|} \frac{Prec\ k(c1, \dots, cn)}{Q} \quad (2.1)$$

Le taux global de précision moyenne pour l'ensemble des requêtes (noté $|Q|$) est plutôt positif avec 91% des valeurs correctement classées. Parmi les 45 requêtes testées, 16 requêtes obtiennent un taux de précision inférieur à 100% selon la répartition suivante :

	Taux de précision				Totaux
	50-65%	65-85%	85-99%	100%	
Nombres de requêtes	2	6	8	29	45
%	4%	13%	17%	64%	100%

Tableau 7.3 - Répartition et nombre de requêtes selon le taux de précision

Ainsi, ce taux de précision moyenne par requête (noté (R)) nous permet d'identifier les requêtes qui obtiennent des taux de précision à améliorer.

- Evaluation par critères

Une analyse complémentaire vise à mesurer le taux de précision moyenne par critères selon le même principe de calcul. Une évaluation globale des résultats est effectuée selon une valeur moyenne des taux de précision (ou *Mean Average Precision*) par critères pour l'ensemble des requêtes.

Selon la formule 2.2, le principe de calcul du taux de précision moyenne repose sur une agrégation de l'ensemble des taux de précision des critères (noté $\{c1, \dots, cn\}$) pour un critère spécifique (noté C) :

$$Prec_{moy}(C) = \sum_{k=1}^{|Q|} \frac{Prec\ k(c1, \dots, cn)}{C} \quad (2.2)$$

Ainsi, cette mesure d'évaluation vise à identifier les critères des requêtes qui peuvent influencer la capacité du traitement à ne pas classer correctement les résultats. Parmi les 45 requêtes testées (noté $|Q|$), les résultats obtenus sont plutôt positifs avec un taux global de précision moyenne de 84,86 %.

Le tableau suivant détaille le taux de précision par critères de décision selon un taux de précision en % soit :

	Critères	Taux de précision (%)
$C1$	Mode organisationnel	98.89
$C2$	Type de regroupement	80.13

C3	Statut	81.30
C4	Secteur d'activité	97.23
C5	Type d'établissement	80.00
C6	Type de personne	62.00
C7	Historique de formation	91.67
C8	Taille	99.72
C9	Mixité	42.50
C10	Département	100.00
C11	Région	100.00

Tableau 7.4 - Taux de précision moyenne par critères pour l'ensemble des requêtes

Le détail des taux de précision et des précisions moyenne par critères et par requêtes sont présentés en Annexe 17.

A partir des résultats obtenus, plusieurs observations peuvent être constatées :

-Le taux de précision du critère « Mode organisationnel » est quasi-maximal. Des erreurs d'affectation de personnes sont constatées pour une seule requête (requête n°25). Pour ce cas précis, le critère du mode « Inter-organisationnel » génère des groupes de formation dont les personnes travaillent dans la même organisation. On peut en déduire que ces erreurs d'affectation est principalement liées à la modalité de taille minimale formulée (2-4 personnes) et à la restriction de l'espace de recherche.

-Parmi l'ensemble des critères, le critère « Mixité » est le moins performant avec un taux de précision équivalent à 42,5%. Parmi la collection des requêtes, le critère « Mixité » est spécifié dans 10 requêtes selon la répartition suivante :

-8 requêtes incluent la modalité « H=F »¹³³

-1 requête inclut la modalité « H>F »¹³⁴

-1 requête inclut la modalité « F>H »¹³⁵.

Cette baisse du taux de précision concerne essentiellement les requêtes selon la modalité « H=F ». Elle est principalement justifiée liée à la modalité de la taille de groupe choisie. En effet, parmi les 8 requêtes testées, 4 requêtes incluent une modalité de taille « 5-6 » qui génèrent par défaut des groupes de taille impaire (5 personnes). Par conséquent, le critère de parité ne peut pas être respecté et nous oriente vers un défaut de paramétrage de l'AG.

Pour la requête avec un critère « H>F », 9 groupes incluent 80% d'hommes et un seul groupe est exclusivement composé d'hommes.

Pour la requête avec un critère « F>H », la totalité des groupes incluent majoritairement des femmes à la hauteur de 62,5%. Ainsi, le critère de la majorité est respecté. En comparaison, un écart est à noter dans la distribution des genres avec une proportion plus élevée pour les hommes que pour les femmes.

-Un écart du taux de précision est également observé selon la technique de filtrage utilisée. Comme décrit dans le chapitre 6 ([Voir section 6.3.2](#)), l'utilisateur peut choisir les valeurs des paramètres selon deux types de filtrages : le filtrage par liste de champs ou le filtrage par pondération. Parmi la collection des requêtes, 9 requêtes spécifient une valeur (tous critères confondus) selon une technique de filtrage par pondération. Cette méthode induit une baisse du

¹³³ Nombre d'hommes équivalent au nombre de femmes

¹³⁴ Nombre d'hommes supérieur au nombre de femmes

¹³⁵ Nombre de femmes supérieur au nombre d'hommes

taux de précision pour les critères suivants : « Statut », « Type de personne » et « Type d'établissement ». A titre exemple, le taux de précision du critère « Statut » spécifié selon un filtrage de pondération est de 56% tandis qu'il est de 97% quand la valeur est émise selon une technique de filtrage par liste de champs. Par conséquent, nous pouvons déduire que la fonctionnalité liée à la technique de filtrage par pondération n'est pas correctement implémentée et lié à un défaut de paramétrage des préférences au sein du code-source. Ce constat est également observé lors de l'évaluation de la précision moyenne par requête. Parmi la collection du jeu de requêtes, 10 requêtes incluant une valeur par pondération directe, la précision moyenne pour l'ensemble de ces requêtes est de 85% tandis que la précision moyenne pour l'ensemble des requêtes formulé exclusivement selon des valeurs par liste de champs est de 95%.

9.1.4 Evaluation des paramètres de l'AG

Au niveau du paramétrage de l'AG, l'évaluation se répartit de la manière suivante :

- Score d'évaluation
- Nombre d'itération
- Code-source

- Evaluation du score d'évaluation

Le score d'évaluation est fixé pour chaque groupe de formation selon la fonction objective (ou fitness) préalablement déterminé ([Voir section 7.3.2](#)). Chaque requête multicritère de la collection de test génère une liste ordonnée de groupes de formations avec un score d'évaluation converti selon une échelle de valeur comprise en 0 et 5. Pour l'ensemble des requêtes testées, une analyse globale des scores d'évaluation est récapitulée dans le tableau suivant :

Nombre de requêtes	Liste-résultat		Scores d'évaluation			
	Id unique	Nombre de groupe P@10	min	moy	max	écart-type
45	100%	100%	0,430	2,676	5	1,324

Tableau 7.5 - Score minimal, moyen, maximal et écart-type des listes de résultats

L'analyse descriptive des résultats révèle que chaque individu affecté au sein d'un groupe de formation correspond à un identifiant unique. Ainsi, aucun doublon n'est constaté au sein de chaque groupe parmi la liste proposée.

La précision à dix (noté P@10) évalue la capacité de l'AG à générer une liste de résultats pertinentes pour les 10 premiers groupes renvoyés. A partir des résultats obtenus, une observation majeure est à souligner : chaque requête génère exclusivement une liste de résultats avec des scores ex-æquo (100%). Cette proportion de scores ex-aequo induit un défaut d'évaluation lors du traitement de l'AG. A partir de ces résultats, on peut déduire que le calcul du score d'évaluation n'est pas suffisamment discriminant pour départager les participants potentiels parmi le nombre total d'individus trouvés. Ce problème de redondance des résultats est particulièrement observé dans l'utilisation des AG traitant des problèmes de regroupement. Initié par (Falkenauer, E. 1998), l'un des problèmes soulevés est la génération des solutions identiques liée à la stratégie de permutation de la méthode élitiste. L'AG se retrouve « coincé » avec des « parents » qui génère des regroupements (ou descendances) équivalents. Ce défaut de paramétrage peut être pallié en paramétrant des opérateurs génétiques plus complexes (mutation plus élevé, point de croisement multiple...). Au niveau de l'ergonomie de l'interface-utilisateur, ces scores d'évaluation ex-aequo présentent le désavantage de générer une liste de résultat qui ne peut pas être ordonnée d'une façon optimale. Malgré des scores équivalents, cette présentation des résultats contraints l'utilisateur à faire défiler la page pour consulter le dernier groupe de formation potentiel présenté

en fin de liste. Ce défaut d'équité lors de la sélection des réponses nommé « biais d'ordonnement » théorisé par (Becker S., 1954) démontre que la probabilité de sélectionner des items est plus faible quand ceux-ci sont classés en bas liste.

Des variations sont également à souligner concernant les scores d'évaluation attribués en fonction du nombre total d'individus. La collection de requêtes de test présente des écarts considérables selon les critères formulés (intervalle de 11 à 520 individus total). Il est constaté que la dispersion maximale et minimale des valeurs (ou l'écart min_max¹³⁶) est déterminée en fonction de la taille de la population explorée. Une variation progressive des scores d'évaluations attribués est particulièrement observée en fonction du nombre total de participants extraits. Les requêtes résultant des scores d'évaluation au-delà de la moyenne globale (>2.6 sur une échelle de 0 à 5) explorent un nombre total d'individus inférieurs à 90 personnes en moyenne tandis que les requêtes résultats affichant un score faible (<1,73 sur une échelle de 0 à 5) sont principalement des requêtes qui parcourent un espace de recherche plus large. La figure ci-joint met en évidence que le score d'évaluation est dépendant du nombre total d'individus trouvés où l'on peut observer que les requêtes (n°16 et n°22) obtiennent un score d'évaluation faible en parallèle d'un nombre total de personnes trouvés maximal (>500).

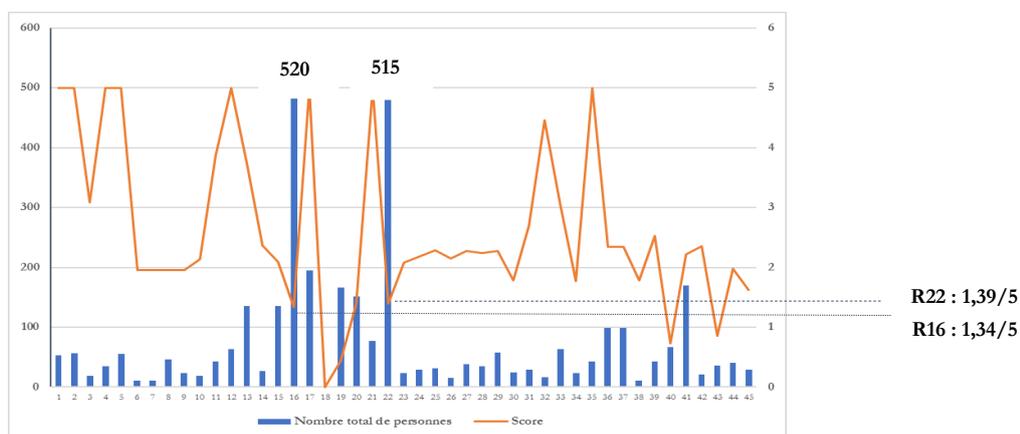


Figure 8.1 - Courbe des scores d'évaluations avec le nombre total de personnes trouvées (ordonnées) et requêtes associées en diagramme en barre (abscisses)

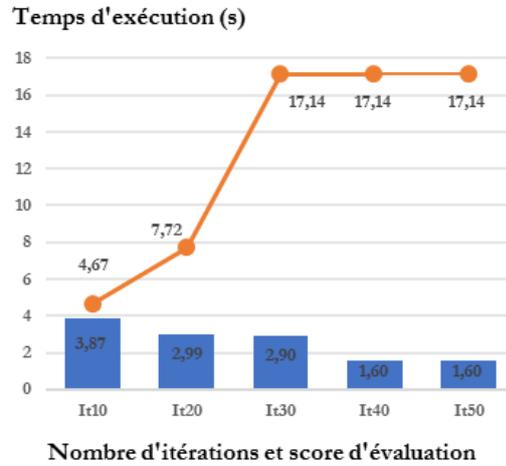
L'Annexe 16 détaille le score en pourcentage des identifiants uniques, le taux de précision à 10 et les scores d'évaluation attribués par requêtes.

- Evaluation du nombre d'itérations

Au niveau des paramétrages de l'AG, le temps d'exécution est également dépendant du nombre d'itérations fixé préalablement lors de la phase d'initialisation. Une série de tests préliminaire a été effectuée selon un nombre d'itérations de départ fixé à 10 jusqu'à 50. A partir des résultats obtenus, ce critère d'itération a été fixé à 10 compte-tenu que le score d'évaluation ne s'améliore pas d'une manière significative et implique un temps d'exécution plus élevé. A titre d'exemple, les figures suivantes présentent le score d'évaluation et le temps d'exécution associés selon le nombre d'itération :

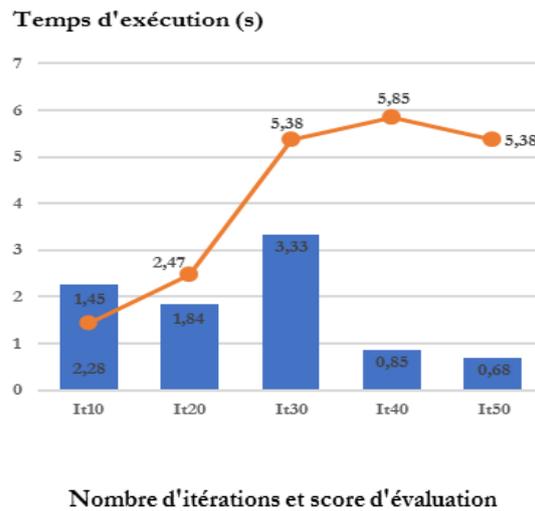
¹³⁶ Ecart obtenu entre le score obtenu minimal et le score obtenu maximal

Taille N	43
Critères	Valeurs
C1	Inter-organisationnel
C2	Homogène
C4	CH/CHU/CHS
C5	Type d'établissement
C8	2-4



(a)

Taille N	38
Critères	Valeurs
C1	Inter-organisationnel
C2	Homogène
C3	Non-Cadre
C4	Culture
C7	Marché Public
C8	5-6



(b)

Figure 8.2 - Résultats de requêtes selon un mode « Inter-Organisationnel » avec courbe du temps d'exécution (ordonnées) et diagramme à barre du nombre d'itération de 10 à 50 et le score d'évaluation associé (abscisses) avec (a) requête n°11 incluant 5 critères (b) requête n°27 incluant 6 critères

Le nombre d'itération minimal s'impose également par la taille de la population fortement réduit liée au filtrage des requêtes en comparaison avec la taille de la population initiale. Cette première série de test nous démontre une relation de proportionnalité du nombre d'itération avec le temps d'exécution mais également avec la taille de la population.

La liste des requêtes avec le nombre d'itération en comparaison avec le temps d'exécution et le score d'évaluation associé est présentée en Annexe 19.

- Evaluation du code-source

L'évaluation du code-source vise à vérifier l'exécution des paramètres fixés par défaut lors du codage de l'AG décrit dans le chapitre 7 ([Voir section 7.3.1](#)) soit la parité des genres et la taille des groupes fixé à 5-6 personnes.

Parité du groupe. Comme décrit dans le chapitre précédent, il s'est avéré nécessaire d'implémenter une partie du code-source en « dur » pour éviter de générer des groupes composés exclusivement d'un genre particulier (homme ou femme). L'objectif est de générer des groupes de formation paritaire selon un nombre de femmes et d'hommes équivalents. Parmi la collection des 45 requêtes existantes, seuls les requêtes où le critère « Mixité » n'est pas spécifié et les groupes comportant un nombre pair de personnes ont été évaluées correspondant à un total de 17 requêtes. Les résultats sont plutôt positifs avec un score global de 92%.

Taille du groupe. Parmi la collection des 45 requêtes, 8 requêtes ne spécifient pas une modalité de taille de groupe précise. Par conséquent, les résultats de ces requêtes ont été analysés. Ils sont particulièrement négatifs avec 75% des requêtes qui génèrent une liste de résultats incluant des groupes de plus de 16 personnes au lieu de 5 à 6 personnes initialement paramétré par défaut.

9.1.5 Exemples de résultats

Cet exemple d'une requête détaille un cas d'utilisation en détaillant les étapes de la requête multicritère, le détail des taux de précision par groupe de formation et les scores brutes des itérations.

L'utilisateur spécifie une requête Q_1 selon des critères spécifiques ($C_1, C_2, C_3, C_4, C_5, C_6$) associées aux valeurs suivantes :

	Critères	Valeurs
C1	Mode organisationnel	Inter
C2	Type de regroupement	Homogène
C3	Statut	Non-cadre
C4	Secteur d'activité	Université/GE ¹³⁷ /COMUE ¹³⁸
C7	Historique	Social
C8	Taille	5-6

Tableau 7.6 - Valeurs des critères de la requête Q_1

En réponse à la requête Q_1 , une liste de résultats de 10 groupes de formation potentiels est générée selon un nombre total de 35 individus.

Le tableau suivant détaille le taux de précision pour chaque critère. Les valeurs entre crochets représentent le rang du groupe dans la liste restituée. Le taux de précision est de 100% pour l'ensemble des critères à l'exception du critère « Historique de formation » qui présente un taux de précision variable selon les groupes. Compte tenu que le critère « Type d'établissement » n'a pas été précisé lors de la formulation de la requête, les valeurs « Grande Ecole » et « Université » sont représentés.

Le critère « Taille » selon une modalité de 5-6 génère par défaut un groupe composé de 5 personnes.

¹³⁷ Grande Ecole

¹³⁸ COMMunauté d'Universités et Etablissements

		[G1]	[G2]	[G3]	[G4]	[G5]	[G6]	[G7]	[G8]	[G9]	[G10]
	ID unique	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
C1	Mode organisationnel	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
C2	Type de regroupement	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
C3	Statut	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
C4	Secteur	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
C7	Historique	40.0	60.0	80.0	60.0	60.0	40.0	20.0	60.0	40.0	60.0
C8	Taille	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Tableau 7.7 - Liste des groupes restitués pour la requête Q1

Dans le cas de cette requête, il est à noter qu'un score d'évaluation ex-aequo est associé pour l'ensemble des résultats. A titre d'exemple, la sortie des résultats des itérations lors du traitement de l'AG propose les scores bruts associés suivants :

It0	It1	It2	It3	It4	It5	It6	It7	It8	It9
1438	1324	1278	1086	1084	1012	1020	1008	1024	1004

Tableau 7.8 - Détails des scores bruts par itérations pour la requête Q1

Par la suite, l'ensemble des scores brutes par itérations est converti selon un score sur une échelle de 0 à 5 en prenant en compte le score minimal et le score maximal $(1 - (1004/1438 * 5))$. Le calcul résulte d'un score d'évaluation d'environ de 2,49/5.

Selon l'index de la collection des requêtes présenté en Annexe 15, ce cas d'utilisation correspond à la requête n°28.

Il est à noter que le principe de traitement de type heuristique de l'AG implique que ces résultats peuvent différer lors de nouvelles exécutions de ces requêtes notamment au niveau de l'affectation des individus. Contrairement aux méthodes déterministes qui ne font pas appel au hasard, les méthodes heuristiques ne garantissent pas la répétition des résultats d'une manière exacte lors d'une nouvelle série d'expérimentation malgré un jeu de données identiques.

9.2 Tests fonctionnels et techniques

Les tests fonctionnels et techniques consistent à vérifier la conformité des spécifications fonctionnelles et les performances techniques du prototype d'application préalablement déterminées dans le chapitre 6 ([Voir section 6.1.2](#)).

9.2.1 Cas d'utilisation

Les tests fonctionnels et techniques sont menés à partir de cas d'utilisations. Ils se concentrent sur une séquence d'opérations réelles basée sur la simulation d'un scénario

prédéterminé. Les résultats visent à identifier les écarts éventuels afin de réaliser les correctifs nécessaires. Le périmètre des tests fonctionnels et techniques se limite aux interfaces-utilisateurs implémentées au sein de la première version du prototype dont les fonctionnalités « Recherche » et « Planification ». Le nombre total s'élève à 75 cas d'utilisations selon la répartition suivante :

- Tâches de recherche simple : 30 cas d'utilisations
- Tâches de planification : 45 cas d'utilisations

9.2.2 Mesures d'évaluation

Parmi les mesures existantes pour l'évaluation des logiciels, la norme ISO 25000¹³⁹ regroupe une série de normes relatives à la qualité logicielle. Notre approche d'évaluation s'appuie plus précisément sur la norme ISO/IEC 25022 relative aux mesures de qualité du fonctionnement récapitulé dans le tableau suivant :

Mesures de qualité	Nom	Fonction de mesure
Efficacité	Taux d'achèvement des tâches	$X = A/B$ A=Nombre de tâches achevées/ B=Nombre total de tâches
Efficienc	Temps d'exécution d'une tâche	$X = A/T$ A=Nombre d'objectifs atteints/T=Temps d'exécution

Tableau 7.9 - Mesures d'efficacité et d'efficienc de la norme ISO/IEC 25022

Ainsi, la mesure d'efficacité à travers le taux d'achèvement vise à vérifier le taux de réalisation correcte d'une tâche. La mesure d'efficienc à travers le temps d'exécution consiste à évaluer le temps de réponse du système.

9.2.3 Protocole de tests

Le protocole de tests vise principalement des interfaces-utilisateurs qui ne sont pas implémentés comme des pages publiques. Par conséquent, les tests d'exécution nécessitent une authentification sécurisée et menés manuellement par un seul utilisateur ayant des droits d'accès comme étant « administrateur ». Les requêtes sont exécutées l'une après l'autre à partir d'un poste de travail fixe¹⁴⁰ sans opérations simultanées et un réseau internet sans fil au niveau du « Front-End ». Ainsi, les enregistrements sont effectués en utilisant l'outil de développement du navigateur¹⁴¹. Les résultats obtenus visent à identifier les éventuels goulots d'étranglements.

La période de profilage est identique pour l'ensemble des cas d'utilisations. Elle est déterminée de la manière suivante :

- Page de départ : page d'accueil
- Page d'arrivée : affichage complet des résultats et du téléchargement du dernier composant graphique.

Ainsi, chaque enregistrement des cas d'utilisation génère l'intervalle de profilage global associé à une répartition détaillée des temps de chargement.

¹³⁹ SQuaRE (Ingénierie des systèmes et du logiciel – Exigences de qualité et évaluation des systèmes et du logiciel)

¹⁴⁰ Marque : Dell precision T1700, CPU : Intel CORE i5 2.27 GHz, Mémoire : 4GB

¹⁴¹ Outil de développement « Devtools » du navigateur Chrome. Protocole : <https://chromedevtools.github.io/devtools-protocol>

A titre d'exemple, l'outil d'analyse de performance propose les sorties suivantes pour une tâche de recherche simple et une tâche de planification :

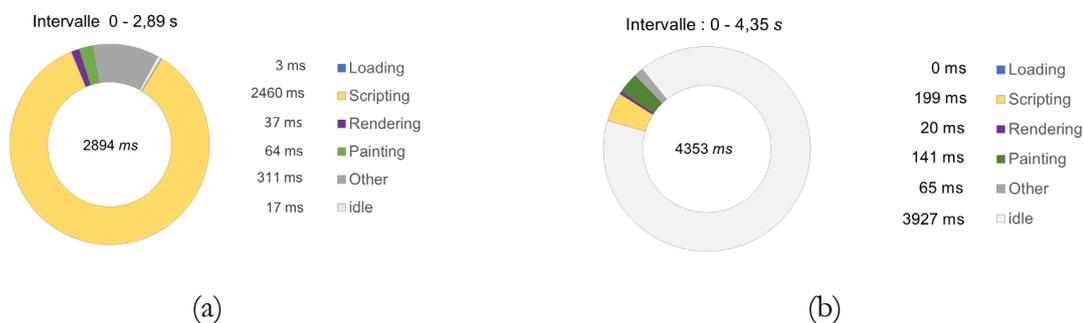


Figure 8.3 - Exemple d'un enregistrement avec répartition des temps selon les mesures de performances « Loading » (Chargement), « Scripting » (Script), « Rendering » (Rendu), « Painting » (Graphisme), « Other » (Autre), « Idle » (Inactivité) pour (a) tâche de recherche et (b) tâche de planification

Il est à noter que ces exemples d'enregistrements démontrent des répartitions différentes selon la nature de la tâche d'action. La majorité du temps de chargement pour la tâche de recherche est liée au « Scripting » relative à l'exécution du JavaScript qui doit être par conséquent optimisé pour améliorer les délais d'affichage. Le temps de chargement pour la tâche de planification est « idle » considéré comme étant un « temps d'inactivité » du système correspondant au temps de traitement de l'AG du côté « Back-End ».

9.2.4 Résultats

Les résultats obtenus pour le taux d'achèvement des tâches sont très satisfaisants pour l'ensemble des cas d'utilisation (98%). En effet, la quasi-majorité des cas d'utilisations ne fait apparaître aucun échec de téléchargement à l'exception de deux cas d'utilisations.

Les cas d'utilisations et les résultats associés sont présentés en Annexe 18 dont un extrait est présenté dans le tableau suivant :

ID	Interface-utilisateur		Filtrage	Mesures		
	Niveau 1	Niveau 2		Nombre de critères	Efficacité	Efficiace
					Taux d'achèvement	Temps d'exécution (ms) ¹⁴²
1	Recherche	Personne	2	100%	0-3802	
				
11	Recherche	Organisa-tion	2	100%	0-2268	
				
21	Recherche	Formateur	1	100%	0-2286	
				
25	Recherche	Evènement	3	100%	0-2408	
					
30	Planification		4	100%	0-4354	
					

¹⁴² Unité de temps en millisecondes

75	Planification		5	100%	0-3765
----	---------------	--	---	------	--------

Tableau 7.10 - Extraits des résultats des tests selon des mesures d'efficacité et d'efficience

Pour rappel, les requêtes liées aux tâches de recherche simple liée à la fonctionnalité « Recherche » reposent sur un système de mise en correspondance avec une base de données à base de graphe résultant d'items individuels. Les tâches de recherche liées à la fonctionnalité « Planification » repose sur un traitement supplémentaire *via* un AG résultant d'items groupaux. Par conséquent, l'évaluation du temps d'exécution est distinguée compte tenu du traitement spécifique appliqué.

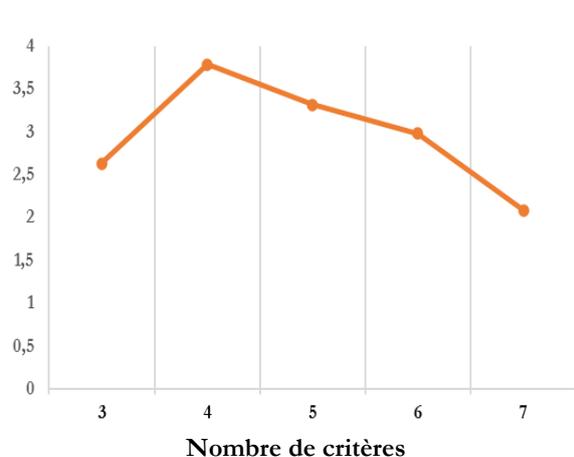
- Tâche de recherche simple

L'évaluation du temps d'exécution des tâches de recherche simple se base sur 30 cas d'utilisation qui se distinguent selon le nombre de critères (type d'entités, nombre de filtres, nombre de résultats...). Le temps d'exécution pour l'ensemble des cas d'utilisations est de 2,8s en moyenne. Des échecs de téléchargement sont constatés au niveau du chargement des listes de valeurs (« Type d'établissement » et « Secteur d'activité ») lors de la recherche d'une entité « Organisation ». Des difficultés mineures sont constatées pour les champs en saisie libre : la valeur d'un champ est sensible à la casse c'est-à-dire que la saisie de la valeur doit être exacte (par exemple, la saisie du nom « Université de Lille » ne retournera aucun résultat compte tenu que l'entité est indexée au sein de la base de données selon la valeur « Université Lille 2 »).

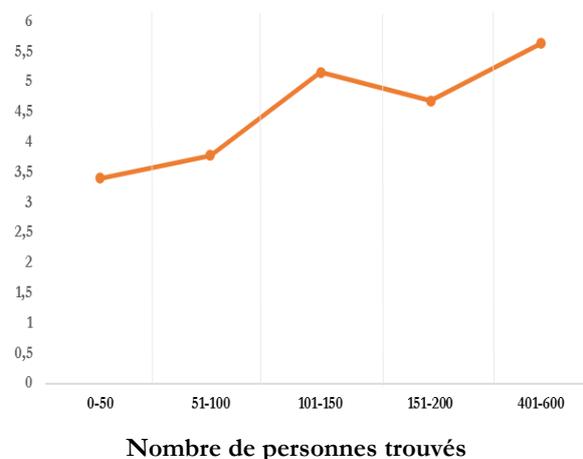
- Tâche de planification

L'évaluation du temps d'exécution des tâches de planification fait l'objet d'une évaluation spécifique liée à l'implémentation de l'AG. En effet, l'objet principal du prototype d'application est de minimiser le temps accordé lors de la planification des sessions de formation. Par conséquent, le temps de calcul et la génération de la liste de résultats ne doit pas excéder un certain seuil d'attente pour le gestionnaire de formation ([Voir section 7.3.1](#)). Cette étape se base sur la collection de requêtes utilisées lors de l'évaluation des résultats de l'AG.

Pour l'ensemble de requêtes, le temps de recherche est de 3.7s en moyenne. Il apparait que plusieurs paramètres influencent le temps d'exécution selon le type de requête.

Temps d'exécution (s)

(a)

Temps d'exécution (s)

(b)

Figure 8.4 - (a) Courbe du temps d'exécution (ordonnées) selon le nombre de critères des requêtes (abscisses) et (b) Courbe du temps d'exécution (ordonnées) selon le nombre de personnes trouvés (abscisses)

En premier lieu, nous pouvons observer que le nombre de critères de la requête n'influence pas le temps d'exécution. Des observations complémentaires font apparaître que le nombre total de personnes trouvées influence le temps d'exécution. La courbe du temps d'exécution a tendance à augmenter à partir d'une tranche de plus de 100 personnes. Parmi la collection de requête de test, le nombre de personnes trouvé est de 76 personnes en moyenne. Il apparaît que le temps d'exécution minimale (1,26s) s'effectue sur un nombre total de personnes minimal (11 personnes) tandis que le temps d'exécution maximale (6,53s) s'effectue sur un nombre total s'élevant à 520 personnes.

Au niveau de l'ergonomie, des défauts mineurs sont également constatées au niveau de la liste des champs (exemple : liste d'affichage du champs « Département » n'est pas dépendante du choix du champ « Région »).

9.3 Tests-utilisateurs

Les tests-utilisateurs consistent à confronter les utilisateurs au prototype d'application en simulant des tâches opérationnelles et vérifier l'utilisabilité des interfaces indépendamment des résultats.

9.3.1 Méthodologie

Dans le cadre de notre application, la version intermédiaire du prototype d'application est suffisamment fonctionnelle pour mener des tests-utilisateurs. La méthodologie se déroule selon les étapes suivantes :

- (1) Définition des objectifs d'évaluation
- (2) Description des évaluateurs
- (3) Définition des mesures d'évaluation
- (4) Rédaction et passation du questionnaire d'évaluation

- Objectifs d'évaluation

Les objectifs de l'évaluation des interfaces-utilisateurs peuvent être synthétisés de la manière suivante :

- Evaluer la satisfaction globale ;
- Evaluer l'utilité et la facilité d'utilisation des interfaces « Recherche » et « Planification » ;
- Evaluer l'interface de visualisation interactive.

- Evalueurs

Les profils des évaluateurs sont des personnes internes à l'entreprise. Le principal avantage est que ce type de profil utilise quotidiennement l'outil existant et mener ainsi une comparaison avec le projet de prototype d'application proposé. Le tableau suivant synthétise les caractéristiques des évaluateurs :

Utilisateur	Age	Sexe	Statut
Utilisateur 1	30	F	Salarié
Utilisateur 2	38	F	Salarié
Utilisateur 3	55	M	Associé
Utilisateur 4	48	M	Associé
Utilisateur 5	34	M	Salarié
Utilisateur 6	32	F	Salarié

Tableau 7.11 – Nombre et caractéristiques des évaluateurs

Il est à noter que ce type de profil présente l'inconvénient de ne pas être des professionnels RH spécialisés dans la formation continue auquel le prototype d'application est destiné ([Voir section 6.1.3](#)). Cette exigence méthodologique implique des contraintes expérimentales lourdes qui impliqueraient un recrutement externe des évaluateurs. Par conséquent, cette évaluation consiste principalement à bénéficier d'un premier retour d'expérience des interfaces-utilisateurs implémentées au préalable d'une enquête à plus large échelle lors de la version finale du prototype d'application.

- Mesures d'évaluation

Dans la continuité des tests fonctionnels et techniques, les mesures d'évaluation s'appuient également sur la norme ISO/IEC 25022 qui vise à évaluer la qualité de l'expérience utilisateur.

Mesures de qualité	Description	Fonction de mesure
Satisfaction générale	Satisfaction globale de l'utilisateur	$X = \sum (A_i)$ $A_i = \text{Réponse à une question}$
Utilité	Satisfaction des fonctionnalités	$X = \sum (A_i)$ $A_i = \text{Réponse à une question relative à une fonctionnalité}$
Expérience utilisateur	Plaisir de l'utilisateur	$X = A$ $A = \text{échelle de valeur}$
Ergonomie	Confort physique	$X = A$ $A = \text{échelle de valeur}$

Tableau 7.12 - Mesures de satisfaction de la norme ISO/IEC 25022

- Questionnaire d'évaluation

Le questionnaire comporte 17 questions fermées avec la possibilité d'émettre un avis complémentaire en réponse libre pour chacune d'entre elles. La conception du questionnaire est regroupée selon les thématiques suivantes :

- Evaluation globale ;
- Evaluation du tableau de bord ;
- Evaluation de la fonctionnalité « Recherche » ;
- Evaluation de la fonctionnalité « Planification ».

Le questionnaire est administré en ligne à l'ensemble des évaluateurs avec un compte d'accès personnalisé. Ce mode d'administration présente l'avantage de rendre l'évaluateur

autonome dans la participation à l'enquête. Le temps de réponse accordé s'étale sur une période de trois semaines.

Le détail du questionnaire est présenté en Annexe 20.

9.3.2 Résultats

Le questionnaire a été complété par six évaluateurs parmi les huit évaluateurs sollicités. Le taux de remplissage est maximal compte tenu de l'obligation des réponses pour l'ensemble des questions.

La satisfaction globale est positive mais partagée avec des réponses « Bonne » (50%) voire « Très bonne » (50%).

La prise en main du prototype d'application est aisée avec une facilité d'utilisation jugée « Très facile » (50%)¹⁴³. L'ensemble des simulations des tâches de recherche de données ont été réalisées sans aucune erreurs compte tenu des réponses correctes apportées. La facilité d'utilisation est jugée « Très facile » (66.7%) pour la recherche simple d'une personne et pour la recherche simple d'un formateur (83.3%). Aucune réponse négative n'est constatée pour ces deux niveaux d'analyse.

Au niveau de l'utilité des fonctionnalités, la planification de sessions de formation se démarque (83.3%) en comparaison avec la recherche simple de données (66.7%). En comparaison avec l'outil existant¹⁴⁴, le prototype d'application est jugé plus pertinent pour la gestion de la formation continue (83.3%) avec comme principal avantage une distinction des données avec l'activité de conseil.

Au niveau de l'ergonomie, l'évaluation du tableau de bord obtient un taux de satisfaction partagé (50%) et fait l'objet de multiples commentaires : manque d'informations commerciale (catalogue de formation) financières (facturation, prix...) et un manque de flexibilité dans l'agencement des vues. Néanmoins, il est à noter que l'implémentation de l'interface du tableau de bord est partielle impliquant une évaluation approximative.

L'interface-utilisateur dédié pour la planification d'une session de formation est jugée « Bonne » pour la totalité des évaluateurs avec des failles techniques relevés au niveau des liste des champs. L'utilisation d'un calendrier est également jugée plus fonctionnelle pour la saisie de critères temporels.

Au niveau de la visualisation interactive, la présentation des résultats sous forme de représentation graphique est préférée (66,7%). Toutefois, la forme textuelle reste appréciée (33.3%). Parmi les commentaires, il est également observé que la représentation graphique n'est pas adaptée en cas de volume important des données. Ce manque de lisibilité du graphe implique de laisser à disposition les résultats sous forme de tableur en alternative pour l'utilisateur. La fonction « *Focus+Context* » pour détailler les nœuds du graphe est utilisé intuitivement par l'ensemble des utilisateurs sans indication préalable.

Les commentaires en réponse libre proposent des axes d'amélioration du prototype dont principalement des besoins de personnalisation. En effet, il apparaît que les réponses diffèrent selon les profils des évaluateurs : le manque d'information de nature économique (comptable et financière) et commerciale (prospection) sont mis en évidence par des utilisateurs qui ont un pouvoir décisionnel au sein de l'entreprise (Associé). Les utilisateurs ayant un statut de salariés ont davantage axé leurs réponses sur les aspects fonctionnels, ergonomiques et techniques du prototype d'application.

L'ensemble des réponses par utilisateurs est détaillé dans l'Annexe 21.

¹⁴³ Réponses : « Plutôt facile » (35%), « Normal » (15%)

¹⁴⁴ CRM (Customer Relation Management)

9.4 Conclusion

Ce dernier chapitre présente plusieurs séries d'expérimentations distinguées selon deux niveaux d'évaluation.

Au niveau de l'évaluation de l'AG, le taux de précision des résultats selon une collection de requêtes préalablement définie est positif avec des variations observées selon les critères spécifiés. Néanmoins, ces résultats sont à nuancer compte tenu de la proportion des scores d'évaluation ex-aequo observés lors de la génération de la liste des groupes de formation potentiels. Ce constat nous oriente à affiner la fonction objective de l'AG et des paramétrages associés. De même, certains paramétrages fixés par défaut (taille du groupe) nécessitent un réajustement du code-source.

Au niveau de la performance du prototype d'application, les tests fonctionnels et techniques permettent de valider la première version fonctionnelle du prototype d'application. Le taux d'efficacité et le taux d'efficacité sont particulièrement corrects compte tenu que les cas d'utilisation ont permis à l'utilisateur d'atteindre le résultat prévu selon un temps d'exécution raisonnable.

Au niveau des tests utilisateurs, les résultats présentent des retours d'expériences-utilisateurs qui peuvent faire l'objet de nombreuses améliorations lors de l'étape suivante du cycle de développement (profils des utilisateurs, entretien qualitatif...).

A ce stade du projet, l'automatisation des tests n'est pas nécessaire compte tenu du nombre minimal d'interfaces-utilisateurs à évaluer mais peuvent faire l'objet de tests automatisés lors du passage à l'échelle du prototype.

Chapitre 10. Conclusion et Perspectives

En conclusion, ce chapitre synthétise les contributions et les perspectives de développement proposées.

10.1 Synthèse

Elaboré dans un contexte industriel, ce projet de recherche opérationnelle propose un système d'aide à la gestion et à la planification de sessions de formation dans le domaine de la formation continue. Selon une approche pluridisciplinaire, plusieurs contributions sont proposées et synthétisées de la manière suivante :

- Présentation et visualisation de données complexes

Ce prototype d'application propose une agrégation et une représentation des données complexes instrumentées à travers des tableaux de bord multiples. En complément, une représentation de type nœud-lien associée à des techniques de visualisation interactives (vue globale, zoom, filtre, détails à la demande...) permettent une exploration intuitive des données à disposition pour l'utilisateur. Dans notre cas d'application, une attention particulière est accordée à l'identité professionnelle des acteurs du fait qu'elle influence les relations inter-groupes d'une manière déterminante. Théorisé par (Bourgeois E. 2001), les relations professionnelles peuvent faire l'objet de symétrie ou d'asymétrie des relations sociales notamment au sein du groupe de formation pour adulte où s'articule trois acteurs : le participant, l'organisation (ou employeur) et le formateur. Ainsi, le groupe de formation pour adulte devient un objet de recherche qui se différencie du groupe d'apprentissage (ou scolaire) selon diverses caractéristiques (mode organisationnel, taille, mixité, type de regroupement...) mais également met en exergue des dynamiques identitaires et des enjeux socio-professionnels. Notre contribution consiste à prendre en considération cette spécificité de la situation formative d'une manière descriptive à travers une représentation graphique simple de type nœud-lien. Ainsi, la restitution des résultats permet de distinguer d'une manière instantanée le statut, la typologie des relations et le voisinage socio-professionnel des différents acteurs de la formation. A partir des tests-utilisateurs menés, les premiers résultats obtenus ont mis en évidence l'utilité et la facilité d'utilisation des techniques de visualisation pour l'interprétation des données.

- Planification semi-automatisée du groupe de formation

Le développement d'un outil de planification semi-automatisée des sessions de formation en présentiel caractérise la contribution majeure de ce projet de recherche. Afin de résoudre ce problème d'optimisation de type combinatoire, notre démarche consiste à appliquer une méthode méta-heuristique et plus précisément un algorithme génétique. La principale difficulté est la définition de la fonction objective (ou *fitness*) et les choix des paramètres pertinents associés (sélection, croisement, mutation, remplacement) durant la phase d'initialisation de l'AG. En effet, cette étape s'avère déterminante afin de générer des résultats de qualité mais implique une connaissance du problème de décision. Dans ce cas d'application, la fonction objective (ou *fitness*) de l'AG est orienté selon les résultats d'une analyse statistique liée à l'activité de formation. Cette approche analytique a permis de justifier les caractéristiques d'un groupe de formation qui conditionnent favorablement les interactions sociales entre les membres. Ainsi, la planification des groupes de formation est paramétrée selon un type de regroupement homogène associé à des paramètres élémentaires (méthode élitiste, codage binaire, croisement unique...).

Abordé sous l'angle de la recherche opérationnelle et non d'un point de vue mathématique, cette méthode d'optimisation s'avère particulièrement adaptée pour répondre à un problème de planification en explorant et en générant rapidement des groupes de formation potentielles. Les résultats de l'AG ont été évalués selon le taux de précision en fonction d'une collection de requêtes multicritères choisis en fonction de la complexité combinatoire. De plus, une attention particulière a été accordée au jeu de données afin qu'il soit réaliste et volumineux pour garantir une exploration suffisamment large inhérente à l'utilisation de l'AG. Cette approche nous permet d'obtenir des résultats avec un taux de précision plutôt positif et généré selon un temps de traitement raisonnable (environ 3,7s en moyenne). Néanmoins, ces résultats sont à nuancer compte tenu des variations observées des taux de précision selon les critères spécifiés. De plus, la proportion des scores ex-aequo lors de la génération des listes des résultats révèle un défaut d'évaluation qui peut faire l'objet d'amélioration au niveau du paramétrage de l'AG.

Ainsi, l'objectif principal de ce projet de recherche opérationnelle a permis de mettre en évidence une démarche méthodologique et l'intérêt des méthodes méta-heuristiques dans le cadre spécifique de la formation continue. Bien que partielle, le développement du prototype d'application a permis de répondre à des besoins opérationnels et de suppléer l'activité humaine dans un domaine d'application spécifique.

Dans un contexte industriel, l'informatisation et l'utilisation des techniques de l'IA et plus précisément les algorithmes évolutionnistes peuvent s'étendre à d'autres missions de GRH (aide au recrutement, gestion des carrières et des compétences, suivi de l'absentéisme...). De même, l'analyse statistique des données (ou *HR Analytics*) devient incontournable particulièrement au sein des grandes organisations qui doivent assurer la gestion courante et stratégique de son personnel et de son activité. En parallèle, cette approche implique des réflexions transversales d'ordre éthique et juridique lors du traitement des données personnelles qui peuvent être liées à des problématiques sensibles (confidentialité des données personnelles, égalité professionnelle, handicap et l'employabilité des jeunes et des seniors...).

De plus, cette démarche méthodologique peut être transposée à la gouvernance des universités françaises qui intègrent progressivement la formation continue parmi leurs missions pédagogiques. Relative au domaine de l'analyse de l'apprentissage (ou *Learning Analytics*), un domaine de recherche nommée « *Institutional Analytics* » vise à optimiser le fonctionnement organisationnel destinés aux décideurs et personnels universitaires mais qui reste principalement étudié dans les universités anglo-saxonnes.

10.2 Perspectives

Dans le prolongement des contributions citées précédemment, plusieurs axes d'améliorations peuvent être proposés en privilégiant tout autant le traitement et la présentation des résultats. Ils peuvent être distingués de la manière suivante :

-Au niveau du traitement, le paramétrage de l'AG peut être optimisé pour affiner les scores d'évaluation obtenus. En complément, une analyse qualitative consisterait à valider notre proposition en quantifiant les interactions sociales entre les participants du groupe de formation en situation réelle.

-Au niveau des interfaces-utilisateurs, une approche complémentaire consisterait à personnaliser les vues du tableau de bord principal selon les profils des utilisateurs.

- Optimisation du traitement de l'AG

L'évaluation des résultats et plus précisément les scores d'évaluation ex-aequo obtenus mettent en évidence un défaut de paramétrage de l'AG. Ce problème de redondance des résul-

tats est un problème fréquent lors de l'application d'un AG classique (Falkenauer, E. 1998) repris par (Brown, E. C., 2005). Ce défaut d'évaluation peut être résolu en utilisant des opérateurs génétiques plus complexes lors de la phase d'initialisation (point de croisement multiple, critère d'arrêt flexible, codage...). Dans la continuité de l'AG, des méthodes similaires peuvent être également envisagées dont l'algorithme génétique de regroupement (ou *Grouping Genetic Algorithm*). Cette technique a été créée spécifiquement pour les problèmes d'optimisation contraints impliquant un codage groupal et non individuel. Plus récemment, cette méthode de résolution a été appliquée pour la constitution d'équipes de travail en entreprise (Agustín-Blas L., 2011) et dans un cadre universitaire (Mutingi, M. 2017) mais reste peu étudié dans la littérature scientifique.

Dans notre cas d'application, le paramétrage de l'AG peut être amélioré en affinant l'évaluation du score d'évaluation du groupe de formation. Basé sur un nombre de caractéristiques limité des participants (statut, type de formation, secteur d'activité et historique de formation), des caractéristiques supplémentaires peuvent être intégrées comme l'attribut « Profession » ainsi que les préférences d'ordre spatio-temporelles des participants. Cette perspective impliquerait d'utiliser un codage réel permettant ainsi une représentation des valeurs non-numériques au lieu d'un codage binaire initialement utilisé. Ces propositions peuvent être synthétisés de la manière suivante :

Ontologie de domaine. Parmi les caractéristiques des participants utilisés, l'attribut « Profession » n'a pas pu être exploité du fait de la variation terminologique observée. En effet, les données à disposition ont permis de constater que les participants exercent des responsabilités et des missions équivalentes mais ne font pas l'objet d'une terminologie professionnelle commune. Parmi les solutions existantes, une ontologie de domaine défini par (Gruber T., 1995) comme un outil de conceptualisation d'un domaine permet de normaliser et de traiter les multiples terminologies en appliquant un référentiel commun. La tâche de normalisation consisterait à relier des valeurs catégorielles à un concept commun pour traiter la synonymie des noms de profession. Par exemple, les valeurs telles que « *Chargé de la commande publique* » et « *Responsable des achats publics* » seraient indexées selon une référence commune afin de formaliser la similarité sémantique des termes. La conception de ce type d'ontologie peut se baser sur des référentiels métiers existants et spécialisé selon un secteur d'activité^{145 146}. Cette approche présente l'avantage de bénéficier d'une structuration des données préétablis facilitant ainsi l'indexation des concepts. De plus, la modélisation de l'ontologie peut s'étendre à d'autres caractéristiques professionnelles du participant (ancienneté, diplômes obtenus, type de contrat, niveau de salaire...). Ainsi, la sélection des participants sera affinée lors du calcul d'évaluation pour maximiser l'homogénéité des groupes de formations.

Dans un cadre d'application plus large, des travaux de recherche abordent cette méthode de modélisation des connaissances dans le domaine des ressources humaines notamment pour optimiser le recrutement en ligne (Alfonso-Hermelo, D., 2019) (Kessler, R., 2017). Au-delà de ce domaine d'application, des travaux scientifiques associant une ontologie à l'algorithme génétique ont été expérimentés pour traiter les variables catégorielles et ainsi améliorer la précision des résultats (Lv, G., 2016), (Sharma, M., 2019).

Elicitation de préférences spatio-temporelles. Dans le cadre de la formation continue, l'action de formation en présentiel reste soumise à des contraintes logistiques. Lors de la formulation de la requête, l'interface de recherche propose des critères spatio-temporels à titre indicatif mais ne peuvent faire l'objet d'un traitement compte tenu des valeurs manquantes associées. Une élicitation des préférences d'une manière explicite à travers les questionnaires de satisfaction vise à obtenir des informations supplémentaires pour chaque participant dont la disponibilité calendaire (période durant ou hors vacances scolaires, jours consécutifs, durée de formation) et les péri-

¹⁴⁵ ROME : Répertoire Opérationnel des Métiers et des Emplois

¹⁴⁶ CNFPT : Centre National de la Fonction Publique Territoriale

mètres géographiques (Région, Département). Initialement appliqué dans les SIG¹⁴⁷, les critères spatio-temporels nécessitent une modélisation et un traitement spécifique qui peut s'adapter à d'autres domaines d'applications (Touzani, M., 2016). Ainsi, les participants susceptibles d'accepter le projet de formation seraient mieux ciblés lors de l'exploration de la population initiale.

- Analyse qualitative des résultats

Dans notre cas d'étude, notre proposition est évaluée essentiellement selon une approche informatique et n'est pas expérimentée selon une situation de formation réelle. L'analyse qualitative viserait à tester un groupe de formation générée par l'AG en analysant les interactions sociales entre le(s) participant(s) et le formateur. Ainsi, la démarche expérimentale consisterait à vérifier la valeur ajoutée d'une composition semi-automatisée du groupe de formation en comparaison avec une méthode de composition manuelle initialement utilisée. Cette approche qualitative nécessite une démarche méthodologique et des conditions expérimentales spécifiques (nombre minimal de participants, recueil des consentements des personnes...) mais demeure une perspective notable pour valider notre contribution de recherche d'une manière plus avancée.

Parmi les méthodes existantes, la méthode IPA (ou *Interaction Process Analysis*) proposée par (Bales R.F, 1950) quantifie les interactions sociales au sein d'un groupe qui distingue deux domaines d'interaction : le domaine socio-affectif (domaine émotionnel positif ou négatif) et le domaine de la nature de la tâche (questions ou réponses). D'une part, ces multiples dimensions socio-cognitives permettent d'analyser les dynamiques interactionnelles dont la symétrie et l'asymétrie des relations socio-professionnelles dans le groupe de formation d'adultes. D'autre part, une analyse complémentaire des phénomènes sociaux dont la réciprocité et l'intensité des interactions sociales (entraide, affinité, tension...) peuvent faire l'objet d'hypothèses de recherche pour mieux comprendre le fonctionnement et la régulation groupale qui reste peu étudiée dans le domaine de la formation pour adulte.

- Tableaux de bord et visualisation adaptative

Cette version fonctionnelle du prototype d'application propose des tableaux de bord multiples incluant des indicateurs prédéfinis et statiques selon les besoins d'un utilisateur spécifique (plus précisément le gestionnaire de formation). Les résultats préliminaires obtenus lors des tests-utilisateurs ont permis de mettre en évidence des besoins qui peuvent différer selon le profil des utilisateurs : les décideurs souhaitent des indicateurs de performance de nature financière et économique tandis que les utilisateurs restants qui ont un statut de salariés ont davantage axés leurs appréciations selon des considérations ergonomiques.

Les perspectives de développement consisteraient à proposer une personnalisation des vues et des indicateurs selon les profils des utilisateurs intervenant dans l'environnement de la formation continue (décideur, formateur, gestionnaire/administratif de la formation...). Cette étape consisterait à approfondir les tests-utilisateurs initiaux et à recueillir d'une manière explicite les besoins et les tâches spécifiques à travers des entretiens qualitatifs. Parmi les travaux existants, cette perspective peut s'appuyer sur une méthodologie brevetée par (Chowdhary, P. 2014), repris par (Dabbebi, I. 2019) qui propose un processus de génération automatique de tableaux de bord d'apprentissage personnalisée dans le domaine spécifique de la formation.

Au niveau de la visualisation, la restitution des résultats se limite à une représentation graphique standard avec une seule arête. Au niveau de la conception visuelle, des améliorations peuvent être envisagées comme par exemple une identification visuelle supplémentaire pour affiner la typologie du nœud du participant (historique de l'activité de formation du participant,

¹⁴⁷ Système d'Information Géographique

statut particulier...). Néanmoins, ce niveau d'abstraction peut s'avérer limité pour visualiser les relations intrinsèques au sein d'un réseau professionnel en dehors du cadre formalisé de la formation continue (exemple l'historique et la mobilité professionnelle). Récemment, des travaux de recherche ont mis en évidence la temporalité des réseaux sociaux qui nécessitent une modélisation *via* un graphe multicouche (ou multiniveaux) pour s'adapter à l'évolution des entités dans un contexte professionnel (Basole, R. C., 2018), Bidart, C., 2018)

Chapitre 11. Bibliographie

Abnar, S., Orooji, F., & Taghiyareh, F. (2012). An evolutionary algorithm for forming mixed groups of learners in web based collaborative learning environments. In *2012 IEEE international conference on technology enhanced education (ICTEE)* (pp. 1-6).

Acquatella, F. (2016). Le COOC, un autre visage du MOOC. *Distances et médiations des savoirs. Distance and Mediation of Knowledge*, (14).

Agustín-Blas, L. E., Salcedo-Sanz, S., Ortiz-García, E. G., Portilla-Figueras, A., Pérez-Bellido, Á. M., & Jiménez-Fernández, S. (2011). Team formation based on group technology: A hybrid grouping genetic algorithm approach. *Computers & Operations Research*, 38(2), 484-495.

Atif, L. (2017). P©, une approche collaborative d'analyse des besoins et des exigences dirigées par les problèmes : le cas de développement d'une application Analytics RH (Thèse de doctorat), Université de Paris Dauphine.

Angrave, D., Charlwood, A., Kirkpatrick, I., Lawrence, M., & Stuart, M. (2016). HR and analytics: why HR is set to fail the big data challenge. *Human Resource Management Journal*, 26(1), 1-11.

Anzieu, D., & Martin, J. Y. (1971 reedité en 1994). La dynamique des groupes restreints. P.U.F Presses universitaires de France, Paris, 10^{ème} édition.

Alfonseca, E., Carro, R. M., Martín, E., Ortigosa, A., & Paredes, P. (2006). The impact of learning styles on student grouping for collaborative learning: a case study. *User Modeling and User-Adapted Interaction*, 16(3-4), 377-401.

Alfonso-Hermelo, D., Langlais, P., & Bourg, L. (2019). Automatically learning a human-Resource ontology from professional social-network data. In *Canadian Conference on Artificial Intelligence* (pp. 132-145). Springer, Cham.

Auber David, Beaudouin-Lafon Michel (2017). Visulisation de données massives. In *Les Big Data à découvert*, sous la direction de Mokrane Bouzeghoub et Rémy Mossery, CNRS Editions, p152-153. In Bouzeghoub, M., & Mosseri, R. (Eds.). (2017). *Les Big Data à découvert*.

Babar, M. A., & Kitchenham, B. (2007). The impact of group size on software architecture evaluation: a controlled experiment. In *First International Symposium on Empirical Software Engineering and Measurement (ESEM 2007)* (pp. 420-429). IEEE.

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics* (pp. 61-75). Springer, New York, NY.

Back, T. (1996). *Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms*. Oxford university press.

Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 44-54). ACM.

Bardin, L. (1998). L'analyse de contenu. Paris : Presses Universitaires de France.

- Basole, R. C., Srinivasan, A., Park, H., & Patel, S. (2018). ecovight: Discovery, exploration, and analysis of business ecosystems using interactive visualization. *ACM Transactions on Management Information Systems (TMIS)*, 9(2), 1-26.
- Bataille, M. (2000). Représentation, implicitation, implication; des représentations sociales aux représentations professionnelles. *Représentations sociales et éducation*, 165-189.
- Bales, Robert F. "Interaction process analysis; a method for the study of small groups." (1950). Cambridge : Addison-Wesley.
- Balleux, A. (2000). Évolution de la notion d'apprentissage expérientiel en éducation des adultes: vingt-cinq ans de recherche. *Revue des sciences de l'éducation*, 26(2), 263-286.
- Baudoin, J. M. (2001). La dimension du groupe, seconde et primordiale: histoire de vie et recherche-formation. C. Solar (éd.), *Le groupe en formation, Bruxelles, De Boeck*, 35-56.
- Baudoin, E., Diard, C., Benabid, M., & Cherif, K. (2019). *Transformation digitale de la fonction RH* (Vol. 1). Dunod.
- Beauguitte, L. (2013). L'analyse des graphes bipartis. Synthèse de groupe de travail FMR (Flux, Matrices, Réseaux).
- Beheshitha, S. S., Hatala, M., Gašević, D., & Joksimović, S. (2016, April). The role of achievement goal orientations when studying effect of learning analytics visualizations. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 54-63).
- Benchettara, N. (2011). *Prévision de nouveaux liens dans les réseaux d'interactions bipartis: Application au calcul de recommandation* (Thèse de doctorat)-Université Paris 13.
- Bentayeb, N. (2010). Une évaluation contextuelle de la collaboration: cas du Centre jeunesse de Montréal-Institut universitaire (Thèse de doctorat) ENA (École Nationale d'Administration publique).
- Bennour, M., Addouche, S., & El Mhamedi, A. (2012). RCPSP sous contraintes de compétences dans un service de maintenance. *European Journal of Information Systems*, 46(8), 877-907.
- Bertin, J. (1977). *La graphique et le traitement graphique de l'information*. Flammarion, 1977. 273 p.
- Besse, P. (2020). Détecter, évaluer les risques des impacts discriminatoires des algorithmes d'IA. Contribution au séminaire conjoint: Défenseur des Droits & CNIL.
- Betton, E. (2013). Pour une nouvelle légitimité de la pédagogie en formation des adultes. *Education permanente*, (197), pp-161.
- Bevan, N., Carter, J., Earthy, J., Geis, T., & Harker, S. (2016). New ISO standards for usability, usability reports and usability measures. In *International Conference on Human-Computer Interaction* (pp. 268-278). Springer, Cham.
- Beyer, H. G., & Schwefel, H. P. (2002). Evolution strategies—A comprehensive introduction. *Natural computing*, 1(1), 3-52.

- Bhalgat, K. H. (2019). *An exploration of how Artificial Intelligence is impacting Recruitment and Selection process* (Thèse de doctorat) Dublin Business School.
- Bidart, C., & Grossetti, M. (2018). Introduction: les temporalités entrecroisées des réseaux sociaux. *Temporalités. Revue de sciences sociales et humaines*, (27).
- Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *US Department of Education, Office of Educational Technology*, 1, 1-57.
- Bigrat, F., Laurent, M., & Méadel, C. (2019). Valorisation des données pour l'amélioration des plateformes d'hébergement de MOOC. In *Environnements Informatiques pour l'Apprentissage Humain*.
- Binitha, S., & Sathya, S. S. (2012). A survey of bio inspired optimization algorithms. *International journal of soft computing and engineering*, 2(2), 137-151.
- Bistodeau A, (2019). Analyse de l'apprentissage pour soutenir l'apprentissage, la persévérance et la réussite en formation en ligne au postsecondaire. In *Pratiques et innovations à l'ère du numérique en formation à distance: Technologie, pédagogie et formation*. PUQ-268 pp
- Bistorin, O. (2007). *Méthodes et outils d'aide à la conception des processus opérationnels des systèmes de formation* (Thèse de doctorat)-Université de Metz.
- Black, J. S., & van Esch, P. (2020). AI-enabled recruiting: What is it and how should a manager use it?. *Business Horizons*, 63(2), 215-226.
- Bodily, R., & Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 309-318).
- Bodily, R., Kay, J., Alevan, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018). Open learner models and learning analytics dashboards: a systematic review. In *Proceedings of the 8th 11 international conference on learning analytics and knowledge* (pp. 41-50). ACM.
- Bos, M. C. (2007). Du réseau à la communauté d'apprenants. Quelle dynamique du lien social pour "faire oeuvre" sur Internet ? (Thèse de doctorat)-Université de Provence-Aix-Marseille I.
- Bossard, A., Cataldi, M., Lamolle, M., & Le Duc, C. (2016). Évaluation et Prédiction de la Centralité de Groupes de Recherche dans un Réseau de Collaborations Scientifiques. In *EGC* (pp. 407-418).
- Boudreau J., Ramstad P., "Strategic HRM measurement in the 21st century: from justifying HR to strategic talent leadership", *HRM in the 21st Century*, M. Goldsmith, R.P. Gandossy, M.S. Efron (Eds.), John Wiley, New York, NY, 2003, p. 79-90
- Boudreau, J. W., & Ramstad, P. M. (2005). Talentship and the evolution of human resource management: From professional practices to strategic talent decision science. *Human Resource Planning Journal*, 28(2), 17-26.

- Bouffard, G. (2015). L'apprentissage par les pairs : l'apport d'Eric Mazur à la pédagogie. *Pédagogie collégiale Vol. 27, no 2, hiver 2014*, p. 29-33.
- Boulanger, P. M. (2004). Les indicateurs de développement durable: un défi scientifique, un enjeu démocratique. *Les séminaires de l'Idri*, 12, 24.
- Bourdeau, J., & Grandbastien, M. (2011). La modélisation du tutorat dans les systèmes tutoriels intelligents. *STICEF (Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation)*, 18, 14-pages.
- Bourgeois E., Nizet J. (1997). Apprentissage et formation des adultes. Paris : Presses Universitaires de France, 232 pages.
- Bourgeois, E., & Nizet, J. (1999). Apprentissage et formation des adultes (2nd édition): Chap. *VIII Interactions sociales et apprentissage*, p. 155-200.
- Bourgeois, É. & Frenay, M. (2001). Chapitre 5. Apprendre en groupe. Rôle de l'asymétrie et de l'argumentation. Dans : Claudie Solar éd., *Le groupe en formation des adultes : Comprendre pour mieux agir* (pp. 99-114). Louvain-la-Neuve, Belgique: De Boeck Supérieur.
- Bourgeois, E. Nizet J. (2005). Apprentissage et formation des adultes. Paris : PUF
- Bouzayane, S., Saad I., Kassel, G., & Gargouri, F. (2017). Recommandation basée sur l'aide multicritère à la décision pour personnaliser l'échange d'information. *Ingenierie des Systemes d'Information*, 22(6), 71.
- Boyer, A. (2019). Quelques réflexions sur l'exploration des traces d'apprentissage. *Distances et médiations des savoirs. Distance and Mediation of Knowledge*, (27).
- Brown, E. C., & Sumichrast, R. T. (2005). Evaluating performance advantages of grouping genetic algorithms. *Engineering Applications of Artificial Intelligence*, 18(1), 1-12.
- Bruillard, É. (2011). Le déploiement des ENT dans l'enseignement secondaire: entre acteurs multiples, dénis et illusions. *Revue française de pédagogie. Recherches en éducation*, (177), 101-130.
- Brundseaux, M. F., Jérôme, F., & Georges, F. (2009). Elaboration d'un référentiel des compétences du formateur d'adultes dans une perspective de développement professionnel. In 21e colloque de l'ADMEE (Association pour le Développement des Méthodologies d'Evaluation en Education)-Europe.
- Bruno, I. (2015). Défaire l'arbitraire des faits. De l'art de gouverner (et de résister) par les «données probantes». *Revue française de socio-Economie*, (2), 213-227.
- Buckley, C., & Voorhees, E. M. (2017). Evaluating evaluation measure stability. In *ACM SIGIR Forum* (Vol. 51, No. 2, pp. 235-242). New York, NY, USA: ACM.
- Burt Ronald, S. (2001). The Social Capital of Structural Holes. *New Directions in Economic Sociology*, edited by Guillen Mauro F., Collins Randall, England Paula, Meyer Marshall., editors. New York: Russell Sage, 201-250.

- Burt, R. S. (2017). Structural holes versus network closure as social capital. In *Social capital* (pp. 31-56). Routledge.
- Butera, F., Darnon, C., Buchs, C., & Muller, D. (2006). Les méfaits de la compétition: comparaison sociale et focalisation dans l'apprentissage. *Bilans et perspectives en psychologie sociale*, 15-44.
- Card, S.K, Mackinlay, J.D, and Shneiderman, B. (1999). « Information Visualization ». *Readings in Information Visualization: Using Vision to Think*, 1999, Morgan Kaufmann Publishers edition.
- Carugati, F., de Paolis, P., & Mugny, G. (1980). Conflit de centrations et progrès cognitif, III: régulations cognitives et relationnelles du conflit socio-cognitif. *Bulletin de Psychologie*.
- Cen, L., Ruta, D., Powell, L., & Ng, J. (2014). Does gender matter for collaborative learning?. In *Teaching, Assessment and Learning (TALÉ), 2014 International Conference on* (pp. 433-440). IEEE.
- Cercle SIRH (2017). Le SIRH. Enjeux, bonnes pratiques et innovation. 448 pages.
- Champy, P., & Etévé, C. (2011). *Dictionnaire encyclopédique de l'éducation et de la formation : 3^{ème} version* Nathan, 1108 p.
- Charleer, S., Klerkx, J., & Duval, E. (2014). Learning dashboards. *Journal of Learning Analytics*, 1(3), 199-202.
- Chen, C. M., & Kuo, C. H. (2019). An optimized group formation scheme to promote collaborative problem-based learning. *Computers & Education*, 133, 94-115.
- Chennaoui, A. (2015). *Modèle multi-agents d'aide à la décision pour la gestion des services préhospitaliers d'urgence* (Thèse de doctorat, École de technologie supérieure).
- Chenevotot-Quentin, F., Grugeon-allys, B., Pilet, J., & Delozanne, E. (2012). De la conception à l'usage d'un diagnostic dans une base d'exercices en ligne. Actes Espace mathématique franco-phonie.
- Chniter, M., Abid, A., & Kallel, I. (2018). Towards a Bio-inspired ACO Approach for Building Collaborative Learning Teams. In *2018 17th International Conference on Information Technology Based Higher Education and Training (ITHET)* (pp. 1-8). IEEE.
- Chou, P. N., & Chang, C. C. (2018). Small or large ? The effect of group size on engineering students' learning satisfaction in project design courses. *EURASIA Journal of Mathematics, Science and Technology Education*, 14(10), em1597.
- Chowdhary, P. R., Pinel, F. A., Palpanas, T., & Chen, S. K. (2014). *U.S. Patent No. 8,843,883*. Washington, DC: U.S. Patent and Trademark Office.
- Christodoulopoulos, C. E., & Papanikolaou, K. (2007). Investigation of group formation using low complexity algorithms. In *Proc. of PING Workshop* (pp. 57-60).
- Ciguene, R., Joiron, C. & Dequen, G. (2019). Automatically generating assessment tests within higher education context thanks to genetic approach. In *Bioinspired Heuristics for Optimization* (pp. 269-282). Springer, Cham.

- Clarens V., Formation en entreprise : une fonction partagée ?. In « Actualité de la formation permanente », Centre Inffo, Saint-Denis-La Plaine, n° 213, pp 213-215 2008.
- Cleverdon, C. (1967, Juin). The Cranfield tests on index language devices. In *Aslib proceedings*. MCB UP Ltd.
- Clivillé, V., De Oliva, V., & y Bassa, J. M. (2018). Retour d'expérience sur la réalisation et la pratique d'un MOOC en Aide MultiCritère à la Décision. In 88e Journées du Groupe de Travail Européen «Aide Multicritère à la Décision».
- Coron, C. (2019). Quantification en ressources humaines : usages et analyses. ISTE Group. 224 p.
- Cossette, M., Lépine, C., & Raedecker, M. (2014). Mesurer les résultats de la gestion des ressources humaines: principes, état des lieux et défis à surmonter pour les professionnels RH. *Gestion*, 39(4), 44-54.
- Cossette, M. (2019). *L'analytique ressources humaines*. Editions JFD.
- Cram, D., Jouvin, D., & Mille, A. (2007). Visualisation interactive de traces et réflexivité: application à l'EIAH collaboratif synchrone eMédiathèque. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, 14(1), 491-530.
- Creusefond, J. (2017). *Caractériser et détecter les communautés dans les réseaux sociaux* (Thèse de doctorat) Université de Normandie.
- Cristol, D. (2017). Les communautés d'apprentissage : apprendre ensemble. *Savoirs*, (1), 10-55.
- Crossley, N. (2010). The social world of the network. Combining qualitative and quantitative elements in social network analysis. *Sociologica*, 4(1).
- Crouch, CH et Mazur, E. (2001). Instruction par les pairs : dix ans d'expérience et de résultats. *Journal américain de physique*, 69 (9), 970-977.
- Cuenca, E. (2018). Visualisation de données dynamiques et complexes : des séries temporelles hiérarchiques aux graphes multicouches (Thèse de doctorat). Université de Montpellier.
- Cruz, W. M., & Isotani, S. (2014). Group formation algorithms in collaborative learning contexts: A systematic mapping of the literature. In CYTED-RITOS International Workshop on Groupware (pp. 199-214). Springer, Cham.
- Dabbebi, I. (2019). Conception et génération dynamique de tableaux de bord d'apprentissage contextuels. (Thèse de doctorat). Université du Mans.
- Daniel, B. K. (2016). *Big data and learning analytics in higher education*. New York : Springer.
- Dascalu, M. I., Bodea, C. N., Lytras, M., De Pablos, P. O., & Burlacu, A. (2014). Improving e-learning communities through optimal composition of multidisciplinary learning groups. *Computers in Human Behavior*, 30, 362-371.
- Dalsgaard, C. (2006). Social software : E-learning beyond learning management systems. *European Journal of Open, Distance and e-learning*, 9(2).

- David, A. & Damart, S. (2011). Bernard Roy et l'aide multicritère à la décision. *Revue française de gestion*, 214(5), 15-28.
- Davoodi, A., Kardan, S., et Conati, C. (2013). Comprendre le comportement d'interaction des utilisateurs avec un jeu éducatif intelligent : Prime Climb. Dans *AIED 2013 Workshops Proceedings Volume 2 Scaffolding in Open-Ended Learning Environments (OELE)* (p. 9).
- Decamps, S., & Depover, C. (2011). La perception du tutorat par les acteurs de la formation à distance. *Le tutorat en formation à distance*, 109-124.
- Degenne, A., & Forse, M. (1994). Les réseaux sociaux, une approche structurale en sociologie. Paris : Ed. Armand Colin.
- Delozanne, É., Prévité, D., Grugeon, B., & Chenevotot, F. (2008). Automatic multi-criteria assessment of open-ended questions : A case study in school algebra. In *International Conference on Intelligent Tutoring Systems* (pp. 101-110). Springer, Berlin, Heidelberg.
- De Paolis, P. (1982). Marquage social et développement cognitif. In *conference «New Perspectives in the Experimental Study of the Social Development of the Intelligence», University of Geneva, Switzerland*.
- Depover, C., Quintin, J. J., & De Lièvre, B. (2003). Un outil de scénarisation de formations à distance basées sur la collaboration. In *Environnements Informatiques pour l'Apprentissage Humain 2003* (pp. 469-476). ATIEF; INRP.
- Deschenaux, F. (2007). Guide d'introduction au logiciel QRSNvivo7. *Les cahiers pédagogiques de l'Association pour la recherche qualitative*. Bibliothèque National du Québec
- Dewey, J. (1916/1973). *Democracy and Education*. New York, Macmillan Co. Edited by J.J. McDermott (1973), *The Philosophy of John Dewey*. Volume 2. *The Lived Experience* (pp. 494-523). Chicago, University of Chicago Press
- Dijkkamp, J. (2019). *The recruiter of the future, a qualitative study in AI supported recruitment process* (Master's thesis, University of Twente).
- Dillenbourg, P., Poirier, C., & Carles, L. (2003). Communautés virtuelles d'apprentissage : e-jargon ou nouveau paradigme. A. Taurisson et A. Sentini. *Pédagogies. Net*. Montréal, Presses, 11-47.
- Dimitracopoulou, A., & Bruillard, E. (2006). Enrichir les interfaces de forums par la visualisation d'analyses automatiques des interactions. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, 13(1), 345-397.
- Drachsler, H., Verbert, K., Santos, O. C., & Manouselis, N. (2015). Panorama of recommender systems to support learning. In *Recommender systems handbook* (pp. 421-451). Springer, Boston, MA.
- Draghici, C. (2005). Modélisation et conception d'algorithmes pour la planification automatique du personnel de compagnies aériennes (Thèse de doctorat). Institut National de Sciences Appliquées de Toulouse.

- Dréo, J., Pétrowski, A., Siarry, P., & Taillard, E. (2003). *Métabeuristiques pour l'optimisation difficile* (p. 356). Eyrolles.
- Dos Santos, S., & Brodlie, K. (2004). Gaining understanding of multivariate and multidimensional data through visualization. *Computers & Graphics*, 28(3), 311-325.
- Doise, W. (2003). 10. Attitudes et représentations sociales. *D. Jodelet, Les représentations sociales*, 240.
- Du, N., Wu, B., Pei, X., Wang, B., & Xu, L. (2007). Community detection in large-scale social networks. In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis (pp. 16-25). ACM.
- Dubar C., (2000), *La socialisation. Construction des identités sociales et professionnelles*, Armand Colin, collection U (3e édition), 256p.
- Dubois, J. M., Marion-Vernoux, I., & Noack, E. (2016). Le dispositif d'enquêtes : Défis, un nouveau regard sur la formation en entreprise. *Céreq* (Centre d'études et de recherches sur les qualifications) n° 344.
- Dujarier, M. A. (2010). L'automatisation du jugement sur le travail. Mesurer n'est pas évaluer. *Cahiers internationaux de sociologie*, (1), 135-159.
- Duprez, J. M., & Barbut, M. (2007). Apprécier l'homogénéité dans la formation de paires. *Revue française de sociologie*, 48(2), 333-367.
- Dupriez, V., & Draelants, H. (2004). Classes homogènes versus classes hétérogènes : les apports de la recherche à l'analyse de la problématique. *Revue française de pédagogie*, 148(1), 145-165.
- Durkheim, É. (1898). Représentations individuelles et représentations collectives. *Revue de Métaphysique et de Morale*, 6(3), 273-302.
- Dussarps, C. (2014). *Dimension socio-affective et abandon en formation ouverte et à distance*-Thèse de doctorat, Université de Bordeaux.
- Duval, E. (2011). Attention please! : learning analytics for visualization and recommendation. *LAK*, 11, 9-17.
- Dwivedi, P., & Bharadwaj, K. K. (2015). eLearning recommender system for a group of learners based on the unified learner profile approach. *Expert Systems*, 32(2), 264-276.
- Dwivedi, S., & Roshni, V. K. (2017). Recommender system for big data in education. In *2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH)* (pp. 1-4). IEEE.
- Dwivedi, P.; Kant, V. & Bharadwaj, K. K. (2018). Learning path recommendation based on modified variable length genetic algorithm. *Education and Information Technologies*, 23(2), 819-836
- Edgeworth, 1881 Edgeworth F-Y., *Mathematical Physics*, P. Keagan, London, 1881.
- Elias, T. (2011). Learning analytics. *Learning*, 1-22.

- Esteban, A.; Zafra, A. & Romero, C. (2018). A Hybrid Multi-Criteria Approach Using a Genetic Algorithm for Recommending Courses to University Students. International Educational Data Mining Society.
- Fablet, D. (2004). Les groupes d'analyse des pratiques professionnelles : une visée avant tout formative. *Connexions*, (2), 105-117.
- Falkenauer, E. (1998). *Genetic algorithms and grouping problems*. John Wiley & Sons, New-York.
- Fericelli, A. M. (1978). *Théorie statistique de la décision : application à la gestion des entreprises*. Paris : Economica.
- Feschet, P., & Lairez, J. (2016). Chapitre 3. Comprendre et utiliser des indicateurs dans une évaluation multicritère. In *Agriculture et développement durable* (pp. 59-90). Educagri éditions.
- Felfernig, A., Boratto, L., Stettinger, M., & Tkalčič, M. (2018). *Group recommender systems: An introduction*. Springer International Publishing.
- Festinger, L. (1949). The analysis of sociograms using matrix algebra. *Human relations*, 2, 153-158.
- Festinger, L. (1954). Une théorie des processus de comparaison sociale. *Relations humaines*, 7 (2), 117-140.
- Filius, R. M., de Kleijn, R. A., Uijl, S. G., Prins, F. J., van Rijen, H. V., & Grobbee, D. E. (2018). Strengthening dialogic peer feedback aiming for deep learning in SPOCs. *Computers & Education*, 125, 86-100.
- Fiol, M. (2006). Contrôle de gestion et cohérence organisationnelle : un rendez-vous manqué, *27ème congrès de l'Association Francophone de Comptabilité*, Tunis.
- Fisher Gustave-Nicolas, Les concepts fondamentaux de la psychologie sociale, Paris, Dunod, 2005 (1997), 278 pp.
- Fitz-Enz, J. (2010). The new HR analytics. *Predicting the Economic Value of Your Company's Human Capital Investments*.
- Forsé, M. (2008). Définir et analyser les réseaux sociaux. *Informations sociales*, (3), 10-19.
- Fortin, L., Filiault, M., Plante, A., & Bradley, M. F. (2011). Recension des écrits sur le regroupement homogène ou hétérogène des élèves. *Sherbrooke, Québec: Chaire de recherche de la Commission scolaire de la Région-de-Sherbrooke sur la réussite et la persévérance scolaire*.
- Fortunato. Community detection in graphs. *Physics Reports*, 486(3) :75–174, 2010.
- Fougères, A. J. (2001). Un système tutoriel intelligent adapté à la formation de régulateurs de trafic. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, 8(1), 141-147.
- Franchini, L. (2000). *Aide à la décision pour la gestion des opérateurs en production: modélisation, planification et évaluation* (Thèse de doctorat), Université de Toulouse.

- Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11), 1129-1164.
- Gadéa, C., & Demazière, D. (2009). Sociologie des groupes professionnels. Acquis récents et nouveaux défis. *Lectures, Publications Reçues*.
- Gallupe, R. B., Dennis, A. R., Cooper, W. H., Valacich, J. S., Bastianutti, L. M., & Nunamaker Jr, J. F. (1992). Electronic brainstorming and group size. *Academy of Management Journal*, 35(2), 350-369.
- García-Arroyo José, Amparo Osca. (2019) Big data contributions to human resource management : a systematic review. *The International Journal of Human Resource Management*, pages 1-26.
- Garey, M. R. (1979). A Guide to the Theory of NP-Completeness. *Computers and intractability*.
- Garrot, É. (2007). Assistance au tuteur: prototype d'un système pour l'adaptation de situations d'apprentissage aux apprenants. *Revue des Sciences et Technologies de l'Information-Série TSI: Technique et Science Informatiques*, 26(6), 723-750.
- Gérard, B., D'hont, L., Oboeuf, A. (2013). Repérer les relations socio-affectives dans les groupes organisés. Le «détournement» de la grille d'observation de Bales». *Recherches Qualitatives*, 32(2), 103-128.
- Gentili, F. (2005). Comment définir l'identité professionnelle ?. Dans : F. Gentili, *La rééducation contre l'école, tout contre: L'identité professionnelle des rééducateurs en question* (pp. 17-57). Toulouse, France.
- Gilliot, Jean-Marie, Sébastien Iksal, Daniel Medou, and Inès Dabbebi. "Conception participative de tableaux de bord d'apprentissage." In *IHM'18: 30e Conférence Francophone sur l'Interaction Homme-Machine*, pp. pp-119. 2018.
- Ginns, P., & Ellis, R. (2007). Quality in blended learning: Exploring the relationships between on-line and face-to-face teaching and learning. *The Internet and Higher Education*, 10(1), 53-64.
- Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12), 7821-7826.
- Gilbert, F. (2012). *Méthodes et modèles pour la visualisation de grandes masses de données multidimensionnelles nominatives dynamiques*-Thèse de doctorat, Université de Bordeaux.
- Goguelin, P. (1994). *La Formation continue des adultes*. Paris cedex 14, France: Presses Universitaires de France.
- Goldberg, D. E., & Holland, J. H. (1989). Genetic algorithms and machine learning. Reading MA Addison Wesley.
- Gorry, G. A., & Scott Morton, M. S. (1971). A framework for management information systems.
- Granovetter, M. S. (1977). The strength of weak ties. In *Social networks* (pp. 347-367).

- Gras, B. (2019). Éthique des Learning Analytics. *Distances et médiations des savoirs. Distance and Mediation of Knowledge*, (26).
- Gruber, TR (1995). Vers des principes de conception d'ontologies utilisées pour le partage des connaissances ?. *Revue internationale d'études homme-ordinateur*, 43 (5-6), 907-928.
- Grugeon-Allys, B. (2016). Modéliser le profil diagnostique des élèves dans un domaine mathématique et l'exploiter pour gérer l'hétérogénéité des apprentissages en classe: une approche didactique multidimensionnelle. *e-JIREF*, 2(2), 63-88.
- Guillemot, D., & Melnik-Olive, E. (2017). Introduction: La formation continue en contexte: l'entreprise et ses salariés au cœur des enjeux. *Formation emploi. Revue française de sciences sociales*, (137), 7-14.
- Hachour, H. (2015). De la fouille à la visualisation de données: un processus interprétatif. *I2D Information, donnees documents*, 52(2), 42-43.
- Hadji, C.(2012). Faut-il avoir peur de l'évaluation ? Bruxelles: de Boeck. *Mesure et évaluation en éducation*, 36(1), 111-117.
- Hämäläinen, W., Kumpulainen, V., & Mozgovoy, M. (2015). Évaluation des méthodes de clustering pour les systèmes d'apprentissage adaptatifs. Dans *les applications de l'intelligence artificielle dans l'enseignement à distance* (p. 237-260). IGI Global.
- Harary, F., Norman, R. Z., & Cartwright, D. (1965). Structural models : an introduction to the theory of directed graphs. New York : Wiley.
- Haythornthwaite, C. (2002). Strong, weak, and latent ties and the impact of new media. *The information society*, 18(5), 385-401.
- Hemamou, L., Felhi, G., Vandenbussche, V., Martin, J. C., & Clavel, C. (2019, July). HireNet: A Hierarchical Attention Model for the Automatic Analysis of Asynchronous Video Job Interviews. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 573-581).
- Herodotou, C., Rienties, B., Verdin, B., & Borooa, A. (2019). Predictive Learning Analytics 'At Scale': Guidelines to Successful Implementation in Higher Education. *Journal of Learning Analytics*, 6(1), 85—95. <https://doi.org/10.18608/jla.2019.61.5>
- Hmelo-Silver, C. E., Rosé, C. P., & Levy, J. (2014). Fostering a Learning Community in MOOCs. In *LAK Workshops*.
- Ho, H. N., Rabah, M., Nowakowski, S., & Estrailier, P. (2015, June). Personnalisation interactive de parcours pédagogiques par une méthode de décision multicritère à base de traces. In *7ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (ELAH 2015)* (pp. 429-431).
- Holland John H.. *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, 1975.
- Holzer, A., Voznuik, A., Gillet, D., Rodríguez-Triana, M. J., Schwendimann, B. A., Prieto, L. P., & Dillenbourg, P. (2016, April). Understanding learning at a glance: A systematic literature review of learning dashboards. In *6th International Conference on Learning Analytics & Knowledge (LAK'16)*.

- Horn, J. (1997). Multicriteria Decision Making and Evolutionary Computation in T. Bäck, DB Fogel, Z. Michalewicz (Eds.), *Handbook of Evolutionary Computation*.
- Huguet, P., Charbonnier, E., & Monteil, J. M. (1999). Productivity loss in performance groups: People who see themselves as average do not engage in social loafing. *Group dynamics: Theory, research, and practice*, 3(2), 118.
- Hurter, C. (2010). Caractérisation de visualisations et exploration interactive de grandes quantités de données multidimensionnelles (Thèse de doctorat).
- Jameson, A. (2004). More than the sum of its members: challenges for group recommender systems. In *Proceedings of the working conference on Advanced visual interfaces* (pp. 48-54). ACM.
- Jameson, A., & Smyth, B. (2007). Recommendation to groups. In *The adaptive web* (pp. 596-627). Springer, Berlin, Heidelberg.
- Jatobá, M., Santos, J., Gutierrez, I., Moscon, D., Fernandes, P. O., & Teixeira, J. P. (2019). Evolution of Artificial Intelligence Research in Human Resources. *Procedia Computer Science*, 164, 137-142.
- Jobert, G. (2013). Le formateur d'adultes : un agent de développement. *Nouvelle revue de psychosociologie*, (1), 31-44.
- Joseph, N., Pradeesh, N., Chatterjee, S., & Bijlani, K. (2017). A novel approach for group formation in collaborative learning using learner preferences. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1564-1568). IEEE.
- Kalles, D. & Pierrakeas, C. (2006). Analyzing student performance in distance learning with genetic algorithms and decision trees. *Applied Artificial Intelligence*, 20(8), 655-674.
- Kast, R. (1993). *La théorie de la décision* (pp. 61-64). Paris: La Découverte.
- Kay, R., Li, J., & Markovich, L. (2016). Examining Blended Learning for Adult Learners with Special Needs. In *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (pp. 711-718). Association for the Advancement of Computing in Education (AACE).
- Keim, D., Andrienko, G., Fekete, J. D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In *Information visualization* (pp. 154-175). Springer, Berlin, Heidelberg.
- Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F. (2010). Mastering the information age: solving problems with visual analytics.
- Kessler, R., & Lapalme, G. (2017). Agohra : Génération d'une ontologie dans le domaine des ressources humaines. *Traitement Automatique des Langues*, 58(1), 39-62.
- Kirkpatrick, D. L. (2009). *Implementing the four levels: A practical guide for effective evaluation of training programs*.
- Koschmann, T. (1992). Computer Support for Collaborative Learning: Design, Theory, and Research Issues. *ACM SIGCUE Outlook*, 21(3), 1-2.

- Konert, J., Burlak, D., & Steinmetz, R. (2014). The group formation problem: an algorithmic approach to learning group formation. In *European Conference on Technology Enhanced Learning* (pp. 221-234). Springer, Cham.
- Kooloos, J. G., Klaassen, T., Vereijken, M., Van Kuppeveld, S., Bolhuis, S., & Vorstenbosch, M. (2011). Collaborative group work: Effects of group size and assignment structure on learning gain, student satisfaction and perceived participation. *Medical Teacher*, 33(12), 983-988.
- Koza, J. R. (1994). Genetic programming as a means for programming computers by natural selection. *Statistics and computing*, 4(2), 87-112.
- Kozlowski, S. W. J. Bell (2013). «Work groups and teams in organizations». *Handbook of Psychology: Industrial and Organizational Psychology*. Wiley, London, 12, 333-375.
- Kozlowski, SW et Bell, BS (2008). Apprentissage, développement et adaptation en équipe. *Apprentissage en groupe*, 15-44.
- Knowles, M. (1970). The modern practice of adult education: Andragogy versus pedagogy. New-York (NY): Association Press.
- Knowles, M. (1990). L'apprenant adulte, vers un nouvel art de la formation. Paris: Les Éditions d'Organisation.
- Kučak, D., Juričić, V., & Đambić, G. (2019). Application of genetic algorithmes in higher education aera.. *Annals of DAAAM & Proceedings*, 30.
- Kun-fa, L., Jing-chun, C., & Yan-xi, W. (2019). Big Data Informatization Applied to Optimization of Human Resource Performance Management. In *Proceedings of the 2019 2nd International Conference on Information Management and Management Sciences* (pp. 12-17). ACM.
- Laferrière, T. (2013). TEN-7007—Environnement d'apprentissage en réseau.
- Lazarsfeld, P. F., & Merton, R. K. (1954). Friendship as a social process: A substantive and methodological analysis. *Freedom and control in modern society*, 18(1), 18-66.
- Lazega, E. (1994). Analyse de réseaux et sociologie des organisations. *Revue française de sociologie*, 293-320.
- Lazega, E. (2014). *Réseaux sociaux et structures relationnelles: «Que sais-je?» n° 3399*. Que sais-je.
- Le Boucher, C. (2015). Facteurs de pérennisation d'un réseau de formation par les pairs: le cas des Réseaux d'Echanges Réciproques de Savoirs (Thèse de Doctorat, Université de Rennes).
- Le Gall, A. C. (2009). Panorama des méthodes d'analyse multicritère comme outils d'aide à la décision. *Rapport D'étude ONEMA-INNERIS*.
- Lemire, L., Charest, É., Martel, G., & Larivière, J. (2015). *La planification stratégique des ressources humaines*. PUQ.
- Leon Blanco, M. (2019). Analyse, mesure et optimisation de la performance du système d'infor-

mation d'une entreprise en croissance et de ses départements clés.

Leony, D., Pardo, A., de la Fuente Valentín, L., de Castro, D. S., & Kloos, C. D. (2012). GLASS: a learning analytics visualization tool. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 162-163).

Le Louarn, J. Y. (2008). *Les tableaux de bord ressources humaines: le pilotage de la fonction RH*. Wolters Kluwer France.

Lemieux, V., & Ouimet, M. (2004). *L'analyse structurale des réseaux sociaux*. De Boeck Supérieur.

Le Riche, R., Schoenauer, M., & Sebag, M. (2007). Un état des lieux de l'optimisation évolutionnaire et de ses implications en sciences pour l'ingénieur. In *Modélisation numérique. 2, Défis et perspectives*. Traité MIM, série Méthodes numériques et éléments finis, Hermès, pp.187-259.

Leung, Y. K., & Apperley, M. D. (1994). A review and taxonomy of distortion-oriented presentation techniques. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 1(2), 126-160.

Levine, J. M., & Moreland, R. L. (1998). Small groups.

Lewin, K. (1959). *Psychologie et dynamique*, Presse Universitaire de France.

Liao, SH, Chu, PH et Hsiao, PY (2012). Techniques et applications d'exploration de données - Une décennie d'examen de 2000 à 2011. *Systèmes experts avec applications*, 39 (12), 11303-11311.

Lin, Y. T., Huang, Y. M., & Cheng, S. C. (2010). An automatic group composition system for composing collaborative learning groups using enhanced particle swarm optimization. *Computers & Education*, 55(4), 1483-1493.

Longueville, B., & Gardoni, M. (2003). A survey of context modeling: approaches, theories and use for engineering design researches. In *DS 31: Proceedings of ICED 03, the 14th International Conference on Engineering Design, Stockholm* (pp. 437-438).

Lopez, M. I., Luna, J. M., Romero, C., & Ventura, S. (2012). Classification via clustering for predicting final marks based on student participation in forums. *International Educational Data Mining Society*.

Lu, H. (2018). Construction of spoc-based learning model and its application in linguistics teaching. *International Journal of Emerging Technologies in Learning (IJET)*, 13(02), 157-169.

Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., & Yang, S. J. (2018). Applying learning analytics for the early prediction of Students' academic performance in blended learning. *Journal of Educational Technology & Society*, 21(2), 220-232.

Lv, G., Hu, C. et Chen, S. (2016). Recherche sur le système de recommandation basé sur l'ontologie et l'algorithme génétique. *Neurocomputing*, 187, 92-97.

Magnisalis, I., Demetriadis, S., & Karakostas, A. (2011). Adaptive and intelligent systems for collaborative learning support: A review of the field. *IEEE transactions on Learning Technologies*, 4(1), 5-20.

- Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., & Koper, R. (2011). Recommender systems in technology enhanced learning. In *Recommender systems handbook* (pp. 387-415). Springer, Boston, MA.
- Mansour, N., & Gaha, C. (2004). Contribution des pratiques GRH au management des savoirs: Quelques voies de réflexion.
- Marchand L, « Un changement de paradigme, pour un enseignement universitaire moderne », *Distances*, vol.2, n°2, 1998, p. 7-26
- Maskit, D., & Hertz-Lazarowitz, R. (1986). Adults in Cooperative Learning: Effects of Group Size and Group Gender Composition on Group Learning Behaviors (A Summary).
- Martin, J.P, Savary, E. (1996). *Formateur d'adulte. Se professionnaliser. Exercer au quotidien*. Bruxelles, Lyon : Evo Chronique sociale.
- Márquez, J. O. Á., & Ziegler, J. (2015). Preference elicitation and negotiation in a group recommender system. In *IFIP Conference on Human-Computer Interaction* (pp. 20-37). Springer, Cham.
- Marler, J. H., & Boudreau, J. W. (2017). An evidence-based review of HR Analytics. *The International Journal of Human Resource Management*, 28(1), 3-26.
- Matias, J. B.; Fajardo, A. C. & Medina, R. P. (2018) A Hybrid Genetic Algorithm for Course Scheduling and Teaching Workload Management. In 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM) (pp. 1-6)
- Matthews, T., Whittaker, S., Moran, T., & Yuen, S. (2011). Collaboration personas: A new approach to designing workplace collaboration tools. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2247-2256).
- Maqtary, N., Mohsen, A., & Bechkoum, K. (2019). Group formation techniques in computer-supported collaborative learning: A systematic literature review. *Technology, Knowledge and Learning*, 24(2), 169-190.
- Mazur, E. (1997). Enseignement par les pairs: amener les élèves à réfléchir en classe. In *AIP Conference Proceedings* (Vol. 399, n ° 1, pp. 981-988). Institut américain de physique.
- Mayen, P. & Mayeux, C. (2003). Expérience et formation. *Savoirs*, 1(1), 13-53.
- McCain, D.V. (2005). *Evaluation Basics*. ASTD Press, Alexandria.
- McCarthy B. 1985. *Teaching to Learning Styles*. Barrington: Excel Inc. Illinois.
- McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of consulting and clinical psychology*, 51(6), 882.
- McPherson, J. M., & Smith-Lovin, L. (1987). Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *American sociological review*, 370-379.

- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1), 415-444.
- Mesquita, C., & Lopes, R. P. (2018). Assessing students' participation under different group formation strategies. In *10th International Conference on Education and New Learning Technologies* (pp. 9802-9811).
- Mercklé, P. (2011). *Sociologie des réseaux sociaux*. Paris : La Découverte.
- Metz, C. (1971). Réflexions sur la "Sémiologie graphique" de Jacques Bertin. In *Annales. Histoire, Sciences Sociales* (Vol. 26, No. 3-4, pp. 741-767). Cambridge University Press.
- Micari, M., & Drane, D. (2011). Intimidation in small learning groups: The roles of social-comparison concern, comfort, and individual characteristics in student academic outcomes. *Active Learning in Higher Education*, 12(3), 175-187.
- Micari, M., & Pazos, P. (2014). Worrying about what others think: A social-comparison concern intervention in small learning groups. *Active Learning in Higher Education*, 15(3), 249-262.
- Michailidis, M., (2018). Hie défis de l'IA et de la blockchain sur les pratiques de recrutement RH. *Chypre Review* , 30 (2).
- Mikeli, A., Apostolou, D., & Despotis, D. (2015). A new recommendation technique for interval scaled multi-criteria rating systems incorporating intensity of preferences. *Intelligent Decision Technologies*, 9(3), 283-294.
- Mizzaro, S. (1996). A cognitive analysis of information retrieval. *Information Science: Integration in Perspective—Proceedings of CoLIS2*, 233-250.
- Moreno, J. L. (1934). Who shall survive?: A new approach to the problem of human interrelations.
- Moreno, J., Ovalle, D. A., & Vicari, R. M. (2012). A genetic algorithm approach for group formation in collaborative learning considering multiple student characteristics. *Computers & Education*, 58(1), 560-569.
- Mortensen, M., Doherty, N. and Robinson, S. (2015). 'Operational research from taylorism to terabytes: a research agenda for the analytics age'. *European Journal of Operational Research*, 241: 3,583– 595.
- Moscovici, S. (2003). 2. Des représentations collectives aux représentations sociales: éléments pour une histoire. In *Les représentations sociales* (Vol. 7, pp. 79-103). Presses Universitaires de France.
- Mucchielli R. 1991. *Les méthodes actives dans la pédagogie des adultes*. Paris : ESF éditeur, 8ème édition
- Mucchielli, R. (2016). *Les méthodes actives: Dans la pédagogie des adultes*. ESF Sciences Humaines.
- Mucchielli, R. (2019). *Le travail en équipe: clés pour une meilleure efficacité collective*. ESF Sciences Humaines.

- Mutingi, M. et Mbohwa, C. (2017). Formation d'équipes multicritères à l'aide de l'approche d'algorithme génétique de regroupement flou. Dans *Grouping Genetic Algorithms* (pp. 89-105). Springer, Cham.
- Nadeau, R., & Landry, M. (1986). *L'aide à la décision: nature, instruments et perspectives d'avenir* (Vol. 2). Presses Université Laval.
- Newman, M. E., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2), 026113.
- Negre Elsa. Système de recommandation : introduction. ISTE Group. 1 juin 2015. 82 pages.
- Nieto, Y., Gacía-Díaz, V., Montenegro, C., González, C. C., & Crespo, R. G. (2019). Usage of machine learning for strategic decision making at higher educational institutions. *IEEE Access*, 7, 75007-75017.
- Noaman, A. Y., & Ahmed, F. F. (2015). ERP systems functionalities in higher education. *Procedia Computer Science*, 65, 385-395.
- Odo, C., Masthoff, J., et Beacham, N. (2019). Formation de groupe pour l'apprentissage collaboratif. Dans *International Conference on Artificial Intelligence in Education* (pp. 206-212). Springer, Cham.
- Olakanmi, O. A., & Vassileva, J. (2017). Group matching for peer mentorship in small groups. In *CYTED-RITOS International Workshop on Groupware* (pp. 65-80). Springer, Cham.
- Oliveri, N. & Rasse, P. (2017). Les Mooc et leurs dérivés, ou l'imaginaire des technologies pédagogiques. *Hermès, La Revue*, 78(2), 110-117.
- O'neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.
- Osman, I. H. (1999). A unified-metaheuristic framework. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* (pp. 11-12). Springer, Berlin, Heidelberg.
- Ouakrat, A., & Mésangeau, J. (2016). Resocialiser les traces d'activités numériques: une proposition qualitative pour les SIC. *Revue française des sciences de l'information et de la communication*, (8).
- Pareto V., Cours d'économie politique, Rouge, Lausanne, 1896.
- Papamitsiou, Z., & Economides, A. A. (2018). Recommendation of educational resources to groups: A game-theoretic approach. In *2018 IEEE Global Engineering Education Conference (EDU-CON)* (pp. 760-767). IEEE.
- Paquay, L., Crahay, M., & De Ketele, J. M. (2006). L'analyse qualitative en éducation: des pratiques de recherche aux critères de qualité.
- Peraya, D. (2019). Les Learning Analytics en question. Panorama, limites, enjeux et visions d'avenir. *Distances et médiations des savoirs. Distance and Mediation of Knowledge*, (25).

- Peraya, D. (2019). Les Learning Analytics: contraintes méthodologiques et «gouvernance» éthique des données. *Distances et médiations des savoirs. Distance and Mediation of Knowledge*, (26).
- Pertinant Guillaume, Richard Sébatien, Storhaye Patrick, (2017). Analytique RH : Démarche, bénéfices, défis, Edition Management et Société (EMS), Caen.
- Petit, A. (1993). *Gestion stratégique et opérationnelle des ressources humaines*.
- Piasser, A. (2014). Représentations professionnelles. *Dictionnaire des concepts de la professionnalisation (1re éd., p. 273-281)*. Bruxelles: De Boeck Supérieur.
- Picard, D. & Marc, E. (2013). Introduction. Dans : Dominique Picard éd., *L'École de Palo Alto* (pp. 3-6). Presses Universitaires de France.
- Poon, L. K., Kong, S. C., Yau, T. S., Wong, M., & Ling, M. H. (2017). Learning analytics for monitoring students participation online: Visualizing navigational patterns on learning management system. In *International Conference on Blended Learning* (pp. 166-176). Springer, Cham.
- Postic, M. (2015). La relation éducative. Presses universitaires de France.
- Poizat, G., & Durand, M. (2017). Réinventer le travail et la formation des adultes à l'ère du numérique: état des lieux critique et prospectif. *Raisons éducatives*, (1), 19-44.
- Pollalis, Y. A., & Mavrommatis, G. (2009). Using similarity measures for collaborating groups formation: A model for distance learning environments. *European Journal of Operational Research*, 193(2), 626-636.
- Power, D. J. (2008). Decision support systems: a historical overview. *Handbook on decision support systems 1*, 121-140.
- Power, D. J. (2014). Using Big Data for analytics and decision support. *Journal of Decision Systems*, 23(2), 222-228.
- Prabhakar, S. (2017). Reciprocal Recommendation System and Formation of Learning Groups in Massive Open Online Courses (MOOCs).
- Prat N., Comyn-Wattiau I., Akoka J. (2015), A Taxonomy of Evaluation Methods for Information Systems Artifacts, *Journal of Management Information Systems*, vol. 32, n° 3, p. 229-267.
- Prévit, D. (2008). Génération d'exercices et analyse multicritère automatique de réponses ouvertes: PépiGen, un système auteur en algèbre élémentaire (Thèse de doctorat, Le Mans).
- Putro, B. L., Rosmansyah, Y., & Suhardi, S. (2020). An intelligent agent model for learning group development in the digital learning environment: A systematic literature review. *Bulletin of Electrical Engineering and Informatics*, 9(3), 1159-1166.
- Rafiq, K. R. M., Hashim, H., & Yunus, M. M. (2019). MOOC for Training: How Far It Benefits Employees?. In *Journal of Physics: Conference Series* (Vol. 1424, No. 1, p. 012033). IOP Publishing.

- Raynal, F., & Rieunier, A. (2018). *Pédagogie, dictionnaire des concepts clés: apprentissages, formation, psychologie cognitive*. ESF Sciences Humaines.
- Redjem, R. (2013). *Aide à la décision pour la planification des activités et des ressources humaines en hospitalisation à domicile* (Thèse de doctorat)-Université Jean Monnet-Saint-Etienne.
- Renauld R et Roy, D (2010). Modélisation du système de pilotage d'un organisme de formation. In *8e Conférence Internationale de MODélisation et SIMulation-MOSIM'10*.
- Resta, P., & Laferrière, T. (2007). Technology in support of collaborative learning. *Educational Psychology Review*, 19(1), 65-83.
- Reix, R. (2004). *Systèmes d'information et management des organisations* (Vuibert).
- Rey, JP. Le groupe-Ed. Revue EPS, 2000 - 123 pages
- Riel, M., & Polin, L. (2004). Learning communities: Common ground and critical differences in designing technical support. *Designing for virtual communities in the service of learning*, 16-52.
- Roy, B. (1960). *Contribution de la théorie des graphes à l'étude des problèmes d'ordonnement*. Société d'Economie et de Mathématique Appliquées.
- Roy, B. (1971). Problems and methods with multiple objective functions. *Mathematical programming*, 1(1), 239-266.
- Rheingold, H., Corneli, J., Danoff, C., Pierce, C., & MacDonald, L. (2015). *Peeragogy Handbook: A guide for Peer-Learning and Peer Production*.
- Russell, S.J. and Norvig, P. (2016) *Artificial Intelligence: A Modern Approach*. Pearson Education Limited, Malaysia.
- Rygielski, C., Wang, J. C., & Yen, D. C. (2002). Data mining techniques for customer relationship management. *Technology in society*, 24(4), 483-502.
- Sadeghi, H., & Kardan, A. A. (2016). Toward effective group formation in computer-supported collaborative learning. *Interactive learning environments*, 24(3), 382-395.
- Sadok, W., Angevin, F., Bergez, J. E., Bockstaller, C., Colomb, B., Guichard, L., ... & Doré, T. (2008). Ex ante assessment of the sustainability of alternative cropping systems: implications for using multi-criteria decision-aid methods. A review. *Agronomy for Sustainable Development*, 28(1), 163-174.
- Salles, M. (2015). *Décision et système d'information* (Vol. 2). ISTE Group.
- Sauvé, L. (2014). Des dispositifs en ligne pour personnaliser l'apprentissage tout au long de la vie: quelques recommandations. *Distances et médiations des savoirs. Distance and Mediation of Knowledge*, 2(5).
- Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., ... & Dillenbourg, P. (2016). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies*, 10(1), 30-41.

- Scholtz, B., Calitz, A. P., & Jonamu, B. (2016). A Framework for Environmental Management Information Systems in Higher Education. In *Information Technology in Environmental Engineering* (pp. 29-40). Springer, Cham.
- Sclater, N., Peasgood, A., & Mullan, J. (2016). Learning analytics in higher education. *London: Jisc. Accessed February, 8, 2017.*
- Sharma, M., Singh, G., et Singh, R. (2019). Conception de cadres PNL basés sur l'AG et l'ontologie pour l'exploration d'opinions en ligne. *Brevets récents sur l'ingénierie*, 13 (2), 159-165.
- Shukla, M., & Malviya, A. K. (2019). Modified classification and prediction model for improving accuracy of student placement prediction. In *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE)*.
- Sie, R. L., Ullmann, T. D., Rajagopal, K., Cela, K., Bitter-Rijkema, M., & Sloep, P. (2012). Social network analysis for technology-enhanced learning: review and future directions.
- Siemens, G. (2010). What are learning analytics. *Retrieved March, 10, 2011.*
- Siemens, G. (2011). Learning and academic analytics. *Learning and Knowledge Analytics*, 5.
- Siemens, G., & Baker, R. S. J. d. (2012). Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge* (p. 252–254). New York, NY, USA: ACM.
- Simmel G., (1908 réédité en 1999). *Sociologie, études sur les formes de la socialisation*, Paris, PUF.
- Smith, B. L., & MacGregor, J. T. (1992). What is collaborative learning. *Towards the Virtual University: International Online Learning Perspectives*, 217-232.
- Solar, C. (2011). La groupe de formation In Carré, P., & Caspar, P. (2011). *Traité des sciences et des techniques de la formation* (p. 648). Paris: Dunod.
- Srisamutr, A.; Raruaysong, T. & Mettanant, V. (2018). A Course Planning Application for Undergraduate Students Using Genetic Algorithm. In 2018 Seventh ICT International Student Project Conference (ICT-ISPC) (pp. 1-5).
- Storhay Patrick (2013). *Le SIRH : enjeux, facteurs de succès et perspectives*. Ed. Dunod, 240p.
- Storn, R., & Price, K. (1997). Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4), 341-359.
- Sukstrienwong, A. (2012). Genetic algorithm for forming student groups based on heterogeneous grouping. In *Recent advances in information science: Proceedings of the 3rd European conference of computer science* (pp. 92-97).
- Temperman, G. (2013). *Visualisation du processus collaboratif et assignation de rôles de régulation dans un environnement d'apprentissage à distance* (Thèse de doctorat, Université de Mons).

- Tchounikine, P. (2002). Pour une ingénierie des Environnements Informatiques pour l'Apprentissage Humain.
- Tchounikine, P., & Tricot, A. (2011). Environnements informatiques et apprentissages humains. *Informatique et Sciences Cognitives: influences ou confluences*, 167-200.
- Tournois, J., Mesnil, F., & Kop, J. L. (2000). Autoduperie et hétéroduperie: un instrument de mesure de la désirabilité sociale. *Revue européenne de psychologie appliquée*, 50(1), 219-232.
- Touzani, M., Ponsard, C., Laurent, A., Rouge, T. L., & Quinqueton, J. (2016, May). Vers une modélisation et une analyse des exigences spatio-temporelles. In *INFORSID* (pp. 51-66).
- Thammano, A., & Moolwong, J. (2010). A new computational intelligence technique based on human group formation. *Expert Systems with Applications*, 37(2), 1628-1634.
- Van Maanen, J., & Barley, R. S. (1984). Occupational Communities: Culture and Control in Organizations. *Research in Organizational Behaviour*.
- Van Wijk, J. J. (2005). The value of visualization. In *VIS 05. IEEE Visualization, 2005*. (pp. 79-86). IEEE.
- Vayer, P., & Roncin, C. (1987). *L'enfant et le groupe: la dynamique des groupes d'enfants dans la classe*. FeniXX.
- Vayre, J. S. (2015). Les tableaux de bord sur données massives, pour un nouveau management de l'innovation?. *Innovations*, (2), 101-121.
- Vedapradha, R., Hariharan, R., & Shivakami, R. (2019). Artificial Intelligence: A Technological Prototype in Recruitment. *Journal of Service Science and Management*, 12(3), 382-390.
- Veneziano, T., Meyer, P., & Bisdorff, R. (2010). Analyse inverse robuste à partir d'informations préférentielles partielles. *Actes Roadef 2010*, 75-88.
- Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10), 1500-1509.
- Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, 122, 119-135.
- Viberg, O., Hatakka, M., Bälter, O., et Mavroudi, A. (2018). Le paysage actuel de l'analyse de l'apprentissage dans l'enseignement supérieur. *Ordinateurs dans le comportement humain*, 89, 98-110.
- Von Landesberger, T., Kuijper, A., Schreck, T., Kohlhammer, J., van Wijk, J. J., Fekete, J. D., & Fellner, D. W. (2011, September). Visual analysis of large graphs: state-of-the-art and future research challenges. In *Computer graphics forum* (Vol. 30, No. 6, pp. 1719-1749). Oxford, UK: Blackwell Publishing Ltd.
- Voyer, P. (2011). *Tableaux de bord de gestion et indicateurs de performance: 2e édition*. PUQ.
- Vuillemot, R. (2010). *Un cadre de conception pour la visualisation d'information interactive* (Thèse de doctorat, Lyon, INSA).

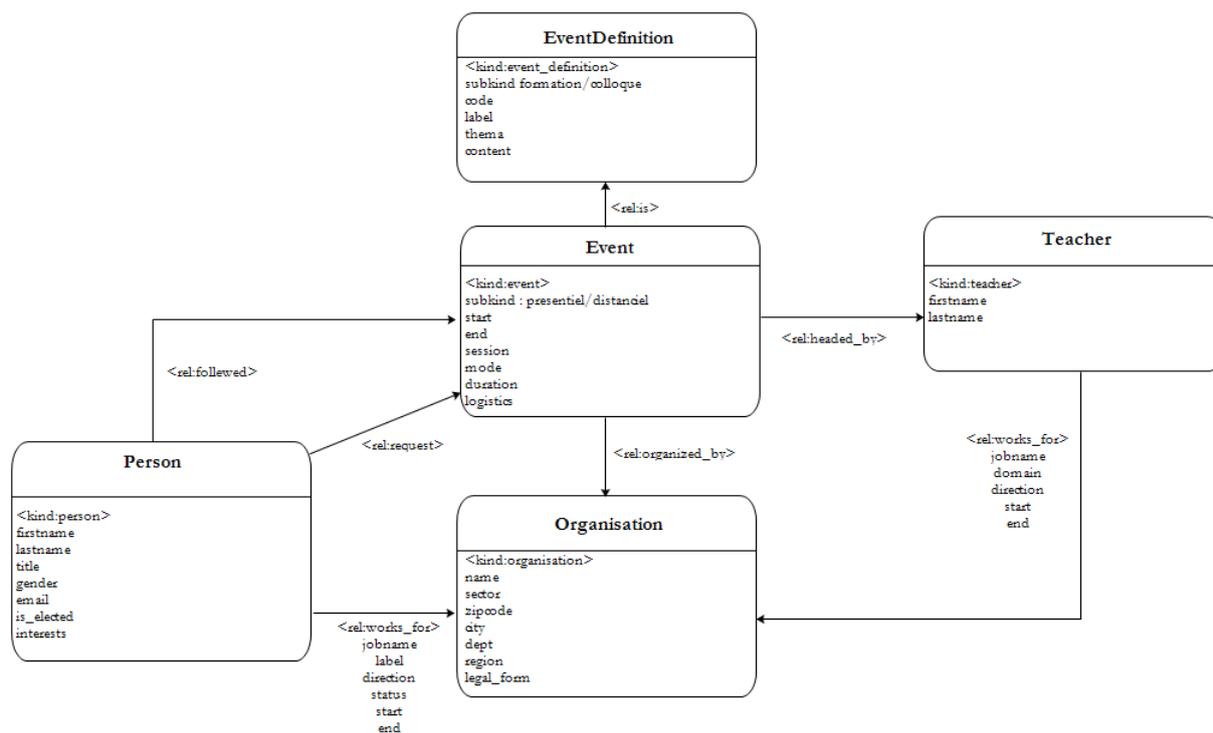
- Webber, C. G., & Lima, M. D. F. W. D. (2012). Evaluating automatic group formation mechanisms to promote collaborative learning-a case study. *International Journal of Learning Technology*, 7(3), 261-276.
- Wenger, E. (2005). *La théorie des communautés de pratique*, Presses Université Laval.
- Wenger, E. (2011). *Communities of practice : Learning, meaning , and identity*. Cambridge , UK : Cambridge University Press .
- Xing, W.; Guo, R.; Petakovic, E. & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory.
- Yao, J., & Wang, Y. (2019). Practice of New Employee Training Mode in School-enterprise Cooperative Enterprises based on" SPOC Method". In *3rd International Conference on Economics and Management, Education, Humanities and Social Sciences (EMEHSS 2019)*. Atlantis Press.
- Yahia, N. B., Rebhi, W., Saoud, N. B. B., & Ghézela, H. B. (2015). Approche hybride de recommandation des nouvelles collaborations. In *INFORSID* (pp. 285-299).
- Yanhui, D., Dequan, W., Yongxin, Z., & Lin, L. (2015). Un système de recommandation de groupe pour l'étude de cours en ligne. En 2015, *7e Conférence internationale sur les technologies de l'information en médecine et en éducation (ITME)* (pp.318-320). IEEE.
- Yang, X. S. (2020). *Nature-inspired optimization algorithms*. Academic Press.
- Yannibelli, V. D., Armentano, M. G., Berdun, F. D., & Amandi, A. A. (2016). A steady-state evolutionary algorithm for building collaborative learning teams in educational environments considering the understanding levels and interest levels of the students.
- Yemi, A. & Oludele O., (2019) A. Performance Evaluation of Genetic Algorithm Tuned With Particle Swarm Optimization for Solving University Examination Timetabling Problem
- Yukselturk, E.; Ozekes, S. & Türell, Y.K. (2014). Predicting Dropout Student: An Application of Data Mining Methods in an Online Education Program. *Computers in Human Behavior*
- Zahedifard, M.; Attarzadeh, I.; Pazhokhzadeh H. & Malekzadeh, J. (2015). Prediction of students' performance in high school by data mining classification techniques, *International Academic Journal of Science and Engineering*
- Zaraté, P. (2005). *Des Systèmes Interactifs d'Aide à la Décision aux Systèmes Coopératifs d'Aide à la Décision: Contributions conceptuelles et fonctionnelles* (Thèse de doctorat).
- Zheng, Y., Li, C., Liu, S., & Lu, W. (2018). An improved genetic approach for composing optimal collaborative learning groups. *Knowledge-Based Systems*, 139, 214-225.
- Zhou, A., Qu, B. Y., Li, H., Zhao, S. Z., Suganthan, P. N., & Zhang, Q. (2011). Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm and evolutionary computation*, 1(1), 32-4

Annexe 1. Type de variables selon la catégorisation des données (Personne physique/morale-Activité de formation)

Personne physique et morale					
Thématique	Type de variable		Thématique	Type de variable	
Participant	Id	Numérique	Organisation	Nom	Catégoriel
	Type	Catégoriel		Secteur d'activité	Catégoriel
	Titre	Catégoriel		Type d'établissement	Catégoriel
	Genre	Catégoriel		Département	Numérique
	Nom de famille	Catégoriel		Région	Numérique
	Prénom	Catégoriel	Thématique	Type de variable	
	Fonction	Catégoriel	Formateur	Id	Numérique
	Statut	Catégoriel		Nom de famille	Catégoriel
		Catégoriel		Prénom	Catégoriel
	Type d'établissement	Catégoriel		Expertise	Catégoriel
	Direction	Catégoriel		Fonction	Catégoriel
	Département	Numérique		Type	Catégoriel
	Région	Numérique			

Activité de formation					
Thématique	Type de variable		Type de variable		
Session de formation	Date de début	Numérique	Mode	Catégoriel	
	Date de fin	Numérique	Formateur	Catégoriel	
	N° session	Numérique	Organisateur	Catégoriel	
	Thématique de formation	Catégoriel	Taille	Numérique	
	Intitulé de formation	Catégoriel	Durée (jours)	Numérique	
Evaluation	Score de satisfaction global	Numérique	Score de satisfaction (organisation, contenu pédagogique, animation)	Numérique	

Annexe 2. Représentation graphique de la modélisation conceptuelle



Annexe 3. Détails des modules techniques de développement

Modules	Descriptif
Framework web (Django ¹⁴⁸)	<p>Django est un cadre de développement web open-source basé sur le langage Python qui utilise une architecture logicielle MVC (Modèle/Vue/Contrôleur). La structure de l'application est facilitée compte tenu que la base de données peut être gérée via des simples classes python appelées modèles.</p> <p>Les vues, fondées sur les classes, sont des exécutables qui acceptent et renvoient une requête HTTP entrantes, le contrôleur choisit la vue à exécuter en fonction de l'URL associée. En fin d'exécution, un fichier HTML est renvoyé via un fichier de template qui affiche le résultat de la requête.</p> <p>Liens : https://www.djangoproject.com/</p>
Base de données orienté graphe (Grapheek)	<p>Le modèle de données orienté graphe repose sur une solution open-source¹⁴⁹, codé en Python, qui permet de stocker et rechercher des nœuds et des relations au sein de graphe. Parmi les différents modèles NoSql, les systèmes orientés graphes (appelé <i>Graph data model system</i>) utilisent une technique nommé « index free-adjancy » qui consiste que chaque nœud du graphe est en relation directe avec les nœuds adjacents. Ainsi, la liaison entre les nœuds par le biais de la relation permet d'éviter de consulter l'index et évite l'utilisation de jointures complexes des modèles relationnels.</p> <p>Liens : https://pypi.python.org/pypi/grapheekdb</p>
Librairie de graphe (Networkx)	<p>NetworkX est une bibliothèque Python pour la création, la manipulation et l'étude des réseaux complexes. La librairie contient un grand nombre de générateurs de graphes (orienté ou non orienté, graphe simple ou multi-graphe), une base d'algorithmes (plus courts chemins, connectivité, arbres de recouvrement, isomorphismes, flots, etc.) et gère également les graphes labellisés.</p> <p>Liens : https://networkx.github.io/</p>
Librairies de visualisation (Sigma)	<p>Sigma est une librairie open-source JavaScript dédiée à la visualisation et l'exploration de graphes de type statique ou générés dynamiquement.</p> <p>Liens : http://sigmajs.org/</p>
Framework CSS (Bootscrap)	<p>Bootstrap est un framework CSS qui permet de créer des interfaces standardisées (graphisme, animation et interactions avec le navigateur) et de simplifier l'implémentation des sites et d'applications web et la compatibilité selon les différents navigateurs.</p> <p>Liens : https://getbootstrap.com/</p>
Conteneur (Docker)	<p>Docker est une plateforme open-source qui intègre des logiciels dans des unités normalisées appelées conteneurs pour garantir le fonctionnement et la facilité de déploiement de site d'application.</p> <p>Liens : https://www.docker.com/</p>

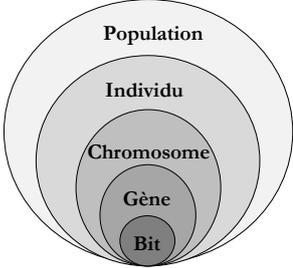
148 Version 1.7

149 GrapheekDB

Annexe 4. Mode d'interactivité et détails des valeurs par critères de recherche

Critères	Mode d'interactivité	Valeurs	
Mode organisationnel	Liste de champs	Inter-organisationnel Intra-organisationnel	
Organisation	Liste de champs	Liste nominative	
Type de regroupement	Liste de champs	Homogène Hétérogène	
Secteur d'activité	Liste de champs/ Pondération	CH/CHU/CHS Collectivité territoriale Culture Etat Habitat social	Médico-social Privé Région Syndicat Université/COMUE/GE
Type d'établissement	Liste de champs/ Pondération	Association Autre Centre Hospitalier Centre Hospitalier Uni- versitaire Centre hospitalier Spécia- lisé COMUE Département EHPAD EPCI	Etablissement culturel Grande Ecole Ministère OPH Organisme de recherche Région Syndicat Unicancer Université Ville
Statut	Liste de champs/ Pondération	Non-cadre Cadre Cadre+	
Typologie de personne	Liste de champs/ Pondération	Participant Colloque Elu	
Taille du groupe	Liste de champs	2 à 4 5 à 6 7 à 9	10 à 14 15+
Mixité	Liste de champs	H>F F>H F=H	
Thématique de formation	Liste de champs	Marché/Commande public Fiscalité/TVA Social/Paie	
Historique de la formation	Liste de champs/ Pondération	Marché/Commande public Fiscalité/TVA Social/Paie	
Formateur	Liste de champs	Liste nominative	
Durée	Liste de champs	1 à 10 jours	
Mois	Liste de champs	Janvier à Décembre	
Région	Liste de champs	Liste nominative	
Département	Liste de champs	Liste nominative	

Annexe 5. Lexique comparée des AG

	<p>Population. La population est l'ensemble des individus et des solutions considérées à un moment donné. La population initiale est le processus de génération définissant les premiers individus qui serviront à alimenter la première génération de l'algorithme.</p>
	<p>Individu. Un individu représente l'unité élémentaire et une solution possible au sein de la population.</p>
	<p>Chromosome (ou génotype). Un chromosome (ou individu) incluant un ensemble de gènes est une représentation d'une solution possible parmi les solutions potentielles. En informatique, un chromosome est décrit comme une chaîne de bits.</p>
	<p>Gène. Un gène est un élément atomique permettant le codage de la représentation du chromosome (e.g. un bit dans une chaîne de bits).</p>
	<p>Allèle. Valeur d'un gène. Equivalent à un bit (<i>binary digit</i>)</p>
<p>Niveaux d'organisation</p>	

Codage. Représentation des variables et des valeurs des individus. Trois types de codages existants :

Codage binaire : Chaque séquence de bits (ou chromosomes) est représentée par un entier long (32 bits) et dispose du même alphabet binaire $\{0, 1\}$

Codage réel : Chaque séquence de bits (ou chromosomes) est représenté par des valeurs réelles

Croisement. Opération qui consiste à obtenir de nouveaux chromosomes (ou individus) composés des gènes des chromosomes d'entrée. Les trois opérateurs les plus utilisées sont : le croisement à un point, le croisement multipoints et le croisement uniforme.

Fonction d'adaptation (ou fonction objective, ou *fitness* en terminologie anglo-saxonne). Fonction qui détermine la capacité de l'individu à survivre et à se reproduire

Génération. Une génération N est l'ensemble des chromosomes obtenus lors de la N ème itération.

Mutation. Opération qui consiste à appliquer une variation aléatoire aux gènes. Ainsi, un nouveau chromosome est obtenu en modifiant un ou plusieurs gènes (ou caractéristiques) d'un chromosome (ou d'un individu) déjà existant.

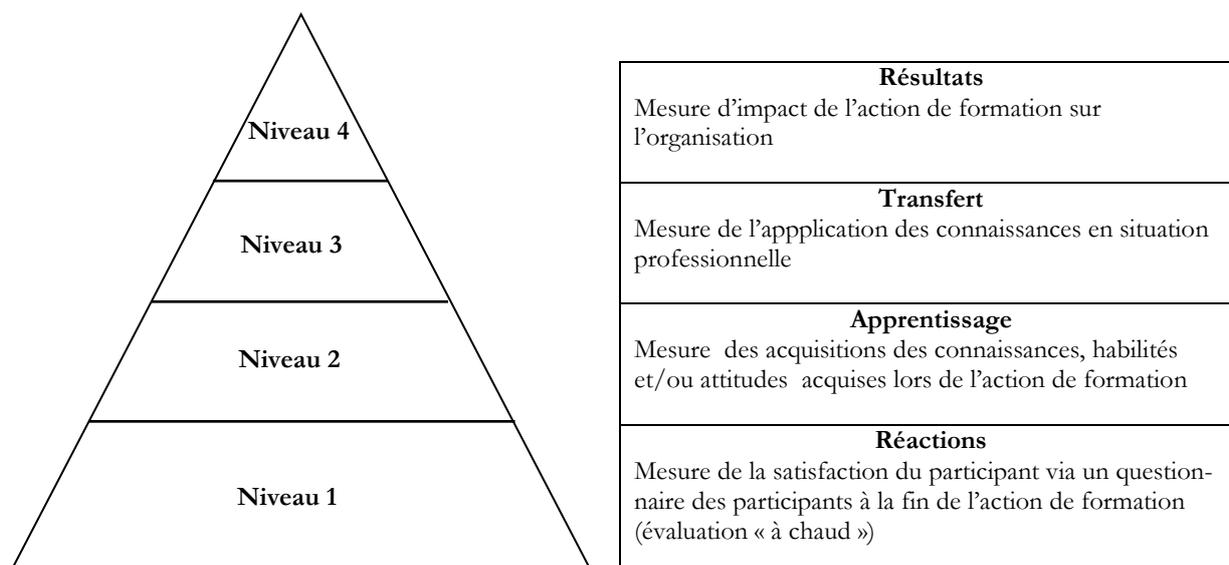
Sélection. Sélection des individus qui serviront de parents aux opérateurs génétiques (recombinaison et mutation). Parmi les méthodes existantes :

Sélection élitiste. Méthode consistant à conserver les meilleurs individus de la génération précédente.

Sélection par tournoi. Méthode consistant à comparer une paire d'individus choisie au hasard où le meilleur des deux est conservé.

Sélection par la roulette (RWS : *Roulette Wheel Selection*). Méthode consistant à affecter une probabilité de sélection à chaque individu.

Annexe 6. Modèle d'évaluation de Kirpatrick



Source : Kirkpatrick, D. L. (1994). Evaluating training programs: the four levels. San Francisco: Berrett-Koehler.

Annexe 7. Modèle du questionnaire de satisfaction

Catégorisation : Identité professionnel/Satisfaction globale/Organisation de la formation/Contenu de la formation/Animation de la formation/

Nombre total de questions : 20

Question(s) fermée(s) : Case à cocher

Echelle de mesure : Type Likert sur une échelle de 5 items (0 : Pas du tout, 1 : Peu satisfait, 2 : Satisfait, 3 : Assez satisfait, 4 : Très satisfait)

Questions(s) ouverte(s) :

Nom/Prénom/Fonction/ Age/ Ancienneté/Statut/Organisation/Point positif-point négatif

Taux de remplissage : environ 93 % (93,8%)

Identité professionnelle			
Questions	Variables	Format de réponse	
Q1	Nom/Prénom	Ouverte	Texte
Q2	Genre	Fermée	Unique
Q3	Fonction	Ouverte	Texte
Q4	Age	Ouverte	Numérique
Q5	Ancienneté	Ouverte	Numérique
Q6	Statut	Fermée	Unique
Q6.1	Statut/Direction d'entité	Fermée	Unique
Q7	Organisation	Ouverte	Texte
Q8	Financement	Fermée	Unique
Q9	Objectif(s) globale(s) de la formation	Fermée	Multiple
Q10	Initiative	Fermée	Unique
Satisfaction globale			
Q11	Satisfaction globale	Fermée	Echelle de Likert
Q13	Point positif/point négatif	Ouverte	Texte
Organisation de la formation			
Q14	Conditions matérielles	Fermée	Echelle de Likert
Q15	Composition du groupe (homogène/hétérogène)	Fermée	Echelle de Likert
Q16	Durée	Fermée	Echelle de Likert
Contenu pédagogique de la formation			
Q17	Contenu et support pédagogique	Fermée	Echelle de Likert
Q18	Transfert	Fermée	Echelle de Likert
Animation de la formation			
Q19	Qualité du formateur	Fermée	Echelle de Likert
Q20	Collaboration (Echanges/Interaction avec le groupe)	Fermée	Echelle de Likert

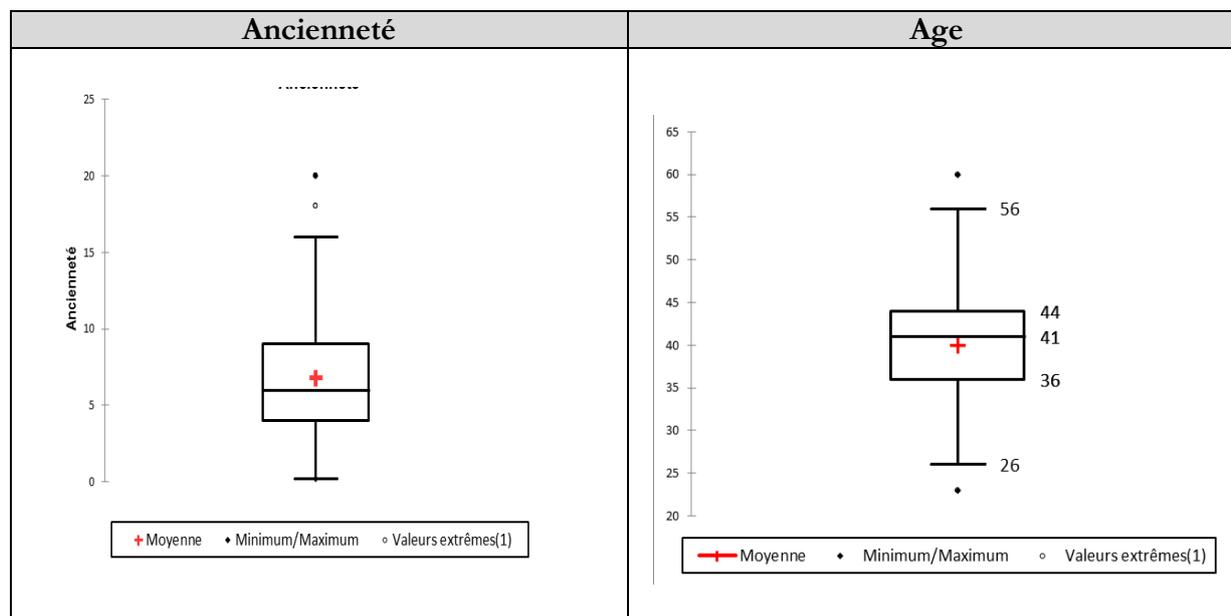
Annexe 8. Statistique descriptive des groupes de formation

Variables	Modalité	Nombre de sessions de formation	Part %
Mode organisationnel	Inter	34	63,0%
	Intra	20	37,0%
Composition	Hétérogène	14	24,6%
	Homogène	40	74,1%
Durée	1	27	50,0%
	2	18	33,3%
	3	9	16,7%
Mixité	F	9	16,7%
	F=H	8	15,1%
	F>H	25	46,3%
	H	2	3,7%
	H>F	10	18,5%
Taille	2 à 4	16	29,6%
	5 à 6	17	31,5%
	7 à 9	8	14,8%
	10 à 14	9	16,7%
	15 et plus	4	7,4%
Genre du formateur	F	34	63,0%
	H	20	37,0%
Thématique	Fiscalité locale	3	5,6%
	Marché Public	9	16,7%
	Social	7	13,0%
	TVA/Fiscal	35	64,8%

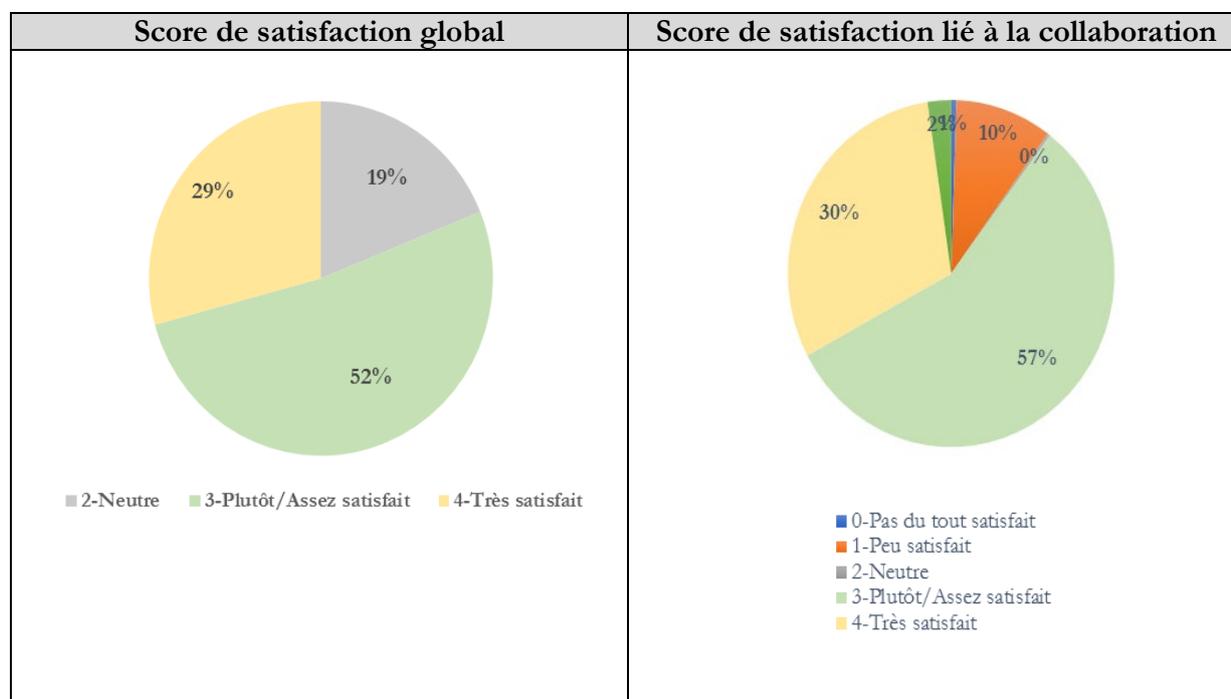
Annexe 9. Statistique descriptive des participants (Genre, Statut professionnel, Secteur d'activités)

Variables	Modalités	Effectifs	Pourcentages
Genre	Femme	239	61,8%
	Homme	143	37%
	Valeurs manquantes	5	1,3%
	Ensemble	387	100%
Cadre	Oui	316	81,7%
	Non	32	8,3%
	Indéterminé	39	10,1%
	Ensemble	387	100%
Cadre + (ou Directeur d'entité)	Non	183	47,3%
	Oui	115	29,7%
	Valeurs manquantes	50	12,9%
	Indéterminé	39	10,1%
	Ensemble	387	100%
Equipe	Non	32	8,3%
	+de 20 personnes	1	0,3%
	5 à 20 personnes	40	10,5%
	Moins de 5 personnes	61	15,8%
	Valeurs manquantes	214	55,3%
	Indéterminé	39	10,1%
	Ensemble	387	100%
	Secteurs d'activités	CH/CHU/CHS	137
Collectivités territoriales		104	27%
Etat		2	1%
Médico-social		17	4%
Privé		14	4%
Université/COMUE/GE		57	15%
Valeurs manquantes		32	8%
Ensemble		387	100%

Annexe 10. Représentation graphique en boîte à moustache des profils des participants (Ancienneté et Age)



Annexe 11. Représentation graphique en diagramme circulaire du score de satisfaction global et score de satisfaction lié à la collaboration



Annexe 12. Fréquence des formes actives par catégories grammaticales

Nom							
Formes actives	Fréq. >2						
cas	60	taille	5	introduction	3	novice	2
formation	55	présentation	5	intervenant	3	minimum	2
formateur	42	pédagogie	5	interaction	3	métropole	2
durée	33	niveau	5	généraliste	3	lexique	2
temps	22	fonction	5	expérience	3	finance	2
exemple	19	établissement	5	dossier	3	fin	2
groupe	15	complexité	5	domaine	3	documentation	2
support	14	attente	5	disponibilité	3	document	2
journée	13	situation	4	activité	3	dépense	2
sujet	11	service	4	vocabulaire	2	département	2
question	11	réponse	4	travail	2	compte	2
partie	11	recette	4	terrain	2	chapitre	2
exercice	10	rappel	4	terme	2	cadre	2
contenu	10	procédure	4	technicité	2	besoin	2
connaissance	10	précision	4	taux	2	articulation	2
jour	9	information	4	spécificité	2	appréhension	2
notion	8	gestion	4	retard	2	apport	2
échange	8	général	4	rencontre	2	amont	2
base	8	fiscalité	4	régularisation	2	aménagement	2
application	8	explication	4	règle	2	agent	2
expertise	7	collectivité	4	recherche	2	acquisition	2
théorie	6	clarté	4	récapitulatif	2		
secteur	6	calcul	4	professionnel	2		
remise	6	aspect	4	principe	2		
participant	6	stagiaire	3	place	2		
papier	6	rapport	3	parole	2		
mécanisme	6	qualité	3	opération	2		
déduction	6	programme	3	opérateur	2		
Adjectif							
Formes actives	Fréq. >2						
pratique	52	meilleur	5	dynamique	3	foncier	2
concret	31	petit	4	compétent	3	étranger	2
technique	17	pédagogique	4	clair	3	difficile	2
problématique	16	intéressant	4	applicable	3	différent	2
rapide	12	dense	4	supplémentaire	2	appliqué	2
court	10	spécifique	3	seul	2	adéquat	2
théorique	8	réel	3	précis	2	accessible	2
complexe	8	quotidien	3	pragmatique	2	abordable	2
complet	8	nombreux	3	personnel	2	foncier	2
insuffisant	6	important	3	personnalisé	2	étranger	2
faible	6	immobilier	3	long	2	difficile	2
particulier	5	homogène	3	imparti	2	différent	2
nécessaire	5	grand	3	hospitalier	2	appliqué	2

Annexe 13. Lexique des mots-clés pivot pour les requêtes de cooccurrences

Catégorie grammaticale	Mots-clés
Nom	Connaissance, cours, contenu, document, documentation, exemple, exercices, exercices, Formation, formation, niveau, programme, sujet, support, présentation.
	animatrice, formateur, formatrice, intervenant, pédagogue, pédagogie, groupe, groupe, dialogue discussion, échange, interaction, question, participant
Adjectif	durée, temps, journée, taille, lieu
	Concret, technique, rapide, court, théorique, complexe, complet, insuffisant, faible, meilleur, petit, pédagogique, intéressant, nécessaire

Annexe 14. Distribution du jeu de données mixte (Réel+Fictif)

Genre	Nombre id	%	Secteur d'activité	Nombre id	%
F	459	47,81	Collectivité territoriale	370	38,54
H	464	48,33	CH/CHU/CHS	124	12,92
nc	31	3,23	Université/COMUE/GE	119	12,40
vm	6	0,63	Habitat Social	88	9,17
Total général	960	100,00	Culture	66	6,88
Statut	Nombre id	%	Syndicat	59	6,15
Cadre	448	46,67	Privé	53	5,52
Cadre+	169	17,60	Autre	29	3,02
Non-Cadre	269	28,02	État	28	2,92
vm	42	4,38	Médico-Social	23	2,40
nc	31	3,23	vm	1	0,10
Stagiaire	1	0,10	Total général	960	100,00
Total général	960	100,00			
Type	Nombre id	%			
Participant	685	71,35			
Colloque	275	28,65			
Total général	960	100			

nc : non communiqué

vm : valeur manquante

Annexe 15. Collection de requêtes et distribution des critères et des valeurs associées

Jeu de données	N° Req	Nombre de critères	Nombre total de personnes	Mode organisationnel	Type de regroupement	Secteur	Type d'établissement	Statut	Type de personne	Taille	Mixité	Historique	Département	Région
R/F	1	5	53	Inter	Homogène	Culture				10-14				Ile-de France
R/F	2	5	57	Inter	Homogène	Culture		Pond 1		7-9				
R/F	3	5	19	Inter	Homogène	Université/COMUE/GE	Grande Ecole			5-6				
R/F	4	6	35	Inter	Homogène	CH/CHU/CHS	CHU					Fiscal/TVA		
R/F	5	6	56	Inter	Homogène	Collectivité territoriale	Ville	Cadre +		7-9				
R/F	6	3	11	Intra ¹⁵⁰	Hétérogène					5-6				
R/F	7	4	11	Intra	Homogène			Pond1		5-6				
R/F	8	5	46	Inter	Homogène	Collectivité territoriale				7-9				Grand-Est
R/F	9	6	23	Inter	Homogène	Collectivité territoriale		Cadre		7-9				Grand-Est
R/F	10	4	19	Intra ¹⁵¹	Hétérogène					5-6		Social		
R/F	11	5	43	Inter	Homogène	CH/CHU/CHS	CHU			2-4				
R/F	12	6	63	Inter	Homogène	CH/CHU/CHS				10-14	H=F	Fiscal/TVA		
R/F	13	5	136	Inter	Homogène	CH/CHU/CHS	Pondération 1	Pond 2						
R/F	14	6	27	Inter	Homogène	CH/CHU/CHS			Colloque	2-4	H=F			
R/F	15	6	136	Inter	Homogène	CH/CHU/CHS		Pond 2	Pond 1	5-6				
R/F	16	6	520	Inter	Homogène	Collectivité territoriale	Pond 1	Pond2		5-6				
R/F	17	6	195	Inter	Homogène	Collectivité territoriale	Ville	Pond1		7-9				
R/F	18	5	50	Inter	Homogène	Collectivité territoriale	Région							
R/F	19	5	166	Inter	Homogène					10-14	H=F	Fiscal/TVA		
R/F	20	4	152	Inter	Homogène			Cadre +				Fiscal/TVA		
R/F	21	4	77	Inter	Homogène	Collectivité territoriale		Cadre +						
R/F	22	6	515	Inter	Homogène	Collectivité territoriale		Pond1	Pond 2	5-6				
R/F	23	5	23	Inter	Homogène	Collectivité territoriale				5-6			59-Nord	
R/F	24	6	29	Inter	Homogène	Collectivité territoriale		Pond 1		5-6			59-Nord	
R/F	25	5	31	Inter	Homogène	Collectivité territoriale				2-4			29-Finistère	
R/F	26	5	15	Inter	Homogène	Culture			Colloque	5-6				
R/F	27	6	38	Inter	Homogène	Culture		Non-cadre		5-6		Marché Public		
R/F	28	6	35	Inter	Homogène	Université/COMUE/GE		Non-cadre		5-6		Social/paie		
R/F	29	6	58	Inter	Homogène	Université/COMUE/GE			Colloque	5-6	H>F			

¹⁵⁰ Conseil Départemental de la Somme¹⁵¹ Port de Marseille

R/F	30	3	25	Intra	Hétérogène					5-6				
R/F	31	5	29	Inter	Homogène	CH/CHU/CHS		Cadre+		5-6				
R/F	32	6	17	Inter	Homogène	CH/CHU/CHS		Non-cadre		7-9	F>H			
R	33	4	63	Inter	Homogène	CH/CHU/CHS						Fiscal/TVA		
R	34	4	23	Inter	Homogène	CH/CHU/CHS								Haut-de-France
R	35	4	43	Inter	Homogène	CH/CHU/CHS	CHU							
R	36	4	99	Inter	Homogène	CH/CHU/CHS				5-6				
R	37	5	99	Inter	Homogène	CH/CHU/CHS				5-6	H=F			
R	38	5	11	Inter	Homogène	CH/CHU/CHS				5-6				Grand-Est
R	39	6	43	Inter	Homogène	CH/CHU/CHS	CHU			5-6	H=F			
R	40	5	67	Inter	Homogène	Université/COMUE/GE				5-6	H=F			
R	41	6	170	Inter	Homogène			Cadre		5-6	H=F	Social		
R	42	7	21	Inter	Homogène			Cadre		5-6	H=F	Commande publique		Ile de France
R	43	5	36	Inter	Homogène	Collectivité territoriale		Pond1		7-9				
R	44	5	41	Inter	Homogène			Non-cadre		7-9		Fiscal/TVA		
R	45	5	29	Inter	Homogène			Cadre+		7-9		Fiscal/TVA		

Annexe 16. Identifiant unique, précision à dix et score d'évaluation par requête

N°/Requête	Id unique	P@10	Score	N°/Requête	Id unique	P@10	Score
1	100%	100%	5	24	100%	100%	2,19
2	100%	100%	5	25	100%	100%	2,29
3	100%	100%	3,09	26	100%	100%	2,15
4	100%	100%	5	27	100%	100%	2,28
5	100%	100%	5	28	100%	100%	2,24
6	100%	100%	1,96	29	100%	100%	2,28
7	100%	100%	1,96	30	100%	100%	1,78
8	100%	100%	1,96	31	100%	100%	2,7
9	100%	100%	1,96	32	100%	100%	4,46
10	100%	100%	2,14	33	100%	100%	3,05
11	100%	100%	3,87	34	100%	100%	1,77
12	100%	100%	5	35	100%	100%	5
13	100%	100%	3,74	36	100%	100%	2,34
14	100%	100%	2,37	37	100%	100%	2,34
15	100%	100%	2,09	38	100%	100%	1,78
16	100%	100%	1,34	39	100%	100%	2,53
17	100%	100%	5	40	100%	100%	0,73
18	100%	100%		41	100%	100%	2,22
19	100%	100%	0,43	42	100%	100%	2,36
20	100%	100%	1,4	43	100%	100%	0,86
21	100%	100%	5	44	100%	100%	1,98
22	100%	100%	1,39	45	100%	100%	1,62
23	100%	100%	2,08				

Annexe 17. Taux de précision par critères et par requête

N° Requête	Mode Organisationnel	Type de regroupement	Type de regroupement/Sous-critères				Historique	Taille	Mixité	Région	Département	Précision moyenne par requête
			C3	C4	C5	C6						
	C1	C2	Statut	Secteur	Etablissement	Personne	C7	C8	C9	C10	C11	
1	100%	100%		100%				100%		100%		100%
2	100%	100%	100%	100%				100%				100%
3	100%	100%		100%	100%			100%				100%
4	100%	100%	100%	100%	100%		100%					100%
5	100%	100%	100%	100%	100%			100%				100%
6	100%	100%						100%				100%
7	100%	0%	0%					100%				67%
8	100%	100%		100%				100%		100%		100%
9	100%	100%	100%	100%				100%		100%		100%
10	100%	60%	60%				100%	100%				90%
11	100%	100%		100%	100%			100%				100%
12	100%	100%		100%			100%	100%	100%			100%
13	100%	0%	0%	97.77%	0%							50%
14	100%	100%		100%		100%		100%	40%			85%
15	100%	73%	100%	92%	100%	0%		90%				88%
16	100%	67%	100%	100%	0%			100%				89%
17	100%	83%	50%	100%	100%			100%				94%
18	100%	100%		100%						100%		100%
20	100%	100%	100%				100%					100%
21	100%	100%	100%	100%								100%
22	100%	57%	60%	100%		10%		100%				86%

23	100%	100%		100%				100%			100%	100%
24	100%	50%	0%	100%				100%			100%	88%
25	50%	100%		100%				100%			100%	88%
26	100%	100%		100%		100%		100%				100%
27	100%	100%	100%	100%			100%	100%				100%
28	100%	100%	100%	100%			100%	100%				100%
29	100%	100%		100%		100%		100%	100%			100%
30	100%	100%						100%				100%
31	100%	98%	100%	96%				100%				99%
32	100%	100%	100%	100%				100%	100%			100%
33	100%	100%		100%			100%					100%
34	100%	100%		100%						100%		100%
35	100%	100%		100%	100%							100%
36	100%	100%		100%				100%				100%
37	100%	100%		100%				100%	0%			75%
38	100%	100%		100%				100%		100%		100%
39	100%	100%		100%	100%			100%	0%			75%
40	100%	100%		100%				100%				100%
41	100%	100%	100%				100%	100%	0%			80%
42	100%	100%	100%				100%	100%	0%	100%		83%
43	100%	100%	100%	100%				100%				100%
44	100%	58%	100%	15%			0%					53%
45	100%	100%	100%				100%	100%				100%
Précision moyenne par critère	98,89%	89,89%	81,30%	97,23%	80,00%	62,00%	91,67%	100%	42,50%	100%	100,00%	

Annexe 18. Résultats d'évaluation des mesures d'efficacité et d'efficience des tâches de simulation « Recherche » et « Planification »

N°	Interface utilisateur		Filtrage	Résultats	Mesures	
	Niveau 1	Niveau 2			Nombres de critères	Nombres de valeurs trouvés
					Taux d'achèvement	Temps d'exécution (s)
1	Recherche	Personne	2	329	100%	0-38,82s
2	Recherche	Personne	2	4	100%	0-36,72s
3	Recherche	Personne	3	0	100%	0-33,57s
4	Recherche	Personne	3	8	100%	0-57,04s
5	Recherche	Personne	2	77	100%	0-28,95s
6	Recherche	Personne	2	77	100%	0-30,56s
7	Recherche	Personne	2	44	100%	0-30,30s
8	Recherche	Personne	2	44	100%	0-33,59
9	Recherche	Personne	2	1	100%	0-16,39s
10	Recherche	Personne	2	4	100%	0-25,62s
11	Recherche	Organisation	2	1	100%	0-22,68s
12	Recherche	Organisation	1	Echec	0%	0
13	Recherche	Organisation	1	Echec	0%	0
14	Recherche	Organisation	2	97	100%	0-36,88s
15	Recherche	Organisation	3	5	100%	0 – 28.87 s
16	Recherche	Organisation	1	0	100%	0-18,96s
17	Recherche	Organisation	2	68	100%	0-24,67s
18	Recherche	Organisation	2	14	100%	0-28,52s
19	Recherche	Organisation	2	1	100%	0-23,89
20	Recherche	Organisation	2	10	100%	0-23,66
21	Recherche	Formateur	1	4	100%	0-22,86

22	Recherche	Formateur	1	3	100%	0-39,91
23	Recherche	Formateur	2	1	100%	0-25,01
24	Recherche	Formateur	1	1	100%	0 – 15.91 s
25	Recherche	Evènement	2	1	100%	0-24,08
26	Recherche	Evènement	2	6	100%	0-53,71
27	Recherche	Evènement	3	6	100%	0-28,66
28	Recherche	Evènement	2	13	100%	0-0,14
29	Recherche	Evènement	1	1	100%	0-16,87
30	Recherche	Evènement	2	2	100%	0-29,23
31	Planification		5	53	100%	0-43,54
32	Planification		4	57	100%	0-34,78
33	Planification		5	19	100%	0-27,19
34	Planification		6	35	100%	0-41,75
35	Planification		6	56	100%	0-42,03
36	Planification		3	11	100%	0-44,22
37	Planification		4	11	100%	0-34,78
38	Planification		5	46	100%	0-39,42
39	Planification		6	23	100%	0-36,83
40	Planification		5	19	100%	
41	Planification		6	43	100%	0-44,79
42	Planification		6	63	100%	0-46,25
43	Planification		5	136	100%	0-57,85
44	Planification		6	27	100%	0-39,06
45	Planification		6	136	100%	0-53,59
46	Planification		6	520	100%	0-58,41
47	Planification		6	195	100%	0-47,59
48	Planification		6	50	100%	0-37,61
49	Planification		5	166	100%	0-43,32

50	Planification		4	152	100%	0-52,99
51	Planification		4	77	100%	0-29,92
52	Planification		6	515	100%	0-54,20
53	Planification		5	23	100%	0-45,49
54	Planification		6	29	100%	0-30,10
55	Planification		5	31	100%	
56	Planification		5	15	100%	0-36,70
57	Planification		6	38	100%	0-38,34
58	Planification		6	34	100%	0-36,99
59	Planification		6	58	100%	0-45,05
60	Planification		3	25	100%	
61	Planification		5	29	100%	0-27,27
62	Planification		6	17	100%	0-30,24
63	Planification		4	170	100%	0-43,55
64	Planification		3	21	100%	0-29,26
65	Planification		4	36	100%	0-31,22
66	Planification		4	44	100%	0-34,42
67	Planification		5	29	100%	0-37,34
68	Planification		4	11	100%	0-30,42
69	Planification		6	43	100%	0-40,79
70	Planification		6	67	100%	0-38,88
71	Planification		6	170	100%	0-46,49
72	Planification		7	170	100%	0-48,20
73	Planification		5	36	100%	0-44,37
74	Planification		5	40	100%	0-22,44
75	Planification		5	29	100%	0-37,65

Annexe 19. Liste des requêtes par nombre d'itération et temps d'exécution associées aux scores d'évaluations

Req	Nomb total de per- sonnes	It10		It20		It30		It40		It50	
		Scor e	Tim e	Score	Time	Score	Time	Score	Time	Score	Time
1	55	2,02	4,52	3,03	7,60	3,11	17,38	3,23	17,38	2,24	17,38
2	57	5	4,95	2,31	8,00	2,09	16,72	2,32	16,72	0,16	16,72
3	19	3,09	2,50	1,52	4,18	1,49	9,73	1,54	9,73	1,48	9,73
4	35	5	5,72	1,20	10,25	1,00	22,54	1,19	22,54	2,51	22,54
5	56	5	4,38	1,94	7,29	1,86	16,31	4,12	16,31	4,14	16,31
6	11	1,96	1,25	4,63	2,14	2,54	4,93	2,69	4,93	2,64	4,93
7	11	1,96	1,25	4,63	2,14	2,54	4,93	2,69	4,93	2,64	4,93
8	46	1,96	4,55	1,54	7,03	2,24	16,02	1,34	16,02	1,27	16,02
9	23	1,96	4,55	1,54	7,03	2,24	16,02	1,34	16,02	1,27	16,02
10	19	2,14	1,31	3,91	2,20	3,40	4,97	2,37	4,97	2,37	4,97
11	43	3,87	4,67	2,99	7,72	2,90	17,14	1,60	17,14	1,6	17,14
12	63	5	8,93	1,59	13,60	0,35	29,16	0,92	29,16	0,29	29,16
13	136	3,74	19,00	2,39	32,08	2,35	63,46	2,29	63,46	3,08	63,46
14	27	2,37	3,39	3,48	5,63	3,16	12,15	3,05	12,15	2,08	12,15
15	136	2,09	9,97	1,25	16,20	1,10	35,45	1,10	35,45	1,12	35,45
16	520	1,34	12,88	1,68	21,49	1,52	46,91	1,52	46,91	1,52	46,91
17	195	5	8,53	0,07	14,20	0,07	31,47	0,07	31,47	0,07	31,47
18	50	5	6,20	2,96	10,36	3,91	23,05	2,67	23,05	2,64	23,05
19	166	0,43	12,14	3,80	19,38	3,82	42,06	3,79	42,06	3,78	42,06
20	152	1,4	20,31	3,60	34,11	4,60	76,21	4,57	76,21	4,59	76,21
21	77	5	10,48	1,96	18,07	1,39	38,56	0,56	38,56	0,01	38,56
22	515	1,39	12,84	1,68	21,77	1,52	47,65	1,52	47,65	1,52	47,65
23	23	2,08	2,06	3,16	3,44	2,20	7,65	2,29	7,65	1,85	7,65
24	29	2,19	2,09	3,16	3,48	2,20	7,66	2,29	7,66	1,85	7,66
25	31	2,29	1,80	2,01	2,96	2,08	6,46	1,97	6,46	1,89	6,46
26	15	2,15	1,72	3,88	2,98	3,98	6,77	2,72	6,77	3,03	6,77
27	38	2,28	1,45	1,84	2,47	3,33	5,38	0,85	5,38	0,68	5,38
28	35	2,24	2,49	1,61	4,31	1,70	9,58	1,58	9,58	1,91	9,58
29	58	2,28	5,71	2,13	9,53	3,28	20,48	1,19	20,48	1,03	20,48
30	25	1,78	33,27	3,82	56,16	3,89	123,00	3,40	123,00	2,68	123,00
31	29	2,7	3,65	2,75	6,19	2,75	13,58	2,75	13,58	4,01	13,58
32	17	4,46	2,06	2,01	3,44	0,74	7,39	0,26	7,39	0,05	7,39
33	63		10,86	0,33	17,22	0,17	38,85	0,63	38,85	0,45	38,85
34	23	1,77	3,20	4,63	5,31	3,20	11,57	3,80	11,57	3,52	11,57
35	43	5	8,31	0,89	13,81	3,33	30,74	1,20	30,74	2,09	30,74
36	99	2,34	9,97	1,25	17,10	1,10	37,10	1,10	37,10	1,1	37,10
37	99	2,34	9,91	2,59	16,36	2,83	37,02	2,64	37,02	2,92	37,02

38	11	1,78	1,54	3,92	2,56	3,33	5,61	5,03	5,61	4,83	5,61
39	43	2,53	5,16	2,46	8,69	2,46	19,15	2,46	19,15	2,46	19,15
40	67	0,73	7,63	3,62	12,70	3,62	27,12	3,62	27,12	3,61	27,12
41	170	2,22	13,98	2,65	22,85	3,11	51,16	3,09	51,16	3,16	51,16
42	21	2,36	4,12	2,50	6,85	2,50	15,08	2,52	15,08	2,5	15,08
43	36	0,86	14,58	3,69	25,07	4,18	55,34	3,05	55,34	3,77	55,34
44	41	1,98	13,93	7,47	23,34	7,71	51,37	7,79	51,37	8,66	51,37
45	29	1,62	14,92	3,84	25,03	2,67	56,19	0,99	56,19	0,99	56,19

Annexe 20. Questionnaire des tests-utilisateurs

Ordre	Questions	Mesures d'évaluation
1	Tableau de bord	
1.1	Est-ce que les vues (ou fenêtres) relative à l'activité de formation (notifications, lieux géographiques et calendrier des dates des sessions de formations) vous semblent-elles bien organisées ?	Ergonomie/Confort physique
1.2	Les fonctionnalités "Recherche", "Planification" et "Evaluation" du panneau central sont-elles suffisamment accessibles ?	Ergonomie/Confort physique
2	Interface-Recherche	
	Simulation d'une recherche d'une personne	
2.1	Comment jugez-vous la facilité d'utilisation de cette tâche de recherche ?	Expérience utilisateur
2.2	Au sein du graphe, citez les personnes en lien avec GD	
	Simulation d'une recherche d'un formateur soit : Saisie du nom : Bensimon Joseph	Vérification
2.3	Comment jugez-vous la facilité d'utilisation de cette tâche de recherche ?	Expérience utilisateur
2.4	Combien de groupe(s) le formateur Joseph Bensimon a t'il animé ?	Expérience utilisateur
2.5	Pour l'ensemble des ces simulations de recherche, est-ce que la légende visuelle (soit la couleur, la forme des noeuds et des liens) est-elle explicite ?	Ergonomie/Confort physique
3	Interface-Planification	
	Simulation d'un scénario de planification	
3.1	Au niveau de l'interface de recherche, comment appréciez-vous l'agencement des critères selon les colonnes suivantes soit "Groupe", "Formation" et "Logistique" ?	Ergonomie/Confort physique
3.2	Quel est le nombre total de groupes de formation générés ?	Vérification
3.3	Quel est le nombre total de personnes trouvés ?	Vérification
3.4	Evaluation du temps de réponses de la liste des résultats	Ergonomie/Confort physique
3.5	Par défaut, la restitution des résultats présente une liste de groupes par représentation graphique. Un tableur est également accessible en haut à droite de la fenêtre de visualisation. Quel mode de représentation vous parait le plus aisé pour la lecture des données ?	Ergonomie/Confort physique
4	Lex/RH-Satisfaction globale	
4.1	Quelle est votre appréciation globale du prototype pour l'activité de la formation continue ?	Satisfaction globale
4.2	Quelle est votre appréciation globale sur la facilité d'utilisation du prototype ?	Satisfaction globale
4.3	Quelle(s) fonctionnalité(s) vous semble t'elle le plus utile ?	Satisfaction des fonctionnalités/Utilité
4.4	Pour la gestion de l'activité de formation, est-ce que le prototype Lexiane RH a t'elle une valeur ajoutée en comparaison avec le CRM actuel ?	Satisfaction des fonctionnalités/Utilité

Annexe 21. Résultats des tests-utilisateurs

Utilisateurs	Questions	Réponses
Utilisateur 1	1.1	1) Réagencement: le calendrier pourrait être mis en premier plan avec une organisation sur le mois (cf outlook et code couleur en fonction des pôles) à la place de notifications et mettre la notification à la place du réseau (donc à droit du calendrier) 2) la carte de France pourrait être plus interactive (légende) et mis au même niveau que le réseau 3) doublon entre "recherche - planification - evaluation" qui apparaissent en haut et à gauche (pour moi à gauche avec les onglets détaillés est suffisant)
	1.2	Oui
	2.1	Plutôt facile, Indication de l'ID permet une utilisation plus rapide et fiable
	2.2	Réponse correcte
	2.3	Très facile
	2.4	Réponse correcte
	2.5	En partie, Le graphe est-il nécessaire ? Complicé visuellement pour une formation avec plus de 20 participants par ex - prévoir un onglet déroulant (ex CRM : dans une opportunité je vois toutes les missions, on pourrait faire pareil en partant d'un formateur pour voir toutes ses formations, ce qui permet aussi de voir rapidement le nombre de sessions et de faire des filtres par exemple)
	3.1	Bonne, On pourrait rajouter le nom de la formation ou sa référence + date pour pouvoir la planifier
	3.2	Réponse correcte
	3.3	Réponse correcte
	3.4	Instantanée
	3.5	Tableur
	4.1	Bonne
	4.2	Normal
	4.3	Recherche d'entités (personne, organisation, formateur, session de formation), Planification de sessions de formation
4.4	Certainement, mais je préfère le CRM qui me parait plus pratique et complet que des graphes où il faut aller chercher les détails et qui ne peuvent être complets	
Utilisateur 2	1.1	En partie, Carte de France, calendrier et réseau pas actif
	1.2	Oui
	2.1	Plutôt facile
	2.2	Réponse correcte
	2.3	Plutôt facile
	2.4	Réponse correcte
	2.5	En partie, Intra trop épais
	3.1	Bonne
	3.2	Réponse correcte
	3.3	Réponse correcte
	3.4	Instantanée
	3.5	Représentation graphique
	4.1	Très bonne
	4.2	Plutôt facile, Commentaire libre : Pas de bouton de déconnexion
	4.3	Recherche d'entités (personne, organisation, formateur, session de formation), Planification de sessions de formation
4.4	Oui	
Utilisateur 3	1.1	En partie, Pas d'informations sur la facturation, personne à contacter
	1.2	Oui
	2.1	Très facile
	2.2	Réponse correcte
	2.3	Très facile
	2.4	Réponse correcte

	2.5	Oui
	3.1	Bonne
	3.2	Réponse correcte
	3.3	Réponse correcte
	3.4	Correct/Raisnable
	3.5	Tableur
	4.1	Bonne
	4.2	Très facile
	4.3	Recherche d'entités (personne, organisation, formateur, session de formation), Planification de sessions de formation. Prévoir un espace pour la gestion administrative et financière de l'activité de formation (contrat en pièce jointe, personne à contacter, historique des relances...)
	4.4	Oui, Avantage : distinction de la comptabilisation et du CA
Utilisateur 4	1.1	En partie. Ajouter des informations commerciales (suivi des prospects) et les projets de formation en cours de réalisation. Prévoir également un espace dédié au catalogue de formation
	1.2	Oui
	2.1	Très facile
	2.2	Réponse correcte
	2.3	Très facile
	2.4	Réponse correcte
	2.5	En partie, Peu lisible dans le détail Mettre par défaut le graphe filtré soit le formateur et les événements de formation
	3.1	Bonne, Mettre un calendrier pour fixer les dates
	3.2	Réponse correcte
	3.3	Réponse correcte
	3.4	Instantanée
	3.5	Représentation graphique
	4.1	Très bonne, Manque une fonctionnalité dédiée à la commercialisation, les projets de formation en cours de réalisation et les alertes lié au marché publics
	4.2	Très facile
	4.3	Planification de sessions de formation, Gain de temps et possibilité de suivi des projet de formation
4.4	Oui	
Utilisateur 5	1.1	Oui
	1.2	Oui
	2.1	Très facile
	2.2	Réponse correcte
	2.3	Très facile
	2.4	Réponse correcte
	2.5	Oui, En partie, Mettre le lien "partenariat" en rouge-Risque de confusion avec le lien "inter"
	3.1	Bonne, La liste des champs est longue pour rechercher un choix
	3.2	Réponse correcte
	3.3	Réponse correcte
	3.4	Instantanée
	3.5	Représentation graphique
	4.1	Très bonne
	4.2	Très facile
	4.3	Recherche d'entités (personne, organisation, formateur, session de formation)
4.4	Oui, Distinction des participants de la formation avec les collaborateurs lié à l'activité de conseil	
Utilisateur 6	1.1	Oui
	1.2	Très facile, Liste longue+Doublon pour les champs "intitulé de poste", et "Direction"

2.1	Réponse correcte
2.2	Réponse correcte
2.3	Très facile
2.4	Réponse correcte
2.5	Oui
3.1	Bonne, Temps de chargement des listes un peu longues/Liste des champs non alphabétique
3.2	Réponse correcte
3.3	Réponse correcte
3.4	Correct/Raisnable
3.5	Représentation graphique
4.1	Bonne
4.2	Plutôt facile
4.3	Planification de sessions de formation, Pas d'information sur l'historique des participants en activité de conseil
4.4	Oui