



HAL
open science

IA et SHS : un dialogue indispensable

Michel Agnola, Ghislaine Azémard, Samuel da Silva

► **To cite this version:**

Michel Agnola, Ghislaine Azémard, Samuel da Silva. IA et SHS : un dialogue indispensable : Programme ANR ARCHIVAL Valorisation d'archives multimédia : Compréhension automatique multimodale du langage pour de nouvelles interfaces intelligentes de médiation et de transmission des savoirs.. EUTIC 2022 " A l'intersection de l'art, de la science et de la technologie : dialogues entre les hommes et les machines ", Académie Ionienne de Corfu, Grèce, Oct 2022, Corfou, Grèce. hal-04220079

HAL Id: hal-04220079

<https://hal.science/hal-04220079>

Submitted on 27 Sep 2023

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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

IA et SHS : un dialogue indispensable

Michel Agnola, Ghislaine Azemard, Samuel Da Silva
Chaire UNESCO ITEN – FMSH – Consortium ANR Archival

Publication des **conférences internationales EUTIC 2022**

Thématique 2022 : « **A l'intersection de l'art, de la science et de la technologie : dialogues entre les hommes et les machines** »

Académie Ionienne de Corfu, Grèce

Octobre 13-14-15, 2022

<https://avarts.ionio.gr/eutic/en/presentations/837/>

<https://avarts.ionio.gr/eutic/en/speakers/>

Résumé

L'accélération des développements technologiques et les enjeux de la « transition » numérique, impliquent plus que jamais le champ des sciences humaines et sociales, au sein de dynamiques transdisciplinaires, pour le développement de nouveaux modèles sociétaux d'intérêt public. Dans le champ de la recherche documentaire les technologies de compréhension du langage ouvrent des perspectives nouvelles pour la valorisation d'archives multimodales, la médiation culturelle et scientifique, en particulier concernant l'analyse et la mise en relation des fonds documentaires. Mais l'utilisabilité des systèmes développés suppose d'y intégrer l'expertise humaine dans les modèles de conception et de les placer au service de l'intelligence humaine. Comment développer de nouvelles interfaces de recherche documentaire avec les technologies de l'intelligence artificielle au service de l'expérience et de l'intelligence humaine ? Dans le cadre de cette approche transdisciplinaire, associant les cultures et les pratiques en SHS et informatique, un des principaux enjeux du projet Archival a été de développer une méthodologie pour répondre à cette problématique. Nous avons expérimenté et évalué la pertinence de différents algorithmes et méthodes d'analyse du langage.

Abstract

The acceleration of technological developments and the challenges of the digital "transition" involve more than ever the field of humanities and social sciences, within transdisciplinary dynamics, for the development of new societal models of public interest. In the field of documentary research, language comprehension technologies open up new perspectives for the enhancement of multimodal archives, cultural and scientific mediation, particularly with regard to the analysis and linking of documentary collections. However, the usability of the systems developed presupposes that human expertise is integrated into the design models and that they are placed at the service of human intelligence. How can we develop new document search interfaces with artificial intelligence technologies at the service of human experience and intelligence? Within the framework of this transdisciplinary approach, combining cultures and practices in SHS and computer science, one of the main challenges of the Archival project was to develop a methodology to address this issue. We have experimented and evaluated the relevance of different algorithms and methods of language analysis.

Nos modèles socio-culturels et économiques connaissent de profondes mutations sous l'impulsion des avancées scientifiques et technologiques. Le développement des réseaux, de la mobilité avec son offre d'applications à tout faire, la multiplication des supports, le développement de nouvelles interfaces, des « systèmes », de la gestion de données avec les progrès « algorithmiques » et les promesses de l'intelligence artificielle ont changé notre environnement. Les rapports sociaux, les médiations scientifique et culturelles, le monde du travail, les modèles économiques, ou encore, les approches éducatives se transforment. Les avancées scientifiques s'accroissent tandis que les champs scientifiques fusionnent ouvrant de nouvelles « promesses technologiques ». Nous vivons désormais dans un « écosystème numérique » présent dans toutes les strates de nos vies quotidiennes numériques.

Dès lors, il est bien plus difficile d'innover depuis son garage, sa chambre universitaire, ou encore, du fin fond de son bureau personnel comme l'ont fait nombre de précurseurs de l'innovation numérique. Il est plus que jamais nécessaire d'analyser le contexte, les usages, de travailler en premier lieu sur un modèle porteur d'un projet et d'une vision de développement (et avant toute modélisation), pour ensuite déployer de nouvelles méthodes de création, de gestion de projet, et d'expérimentation, afin de développer des innovations qui se donnent les meilleures chances de s'intégrer dans cet écosystème médiatique, ces nouvelles pratiques, et proposer de nouveaux modèles de « transition » porteur de sens.

Dans ce contexte, l'approche transdisciplinaire est fondamentale à la réussite des projets, au premier rang desquels l'apport du champ d'expertise des sciences humaines et sociales. L'enjeu étant celui de la « transition » numérique, la première tâche, en amont de tout développement technologique, est bien d'analyser le contexte et d'élaborer le modèle de « transition » : quelles nouvelles formes de médiation culturelle et scientifique pour un musée pour des fonds d'archives des bibliothèques, quel modèle démocratique pour une plateforme territoriale participative, quel modèle pédagogique pour un projet de eformation, ou encore, quel modèle de management pour une plateforme collaborative en entreprise.

La multiplication des innovations technologiques ouvre un large champ d'implication des sciences humaines et sociales dans les projets de recherche et de développement numériques, à fortiori, dans un environnement. D'une part, la capacité d'innovation numérique (y compris sur des sujets liés au services d'intérêt public) est largement sous le contrôle des grands groupes, notamment des GAFAs qui investissent tous nos champs sociétaux (qui à force de larges ressources forgent des modèles de gouvernance et diffusent leurs modèles économiques). D'autre part, les mutations sociétales, les mouvements de crise économique, politique, écologique, posent plus que jamais le besoin de repenser nos modèles de développement en portant des valeurs fondamentales durables d'intérêt public.

Les moteurs de recherche d'informations et les systèmes de recherche documentaires sont un des champs de recherche à fort enjeu. Le numérique transforme l'accès aux savoirs, les sources éditoriales et fonds d'archives se démultiplient sur les réseaux dans de multiples formats, pourtant ces savoirs sont encore difficilement accessibles auprès des publics, les résultats affichés sous forme de liste, l'agrégation de ces données multimodales de sources différentes est difficile, les interfaces peu attractives.

« Archival » est un projet de recherche pluridisciplinaire et multi-acteurs, qui engage la collaboration de chercheurs issus des Sciences de l'Information et de la Communication, et de l'Informatique. Il est dédié à la valorisation d'archives multimédia avec les dernières innovations de l'intelligence artificielle et du web sémantique. Il exploite les techniques informatiques de compréhension automatique multimodale du langage. Il a pour ambition le développement de nouvelles interfaces intelligentes de médiation et de transmission des savoirs.

Le projet est porté par la Chaire Unesco ITEN (Innovation, Transmission, Edition Numérique) au sein de la FMSH / Université Paris 8 (Fondation de la Maison des Sciences de L'Homme), avec les partenariats d'Orange Lab, du Laboratoire d'Informatique et Systèmes (LIS-Lab, Université Aix-Marseille - CNRS) et l'Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA, Université Rennes 1).

Il est ciblé sur la thématique de l'autogestion à partir de la collection de la revue Autogestion (intérêt du sujet par rapport au contexte socio-économique et aussi afin d'aider à faire émerger les dimensions cachées et les facteurs déterminants de ces mouvements).

Pour la chaire UNESCO ITEN il s'inscrit dans la continuité des recherches menées sur les usages scientifiques, culturels et pédagogiques de l'intelligence artificielle.¹

Constats et problématique

Le numérique transforme l'accès aux savoirs encore difficilement appréhendable par les publics (malgré de nombreuses tentatives pour indexer, structurer textes, vidéos et images selon des référentiels métiers). Les résultats présentés sous forme de listes

-
- ¹ 2014/2017 : « Ma classe 3.0 », conception, développement, évaluation d'une plateforme de parcours numériques d'apprentissage adaptatif avec tuteur virtuel, avec Institut Mines-Télécom Paris Tech, Gutenberg, Domoscio, Hatier.
 - 2017/2019 : « Data et Musée » plateforme interopérable d'agrégation de données des institutions culturelles avec recommandations personnalisées avec l'Institut Mines –Télécom Paris Tech, Orphéo, Tech 4 Team, Guestviews, Kernix, Paris Musées, Réciproque, Centre des Monuments nationaux.
 - 2017/2019 : « DED–Lab », expérimentation d'une interface 3D de médiation culturelle dans un corpus complexe croisant plusieurs représentations : géographique, chronologique et sémantique, et séminaire portant sur les nouvelles interfaces et les modes de navigation dans des corpus complexes avec la BnF, Idéfi CréaTIC et la FMSH.
 - 2016/2018 : Réalisation de 3 projets de médiation culturelle "Archivage, médiation numérique et la du patrimoine culturel immatériel" avec la Maison des cultures du Monde, Sysigambis, Idéfi CréaTIC, le Conservatoire International de Musique de Paris.
 - 2014 - : « Dynamicarte : modélisation de plateformes collaboratives » avec le Labex Arts H2H -Maison Des Cultures du Monde /Centre Français du PCI, Idefi CreaTIC, projet TIGA, CNIS, la FMSH.
 - 2018 : Lauréat au titre de l'Université Paris8 de l'AMI TIGA « Territoires Innovants de Grande Ambition » portée par l'Equipement Public Plaine Commune. (Plan d'Investissement d'Avenir)

fonctionnelles offrent peu de possibilités de navigation, d'exploration des documents et articulent mal les contenus entre eux.

Un des enjeux majeurs de l'usage de l'Intelligence artificielle repose sur ses potentialités à produire des médiations à valeur ajoutée pour l'utilisateur. Dans quelles mesures les méthodes de compréhension par les machines dans la réinterprétation de fonds d'archives thématiques peuvent aider à l'hybridation des formes d'expression des contenus en coprésence au sein d'un même fonds, mais aussi par croisements de fonds archivistiques (qui peinent encore à dialoguer), et favoriser l'émergence de dimensions méconnues et sous-jacentes aux corpus. Selon quelles modalités ensuite les interfaces de médiation des contenus peuvent-elles exploiter des résultats générés par les méthodes actuelles d'Intelligence Artificielle ? La théorie de la « transtextualité » (Genette, 1982), prise au sens large, résonne aujourd'hui avec les liens sémantiques du web de données, la transversalité entre documents : relier un texte à d'autres écrits mais aussi à d'autres représentations (images, vidéos, sons). ²

L'Intelligence artificielle apporte des ressources nouvelles qui dépassent les capacités humaines en particulier en termes de capacité de calcul, de mémorisation, concernant l'analyse et de mise en relation des fonds documentaires, mais l'utilisabilité des systèmes développés suppose d'y intégrer l'expertise humaine dans les modèles de conception et de les placer au service de l'intelligence humaine.

Comment développer de nouvelles interfaces de recherche documentaire avec les technologies de l'intelligence artificielle au service de l'expérience et de l'intelligence humaine ?

Dans le cadre de cette approche transdisciplinaire, associant les cultures et les pratiques en SHS et informatique, un des principaux enjeux du projet Archival a été de **développer une méthodologie** pour répondre à cette problématique.

Approche technologique et objectifs scientifiques

Le projet Archival se positionne sur plusieurs objectifs scientifiques et axes d'innovation :

- déployer une méthode de compréhension de documents par les machines de manière multimodale (paroles, textes, images fixes et vidéos), permettant de générer de manière automatisée de liens inter documents multimodaux par l'analyse linguistique et la représentation sémantique,
- expérimenter le questionnement des fonds en langage naturel,
- développer et structurer dynamiquement une interface intelligente de résultats de requêtes afin de valoriser la pertinence sémantique des contenus de manière attractive, enrichissante et ergonomique pour l'utilisateur.
- Et de façon plus globale, optimiser la chaîne de traitement des documents, et ouvrir des perspectives de collaborations inédites entre chercheurs en SHS, STIC et Informatique, en plaçant l'utilisateur final au cœur du processus d'apprentissage machine.

² Genette, G., *Palimpsestes*, éditions du Seuil, 1982.

Sur un plan technologique et informatique, le projet répond aux objectifs aux constats ambitions et objectifs suivants (nous reprendrons au plus près les formulations des collègues informaticiens).

Les méthodes de compréhension automatique du langage (*Natural Language Understanding*), regroupées souvent sous la bannière de l'IA, permettent de structurer l'information et de générer des cartographies de l'espace sémantique qui raisonnent comme un maillage de l'information. Les méthodes à base d'apprentissage automatique par des réseaux de neurones profonds (méthodes neuronales) permettent d'apprendre directement des représentations (projections dans des espaces numériques) en fonction d'une tâche (détecter des concepts ou entités, classer des documents, etc...) sans avoir à spécifier explicitement les différents niveaux d'analyse nécessaires à sa réalisation.

Les espaces vectoriels sémantiques ainsi construits permettent à la fois de définir des relations entre des mots et des documents à l'intérieur de cet espace, mais aussi d'utiliser ces représentations vectorielles pour générer des sorties prenant en compte le « sens », la « science latente » contenue dans les documents pour pouvoir les exploiter. Les avancées récentes en termes de compréhension automatique du langage ont permis de dépasser la simple extraction de termes (mots clés, entités, triplets RDF) pour viser une meilleure prise en compte du sens d'un document. Parallèlement les méthodes de « compréhension de document par les machines » permettent aussi d'interroger des documents par des questions naturelles pour en découvrir des extraits pertinents.

Un document est alors appréhendé au prisme des relations qu'il entretient avec d'autres documents. Des réseaux de liens entre les documents peuvent ainsi être générés par similarité de mots, de sujets, parce qu'ils s'inscrivent dans la continuité du propos, qu'ils sont liés aux mêmes questions et/ou réponses, ou encore, qu'ils proposent les mêmes liens vers Wikipédia. Un réseau très étendu de liens entre les contenus documentaires des fonds est ainsi généré. Il permet de créer des représentations sémantiques à la fois dans une dimension multimodale (entre les différents sources multimédia) et aussi en fonction de la nature des contenus (notions, lieux, personnages).

Il s'agit alors de travailler sur les manières de structurer une collection de documents hétérogènes en faisant apparaître de manière explicite les liens implicites, de révéler la nature de ces liens et de les valoriser de manière intelligible par la médiation d'interfaces ergonomiques qui garantissent une appropriation réussie des contenus. Ainsi un design d'interfaces attractives, plus didactique pour l'utilisateur, permettra de réinterpréter ces archives dans un univers de sens adapté, mis au service de la transmission. Comment faire dialoguer, "parler" les fonds numérisés d'archives, documentaires et leurs métadonnées ? Comment enrichir le recours à des référentiels (BnF, INA) en exploitant l'analyse des contenus et des relations entre éléments de contenu (liens implicites).

Positionner les limites et les apports de l'intelligence artificielle au regard de l'intelligence « utilisateur »

L'Intelligence artificielle est un champ d'innovation très prometteur et à fort enjeux avec des applications dans tous les domaines. Il porté en particulier par les GAFAs, grands réseaux et sociétés technologiques. Il est valorisé aussi, bien souvent, comme un concept marketing avec des acceptations très différentes, dont l'apport en « intelligence » peut-être parfois discutable.

La notion « d'intelligence » des systèmes mérite d'être délimitée au regard de ce qui définit l'intelligence humaine, une machine peut-elle porter une conscience, des capacités émotionnelles et sensorielles par exemple. Les technologies de l'intelligence artificielles sont avant tout basées sur du « calcul », de la programmation informatique (et dont « le programme » porte aussi une intention, des objectifs, voire une certaine vision du monde).

Il s'agit aussi d'une technologie encore largement émergente avec de multiples verrous à dépasser, qui apporte des ressources nouvelles avec de réelles potentialités de développement, mais avec des résultats en l'état actuel des avancées scientifiques souvent aléatoires, à prendre en compte dans les développements de projets.

L'engouement autour des technologies de l'intelligence artificielle soulève aussi beaucoup de discussions et débats, comme toute innovation susceptible de modifier profondément notre environnement (et celle-ci en particulier puisqu'elle touche à cette notion « d'intelligence »). Les discours autour des dérives potentielles de l'intelligence artificielle portent sur la « personnalisation » de l'IA (y prêter une intention de conscience), la position de l'utilisateur par rapport au système (qui contrôle quoi), la domination du « discours machine » et ses tentations « transhumanistes », les dangers de l'hybridation de l'homme et de la machine comme vision de l'avenir (jusque sur un plan physique : ex. implémentation puces), les risques d'homogénéisation des consciences, de mécanisation et de robotisation, de surveillance omnisciente, d'atteinte aux libertés et à nos identités, ou encore, sur les limites de la vision purement prédictive (la créativité par exemple par définition n'est pas prédictive et est un facteur essentiel de nos développements), sur les valeurs portées par les modèles proposées par les systèmes IA (mondialisation, approches consuméristes,...).³

³ Le cadrage d'une éthique de l'IA fait l'objet de multiples travaux. La commission européenne identifie les indicateurs suivants : facteur humain et contrôle humain, robustesse et sécurité, respect de la vie privée et gouvernance des données, transparence, diversité, non-discrimination et équité, bien-être social et environnemental, responsabilisation. Une étude du CEPEJ relève aussi sur un plan juridique plus global le respect des droits fondamentaux. Les Nations Unies - UNESCO associent le principe d'explicabilité à la notion de transparence. Ils ajoutent les indicateurs : surveillance et décisions humaines, sensibilisation et éducation, et, gouvernance et collaboration multipartites et adaptatives.

- [Lignes directrices en matière d'éthique pour une IA digne de confiance](#), Commission européenne, avril 2019
- Quiniou M., [Mise en place éventuelle d'un mécanisme de certification des outils et services d'intelligence artificielle dans le domaine juridique et judiciaire : Etude de faisabilité](#), Commission européenne pour l'efficacité de la justice (CEPEJ), Conseil de l'Europe, 8 décembre 2020
- [Recommandation sur l'éthique de l'intelligence artificielle](#), Organisation des Nations Unies / UNESCO, 2022

Ces nouvelles technologies apportent aussi de nouvelles ressources qui peuvent largement dépasser les capacités humaines et contribuer à venir enrichir avec efficacité nos capacités d'acquisition de connaissances. Les apports technologiques sont à mesurer dans les capacités de calcul, d'analyse de fonds documentaires multiples, le moissonnage des données et la génération de liens automatisés, d'agrégation de ces différentes bases de données et ressources multimédias dans une même interface, dans ces capacités de mémorisation des données et du parcours utilisateur, ou encore, dans le développement de fonctionnalités à valeur ajoutée de personnalisation et intervention sur les contenus, de partage et d'intégration de l'expérience utilisateur (« machine learning »).

Les technologies de l'intelligence artificielle sont ainsi une opportunité si on les place non pas pour venir remplacer l'intelligence utilisateur mais l'augmenter, avec de la pédagogie, en apportant de l'explicabilité aux réponses et propositions apportées par la machine, en permettant à l'utilisateur de garder le contrôle, de trier, sélectionner, de se construire son propre parcours, le mémoriser, l'enrichir, ou encore, le partager.

Définir un modèle porteur d'une posture « humaniste » avant toute modélisation

Le champ d'expertise des sciences humaines et sociales est fondamental dans ce contexte de mutation des environnements sous l'impulsion des innovations numériques, à fortiori pour les projets qui engagent de profondes mutations dans les pratiques et les organisations.

Il intervient dès les premières étapes des projets et avec tout le référentiel de travaux de recherche et d'études disponibles (voire, d'expérimentations déjà menées) en fonction des problématiques, des objectifs et du secteur d'activité considéré, pour mener un travail d'analyse scientifique du contexte et des problématiques, d'étude et d'enquête, ou encore, développer un état de l'art, expertiser les contraintes technologiques avec les autres partenaires, afin de déterminer un modèle. Chaque modèle est spécifique à chaque projet. Il est un élément référentiel ensuite pour les étapes suivantes de conception, de développement et d'expérimentation.

Ce modèle place aussi une posture par rapport au projet : il porte des valeurs fondamentales, une représentation des publics ciblés, des échanges, ou encore, des formes de médiation envisagées.

La chaire Unesco Iteu a mené de multiples travaux sur la thématique de l'humanisme numérique » et les questionnements fondamentaux posés par les mutations sociétales dans ce nouvel environnement technologique. Un séminaire organisé avec le réseau des chaires Unesco en SCI portait sur le thème : « Comment réinvestir, redéfinir et co-construire la notion d'humanisme à l'ère numérique à travers ses valeurs, ses diversités, à préserver et à créer ? ». La chaire s'est investie dans des champs spécifiques applicatifs de la médiation culturelle, scientifique, du développement territorial, de l'innovation sociale, ou encore, de l'éducation et de la

formation. La chaire co-construit et expérimente dans des dynamiques de partenariat multi-acteurs des recherches, expérimentations de plateformes, de service et de dispositifs innovants (cf. page 2) dans la continuité desquelles s'inscrit le projet Archival.

Cette expertise et cette acquisition de connaissances constitue un socle pour la co-élaboration de projets de recherche et d'expérimentation comme le projet Archival.

Différents profils de chercheurs en sciences humaines et sociales ont été associés en amont à l'élaboration du projet ; sur l'expertise du fond sur la thématique de l'autogestion/. Ils ont réalisé un travail éditorial sur l'interprétation sémantique et les notions clefs, complété par des outils d'analyse automatique pour construire et enrichir la première version d'une ontologie. Des analyses ont été produites concernant les attentes et les modes de consultation des bases documentaires, les formes de médiation dans la recherche documentaire, ainsi sur les enjeux technologiques et les innovations dans ce domaine. Un état de l'art sur les interfaces, les fonctionnalités et les approches utilisateur dans le champ des bases de données documentaires a encore été réalisé. Un travail éditorial a été mené sur les contenus, la définition des cas d'usage et des parcours utilisateurs.

Traduction dans une méthodologie de gestion de projet

L'apport de notre expertise en SHS a été porté à tous les stades du projet à la fois pour l'apprentissage et le développement des algorithmes, et pour le développement d'interfaces attractives, enrichies et adaptées aux besoins utilisateurs, avec notamment :

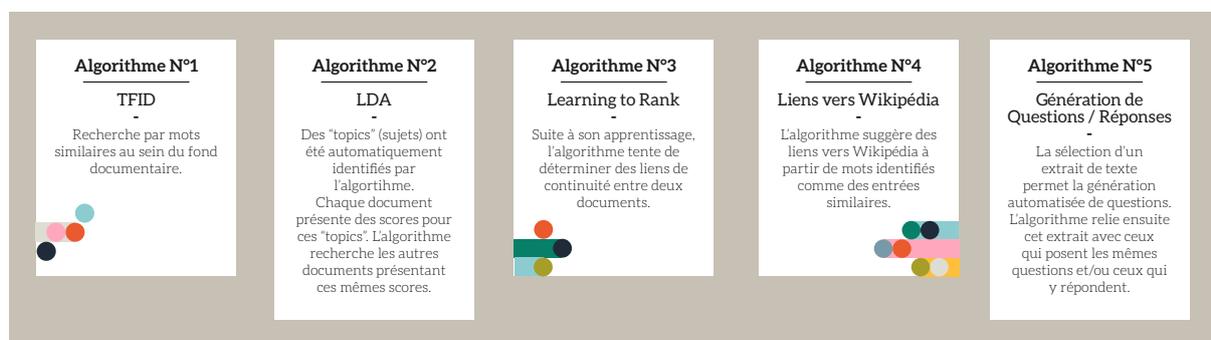
- un travail référentiel d'analyse éditoriale du fond et de la thématique de l'autogestion (détermination des notions fondamentales, identification des grands acteurs, analyse des réseaux, sphères d'influences,...)
- un travail d'annotation et de questionnements préalable du fond par des experts et chercheurs (exploité ensuite notamment pour l'apprentissage des systèmes),
- la production de cas d'usage et de scénarios « utilisateur » sur des sujets ciblés,
- la création d'un thésaurus et ontologie (réalisé à la fois à partir des notions identifiées et du travail d'indexation automatisé), exploité comme une matrice d'intelligibilité pour l'indexation des contenus et la génération de liens,
- un travail sur des règles de pondération des contenus pour la hiérarchisation des contenus, l'ajustement des liens et l'organisation éditoriale des interfaces,
- les enrichissements éditoriaux multiples apportés par l'analyse sémantique des contenus ou les métadonnées des bases documentaires,
- la conception d'une interface enrichie et intelligente qui ouvre de nouvelles formes d'accès aux contenus, positionne les propositions apportées par l'IA comme un support d'aide à la recherche documentaire tout en apportant de la pédagogie sur « l'explicitabilité » des liens suggérés par le système afin d'en favoriser la réappropriation (« faire apparaître de manière explicite les liens implicites, révéler la nature de ces liens et les valoriser de manière intelligible par la médiation d'interfaces ergonomiques qui garantissent une appropriation réussie des contenus »)

- le développement de fonctionnalités avancées de personnalisation, de mémorisation des parcours, des contenus, d'organisation éditoriale, d'annotation, ou encore, de partage.
- les process programmés pour l'expérimentation du projet auprès de publics types et son recadrage.

Environnement et stratégie de développement technique

Un démonstrateur a été réalisé à destination des chercheurs en SHS qui souhaitent travailler sur le corpus multimodal sur la thématique de l'autogestion. Ce corpus est constitué de la revue Autogestion, d'une sélection de documents audiovisuels de Canal U et de l'INA. Il est ouvert à d'autres silos de données par l'utilisation de l'API de la BNF, qui permet de constituer dynamiquement les bibliographies de centaines d'auteurs ou de personnages cités, ainsi que des bibliographies liées à des notions. De la même façon, le prototype est enrichi par l'utilisation des API de Wikidata afin de constituer des fiches biographiques illustrées. Une base de données NOSQL permet d'accepter la diversité de structures de documents propre à toute dimension « big data ».

La lecture automatique du fond textuel par les outils informatiques mis en place par les laboratoires informatiques partenaires du projet permettent d'identifier des mots clés, les personnages et les lieux. Ces éléments identifiés permettent de conditionner les appels aux différentes APIs. Ce fond dynamique constitué est traité pour faire émerger des propositions de lectures multimodales. Ces suggestions sont générées par différents algorithmes. Le principe technique de chaque type d'algorithme de génération de lien est explicité systématiquement à l'utilisateur.



Explication des différents algorithmes exploités pour la suggestion de liens

Comme les outils de la suite bureautique, la maîtrise et la capacité de choisir ces outils algorithmiques deviennent des prérequis afin d'exploiter les fonds documentaires de grande ampleur. Cette volonté d'explicitabilité du rôle des algorithmes s'inscrit dans l'ambition de faire converger les développements informatiques issus des laboratoires, avec une informatique d'usage éprouvée et orientée vers la médiation scientifique et culturelle.

La méthodologie employée a permis la mise en place d'un dialogue interdisciplinaire constant entre les SHS et les informaticiens. L'équipe FMSH intègre des compétences informatiques spécialisés sur les questions de médiations scientifiques et culturelles, en particulier afin d'assurer l'intégration finale du prototype. Les deux laboratoires

informatiques partenaires ont recruté un ingénieur qui par son statut et sa culture a permis de matérialiser les travaux des chercheurs en informatique, vers pour le développement d'un démonstrateur publié, consultable et fonctionnel, dans un dialogue constant avec le référant informatique de l'équipe FMSH. Ces rencontres professionnelles ont permis la mise en place effective du développement du prototype et des échanges productifs pour orienter le projet vers l'utilisateur.

L'attention des chercheurs SHS sur la structure sémantique des textes, l'intertextualité, les rebonds documentaires, a permis aux chercheurs en informatique, d'une part d'éprouver la pertinence des propositions algorithmiques, et d'autre de l'orienter vers une approche utilisateur y compris experte.

Les bibliothèques du futur, les fonds d'archives et leur accès ne pourront se constituer que par la convergence de tous ces types de compétences qui constitue souvent un verrou au développement productif de plateforme d'accès à l'open data d'intérêt public.

Bibliographie

- Authier, M., Lévy, P., *Les arbres de la connaissance*, La Découverte Poche, Essai n°9, 2017.
- Azémard, G., *Humanisme numérique : valeurs et modèles pour demain ?* 3 tomes, Paris, Ed. de l'immatériel 2017.
- Bachimont, B., *L'intelligence artificielle comme écriture dynamique : de la raison graphique à la raison computationnelle*, In J. Petitot & P. Fabbri (Eds.), *Au nom du sens* (pp. 290-319): Paris, Grasset, 2000.
- Bachimont, B., *Ingénierie des connaissances et des contenus : le numérique entre ontologies et documents*. Paris, Hermès, 2007
- Bechet, F., Aloui, C., Charlet, D., Damnati, G., Heinecke, J., Nasr, A., Herledan, F., *CALOR-QUEST : un corpus d'entraînement et d'évaluation pour la compréhension automatique de textes*, soumis à TALN, 2019.
- Bertin, J., *La Graphique et le traitement graphique de l'information*, Flammarion, 1977.
- Bois, R., Vukotić, V., Simon, A.-R., Sicre, R., Raymond, C., Sébillot, P., Gravier, G., *Exploiting multimodality in video hyperlinking to improve target diversity*. MMM 2017.
- Budnik, M., Demirdelen, M., Gravier, G., [A study on multimodal video hyperlinking with visual aggregation](#), IEEE ICME 2018.
- Charlet, D., Damnati, G., [SimBow at SemEval-2017 Task 3: Soft-Cosine Semantic Similarity between Questions for Community Question Answering](#), SemEval 2017.
- Charlet, D., Damnati, G., [Linking written News and TV Broadcast News topic segments with semantic textual similarity](#), NLP 4 Journalism, LREC workshop, 2018.
- Damnati, G., Charlet, D., Denjean, M., [Exploring collections of multimedia archives through innovative interfaces for Digital Humanities](#), Show & Tell Interspeech'16, 2016.
- Dalbera, J.-P., Leblond S., Mouren R., *Enjeux et Spécificités de la médiation numérique*, In. Patrimoine culturel immatériel et numérique, Dir. Severo, M., et Cachat, S., L'Harmattan, Coll. Humanités Numériques, 2017.
- Desprez, H., Dalbera J.-P., *Recherche-création sur les interfaces documentaires prospectives*, Séminaire Interfaces intelligentes, chaire UNESCO ITEN, Paris, 2017.
- Desprez, H., Laborderie, A., *Designer de nouvelles interfaces dynamiques documentaires pour valoriser les données des bibliothèques numériques*, Colloque CIBAHN 2018, Paris, octobre 2018.
- Douehi M., *Pour un humanisme numérique*, Seuil, La librairie du XXIe siècle, 2011.

- Fadili, H. *Une approche de conception collaborative et d'exploitation des modèles ontologiques des données.* Conférences internationale Digital Humanities (DH), Montréal, Canada 2017
- Fadili, H. *Une approche basée sur une stratégie d'apprentissage profond pour la détection du sens dans les données non structurées*, in Ingénierie et management des systèmes d'information, Londres, 2016.
- Ganascia, J-G, *Le mythe de la singularité. Faut-il craindre l'intelligence artificielle ?* Éditions du Seuil, 2017.
- Genette, G., *Palimpsestes*, éditions du Seuil, 1982.
- Jónsson, B., Worring, M., Zahálka, J., Rudinac, S., Amsaleg, L., [Ten Research Questions for Scalable Multimedia Analytics](#). MMM 2016.
- Jung, J., Euzenat, J. *Towards semantic social networks*. Proc. 4th European semantic web conference (ESWC), Innsbruck, pp.267-280, 2007.
- [Lignes directrices en matière d'éthique pour une IA digne de confiance](#), Commission européenne, avril 2019.
- Marzinotto, G., Béchet, F., Damnati, G., A. Nasr, [Sources of Complexity in Semantic Frame Parsing for Information Extraction](#), Int. FrameNet Workshop 2018: Multilingual FrameNets and Constructicons, Japan, 2018.
- Marzinotto, G., Auguste, J., Béchet, F., Damnati, G., Nasr, A., [Semantic Frame Parsing for Information Extraction : the CALOR corpus](#), LREC'18, Japan, 2018.
- Marzinotto, G., Béchet, F., Damnati, G., Favre, B., *Robust Semantic Parsing with Adversarial Learning for Domain Generalization*, proc. of the Industry Track of NAACL HLT 2019, Minneapolis, USA, 2019.
- Marzinotto, G., Heinecke, J., Damnati, G., *MaskParse@ Deskiñ at SemEval-2019 Task 1: Cross-lingual UCCA Semantic Parsing using Recursive Masked Sequence Tagging*, SemEval 2019. Minneapolis, USA, 2019.
- Moravec, H., *L'avenir des robots et l'intelligence humaine*, Éditions Odile Jacob, 2019.
- Moretti, F. *Graphes, cartes et arbres. Modèles abstraits pour une autre histoire de la littérature*. Paris, Les Prairies ordinaires, 2008.
- Neyrat, Fr., *Humanisme, Antihumanisme, Post humanisme*, Éditions Dehors, 2015.
- Nguyen, T., Rosenberg, M., Song, X., Gao, J., Tiwary, S., Majumder, R., & Deng, L. *Ms marco: A human generated machine reading comprehension dataset*. arXiv preprint arXiv:1611.09268, 2016.
- Nguyen, P. A., Li, Q., Cheng, Z.Q., Lu, Y.J., Zhang, H., Wu, X., Ngo, C.W., *VIREO@ TRECVID 2017: Video-to-Text, Ad-hoc Video Search and Video hyperlinking*. TREC Video Retrieval Evaluation Workshop, 2017.
- O'Neil C., *Algorithmes : la bombe à retardement*, Editions Les Arènes, 2018
- Pang, L., Zhang, W., Tan, H. K., Ngo, C.W. *Video hyperlinking: libraries and tools for threading and visualizing large video collection*. ACM international conference on Multimedia, 2012.
- Quiniou M., [Mise en place éventuelle d'un mécanisme de certification des outils et services d'intelligence artificielle dans le domaine juridique et judiciaire : Etude de faisabilité](#), Commission européenne pour l'efficacité de la justice (CEPEJ), Conseil de l'Europe, 8 décembre 2020.
- [Recommandation sur l'éthique de l'intelligence artificielle](#), Organisation des Nations Unies / UNESCO, 2022.
- Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. *Squad: 100,000+ questions for machine comprehension of text*, Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016.
- Wieviorka M., *L'impératif numérique ou La nouvelle ère des sciences humaines et sociales*, CNRS, 2013.