



**HAL**  
open science

## Les réseaux numériques étendus, une alternative aux ontologies ?

Pierre-Michel Riccio

► **To cite this version:**

Pierre-Michel Riccio. Les réseaux numériques étendus, une alternative aux ontologies ?. Roger Bautier, Pierre-Michel Riccio, Geneviève Vidal. Des usages aux pratiques : le Web a-t-il un sens ?, Presses Universitaires de Bordeaux, pp.47-59, 2016, Labyrinthes, 979-10-300-0081-8. hal-01970637

**HAL Id: hal-01970637**

**<https://hal.science/hal-01970637>**

Submitted on 5 Jan 2019

**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.

# Les réseaux sémantiques étendus, une alternative aux ontologies

Pierre-Michel RICCIO

LGI2P, Ecole Nationale Supérieure des mines d'Alès

Ces dernières années, la généralisation des réseaux numériques à haute vitesse a modifié notre façon d'appréhender données, informations et connaissances. L'utilisateur peut désormais accéder facilement à une très grande quantité d'informations, mais il est difficile de trouver l'information utile. La recherche d'informations réalisée en s'appuyant sur des ontologies – c'est-à-dire des modèles de connaissances – donne d'assez bons résultats, mais la mise en œuvre de ce type de dispositif est difficile et coûteuse. Elle reste pour le moment réservée à des domaines très spécialisés.

Aussi, notre idée, à la confluence des sciences de l'information et de la communication et des technologies numériques, est d'appréhender comment construire facilement et à moindre coût des modèles de données ontologiques qui vont permettre de trouver les informations « utiles » dans des corpus scientifiques en conciliant primat de l'intervention des experts en situation et utilisation optimale de la capacité de calcul des ordinateurs ?

Cet article présente les travaux que nous avons conduits sur cette question en associant travaux menés dans le domaine du web sémantique (sciences et technologies de l'information et de la communication - STIC) et autour de l'approche en compréhension des situations (sciences de l'information et de la communication - SIC). Nous présentons la genèse de l'action, l'objectif retenu, puis détaillons la démarche et sa mise en œuvre sur un cas : la construction d'un modèle ontologique de la toxicologie nucléaire environnementale.

## INTRODUCTION

Le développement très rapide du web, qu'il s'agisse des réseaux numériques à haute vitesse ou des données mises à disposition, a modifié notre façon d'appréhender données, informations et connaissances. L'utilisateur, dans ses activités personnelles ou professionnelles, peut désormais accéder facilement à une très grande quantité d'informations.

Au début des années 2000, une question nous intéressait plus particulièrement : comment utiliser la capacité de calcul des ordinateurs pour permettre aux membres d'une communauté scientifique de trouver plus facilement des documents « utiles » ?

## 1. LA GENESE DE L'ACTION

La question n'est pas nouvelle puisque la problématique de la recherche d'informations est née avec l'écriture, et l'idée d'utiliser les calculateurs pour trouver plus facilement un texte, un document, un ouvrage, est apparue avec les premiers ordinateurs.

Les premières applications ont consisté à construire des référentiels sous la forme de thésaurus basés sur la classification proposée par Melvil Dewey en 1876 (Dewey, 1876). Le travail consistait alors à décrire chaque document ou ouvrage à l'aide de descripteurs ou mots-clés issus de ces référentiels ou thésaurus, ce qui présentait l'avantage de limiter les problèmes d'orthographe ou de dénomination. La recherche de documents s'effectuait ensuite à l'aide de requêtes exprimées sous forme booléenne – suite de termes séparés par une combinaison d'opérateurs logiques de type : et, ou – permettant par appariement automatique de produire in fine une liste d'objets.

Les techniques ont évolué. La classification Dewey a été progressivement complétée / enrichie à l'aide de la classification décimale universelle (CDU) (Otlet et La Fontaine, 2001). La description des documents a été élargie à de nouveaux éléments : titre, créateur, sujet, éditeur, contributeur, date, etc. Cet ensemble de métadonnées appelé « Dublin Core » a fait l'objet d'une norme internationale (ISO 15836). De nombreux travaux de recherche ont été initiés pour indexer les documents par analyse automatique du texte intégral (Chauché, 2003). Enfin, les requêtes ont considérablement évolué avec la généralisation de questions en langage naturel et l'émergence d'ontologies, « spécification explicite (formelle) d'une conceptualisation (partagée) » (Gruber, 1993), qui peuvent servir de guide sémantique à l'expression de la requête et de support actionnable pour la recherche d'informations (Ranwez, 2010).

Mais, le volume d'informations accessibles via les réseaux numériques ayant explosé, la recherche non guidée par des ontologies donne un très grand nombre de réponses parmi lesquelles il est difficile de trouver l'information. La recherche d'informations s'appuyant sur les ontologies donne d'assez bons résultats, mais sa mise en œuvre en dehors de domaines très spécialisés reste limitée. Les ontologies restent difficiles et coûteuses à construire et à entretenir.

Aussi, notre idée était d'évaluer s'il est possible de construire rapidement des modèles de données ontologiques permettant pour un usage précis (recherche de documents dans des corpus scientifiques) d'obtenir des résultats d'un bon niveau de qualité tout en conciliant primat de l'intervention humaine (expert en situation) et calculs automatiques (en repoussant les limites habituelles). Cette problématique a constitué la genèse du projet « réseaux sémantiques étendus ».

## 2. LE PROJET « RESEAUX SEMANTIQUES ETENDUS »

Après quelques expérimentations conduites avec des PME, nous avons été contactés en 2004 par le programme incitatif pluridisciplinaire ToxNuc-E initié par le CEA et soutenu par le CNRS, l'Inra et l'Inserm. L'objectif de ce programme était de stimuler l'émergence d'une communauté d'experts et de jeunes chercheurs autour d'un enjeu qui touche à la santé publique et à l'environnement : comprendre les mécanismes d'actions de métaux lourds et de radionucléides aux différents niveaux d'organisation du vivant (Ménager, 2004).

La direction du programme ToxNuc-E nous a proposé de créer un dispositif permettant d'identifier dans de gros volume de données – par exemple l'ensemble des publications liées au domaine de la toxicologie nucléaire environnementale – les documents susceptibles d'intéresser des chercheurs.

Nous avons alors lancé une action de recherche sur laquelle nous avons mobilisé : un doctorant en STIC, un doctorant en SIC, des étudiants de Master et plusieurs permanents du laboratoire.

Au-delà de la question du travail collaboratif sur laquelle nous travaillions déjà par ailleurs, la question essentielle était : quelle instrumentation technologique peut-on imaginer pour accompagner la découverte des travaux et les apprentissages croisés ?

L'objectif étant de donner toute sa place à l'expert (traitement manuel) tout en utilisant au mieux la capacité des calculateurs (traitement automatique), nous avons engagé un travail en sciences de l'information et de la communication pour créer une ontologie de la toxicologie nucléaire environnementale et en sciences et technologies de l'information et de la communication pour imaginer une instrumentation technologique innovante.

## 3. VERS LA CONSTRUCTION D'UNE ONTOLOGIE DE LA TOXICOLOGIE NUCLEAIRE

Les travaux des philosophes et sociologues des sciences comme ceux de Thomas Kuhn (Kuhn, 1972) ont montré que chaque communauté scientifique construit son propre paradigme qui désigne l'ensemble des croyances, valeurs et techniques partagées par le groupe et se traduit par un langage qui favorise la diffusion des travaux.

Aussi, un des principaux verrous à lever pour favoriser l'efficacité collective et en particulier la diffusion des travaux était de faciliter la construction d'un langage commun de la toxicologie nucléaire environnementale.

Le thème de recherche étant nouveau, dans son approche et ses limites, il nous a semblé intéressant d'initier la construction de l'ontologie en nous appuyant sur les documents de synthèse décrivant le programme de travail à moyen terme de chacun de 15 équipes du programme ToxNuc-E.

Aussi, nous sommes partis des 15 documents, résultat du travail d'écriture (c'est-à-dire d'une négociation entre les contributeurs) réalisé par chaque équipe projet composée de spécialistes de différentes disciplines : biologie, chimie, médecine, physique, etc., soit un total de 654 pages ou 263.371 mots.

Le travail a consisté alors à identifier manuellement dans ce corpus l'ensemble du vocabulaire spécifique pour en extraire avec l'appui de trois informateurs clés, chercheurs du programme ToxNuc-E ayant proposé de nous aider, un premier niveau de catégorisation.

Ceci nous a permis de dégager assez rapidement 7 catégories principales : disciplines, organismes, modèles biologiques, toxiques d'intérêt, molécules, outils et type d'études et un ensemble d'environ 1.200 « termes candidats ».

En se référant à la méthode MENELAS (Bouaud et al., 1995) de construction d'une ontologie, théorisée par Bruno Bachimont (Bachimont, 2000), outillée et améliorée par Didier Bourigault (Bourigault et al., 2004) et présentée de façon synthétique par Jean Charlet (Charlet, 2005), nous venons de terminer la première des quatre étapes du processus de construction d'une ontologie.

Le processus de construction d'une ontologie peut être décomposé en quatre étapes :

- choix et analyse de corpus : extraction à partir du corpus d'un ensemble de termes ou signifiés linguistiques (termes candidats),
- normalisation sémantique : normalisation de la signification des termes sous forme d'un arbre de signifiés linguistiques normés (concepts linguistiques ou ontologie régionale),
- engagement ontologique : transformation des concepts linguistiques en treillis de concepts (concepts formels),
- opérationnalisation : représentation de l'ontologie dans un langage de représentation des connaissances (ontologie).

Nous avons ensuite envoyé une première esquisse du travail sous la forme d'un document papier facile à annoter à une trentaine de scientifiques, soit deux par projet. Nous avons accompagné les scientifiques dans la démarche. Le taux de retour des documents a été important : 87 %.

Naturellement le travail le plus difficile a alors commencé : concilier le point de vue des différents spécialistes et s'entendre sur la signification et la position des termes candidats devenant signifiés linguistiques dans l'arbre des concepts linguistiques.

Grandes familles disciplines	Disciplines	Disciplines dérivées	Grandes familles organismes	Grandes familles modèles biologiques	Modèles biologiques
Biologie	Biologie structurale			Procaryotes	Bactéries
	Biologie animale		Algues		Levures
	Biologie cellulaire		Plantes	Eucaryotes	
	Biologie végétale		Bactéries	Protéines	
	Biologie fonctionnelle		Champignons	ADN	
	Enzymologie		Homme	ARN	
	Métabolisme		Levures	Gènes	
	Métabolomique		Mammifères	Poisson zèbre	
	Microbiologie		Microorganismes	Molusque	
	Protéomique		Mollusques bivalves	Organes	Thyroïde
	Radiobiologie				Estomac

Figure 01 – extrait de l'arbre des concepts linguistiques ToxNuc-E

Ceci nous a permis de réaliser la normalisation sémantique, deuxième étape du processus de construction d'une ontologie. L'objectif du travail n'étant pas de construire une ontologie finalisée de la toxicologie nucléaire environnementale, mais d'élaborer un modèle de connaissance nous permettant de faciliter la recherche d'information dans un contexte local, nous avons décidé d'arrêter le processus de construction d'une ontologie à ce stade et de nous contenter de l'arbre des concepts linguistiques ou ontologie régionale (Bachimont, 2000).

Pour mener à bien ce cas (Yin, 1984), nous avons utilisé une approche qualitative et en compréhension (observation, échange de messages, de documents, entretiens téléphoniques et / ou de vive voix). Les informations collectées ont fait l'objet d'une triangulation et le résultat - l'arbre des concepts linguistiques - a été soumis à une assez large palette d'acteurs impliqués dans l'action pour en corriger les erreurs et enrichir les aspects négligés (Huberman et Miles, 1991).

#### 4. MISE AU POINT DES RESEAUX SEMANTIQUES ETENDUS

Dotés d'un arbre des concepts linguistiques de la toxicologie nucléaire environnementale, il nous fallait ensuite imaginer une solution technologique pour favoriser la découverte croisée des connaissances. Ceci nécessitait la mise au point de plusieurs modèles de connaissances, un par thème d'intérêt, pour faciliter la discrimination. Nous avons rapidement évalué que la construction de modèles thématiques complets serait difficile à réaliser de façon manuelle.

Les premières expérimentations que nous avons réalisées de 2001 à 2003, nous avaient permis de découvrir une approche originale dans le traitement automatique des documents. La démarche consistait à utiliser simultanément

plusieurs techniques statistiques pour, à partir de la proximité des termes, de leurs cooccurrences, caractériser des documents.

Mais, considérant que nous appréhendons l'information comme née de la rencontre entre un individu doté d'un système de pertinence, d'une vision du monde, et des données préexistantes (Schutz, 1987), un traitement automatisé même très performant ne pouvait pas se substituer dans le processus de sélection de l'information à l'analyse d'un expert. Il n'était de notre point de vue pas possible de construire les modèles de connaissances thématiques en utilisant uniquement des calculateurs.

Aussi, nous avons décidé de tenter une approche nouvelle : la construction rapide de modèles conceptuels par des experts en situation, puis l'enrichissement de ces modèles conceptuels en prenant appui sur une cartographie de termes issus du traitement automatique d'un ensemble de documents soigneusement sélectionnés.

Exprimé autrement notre démarche consistait à privilégier la construction manuelle de réseaux de concepts d'une taille réduite. Chaque réseau décrivant de façon synthétique un thème d'intérêt. L'idée était de nous inspirer assez directement de la notion de schème proposée par Emmanuel Kant (Eco, 1997).

L'étape suivante consistait à enrichir de façon automatique chacun de ces réseaux de concepts à l'aide d'un ensemble de termes pertinents issu d'un calcul automatique. Le tout ayant pour objet d'obtenir assez rapidement, c'est à dire le temps nécessaire à un expert pour construire le modèle simplifié d'un thème d'intérêt, un modèle conceptuel détaillé. Modèle qui aurait été difficile et long à construire de façon manuelle.

## 5. MISE EN ŒUVRE DE LA DEMARCHE

### 5.1. Construction de « réseaux sémantiques »

Pour faciliter la construction manuelle de modèles conceptuels « schématiques », il nous semblait important d'imaginer un dispositif doté de règles d'utilisation traduisant de fait les règles de construction.

Pour cela nous nous sommes appuyés sur les travaux de Ross Quillian sur les réseaux sémantiques (Quillian, 1968), d'Eleanor Rosch sur les prototypes (Rosch, 1975), de Marvin Minsky sur la représentation des connaissances (Minsky, 1975) et de Grady Booch, James Rumbaugh et Ivar Jacobson sur la construction de modèles objets et le langage unifié de modélisation UML (Booch et al., 2005).

L'ensemble nous a permis de définir que les modèles conceptuels schématiques seraient représentés sous forme de réseaux. Chaque nœud du réseau

« sémantique » matérialisant un concept, et les concepts étant reliés entre eux par des relations.

Mais, il nous semblait nécessaire pour les usages futurs d'organiser le modèle conceptuel. Aussi, nous avons choisi de doter chaque réseau d'un point d'entrée matérialisant de fait le concept central (prototype). Nous avons ensuite opté pour une instanciation de la notion de relation afin que l'utilisateur (expert) puisse plus facilement préciser le type de relation entre les concepts en s'appuyant sur trois types de liens : composition (inclusion), héritage (dérivation) et association (usage). La valeur d'un concept, entendue comme sa contribution relative au thème d'intérêt, décroissant avec l'augmentation de la distance de ce concept au prototype (concept central).

Enfin nous avons adopté quelques règles complémentaires. Un réseau ne devra pas contenir plus de 50 nœuds (concepts). Le point d'entrée (concept central) sera doté d'une valeur de départ de 1. Un poids (compris entre 0 et 1) étant attribué à chaque type de relation (composition, héritage ou association) pour un réseau donné. Le tout permettant de calculer avec une grande précision la valeur d'un concept, sa contribution, dans un réseau. Cette valeur étant obtenue par agrégation du poids des liens reliant ce concept au prototype. Enfin, nous avons décidé de ne pas descendre en dessous d'une valeur de 0,5 pour un concept, ce qui avait pour objectif de limiter de fait la taille des réseaux.

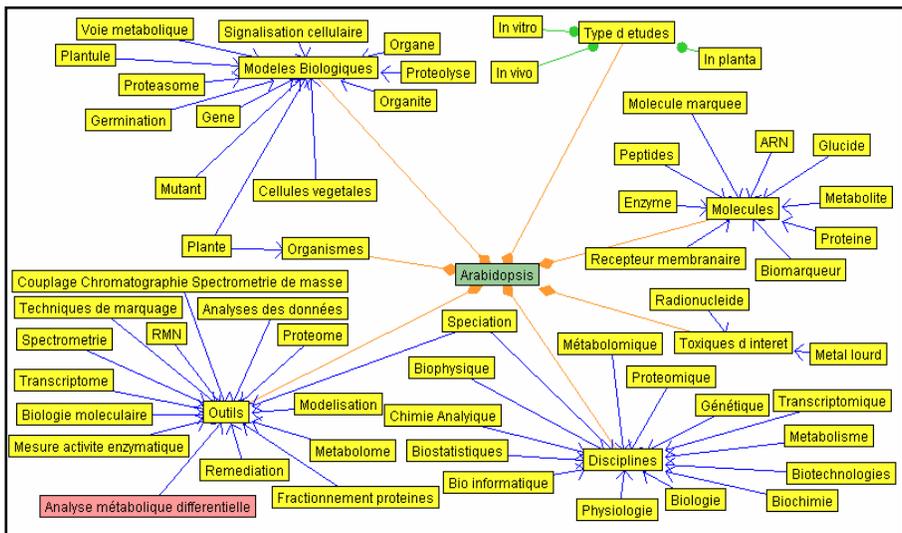


Figure 02 – exemple de réseau sémantique : Arabidopsis

Par exemple : la contribution relative d'un concept de niveau 3 à un thème d'intérêt, via deux relations de composition dotées d'un poids de 0,9, sera de 0,81 : soit  $0,9 \times 0,9$ .

A partir de ces règles nous avons élaboré un prototype d'éditeur de réseaux sémantiques. Cet éditeur permet à un expert de construire facilement le schéma d'un thème d'intérêt. Les réseaux mis au point par deux experts sur le même thème peuvent être différents et représentent de fait la perception d'un individu en situation à un moment donné. La valeur d'un concept est liée à sa position dans le réseau sémantique, aux types de relations utilisées, aux poids choisis pour chaque type de relation et au mécanisme général d'agrégation.

## **5.2. Construction de « réseaux de proximité »**

Le réseau sémantique permettant de décrire de façon simplifiée un thème d'intérêt, il convenait d'imaginer un dispositif pour en étendre la description.

Après avoir évalué plusieurs pistes, nous avons décidé de nous appuyer sur les travaux de Michel Callon sur les mots associés (Callon, 1983) et d'Alain Lelu sur le traitement neuronal de corpus documentaires (Lelu, 1993).

L'idée était de demander à un expert de sélectionner plusieurs articles ou textes pertinents sur un thème d'intérêt à un moment donné. Puis, d'utiliser des techniques de calcul pour identifier de façon automatique les termes et mettre en évidence et les relations entre ces termes : mots-clés considérés comme des « indicateurs de connaissances » (Grivel, 1995).

Pour préparer le calcul nous avons mis au point un prétraitement consistant à ne retenir dans les textes que les termes identifiés dans notre modèle de connaissance de la toxicologie nucléaire environnementale et mémoriser leurs fréquences d'apparition. Puis nous avons implémenté trois types d'algorithmes pour calculer les cooccurrences des termes : mots associés, K-means axiaux, et analyse en composantes principales.

Les trois techniques permettent de construire des grappes ou clusters dans l'espace euclidien et matérialisent de fait la proximité entre les différents termes. Un calcul nous permet ensuite de construire par agrégation un réseau de l'ensemble des termes. Chaque terme est un nœud, et les nœuds sont reliés entre eux par des relations d'association dotées d'un poids spécifique (normalisé entre 0 et 1) issu du calcul. Le réseau peut être très volumineux : les premiers essais nous ont conduits rapidement à des réseaux de plus de 800 nœuds. Un mécanisme complémentaire de seuillage nous permettant d'éliminer les liens d'un poids limité.

## **5.3. Construction des « réseaux sémantiques étendus »**

La troisième et dernière étape consistait à étendre les réseaux sémantiques. Le principe était de compléter les réseaux sémantiques en ajoutant des termes,



Nous avons présenté une série de résultats, de cartes sémantiques, lors d'un séminaire du comité de pilotage du programme rassemblant direction de programme et chefs de projet. Il s'agissait de résultats préliminaires non approfondis, mais les réactions ont été assez positives. Les scientifiques présents ont commentés les différentes cartes, trouvant des explications sur un certain nombre de points qui nous avaient échappés lors d'une première analyse des résultats. Par ailleurs certains points ont été confirmés. Ainsi, l'un des projets semblait manifestement être transversal. Les chercheurs nous ont expliqué que la dimension méthodologique de ce projet expliquait très bien ce positionnement. Certains individus étaient graphiquement éloignés de leurs collègues. Dans un bon nombre de cas les écarts ont pu être expliqués facilement.

Le travail réalisé sur ce projet de recherche est loin d'être achevé. S'il montre que la construction des connaissances peut apporter des éléments intéressants dans le processus de construction du sens, de nombreuses questions restent encore en suspend : Quel est le niveau de qualité des modèles de connaissances ainsi construits ? Quel peut être le rôle de la représentation des connaissances dans le processus plus général de construction du sens ?

Nous avons par contre montré que pour un usage ciblé, ici la recherche de documents scientifiques dans de grands réservoirs de données, il existe une alternative économique à la construction d'ontologies : la construction semi-automatique de réseaux sémantiques étendus.

## BIBLIOGRAPHIE

BACHIMONT B., Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances, In *Ingénierie des connaissances : évolutions récentes et nouveaux défis*, éditeurs CHARLET J. ZACKLAD M., KASSEL G. et BOURIGAULT D., Eyrolles, Paris, 2000.

BOOCH G., RUMBAUGH J. and JACOBSON I., *The Unified Modeling Language User Guide* (2nd edition), Addison-Wesley Professional, 2005.

BOUAUD J., BACHIMONT B., CHARLET J. and ZWEIGENBAUM P., Methodological principles for structuring an "ontology", In *proceeding of the IJCAI'95 workshop on "Basic Ontological Issues in Knowledge Sharing"*, Montréal, Canada, 1995.

BOURIGAULT D., AUSSENAC-GILLES N. et CHARLET J., Construction de ressources terminologiques à partir de textes : un cadre unificateur pour trois études de cas, *Numéro spécial Techniques informatiques et structuration de terminologies de la revue Intelligence Artificielle*, 18(1), 2004.

CALLON M., COURTIAL J.P., TURNER W.A. and BAULIN S., From Translation to Problematic Networks: An Introduction to Co-Word Analysis, *Social Science Information*, 22, 1983, 191-235.

CHARLET J., L'ingénierie des connaissances, entre science de l'information et science de gestion, Actes du colloque de Cerisy « Activité, connaissance, organisation » sous la direction de TEULIER R. et LORINO P., La découverte, Paris, 2005.

CHAUCHE J., PRINCE V., JAILLET S et TESSEIRE M., Classification automatique de textes à partir de leur analyse syntactico-sémantique, Actes du colloque TALN 2003, Batz-sur-Mer, 11-14 juin 2003.

DEWEY M., A Classification and Subject Index for Cataloguing and Arranging the Books and Pamphlets of a Library, Gutenberg Project, <http://www.gutenberg.org/etext/12513>, 1876.

ECO U., Kant et l'ornithorynque, Grasset, Paris, 1997.

GRIVEL L. et FRANCOIS C., Une station de travail pour classer, cartographier et analyser l'information bibliographique dans une perspective de veille scientifique et technique, Les sciences de l'information (sous la direction de Jean-Max Noyer), Presses Universitaires de Rennes, 1995, 81-112.

GRUBER T.R., A translation approach to portable ontology specifications, Knowledge acquisition, 5 (2), 1993, 199-220.

HUBERMAN A.M. et MILES M.B., Analyse des données qualitatives, Edition du renouveau pédagogique (traduction DE BACKER C. et LAMONGIE V.), Bruxelles, Belgique, 1991.

KUHN T.S., La structure des révolutions scientifiques, Flammarion, Paris, 1972.

LELU A., Modèles neuronaux pour l'analyse de données documentaires et textuelles, Thèse de Doctorat, Université Paris VI, mars 1993, 238 p.

MENAGER M.T., Programme Toxicologie Nucléaire Environnementale : comme fédérer et créer une communauté scientifique autour d'un enjeu de société, Colloque Intelligence Collective : partage et redistribution des savoirs, Site de Nîmes de l'Ecole des Mines d'Alès, 29-30 septembre 2004.

MINSKY M., A framework for representing knowledge, The psychology of computer vision, P. Winston editor, Mc Graw-Hill, New-York, 1975, 211-281.

OTLET P. et LA FONTAINE H., Classification décimale universelle : édition abrégée (7ème éd), Editions du Céfal, Liège, 2001, 292 p.

QUILLIAN M.R., Semantic memory, In Semantic Information Processing, Marvin Minsky editor, MIT Press, Cambridge, Massachusetts, 1968, 217-270.

RANWEZ S., RANWEZ V., SY M.F., MONTMAIN J. et CRAMPES M., Utilisation de proximités sémantiques pour améliorer la recherche et le rendu d'information, Actes des 21èmes Journées Francophones d'Ingénierie des Connaissances, Nîmes, 9-11 juin 2010, Presses des Mines, Collection Mathématiques et Informatique, Paris, 2010, 247-258.

RICCIO P.M., Technologie et Organisations, Presses des Mines, 2012, 106 pages.

ROSCH E., Cognitive Representations of Semantic Categories, Journal of Experimental Psychology, 104 (3), 1975, 192-233.

SCHUTZ A., Le chercheur et le quotidien, Méridiens Klincksieck, Paris, 1987.

YIN R.K., Case Study Research, Design and Methods, Sage, London, 1984.