

CONSTRUCTION D'UNE ONTOLOGIE DE DOMAINE FONDEE SUR LE TEXT MINING

Anass EL HADDADI (*, **), Bernard DOUSSET (*), Ilham BERRADA (**), Ismail KASSOU (**)

anass.el-haddadi@irit.fr ; dousset@irit.fr ; iberrada@ensias.ma , kassou@ensias.ma

(*) IRIT, SIG, Université Paul Sabatier, route de Narbonne, 31062 Toulouse cedex 04

(**) ENSIAS, Equipe Al BIRONI, Université Mohamed V – Souissi, B.P. 713 AGDAL, Rabat - Maroc

Mots clés :

Veille Stratégique, Fouille de Texte, Ontologie.

Keywords:

Strategic Watch, Text Mining, Ontology

Palabras clave:

Escudriñar estratégico, la minería de texto, ontología

Résumé:

Les ontologies sont des outils très puissants pour décrire sémantiquement de vastes domaines afin d'en faciliter le traitement automatique. Dans cet article, nous présentons notre approche basée sur la fouille de texte visant à construire une ontologie de domaine, dont la définition des relations de similarité entre les concepts est basée sur des algorithmes non supervisés. Après avoir modélisé le processus de Veille Stratégique, nous analysons les rôles possibles d'une ontologie dédiée à la tâche de veille et nous en détaillons la méthodologie de construction. Enfin, nous présentons les résultats d'analyse en se basant sur l'ontologie développée.

Abstract:

Ontology's are very powerful tools to describe semantically areas in order to facilitate automatic processing. In this article we present our approach based on the Text Mining to build a domain ontology, which the definition of relations of similarity between concepts is based on algorithms unsupervised. Having modeled the process of Strategic Watch, we analyse the possible roles of an ontology dedicated to the task of monitoring and we explain the methodology of construction. Finally, we present the results of analysis based on the ontology developed.

1 Introduction

Dix ans après le début de la libération du secteur des télécommunications au Maroc, les indicateurs sont au vert. Le chiffre d'affaires du secteur a enregistré une forte évolution au cours des dix dernières années, bien que la téléphonie et internet n'évoluent pas au même rythme. Le chiffre d'affaire est passé de 8,5 milliards de dirhams en 1999 à 30 milliards en 2007, l'équivalent de sept pour cent du produit intérieur brut (PIB).

La libéralisation entamée il y a dix ans dans le secteur des télécommunications a permis de passer d'un monopole à un marché concurrentiel, et a aidé le Maroc à devenir un leader régional dans ce secteur. Tout d'abord, on assiste à la convergence des secteurs de l'Informatique, des Télécommunications et des Médias avec un mode de transport fondé sur les réseaux Internet. Ensuite, le marché grand public et professionnel donne le ton en matière de technologie. Les premiers bouleversements sont apparus dans la téléphonie fixe avec les offres de la mobilité restreinte initiée par Maroc connecte. La suite de l'histoire se dénomme le 3G+, une technologie qui commence à investir le marché de la téléphonie mobile en mode nomade, la tendance est donnée. La touche finale viendra des réseaux Internet, qui concéderont la composante mobile à la voix sur Internet : l'écrasante majorité de la téléphonie mobile passera alors sur les réseaux Internet.

Autre fait surprenant, le marché des Nouvelles Technologies de l'Information et des Communications trouve actuellement son innovation dans les secteurs grand public et professionnels. Cet état de fait complexifie les offres à proposer, face à des consommateurs de plus en plus avertis. Ainsi, d'une part le client devient très volatil et très exigeant en maintenant une pression très forte sur les notions de services et de produits offerts, ainsi que sur la concurrence entre les opérateurs de télécommunication et les investissements destinés à améliorer le marché marocain.

Dans ce contexte, il nous semble indispensable de décrire la sémantique et l'organisation des objets du secteur des télécommunications afin de se doter d'une modélisation conceptuelle (non contextuelle et non ambiguë) dont le sens est inscrit dans la structure même du modèle, donc une « ontologie ». Notre hypothèse est que le développement de telles ressources ontologiques permettra de mettre au point des outils performants, fiables et maintenables pour l'aide au partage et l'utilisation de connaissances destinées à mieux gérer et améliorer le secteur de télécommunication. La problématique à laquelle nous nous attaquons est donc la construction de ces ontologies, et le terrain d'expérimentation dans lequel nous nous situons est l'aide au développement d'un outil informatique pour le partage et la gestion de connaissance du secteur des télécommunications.

Le terme ontologie est utilisé depuis le début des années 90 dans le domaine de l'intelligence artificielle, en particulier de l'ingénierie des connaissances et de la représentation des connaissances. Son champ d'application s'est considérablement élargi et il fait désormais partie des objets de recherche courants, notamment dans le secteur de la modélisation des systèmes d'information ou les recherches utilisant les ontologies sont de plus en plus nombreuses [1]. Une ontologie est un système formel dont l'objectif est de représenter un domaine spécifique au moyen d'éléments de base : les concepts définis et organisés les uns par rapport aux autres [2]. La représentation ontologique des connaissances garantit le maintien de la cohérence des axiomes et de l'intégrité du système,

ainsi que l'extensibilité de la représentation sans modification de la structure. Le concept d'ontologie en informatique repose sur la notion d'ontologie formelle développée en philosophie par Edmund Husserl : les ontologies formelles visent au développement d'une logique systématique, formelle et axiomatique afin de classer les entités du monde (objets, événements, etc.) et les catégories qui les modélisent (concepts, propriétés, etc.). L'ingénierie ontologique est une discipline de l'informatique référant aux activités reliées au processus de développement des ontologies ainsi qu'aux méthodes, outils et langages pour développer ces ontologies.

Nous ne prétendons en aucun cas de construire une ontologie universelle de l'industrie des télécommunications mais bien une ontologie de domaine du secteur des télécommunications au Maroc. Dans le cadre de notre recherche, notre objectif est double : il s'agit de construire une ontologie pour une application d'aide à la décision et à la représentation de connaissance et, pour ce faire, de développer une méthodologie de construction d'ontologie à partir de corpus textuels, fondée sur la sémantique différentielle. L'intérêt consiste à mettre au point un processus méthodologique très précis destiné à un ingénieur cognitif de manière à ne faire appel au décideur des télécommunications qu'à des moments précis de la validation.

2 Travaux connexe

Les travaux réalisés au cours de ces dernières années pour la construction d'ontologie à partir de fouille de texte ce sont concentrés sur l'utilisation de technique linguistiques basé sur la cooccurrence de mots et sur des algorithmes d'apprentissage non supervisée comme les k-moyens et la classification hiérarchique et semi supervisé comme « Hidden Markov Models » (HMM) et l'approche par des ensemble d'approximation ou encore « Expectation Maximization » (EM).

[3] a proposé « ENRICH », une méthode qui favorise l'apprentissage organisationnel au sein d'une entreprise. A l'aide des outils et des documents utilisés pour soutenir la collaboration et l'apprentissage, entre les différentes hiérarchies d'une organisation. Deux approches ont été principalement utilisées : une approche théorique qui se base sur la cooccurrence, pour construire l'ontologie. L'autre approche de validation de l'ontologie, qui est soutenu par l'outil WebOnto, qui permet la collaboration de l'édition et la visualisation des modèles par l'intermédiaire d'un navigateur web classique.

[4] spécifie un processus en trois étapes pour développer une ontologie de concepts à partir de documents textuels. Le processus comporte trois étapes : l'apprentissage des concepts, la validation de concept et la gestion de l'ontologie. A partir de World Net qui nous donne une première ontologie, le système applique un ensemble de règles sémantiques pour désambiguïser le sens des cooccurrences des mots et construire une ontologie finale.

[5] ont rapporté une étude comparative de la performance et de l'algorithme d'agglomération hiérarchique de regroupement pour les documents. [6] a proposé un schéma de regroupement hybride qui utilise la décomposition en valeurs singulières (SVD), suivie d'un algorithme de partitionnement k-moyens de text mining. Et [7] a proposé l'utilisation d'un modèle de tolérance des ensembles d'approximation TRSM pour le regroupement des résultats de recherche.

[8] a proposé un algorithme de classification de document semi supervisé basé sur l'approche EM. [9] propose un modèle probabiliste pour le regroupement semi supervisé fondé sur HMM.

3 Utilisation des ontologies pour la veille

Les activités d'analyse prennent en compte des descripteurs issus des documents ainsi que leurs méta-données. De nombreux travaux visent à intégrer de la connaissance dans le processus de veille pour exploiter ces données. Cette connaissance peut être représentée sous forme de ressources de plusieurs natures : ressources terminologiques (vocabulaire contrôlé, thésaurus) ou hiérarchies de concepts.

L'indexation des documents se fait à partir de ce vocabulaire contrôlé. Cette indexation a pour intérêt d'améliorer la classification et la cartographie des données par rapport à un système n'utilisant pas de lexique et basant l'extraction des index sur la recherche de patrons.

Un autre type d'approche consiste à indexer les documents à partir de hiérarchies de concepts. Une hiérarchie de concepts est formée des concepts d'un domaine organisés du plus générique ou plus spécifique. Un concept peut être défini à partir d'un ou plusieurs termes. Cette approche est utilisée dans [10]. Les documents sont annotés à partir de hiérarchies représentant les problèmes rencontrés par un utilisateur de télécommunication. Ces annotations permettent ensuite d'organiser les données et de les catégoriser en fonction des problèmes rencontrés. Cette approche est également utilisée dans [11]. Plusieurs hiérarchies sont alors utilisées pour indexer des documents suivant les différentes facettes du domaine abordé dans un corpus. Dans [12] les ontologies servent de support à la construction de hiérarchies de concepts. Chaque document de la collection est associé à un ensemble de concepts d'une ou plusieurs hiérarchies. Ces hiérarchies sont ensuite utilisées pour l'exploration graphique d'une collection.

Dans la majorité des cas, la connaissance est représentée par des ressources terminologiques. Dans le cas où la connaissance est représentée à un niveau conceptuel (hiérarchie de concept), la seule relation utilisée pour organiser les concepts est la relation de spécificité/généricité. Les travaux présentés dans l'article [13] reposent sur une ontologie dans laquelle le niveau lexical et le niveau conceptuel sont élaborés (concepts définis à partir de plusieurs termes, relations de plusieurs types). Ce type de représentation de connaissance a pour intérêt de permettre un niveau d'analyse plus fin à partir des relations sémantiquement plus riches entre les concepts. Les relations sont également formalisées. Des inférences peuvent alors permettre d'extraire de nouvelles connaissances. Par exemple, si la relation transitive lie les concepts A et B puis les concepts B et C, une inférence permettra d'établir que A et C sont également reliés par R.

3.1 Rôle de l'ontologie pour la tâche de veille

L'exploitation de l'ontologie pourrait être utile au cours de plusieurs phases du processus de veille :

- **Dans la phase de ciblage** : l'utilisation des concepts dans l'ontologie peut aider à éviter l'ambiguïté du contexte de recherche, en permettant de spécifier la formulation de la requête. Une telle ambiguïté peut être liée à des phénomènes linguistiques comme la synonymie ou l'homonymie.
- **Dans la phase de sélection** : l'ontologie peut être utile pour associer les documents aux sujets spécifiques.
- **Dans la phase de diffusion** : selon le centre d'intérêt de l'utilisateur, décrit en utilisant l'ontologie, le système pourra automatiquement envoyer à un utilisateur donné des suggestions pour l'inciter à lire certains documents semblant pertinent pour cet utilisateur.

4 Approche de construction d'ontologie de domaine

4.1 Présentation de corpus d'analyse d'ontologie Télécom

Le développement des ressources servant à construire l'ontologie passe, tout d'abord, par la construction d'un corpus textuel. Dans le but de couvrir, avec le plus d'exhaustivité et d'objectivité possible, l'ensemble de l'activité du secteur des télécommunications, nous nous sommes servi de tout les rapports et études publiés officiellement en 2007 par l'Agence National de Règlementation des Télécommunications (ANRT) qui a pour objectif central de réunir les conditions pour favoriser la dynamique de développement du secteur des télécommunications au Maroc. Au total nous disposons de 14 rapports et études. Ce premier corpus correspond à 134 054 mots. Pour construire l'ontologie à partir du corpus ANRT, nous nous sommes basés sur les résultats fournis par l'analyseur de SPSS « Text Mining pour Clementine 5.0 » (TMC 5.0) qui utilise des technologies linguistiques et d'analyse du langage naturel pour traiter la diversité des données textuelles non structurées à partir de corpus ANRT, et extraire et organiser les concepts-clés. Le module « catégorisation des données textuelles » permet de regrouper ces concepts en catégories qui servent ensuite à la construction de l'ontologie. Puis, le module « analyse des classes » permet un regroupement de concepts liés, généré par des algorithmes de classification non supervisée qui se fondent sur la fréquence d'apparition de ces concepts dans l'ensemble de corpus ANRT et la fréquence d'apparition conjointe des concepts dans un même document du corpus (ou cooccurrence). Chaque concept d'une classe est en cooccurrence avec au moins un autre concept de la classe. Les classes ont pour objectif de déterminer les relations qui existent entre les concepts apparaissant ensemble, alors que les catégories ont pour objectif de regrouper les concepts dans une même entité homogène.

4.2 Méthode de construction d'ontologie Télécom

La méthodologie mise en œuvre permet de décrire les variations du sens des termes considérés dans un même contexte [14]. C'est pourquoi cette méthode considère que le corpus textuel est la source privilégiée permettant de caractériser les notions utiles à la modélisation ontologique et le contenu sémantique qui leur est associé. [15] identifient différentes tâches génériques reliées à la construction d'ontologie dans le contexte de l'extraction semi-automatique de connaissances à partir de corpus textuels non structurés. La figure 1 illustre l'organisation de ces tâches:

1. identification des termes pertinents pour l'ontologie développée;
2. identification des relations de synonymie entre les termes;
3. identification des concepts et de leurs attributs;
4. identification des relations taxonomiques (structure de la hiérarchie) entre concepts;
5. identification des relations non taxonomiques (génériques) entre concepts;
6. identification des règles spécifiant les contraintes sur les propriétés des concepts et les relations entre concepts.

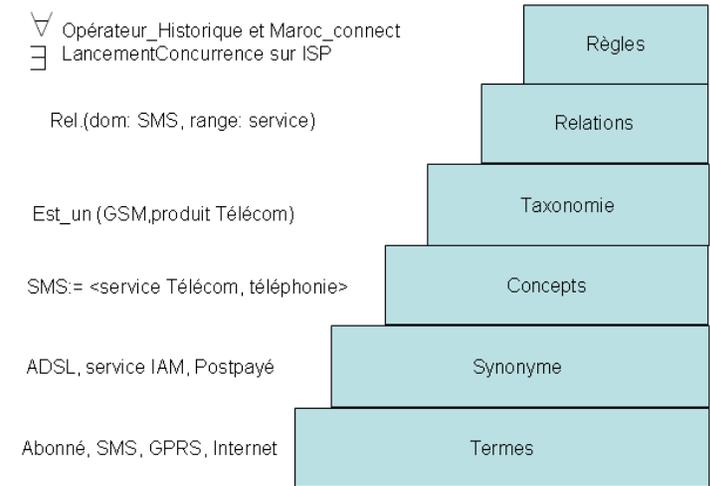


Figure 1 - Tâches liées à la construction d'ontologie

4.3 Construction de l'ontologie phase 1 - définition des catégories et des classes de concepts

Les résultats fournis par l'outil de SPSS TMC 5.0 sur la base du corpus ANRT servent de support dans le choix de catégories de concepts représentatifs du secteur des télécommunications. Nous distinguons deux étapes dans la sélection de ces candidats termes :

- A. Nous parcourons l'ensemble des résultats fournis par TMC 5.0 et choisissons d'étudier, en premier lieu, les concepts dont la fréquence d'apparition dans le corpus est supérieure à 5%. Nous identifions les grands axes conceptuels du domaine des télécommunications dénommés catégorie de concepts, soit la concurrence, l'investissement, le parc des télécommunications, le patrimoine du secteur des télécommunications en tant que secteur d'activité économique, les opérateurs des télécommunications existant sur le marché et enfin les services de télécommunication existant et fixons, pour chaque concept jugé pertinent, un indice de classification correspondant à l'un de ces catégories sur une échelle allant de 1 à 6. Par exemple, nous décidons que l'indice de classification représentant la catégorie concurrence est 6 et fixons également à 6 l'indice de classification pour tous les

candidats termes se rapprochant de cette catégorie conceptuel – e.g. concurrence loyale , concurrence ouverte, litige, etc. Cet indice de classification est un premier moyen mis en oeuvre pour rapprocher les concepts des catégories conceptuels. La figure 2 montre un exemple de catégorie de concept avec un extrait de l'ensemble de concept lié à la catégorie services des télécommunications.

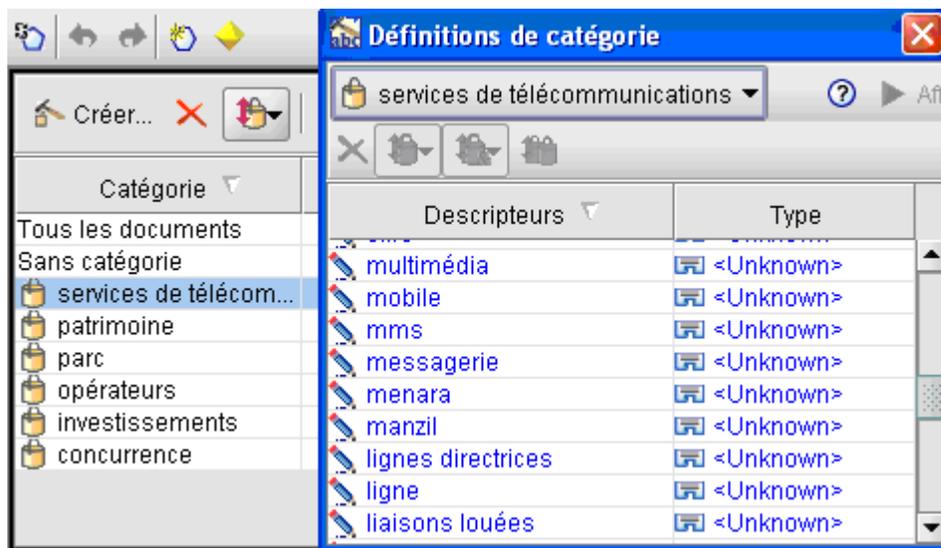


Figure 2 – Les catégories de concepts extraites à partir du corpus ANRT

- B. Nous étudions ensuite les résultats d'analyse distributionnelle définis par le module « analyse des classes » et définissons des groupes de concepts sémantiquement proches les uns des autres pour structurer la hiérarchie de l'ontologie. La figure 3 représente la relation de similarité qui existe entre les concepts liée à Internet, ce schéma illustre la sémantique du concept Internet par rapport à d'autres concepts qui sont liés à un accès au service ADSL avec une ligne liée à un ordinateur et un débit lié au réseau Internet.

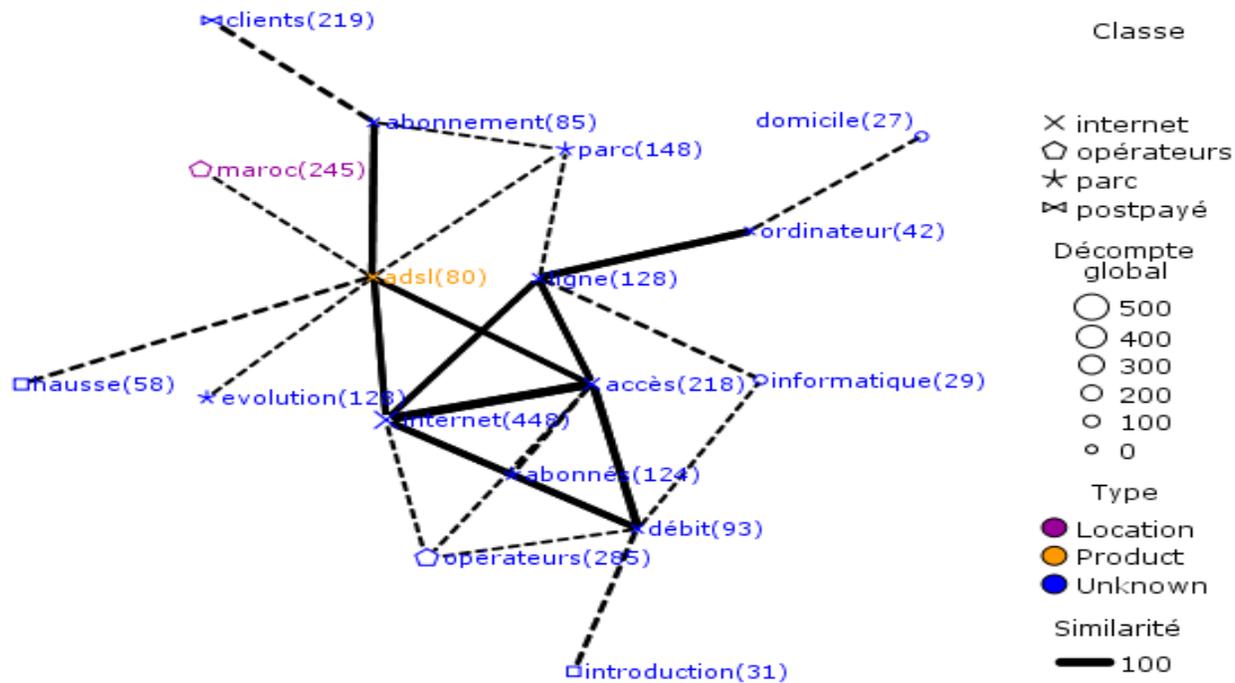


Figure 3 - Relations de similarité entre concepts Internet, opérateurs, parc et Postpayé

4.4 Construction de l'ontologie phase 2 - mise en œuvre des principes différentiels

Pour élaborer cette hiérarchie, il convient d'articuler les candidats termes choisis dans la phase 1 en précisant les principes différentiels qui les définissent. Par exemple, le concept mobilité restreinte et le concept téléphonie fixe sont des concepts frères dont le concept père est téléphonie. Au cours de la construction de l'ontologie on doit décider si le concept doit être primitif ou défini. Ce choix est fait en fonction de la tâche à réaliser, ici une représentation des connaissances du secteur des télécommunications au Maroc. Il existe donc une certaine objectivité. Formellement, un concept primitif est un concept indispensable à la représentation du domaine. Il s'agit d'une « brique » conceptuelle grâce à laquelle il est possible de construire d'autres représentations. Un concept défini est

5 Système de veille stratégique guidée par une ontologie de domaine

Après le développement de l'ontologie de domaine on va l'introduire comme Input du système de veille stratégique Tétralogie de l'I.R.I.T (figure 5) [16], [17], [18]. Qui est un des éléments essentiels de la station bibliométrique "ATLAS" élaborée conjointement grâce aux aides du CEDOCAR et du SGDN et qui fait intervenir de nombreux partenaires français (CNRS, CRRM, IRIT, Ministères,...) afin de proposer sur un même support l'ensemble des méthodes qui sont opérationnelles à l'heure actuelle dans ce domaine.

TETRALOGIE est essentiellement composé de deux modules:

- Un système de manipulation des corpus permettant de gérer les fichiers issus de téléchargements ou d'interrogations sur CD et d'en extraire, par divers outils, des matrices de croisement en tenant compte des spécificités de chaque base et de chaque format.
- Un système d'analyse des informations contenues dans les matrices, qui s'articule autour d'un tableur 3D spécifique et qui fait intervenir des techniques d'analyse des données classiques ou innovantes sur des domaines statiques, bipartis ou évolutifs. Ce système est complété par un module d'aide à l'élaboration des comptes rendus d'analyses facilitant la gestion des résultats numériques et graphiques générés par les diverses méthodes qui sont proposées.

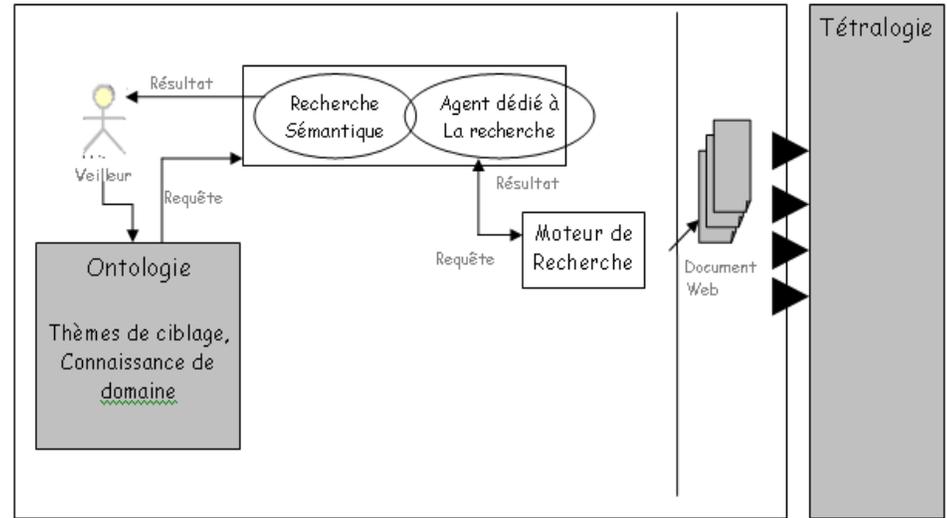


Figure 5 – Le système de veille Tétralogie guidée par une ontologie de domaine

6 Résultat d'analyse

On se basant su l'ontologie télécom, on effectue une recherche sémantique dans la base INSPEC (Figure 6). Et comme résultat on a trouve 9514 articles correspondant à la requête « Morocco, GSM, UMTS, 3G, ADSL » qui sont des concepts de l'ontologie télécom.

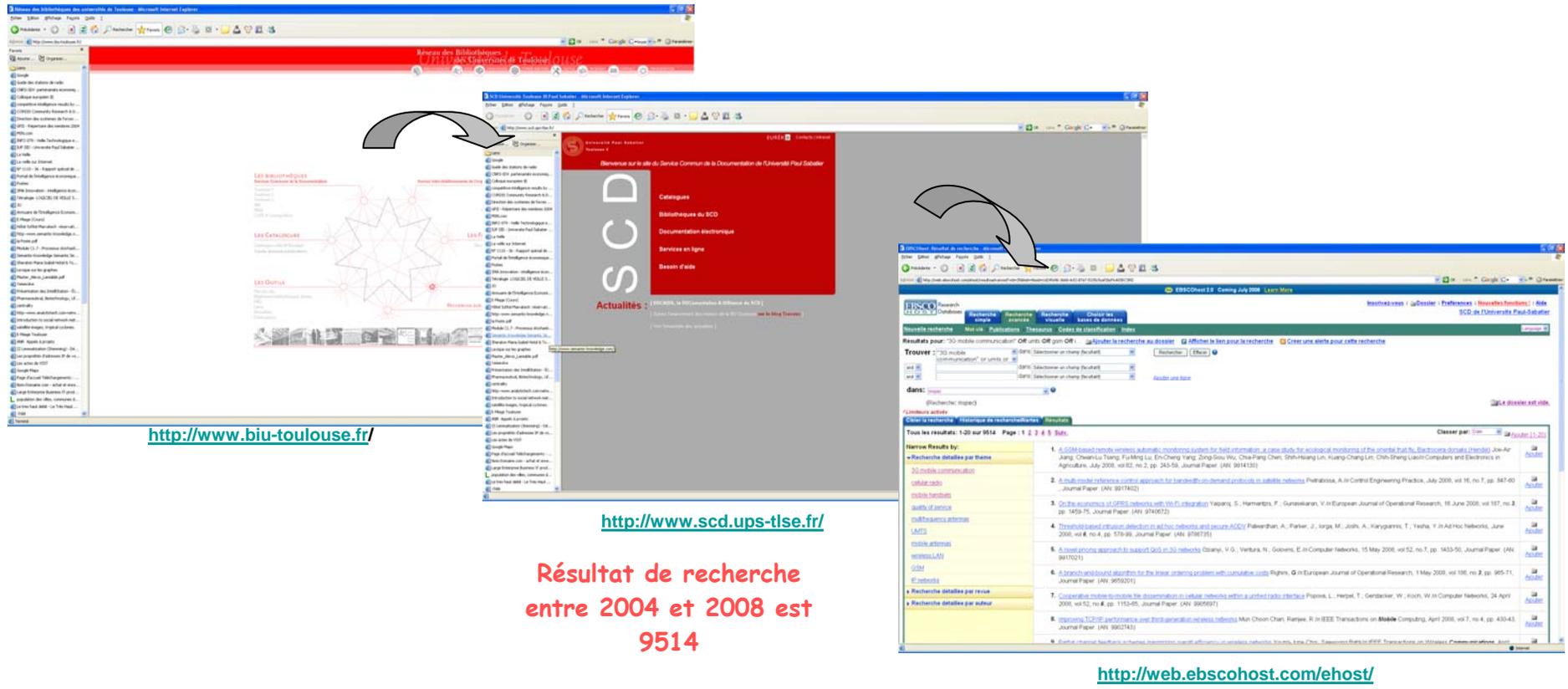


Figure 6 – Résultat de recherche dans INSPEC en se basant sur l’ontologie Télécom

Nous allons exploiter les champs de la recherche en les croisant entre eux pour en faire ressortir des informations. Pour réaliser ce croisement, nous créons une matrice de cooccurrence avec les deux ou les trois champs souhaités en utilisant les options tri absolu, relatif ou par connectivité et retirant les lignes vides. On fait un croisement AU-AU dans l’objectif de suivre les équipes de recherche à travers les collaborations qui existent entre chercheurs. Ces réseaux formels permettent de découvrir les alliances cachées, de détecter les équipes isolées, de saisir les transferts de technologies, etc.

Après extraction des données nous avons récolté plus de 18530 auteurs. Une première simplification du problème a consisté à éliminer tout les auteurs qui n'ont publié qu'une seule fois. Ceci a ramené le nombre d'auteurs aux environs de 2515. Du point de vue recherche de réseaux de collaboration cette première simplification de prêt pas à conséquence puisqu'un auteur qui n'a publié qu'une seule fois ne pourra jamais être l'intermédiaire entre deux auteurs qui ne sont pas déjà en relation.

Une application du tri par connexité nous perme de constate que l'auteur Lorenz a le plus grand nombre de publications en secteur de télécommunication. En plus il est au centre du plus grand équipe de recherche en ce domaine (Figure 7).



Figure 7 – Tableau 2D et Zoom après connexité (Croisement AU-AU)

Pour connaître les relations entre pays en terme de recherche dans le domaine de télécommunication, une classification PA-PA demeure indispensable (figure 9). On constat que le Maroc travail uniquement avec la France. Ainsi, la plupart des chercheurs marocains publient depuis des laboratoires français. A cause des collaborations qui existe entre les différentes universités marocaines et françaises.

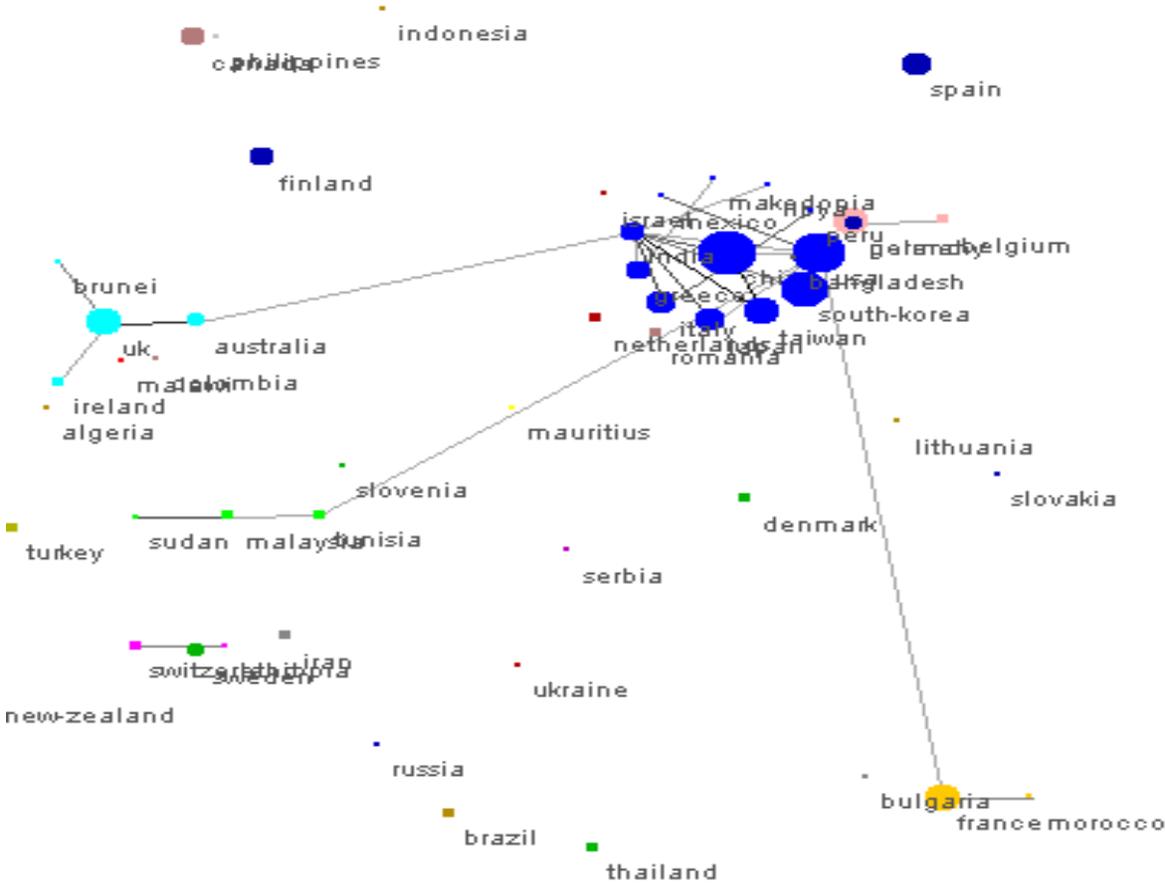


Figure 9 - Croisement PA-PA

Conclusion et perspectives

L'objectif de ce travail et de valider l'intérêt des ontologies pour représenter et gérer les compétences et les connaissances d'un domaine spécifique. Ainsi, la méthodologie mise au point permet à un chercheur, non spécialiste du domaine modélisé, de construire une ontologie à l'aide des techniques de fouille de texte. De plus, nous avons cherché à élaborer une organisation cohérente et robuste des concepts du domaine.

La dernière et indispensable étape de la validation de l'ontologie se fera par l'usage. L'ontologie doit être mise à disposition des acteurs du secteur des Télécommunications au Maroc, via un environnement d'aide à la décision orienté partage et gestion des connaissances.

Nous comptons développer une solution pour l'analyse et la synthétisation de grands volumes de textes pour les opérateurs des Télécommunications. En vue d'extraire des informations qu'il aurait été difficile de trouver sans une analyse automatique et systématique de ces gros volumes de données.

7 Bibliographie

- [1] **STAAB S., STUDER R.**, *Hand book on Ontologies*, Springer, Berlin, Allemagne, 2003.
- [2] **GRUBER T.R.**, *Towards principles for the design of ontologies used for knowledge sharing*. In Roberto Poli Nicola Guarino, editor, International Workshop on Formal Ontology, Padova, Italy, 1993.
- [3] **MULHOLLAND P., ZDRAHAL Z.**, *A Methodological Approach to Supporting Organizational Learning*, Int'l Journal of Human-Computer Studies 55, 2001.
- [4] **VELARDI P., MISSIKOFF M., FABRIANI P.**, *Using Text Processing Techniques to Automatically Enrich a Domain Ontology*, Proceedings of ACM Conference on Formal Ontologies and Information Systems, Ogunquit, Maine, 2002.
- [5] **ZHAO Y., KARYPIS G.**, *Hierarchical Clustering Algorithms for Document Datasets*, Data Mining and Knowledge Discovery, 10, pp. 141-168.
- [6] **KOGAN J., NICHOLAS C., VOLKOVICH V.**, *Text Mining with Hybrid Clustering Schemes*, Workshop on Text Mining, Third SIAM International Conference on Data Mining (SDM 2003), pp.5-16, 2003.
- [7] **NGO C. L., NGUYEN H. S.**, *A Tolerance Rough Set Approach to Clustering Web Search Results*, in LNCS-3202/2004, PKDD-2004, pp. 515, 2004.
- [8] **RIGUTINI L., MAGGINI M.**, *A Semi-Supervised Document Clustering Algorithm Based on EM*, Proceedings of Web Intelligence, WI-2005.
- [9] **BASU S., BILENKO M., MOONEY R. J.**, *A Probabilistic Framework for Semi-supervised Clustering*, Proceedings of the 10th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD-2004), Seattle, WA.
- [10] **GIVEL L., GUILLEMIN-LANNE S., COUPET P., HUOT C.**, *Analyse en ligne de l'information : une approche permettant l'extraction d'information stratégique basée sur la construction de composants de connaissance*, VSST, pp 149-161, from Text: An Overview. In Ontology Learning from Text: Methods, Evaluation and Applications, Frontiers in Artificial Intelligence and Applications 2001

- [11] **AUGE J., ENGLMEIER K., HUBERT G., MOTHE J.**, *Classification automatique de textes basée sur des hiérarchies de concepts*, VSST, pp 291-300, 2001
- [12] **AUSSENAC-GILLES N., MOTHE J.**, *Ontologies as Background knowledge to Explore Document Collections*, RIAO, pp 129-142 2004
- [13] **HERNADEZ N., MOTHE J.**, *Ontologies pour l'aide à l'exploration d'une collection de documents*, VSST, pp 405-416, 2004-
- [14] **BACHIMONT B., ISAAC A., TRONCY R.** *Semantic Commitment for Designing Ontologies: A Proposal*. In A. Gomez-Pérez & V. Benjamins, Eds., 13th EKAW'02, vol 2473 of Lecture Notes in Artificial Intelligence, Springer Verlag,. 114–121, Sigüenza, Espagne.
- [15] **BUITELAAR .P, CIMIANO P., MAGNINI B.**, *Ontology Learning Series*, Vol. 123, IOS Press, pp. 3–12, 2005.
- [16] **DOUSSET B., BENJAMAA T.**, *Trilogie logiciel d'analyse de données, conférence sur les systèmes d'information élaborée : Bibliométrie – Informatique stratégique – veille technologique*, 1988, pp. 36-45
- [17] **DKAKI T.**, *Outils mathématique et méthodes automatiques pour la veille technologique*, Thèse de Doctorat, Université Paul Sabatier-IRIT, France, Aout 1993
- [18] **CHRISEMENT C., DKAKI T., DOUSSET B., MOTHE J.**, *Extraction et synthèse de connaissances à partir de bases de données hétérogènes.*, Ingénierie des systèmes d'information, vol. 5, n° 3, 1997, pp. 367-400.