Une Ontologie pour l'Indexation et la Recherche d'Information Multilingue

Hassina Aliane[1], Souhila Boucham[2]
haliane@hotmail.com, sboucham@usthb.dz
[1] Division Recherche et Développement en Sciences de l'Information, CERIST, Ben Aknoun , Alger.
[2] Laboratoire Systèmes Informatiques, USTHB, Alger, Algérie.

ملخص

نقدم في هذا المقال مقاربة للفهرسة الآلية و استرجاع المعلومات على مدونة ثلاثية اللغة عربي، فرنسي، انجليزي. و يعتمد النظام المقترح نموذج المخططات الدلالية لتمثيل المعلومات كما تتحمل هذه المخططات انطولوجيا لمجال معرفي مختار.
تمثل الوثائق و طلبات المستخدمين على حدسواء وفق نفس النموذج.
تشكل الأنطولوجيا نواة النظام بحيث أنها تستخدم للفهرسة و كذلك للاسترجاع. و يعتمد نظام الفهرسة طريقة المقاطع المتكررة بالإضافة الى قواعد لسانية. أما نظام البحث و الاسترجاع فيعتمد مقارنة المخططات الدلالية للطلبات و الوثائق من أجل الحصول على الوثائق الأكثر تلبية لحاجة المستخدم.

Résumé

Nous proposons dans cet article une approche pour l'indexation et la recherche d'information pour un corpus trilingue: arabe, français et anglais. Le système proposé est fondé sur un formalisme de représentation de connaissances, plus précisément les graphes sémantiques [4] qui supportent une ontologie de domaine. Les documents et les requêtes sont aussi représentés dans ce formalisme. L'ontologie du domaine constitue le noyau du système et est utilisée aussi bien pour l'indexation que pour la recherche. Le système d'indexation utilise une méthode d'extraction qui est basée sur le calcul de segments répétés en utilisant des filtres linguistiques. Quant au système de recherche, il est fondé sur la comparaison de graphes de requêtes et de graphes de documents.

1. Introduction

Dans un système de recherche d'information multilingue, un utilisateur exprime sa requête dans sa langue de travail et obtient en réponse tous les documents pertinents non seulement dans sa langue mais dans toutes les autres langues du corpus.
Par ailleurs, un système de recherche d'information multilingue doit aussi faire face au problème de la représentation du contenu des documents ainsi qu'au problème de l'évaluation de la pertinence. Cette évaluation est plus difficile que dans un système de recherche d'information monolingue, en effet il est difficile de construire une fonction de correspondance avec différents langages pour les documents et la requête [1].
Notre objectif dans ce travail est de développer une ontologie qui supportera l'indexation, la recherche et l'extraction d'information. De là, la conception et le développement de notre système procède en deux étapes : la première construit en collaboration avec un expert humain une ontologie fondée sur le formalisme des graphes sémantiques et qui explicite les concepts du domaine et leurs relations.
Dans une seconde étape, cette ontologie est considérée comme un bootstrap qui initialise le système de connaissances au sens de Pitrat [2], alors, le processus d'indexation est basé sur une méthode linguistique, plus précisément un algorithme d'extraction de segments répétés.
La première partie de l'article présente un état de l'art en recherche d'information et indexation multilingue. La seconde section décrit le noyau de notre système, la troisième section décrit le système d'indexation et enfin la dernière section décrit le système de recherche.

I

2. Etat de l'art en Recherche d'Information Multilingue

Le principe de base de tout système de recherche d'information repose sur la correspondance entre une requête et des documents [3]. La qualité du processus d'indexation est fondamentale pour la qualité du système de recherche. Le processus d'indexation ou la sélection des entités décrivant le mieux les documents devient plus complexe en recherche d'information multilingue et doit transiter par une étape de « traduction » dans le but de représenter aussi bien les documents que les requêtes dans le même espace

d'indexation. Le paragraphe ci-dessous présente les approches connues en recherche d'information multilingue.

2.1 Approches basées sur la traduction des documents

La première approche consiste à traduire tous les documents d'un langage source en un langage cible. Ceci est très coûteux en termes d'effort et de temps. De plus cela conduit à l'augmentation de l'espace de stockage pour chaque document et ses traductions. L'alternative d'utiliser des logiciels de traduction ne donne pas des résultats de qualité suffisante [4].

2.2 Approches basées sur la traduction de la requête

Cette approche est souvent préférée par les chercheurs car traduire des requêtes est plus facile que traduire des documents du moment que les requêtes sont généralement constituées de simples mots de la langue naturelle. Néanmoins, la non prise en compte du contexte dans une analyse basée sur les mots conduit à des interprétations erronées et augment le bruit généré par le système de recherche [4].

2.3 Approches basées sur un langage pivot

Au vu des limites des précédentes méthodes, d'autres approches proposent d'aborder le problème de la recherche d'information multilingue en utilisant un langage pivot. Le langage pivot permet de représenter aussi bien les documents que les requêtes en utilisant des entités d'indexation indépendantes du langage. Cela implique la « traduction » des documents et des requêtes dans le langage pivot, figure 1, [4].

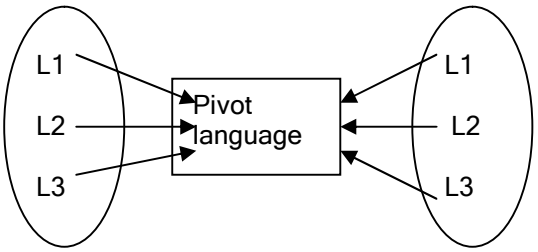


Figure 1. approaches basées sur un langage pivot

2.4 Approches à base de connaissances

L'indexation à base de connaissances désigne la classe de méthodes d'indexation dont l'objectif n'est pas d'identifier l'information contenue dans les documents mais plutôt de caractériser la connaissance associée à ces documents [4]. Elles sont basées sur des formalismes de représentation de connaissances comme les réseaux sémantiques, les graphes conceptuels et les graphes sémantiques [4].

2.5 Les 0ntologies

La formalisation de la connaissance commence par une conceptualisation [5] qui consiste en un ensemble d'objets, concepts et autres entités sur lesquelles la connaissance est exprimée (appelé univers du discours) et les relations entre eux.
Toute base de connaissances est basée sur une conceptualisation. Une spécification explicite de cette conceptualisation est appelée une Ontologie [6].
Le choix d'une ontologie détermine ce qu'un système peut connaître et sur quoi il peut raisonner. Formellement, une ontologie consiste en des termes, leurs définitions et les axiomes les reliant. Chaque ontologie spécifique dépend fortement de l'univers du discours considéré ainsi que des inférences escomptées [6].
Il y a une différence importante entre la notion d'ontologie et le formalisme qui va représenter cette ontologie. Plusieurs formalismes peuvent être utilisés pour exprimer la même structure ontologique : les

frames, le calcul des prédicats, les réseaux sémantiques, les graphes conceptuels, un formalisme étant plus expressif que d'autres [6].

2.6 Approches basées sur l'extraction de l'information

Extraire de l'information à partir de grands corpus textuels constitue l'un des enjeux majeurs pour l'indexation automatique et consiste à implémenter des algorithmes qui identifient automatiquement des éléments significatifs à partir de textes qui véhiculent des notions clés du domaine.
En indexation à base de connaissances et recherche dirigée par les ontologies, les algorithmes d'extraction d'information sont souvent utilisés pour indexer de nouveaux documents étant donnée une base de connaissance initialement existante ou une ontologie. Les résultats du processus d'extraction sont alors proposés pour mettre à jour la base de connaissances.
Les algorithmes statistiques sont les plus populaires parmi les algorithmes d'extraction de l'information, cependant, bien qu'il soit généralement admis que la prise en compte des occurrences simultanées de termes est pertinente, ces méthodes manquent de précision : elles ne peuvent prendre en compte ni le contexte ni la sémantique des termes ni encore les relations qui existent entre les termes.
Les méthodes linguistiques viennent pour apporter des solutions à de tels problèmes en implémentant une analyse morpho- syntaxique et parfois même une analyse sémantique. Néanmoins, construire des analyseurs robustes de la langue naturelle est coûteux et il n'est pas réaliste de construire un analyseur pour chaque langue [7]. Les approches actuelles sont plutôt fondées sur une analyse de surface utilisant des schémas spécifiques pour la sélection des concepts en combinaison avec une méthode statistique [7].

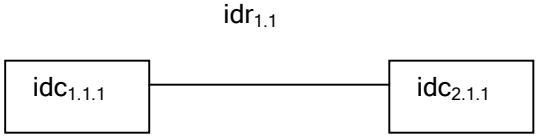
3. Un modèle basé sur les ontologies

Notre approche est centrée autour d'une ontologie du domaine qui est fondée sur le formalisme des graphes sémantiques.
La première caractéristique de notre système consiste en une conception basée sur l'interaction homme-machine. En effet, le rôle de l'expert humain est primordial pour atteindre les objectifs de performance du système et cela en tant que gestionnaire de ressources de connaissances qui peuvent varier en « intelligence » ou « en puissance » [8]. Dans cette étape, des interfaces hautement visuelles sont offertes pour l'expert humain en vue de la construction de son thésaurus sémantique. Le système est indépendant du domaine dans le sens que différents experts peuvent créer différentes bases de connaissances.
L'architecture générale du système est décrite par la Figure 2.

3.1 Le formalisme des graphes sémantiques

Les graphes sémantiques ont été proposés par Roussey [4] pour améliorer la description sémantique des documents dans un contexte multilingue. C'est un formalisme basé sur les graphes conceptuels de Sowa [8]. En effet, les graphes conceptuels se sont montrés aptes à modéliser de nouvelles applications de la RI, comme les systèmes de RI hypermédia ou multimédia. Dans le formalisme des graphes conceptuels, une opération de spécialisation peut être calculée sur les CGs par un opérateur de projection. Dans les graphes conceptuels, un opérateur de projection permet de trouver des composants spécifiques d'un graphe dans un autre graphe. Cette opération est utilisée dans les systèmes de RI pour construire une correspondance entre un graphe de requête et un graphe de document. Lorsqu'un graphe de requêtes est projeté avec succès dans un graphe de document, alors le document est pertinent pour la requête [4] [9]. De cette manière, le processus de recherche est effectué de manière abstraite sur le contenu plutôt que sur les termes.
Néanmoins, l'opérateur de projection présente quelques limites pour les systèmes de RI. D'abord, un système de RI doit pouvoir être en mesure de classer ses réponses par ordre de pertinence alors qu'une projection retourne un résultat booléen. Ensuite, une projection est une opération précise et les documents pertinents retournés sont ceux là qui satisfont exactement la requête ce qui élimine des documents pertinents qui satisfont la requête d'une manière partielle. Ceci a conduit les chercheurs à proposer des extensions au formalisme des graphes conceptuels qui étendent l'opération de projection pour aller au-delà des limites citées.
Les graphes sémantiques proposent d'améliorer la sémantique des documents dans un corpus multilingue en instanciant le modèle des CGS dans le but de construire un formalisme qui serait proche des langages style

thésaurus, ce modèle [4] représente la première tentative pour créer un langage de représentation de documents multilingues basés sur les graphes.
Pour améliorer l'opérateur de projection, un opérateur de projection étendue est construit. Cet opérateur définit une fonction de comparaison entre des graphes sémantiques. Il existe une projection d'un graphe *H* dans un graphe *G* si l'information représentée par *H* est a sémantiquement proche d'une partie de l'information représentée par *G*, on dit alors que H est "comparable" à *G*. [4]
Formellement un graphe sémantique est défini par: *SG*= (*C, A, lab)* [4]
- *C* est l'ensemble non vide des sommets concepts de *SG,*
- $A \subset C$xC l'ensemble des arcs de *SG* , dans le contexte de la RI, un arc *a* représente un couple de concepts *(c, c')* Č *C* ò *C*
- *lab* étiquette les noeuds et les arcs de la manière suivante pour l'exemple de la RI:
a. à tout concept *c* Č *C,* est associée une étiquette qui est une clé numérique *lab(c)* Č T_C (appelée aussi identifiant de *c*.
b. à tout arc *a* Č *A* est associée une étiquette *lab(a)* Č T_R qui est l'identifiant de *a*.
La figure suivante représente un exemple de graphe sémantique avec un seul arc *a* et deux concepts *c, c'*, étiquetés par leur identifiant: *lab(c)= idc*$_{1.1.1}$, *idc*$_{2.1..1}$



3.2 L'Ontologie du domaine

Pour prendre en charge le problème de la langue, nous avons vu plus haut que les systèmes de RI utilisent un langage pivot comme base de l'indexation. Aujourd'hui, les ontologies connaissent un réel succès parmi la communauté des chercheurs en RI. C'est pourquoi nous avons choisi de construire notre langage pivot comme une ontologie de domaine fondée sur le formalisme des graphes sémantiques.
Les graphes sémantiques sont basés sur le formalisme des graphes conceptuels et se focalisent sur les couples de concepts et les relations entre les concepts. Notre ontologie est constituée d'une conceptualisation du domaine appelée support *S* et un ensemble de vocabulaires associés *V*. La conceptualisation formelle de *S* est lisible par un humain en remplaçant les descripteurs formels de *S* par des mots du vocabulaire associé à la langue considérée.
Le support ou conceptualisation du domaine comprend deux hiérarchies: la hiérarchie de types de concepts et la hiérarchie des types de relation. Ces types ne sont pas des termes et donc ne font pas référence à une langue spécifique. Les vocabulaires sont utilisés pour décrire les types dans chacune des langues considérées pour notre corpus: français, anglais, arabe.
Cette ontologie basée sur les graphes sémantiques représente le noyau de notre système. Les experts de domaine créent et mettent à jour le modèle du domaine (avec les vocabulaires associés) à travers des interfaces interactives. Le système d'indexation contribue à enrichir l'ontologie. Alors que la conceptualisation du domaine peut être modifiée et mise à jour uniquement par les utilisateurs experts, les vocabulaires évoluent selon la terminologie des documents du corpus.
Les documents aussi bien que les requêtes sont représentées dans le formalisme des graphes sémantiques qui prend en considération les relations sémantiques.

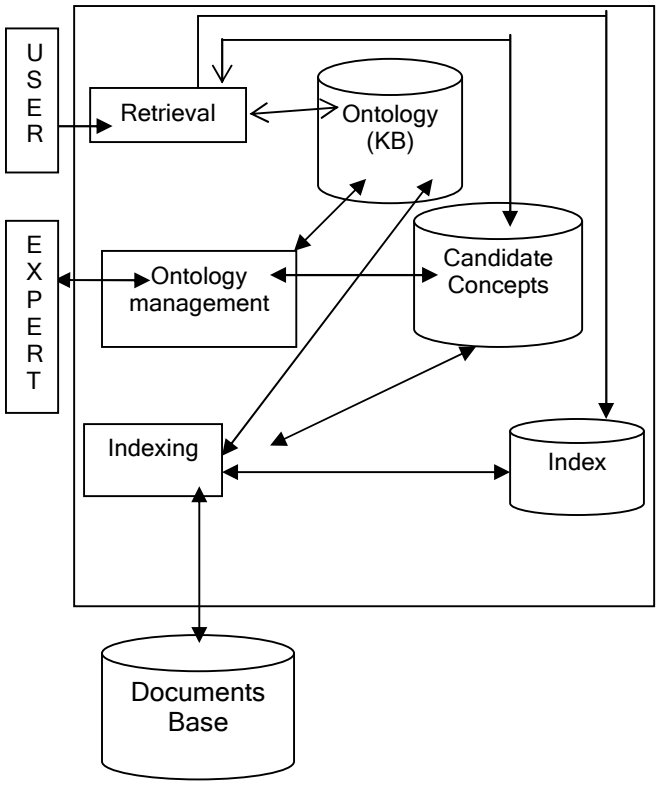


Figure 2. Architecture du système

3.2.1 Le support

La conceptualisation du domaine ou support peut être formellement définie par un triplet $S= (Tc, Tr, \sigma)$ où:
- *Tc* est l'ensemble des types des concepts du domaine, partiellement ordonné par une relation de spécialisation notée ≤.
idc1.1 ≤ *idc1* signifie que le type *idc1.1* est plus spécifique que le type *idc1*. La relation inverse notée ≥ est la relation de généralisation.
- *Tr* est l'ensemble des types de relations partitionné en sous ensembles des types de relations ayant le même nombre d'arguments.
$Tr = Tr_1 \cup Tr_2 \cup \ldots Tr_j \ldots Tr_n$ où Tr_j est l'ensemble des relations de type Tr avec j arguments $j \neq 0$. Deux types sont dits comparables s'il existe une relation de spécialisation ou de généralisation entre eux.
- σ, appelée aussi la signature indique pour chaque argument d'un type de relations, un type de concepts. Ce type correspond au type le plus générique que l'argument de la relation peut avoir. La Figure 3.a montre un exemple d'une hiérarchie de types de concepts, la figure 3.b montre une hiérarchie de types de relations.

3.2.2 La hiérarchie des types de concepts

La hiérarchie des types de concepts n'est autre que la conceptualisation d'un domaine donné, précisément, elle représente une vue d'indexation du domaine. Nous travaillons avec les concepts plutôt que les termes ou les objets atomiques. Le processus de conceptualisation est le suivant:

- identifier le domaine ou les documents (corpus à indexer),
- identifier les concepts ou connaissance du domaine,
- normaliser ces concepts par des types de concepts. Associer des identifiants qui ne sont pas des termes de sorte que nous puissions avoir plus d'un terme pour un concept,
- organiser les types en hiérarchie de types en utilisant les relations "est- un " et "partie de".

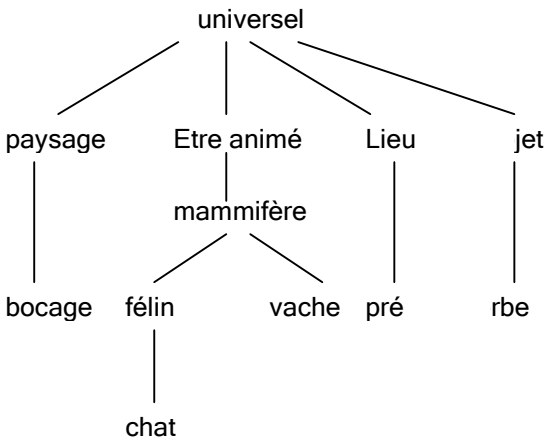


Figure 3.a

3.2.3 La hiérarchie des types de relation

Une fois les concepts du domaine identifiés et formalisés, nous devons identifier les relations qui existent entre ces concepts. Chaque relation possède un identifiant et une définition. Les relations sont organisées en types et les types organisés en hiérarchie.

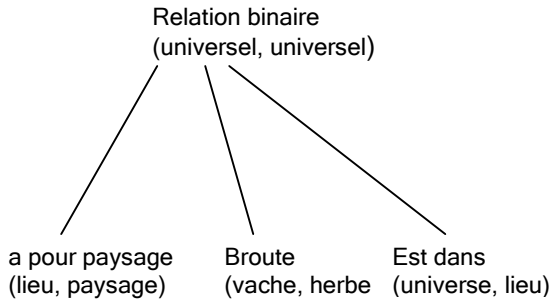


Figure 3.b

3.2.4 Définition formelle de l'ontologie

Nous pouvons à présent donner une définition formelle de l'ontologie. Une ontologie pour *n* langues est un quadruplet $M = (S, V, h_C, h_R)$ avec*:*

- S =(T_C, T_R, σ), un support composé d'un ensemble de types de concepts T_C , d'un ensemble de types de relations T_R et l'application σ qui fait correspondre pour chaque type de relation sa signature.
- V ensemble de vocabulaires partitionné en ensembles de termes appartenant à la même langue naturelle.

$V = V_{fr} \cup V_{ar} \cup V_{eng}$ où V_j est l'ensemble des termes appartenant à une langue $L_j \in$ {fr= French, ar=Arabic, eng = English} j= 1...3.

- h_C est l'ensemble des fonctions $T_C \rightarrow V_j$ qui associe pour chaque type de concept $t_c \in T_C$ au moins un terme du vocabulaire V_j.
- h_R, est l'ensemble des fonctions $T_R \rightarrow V_j$ qui associe à chaque type de relation $t_r \in T_R$ un terme du vocabulaire V_j .

4. L'Indexation

Le système d'indexation est basé sur des algorithmes d'extraction de connaissances à partir de documents textuels. L'approche que nous avons adoptée pour l'extraction est basée sur le calcul de segments (séquences) répétés. Le processus d'indexation distingue deux sortes de connaissances: la connaissance du domaine qui est indépendante de la langue et la connaissance terminologique ou les vocabulaires associés.

Un segment répété est une séquence de mots qui apparaît au moins deux fois dans un texte du corpus [10] [11]. Une séquence ne contient pas de signe de ponctuation et est calculée sur une fenêtre de 2 à 10 mots.
Une fois identifies, les segments répétés sont soumis une procédure de filtrage qui utilise des filtres spécifique (décrits ci dessous). Le choix de cette méthode est motivé par le fait qu'elle est facile à implémenter sur différentes langues. En effet, les filtres et les procédures de segmentation sont élaborés pour une nouvelle langue sans modifier les algorithmes. Le calcul des segments répétés est effectué comme suit:
- Les segments sont calculés sur le texte dans une fenêtre de *n* mots $1 < n < 10$. A chaque segment est associé son nombre d'apparition dans le texte. Examinons par exemple les segments suivants:

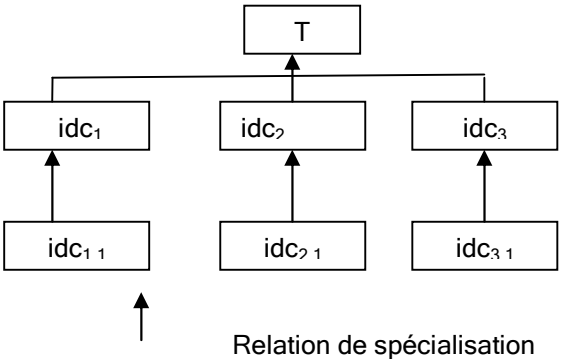
a. computer systems (8)
b. multilingual computer systems (7)
c. multilingual computer systems with (2)
d. the computer systems (8)
e. the multilingual computer systems and (2)

Eliminer les segments c, d, et e ne conduit pas à une perte d'information. L'algorithme de calcul ne prend pas en compte les mots vides. Donc, nous gardons seulement les segments a, b. En plus du filtrage des segments contenant des mots vides, nous avons quelques autres filtres:
- un filtre grammatical: qui identifie pour chaque langue, les déterminants, les conjonctions, les prépositions, les adverbes spécifiques et toutes les formes des verbes auxiliaires. Des listes sont utilisées pour l'Anglais et le Français. Pour l'Arabe, une liste des formes canoniques est utilisée et des algorithmes de dérivation sont implémentés. Par exemple, pour la préposition "min", nous avons à dériver "minho", "famin", wa min" "minhuma"...
- Un filtre verbal: construire une liste filtres de verbes présente quelques difficultés, idéalement, nous pourrons avoir à disposition tous les verbes qui existent dans le corpus (un lexique par exemple).
- Un filtre spécifique au corpus: c'est une liste de certains mots non significatifs du corpus, par exemple, dans un corpus médical, "docteur", "professeur", … ne sont pas significatifs.
Les index sont implémentés comme des graphes sémantiques qui sont construits selon les étapes suivantes:

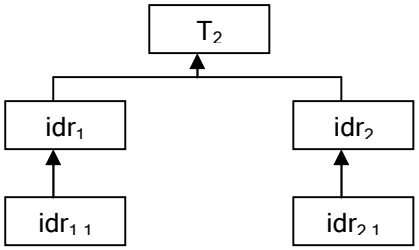
3. Identifier la langue du document,
4. Le processus d'extraction est basé sur la méthode du calcul des segments répétés et produit en sortie une liste de termes candidats ainsi qu'une liste de relations candidates.
5. Cette liste est utilisée pour chercher dans l'ontologie les concepts et les relations correspondant respectivement aux termes candidats et aux relations candidates.
6. Les concepts et les relations extraits à partir des segments et qui ne sont pas trouvés dans l'ontologie sont ajoutés à l'ontologie ainsi qu'à l'index des documents.

5. La Recherche

L'utilisateur exprime sa requête dans l'une des trois langues naturelles. Cette requête est analysée et transformée en un graphe sémantique. De plus, l'ontologie du domaine est utilisée pour étendre la requête initiale dans l'objectif de retrouver plus de documents pertinents. Le processus de recherche consiste alors en une comparaison de graphes pour trouver les documents qui répondent à la requête étendue de l'utilisateur. L'opérateur de projection et l'opérateur de projection étendue sont utilisés (voir plus haut). Le processus de recherche consiste en une projection du graphe de la requête sur les graphes stockés dans l'index. Retrouver de telles projections est coûteux, c'est pourquoi nous avons choisi d'étendre la requête de l'utilisateur.
L'expansion d'une requête Q composée du concept $C_{1.1}$ consiste à ajouter à la requête le père et le fils de $C_{1.1}$ avec des pondérations de similarité VRG et VRS respectivement. VRG et VRS sont deux constantes arbitraires qui calculent la similarité de graphes. Delà, Q peut être représentée par {($C_{1.1,}$ 1), ($C_{1.1,}$ VRG), ($C_{1.1.1,}$ VRS), ($C_{1.1.2,}$ VRS), ($C_{1.1.3,}$ VRS).

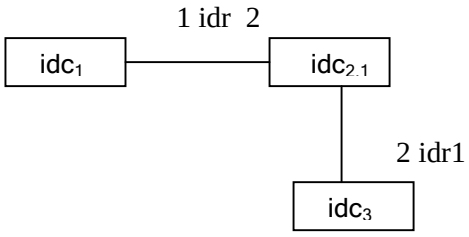


Hiérarchie de types de concepts



Hierarchv of relations

Soit un graphe de requête G:



L'expansion du graphe G de l'exemple donne la requête: Q' = {($idc_{1,}$ 1), ($idc_{1.1,}$ 0.7), ($idc_{2.1,}$ 1), ($idc_{2,}$ 0.9), ($idc_{3,}$ 1), (idr_1 , idc_1 , $idc_{2.1}$, 1), ($idr_{1.1}$, idc_1 , $idc_{2.1}$, 0.7), ($idr_{2.1}$, idc_3 , $idc_{2.1}$, 1), (idr_2 , idc_3 , $idc_{2.1}$, 0.9)}
Cet ensemble est utilisé pour rechercher dans le fichier inverse les documents qui correspondent à chaque élément de cet ensemble.

6. Conclusion

Nous avons présenté dans cet article, une approche fondée sur une ontologie pour la recherche d'information multilingue qui a été implémentée pour l'Arabe, le Français et l'Anglais. Nous avons fondé cette ontologie sur le formalisme de graphes sémantiques introduits par [4] pour améliorer la sémantique des documents dans des systèmes de RI multilingues.
Les étiquettes sur les noeuds des graphes sémantiques sont des numéraux, de sorte qu'ils ne sont associés à aucune langue naturelle. Les documents et les requêtes sont indexés par des graphes sémantiques.
Des algorithmes d'extraction basés sur une approche d'analyse linguistique de surface à savoir la méthode du calcul des segments répétés est utilisée pour indexer les nouveaux documents et mettre à jour l'ontologie.
L'opérateur de projection étendue est utilisé pour la comparaison de graphes. Un algorithme d'expansion de requêtes a été implémenté en vue d'assurer le plus de réponses pertinentes. Des pondérations de similarité permettent de retourner des résultats pondérés plutôt que des résultats booléens. Le système a été développé

en JAVA pour tourner sur des plateformes Windows aussi bien que UNIX. Les documents sont représentés en format XML. Deux types d'interfaces sont fournis pour l'utilisateur expert qui crée, gère et met à jour l'ontologie et pour l'utilisateur final qui recherché des documents. Les interfaces sont trilingues, l'utilisateur pouvant travailler avec la langue de son choix Arabe, Français ou Anglais.

7. References

[1] Hbao- Quoc "Vers une Indexation structurée basée sur des syntagmes nominaux: impact sur la RI en vietnamien et sur la RI multilingue", *thèse de doctorat,* université Joseph Fourrier, Grenoble, 2004.
[2] Pitrat J, *textes, ordinateurs et comprehension,,* Eyrolles editions,1985.
[3] R. Jalam et J. H. Chauchat, "Catégorisation de textes multilingues: quelques solutions" *rapport de recherche,* Laboratoire ERIC, université Lumière, Lyon2, 2004.
[4] C. Roussey, "Une méthode d'indexation sémantique adaptée aux corpus multilingues", thèse de doctorat, LISI-INSA de Lyon, (2000).
[5] Natalya Fridman Noy, Knowledge representation for Intelligent Information Retrieval in Experimental Sciences. *PHD thesis,* northeastern University, Boston. MA, 1997.
[6] T.R. Gruber, "Towards Principles for the design of ontologies used for knowledge sharing. KSL 93-04. Knowledge Systems Laboratory, Stanford University.
[7] H. Aliane, "A knowledge based platform for automatic indexing, information retrieval and knowledge discovery" Proceedings of the IEEE/IMACS conference, Athens, 1999.
[8] M. Chein et M.L. Mugnier, conceptual graphs: fundamental notions, *Revue d'Intelligence Artificielle*, Vol.6, n°4, pp 365-406, 1992.
[9] C. Roussey, "Un modèle de graphe pour la recherche d'information multilingue", *Rapport de recherche,* LISI- INSA 2000.
[10] P. Frath, Sémantique, référence et acquisition automatique de connaissances à partir de textes. *Thèse de doctorat* Université des sciences humaines, Strasbourg, 1997.
[11] R. Oueslati, Aide à l'acquisition de connaissances à partir de corpus, *Thèse de doctorat,* Université Louis Pasteur, Strasbourg, 1999.