
Une métrique pondérée pour la recherche textuelle d'images dans des documents semi-structurés

Mouna Torjmen, Karen Pinel-Sauvagnat, Mohand Boughanem

*IRIT, Université Paul Sabatier
118, route narbonne, 31400, Toulouse*

RÉSUMÉ. La naissance du standard XML et l'utilisation de plus en plus fréquente des images dans les documents électroniques ont soulevé une nouvelle problématique en recherche d'information: la recherche d'images dans des documents semi-structurés.

Cet article s'inscrit dans ce contexte et présente une méthode permettant de calculer une représentation sémantique de l'image en utilisant le texte et la structure des documents. Plus précisément, nous proposons une mesure pour calculer la participation de chaque élément du document dans cette représentation. Cette mesure est inspirée d'une mesure de similarité sémantique entre les concepts d'une ontologie.

Notre méthode est évaluée dans le cadre de la campagne de l'évaluation INEX 2006.

ABSTRACT. The birth of the XML standard and the growing use of images in electronic documents raised an open issue in information retrieval: image retrieval in semi-structured documents.

This article presents a method to evaluate a semantic representation of images using the text and the document structure. More precisely, we propose a measure that evaluates the participation of each element of the document in the image representation. This measure is inspired by a measure of semantic similarity between ontology's concepts.

Our method is evaluated in the context of the INEX 2006 campaign.

MOTS-CLÉS : image, représentation sémantique, xml, contexte

KEYWORDS: image, semantic representation, xml, context

1. Introduction

L'évolution conjointe des besoins utilisateurs et des documents électroniques ne cesse de soulever de nouvelles problématiques dans le domaine de la recherche d'information (RI). Si l'on considère la recherche dans des corpus de documents semi-structurés, la structure des documents permet aux systèmes de recherche d'information (SRI) de se focaliser sur le besoin utilisateur en renvoyant non plus des documents entiers mais des parties de document répondant de manière précise au besoin.

Parallèlement à l'apparition d'information structurelle dans les documents, l'insertion de contenus multimedia, et notamment d'images, a créé de nouveaux besoins concernant leur recherche.

Nos travaux portent sur la recherche multimedia et plus particulièrement sur la recherche d'images dans des documents semi-structurés.

Les documents semi-structurés sont généralement balisés grâce au standard XML (*eXtensible Markup Language*). Une de ses caractéristiques est la séparation du contenu des documents des instructions de présentation. Un document XML peut être représenté par un arbre, où la racine est le document, les noeuds internes sont les noeuds représentant l'élément ou les attributs, et les noeuds feuilles sont les noeuds contenant les valeurs (texte, image,...).

Nous traitons dans ce papier le cas du média image, mais nos propositions sont applicables à tout type de média tels que le son et la vidéo. Notre idée de base pour évaluer la pertinence des images dans un tel contexte est la suivante : nous supposons que les images peuvent être représentées sémantiquement par les autres éléments non multimedia du document XML. Par contre, ces éléments n'ont pas le même degré de participation dans la représentation de l'image. Pour quantifier cette participation, nous nous basons sur l'intuition suivante : plus les éléments non-multimédia sont proches de l'image, plus la probabilité de partager le même sujet avec l'image est élevée, et vice versa. Notre problématique est alors : comment et de combien chaque noeud non-multimédia doit-il participer à la représentation de l'image ?

Afin de répondre à la première question, nous avons considéré l'arbre d'un document XML comme une ontologie très simplifiée, où les noeuds sont les concepts, et la relation hiérarchique entre eux est "*partie de*". Le principe de notre proposition est : au lieu de calculer un score de similarité sémantique entre les concepts d'une ontologie, nous calculons un score de participation de chaque noeud non-multimedia dans la représentation sémantique de l'image.

Cet article s'organise comme suit. La section 2 présente un état d'art en deux parties sur la recherche d'images dans des documents semi-structurés, et sur l'utilisation des ressources sémantiques en recherche d'information, et notamment sur les mesures de similarité utilisées dans les ontologies. Dans la section 3, nous présentons notre approche dédiée à la recherche d'images dans des documents XML qui adapte une mesure de similarité sémantique des concepts dans une ontologie. Des expérimentations et résultats sont discutés dans section 4. Enfin, quelques conclusions et perspectives sont décrites dans section 5.

2. Etat d'art

2.1. Recherche multimedia dans des documents semi-structurés

La croissance des documents multimedia a amené la RI vers des nouvelles problématiques. Nous divisons les approches de la recherche d'images en trois catégories :

- 1) Recherche d'images par contenu (CBIR) : dans ce type de recherche, des techniques d'extraction et d'analyse de caractéristiques de bas niveau sont utilisées.
- 2) Recherche d'images par le contexte : dans ces approches, des informations sur le contexte des images (telles que le texte et la structure) sont utilisées.
- 3) La combinaison des deux premières méthodes.

Nos travaux se focalisent sur la deuxième approche, pour laquelle nous décrivons quelques méthodes issues de l'état de l'art ci-dessous.

Jusqu'en 2005, où la campagne d'évaluation INEX¹ a donné naissance à une nouvelle tâche appelée tâche multimedia, offrant ainsi une plateforme d'évaluation de traitement de requêtes multimedia, peu de travaux se sont intéressés à la recherche d'images dans des documents structurés.

Parmi les premiers travaux proposés utilisant la structure XML pour la recherche d'éléments multimedia, citons celui de (Kong *et al.*, 2005) qui consiste à diviser tout le contenu textuel du document XML multimedia en plusieurs *Region Knowledge*² *RKs* : *Self level RK* : *RK* du nœud multimedia ; *Sibling level RK* : *RK* des nœuds frères du nœud multimedia ; *1st ancestor level RK* : *RK* du premier ancêtre (parent) du nœud multimedia à l'exclusion du texte déjà utilisé ; *2nd ancestor level RK, ..., Nth ancestor level RK*.

L'évaluation de cette méthode a conduit à de bons résultats avec l'utilisation des *RKs* de premier, deuxième et troisième ancêtre, et à de mauvais résultats avec l'utilisation des autres *RKs*. Ces résultats se justifient par deux raisons : (1) la richesse textuelle qui varie d'une *RK* à une autre et (2) la différence entre le vocabulaire utilisé dans les requêtes et celui utilisé pour la description des images.

Les résultats sont certes liés à la collection utilisée mais nous constatons bien l'importance du choix des régions de connaissance et de la structure pour la recherche multimedia.

Un autre travail utilisant XPATH et XLINK est proposé par (Kong *et al.*, 2004). Les auteurs proposent l'utilisation de XPath pour modéliser le document XML comme un arbre de noeuds dans lequel une structure hiérarchique existe, et XLink pour présenter un rapport explicite entre les ressources ou les parties des ressources dans les documents XML. L'intégration de XPath et XLink dans le processus de recherche combine ainsi l'information hiérarchique et l'information de lien dans les documents XML. Les auteurs construisent des liens à partir des caractéristiques

1. Initiative for the Evaluation of XML Retrieval. <http://inex.is.informatik.uni-duisburg.de/>

2. Le contenu textuel de l'objet multimedia et des éléments l'entourant hiérarchiquement.

de bas niveau des images considérées. Grâce à ces liens, les objets multimédia qui ont peu ou pas d'informations textuelles associées peuvent toujours être retrouvés s'ils sont liés à d'autres objets multimédia possédant suffisamment d'informations textuelles. Par conséquent, la liste des résultats peut être augmentée grâce aux objets pertinents additionnels retrouvés grâce aux liens.

Une autre proposition d'utilisation de la structure pour la recherche d'images est faite dans (Torjmen *et al.*, 2007). Elle consiste à utiliser trois sources d'évidences pour calculer un score aux noeuds images. Ces sources sont : les noeuds descendants, les noeuds frères et les noeuds ascendants de l'image. Chaque noeud de ces trois sources a un score de pertinence obtenu par le système XFIRM (Sauvagnat, 2005) basé sur une méthode de propagation de scores. Des expérimentations avec la collection Wikipedia d'INEX³ ont montré que l'utilisation de ces trois sources simultanément amène à des résultats meilleurs que l'utilisation d'une seule source d'évidence.

2.2. Utilisation de ressources sémantiques en RI

Une autre évolution de la RI concerne la recherche sémantique des données, et ceci par l'utilisation des ressources sémantiques et conceptuelles telles que les ontologies.

Face à l'évolution rapide des masses d'information et de l'hétérogénéité des données, le vocabulaire seul des documents peut ne pas suffire à retrouver l'information pertinente, et de nombreux travaux utilisant la sémantique ont été proposés.

Dans (Baziz, 2005), une étude concernant l'utilisation étendue et l'utilisation partielle d'ontologies dans le processus de RI est faite. L'utilisation étendue d'ontologies consiste à projeter le document sur l'ontologie et à construire le meilleur réseau sémantique permettant de représenter au mieux le contenu sémantique de document. L'utilisation partielle d'ontologies consiste quant à elle à étendre la requête initiale en la passant par le réseau conceptuel de l'ontologie.

(Bhagal *et al.*, 2007) ont montré l'intérêt des ontologies dans la RI en général et dans l'expansion de requêtes en particulier. En effet, les ontologies ont été aussi utilisées pour les résumés automatiques, la désambiguïsation du sens des mots, la reformulation de requêtes, la Recherche d'Information multilingue...

(Zargayouna, 2004) a intégré la notion de voisinage sémantique en utilisant une ontologie de concepts auxquels sont reliés les termes de documents. Ceci entre dans l'indexation sémantique des documents XML.

La sémantique en RI a montré son intérêt et la plupart des travaux utilisant des ressources sémantiques se servent d'ontologies. Une mesure de similarité entre les concepts afin d'extraire la distance sémantique entre eux est nécessaire.

3. Collection utilisée dans INEX Multimedia 2006 et 2007

Dans la littérature des ontologies, deux types de mesures de similarité entre les concepts existent :

– Mesures de similarité par le nombre d'arcs (dites aussi mesures basées sur les graphes) : elles sont basées sur la structure arborescente de la taxonomie en s'appuyant sur la théorie "*spreading activation theory*" qui pose l'hypothèse que la hiérarchie des concepts est organisée selon les lignes de similarité sémantique. Par conséquent, des concepts de l'ontologie, liés par des arcs, sont similaires si la distance qui les sépare dans le graphe est faible, la distance dans un graphe étant donnée par le plus court chemin à parcourir le long des arcs pour joindre un concept à partir d'un autre.

– Mesures de similarité par le Contenu Informatif (*CI*) : elles quantifient le fait que plus le contenu informatif partagé par deux concepts est élevé, plus ces concepts sont similaires, et elles s'appuient également sur la structure taxonomique de l'ontologie. Le contenu informatif d'un concept *c* est fonction de la probabilité d'occurrence (probabilité de présence) des composantes de ce concept, laquelle dépend de la fréquence de ses composantes dans l'ensemble de l'ontologie.

Dans nos travaux, nous nous intéressons au premier type de mesure de similarité. Plusieurs mesures de similarités basées sur le nombre d'arcs entre les concepts sont proposées dans la littérature telles que celle de (Rada *et al.*, 1989), celle de (Hirst *et al.*, 1997) et celle de (Wu *et al.*, 1994).

Nous nous intéressons dans cet article à la mesure de (Wu *et al.*, 1994) qui est connue sous le nom de Wu-Palmer et qui prend en compte la position des concepts par rapport à la racine de l'ontologie. Elle est définie comme suit :

$$Sim_{WP}(C1, C2) = \frac{2 * N3}{(N1 + N2 + 2 * N3)} \quad [1]$$

où *N1* et *N2* sont le nombre d'arcs qui séparent *C1* et *C2* de leur ascendant commun le plus spécifique *C*. *N3* est le nombre d'arcs qui séparent *C* de l'élément racine. Le choix de cette mesure est basé sur le fait qu'elle est à la fois la plus simple à implémenter et la plus performante (Lin, 1998).

Nous proposons dans la section suivante une méthode utilisant le contexte textuel et structurel des images des documents et permettant une représentation sémantique des images en fonction des autres éléments non-multimédia. Cette représentation est inspirée de la mesure de Wu-Palmer.

3. Représentation sémantique d'images dans des documents semi-structurés

Notre approche est basée sur l'intuition que chaque noeud textuel porte des informations permettant de représenter sémantiquement chaque image. Par conséquent, chaque noeud doit participer à représenter l'image sémantiquement. La question qui

s'impose derrière cette idée est : comment calculer la participation de chaque noeud textuel dans la représentation sémantique de l'image ?

Un document XML peut être représenté par un arbre où la racine est le document, les noeuds internes sont les noeuds représentant les éléments ou les attributs, et les noeuds feuilles sont les noeuds contenant les valeurs (texte, images...). Cette représentation arborescente nous permet de considérer un document XML comme une ontologie très simplifiée où les noeuds sont les concepts qui sont organisés hiérarchiquement avec la relation *est partie de*. Par exemple, l'élément *section est partie de article* et l'élément *paragraphe est partie de section*.

L'idée est alors de se servir d'une mesure de similarité sémantique entre les termes d'une ontologie pour calculer combien chaque noeud peut participer à la représentation de l'image.

Nous considérons le noeud image comme un concept $C1$, et le noeud à utiliser comme un autre concept $C2$.

(Zargayouna, 2004) a proposé d'utiliser cette mesure dans un système d'indexation de documents XML. Elle a cependant constaté que cette mesure représente une limite car il est possible d'avoir la similarité entre un concept et son fils inférieure à la similarité entre ce concept et son frère alors qu'elle envisage de ramener tous les fils d'un concept avant ses frères.

Considérons l'exemple de la figure 1. Nous calculons pour les noeuds B et F leur représentation pour l'image I .

$$Sim_{WP}(I, B) = \frac{2 * 3}{(1 + 1 + 2 * 3)} = 0.75 \quad [2]$$

$$Sim_{WP}(I, F) = \frac{2 * 4}{(4 + 0 + 2 * 4)} = 0.66 \quad [3]$$

Le noeud frère B participe plus que le noeud fils F .

Pour éviter cela, (Zargayouna, 2004) a proposé de pénaliser les scores des frères en ajoutant une fonction $spec(C1, C2)$ qui calcule la spécificité de deux concepts par rapport au concept le plus bas (*bottom*) (voir Figure 1).

$$spec(C1, C2) = depth_b(C) * distance(C, C1) * distance(C, C2) \quad [4]$$

avec C est l'ancêtre commun le plus spécifique, $depth_b$ est le nombre maximum d'arcs qui séparent C de *bottom* et $distance(C, C_i)$ est la distance en nombre d'arcs entre C et C_i .

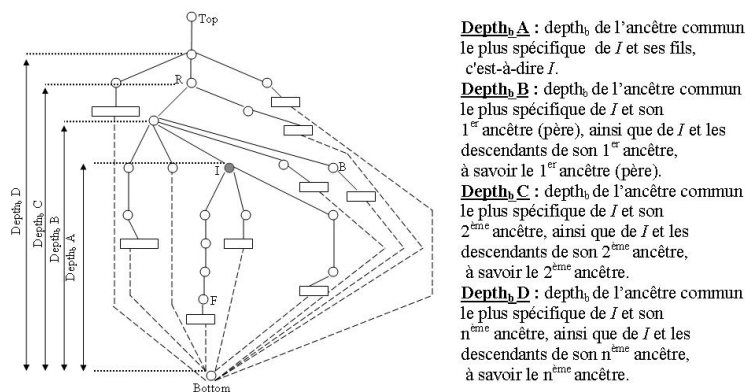


Figure 1. Relations entre les noeuds et bottom

Nous appliquons maintenant cette adaptation sur l'exemple précédent.

$$Sim_{WPSpec}(I, B) = \frac{2 * 3}{(1 + 1 + 2 * 3 + 7 * 1 * 1)} = 0.4 \quad [5]$$

$$Sim_{WPSpec}(I, F) = \frac{2 * 4}{(4 + 0 + 2 * 4 + 6 * 4 * 0)} = 0.66 \quad [6]$$

L'exemple montre l'intérêt de l'adaptation où le noeud fils *F* participe plus que le noeud frère *B*.

Cette mesure a réglé le problème des fils et des frères mais elle ne répond pas encore à nos objectifs. En effet, les noeuds ascendants peuvent participer plus que les noeuds frères dans la représentation de l'image, ce qui nous ne paraît pas une bonne solution. Prenons par exemple le noeud ascendant *R*. Selon l'adaptation avec *spec* (*C1, C2*) :

$$Sim_{WPSpec}(I, R) = \frac{2 * 2}{(2 + 0 + 2 * 2 + 8 * 2 * 0)} = 0.66 \quad [7]$$

et le noeud ascendant *R* participe plus que le noeud frère *B*.

Pour éviter cela, nous proposons de pénaliser les noeuds ascendants en multipliant le dénominateur par $(depth_h)^\beta$ avec $\beta > 1$. Dans cet article, la formule a été évaluée avec $\beta = 2$.

L'utilisation de ce nouveau facteur pénalise les noeuds ascendants par rapport aux noeuds fils et noeuds frères. En effet, plus on monte de niveau, plus le score de participation des noeuds diminue : les noeuds du premier ancêtre de l'image participent plus que les noeuds du deuxième ancêtre, ..., les noeuds de nième-1 ancêtre participent

plus que les noeuds du nième ancêtre (voir Figure 1).

Vérifions maintenant notre adaptation sur l'exemple précédent avec $\beta=2$:

$$Sim_{WPI_m}(I, B) = \frac{2 * 3}{(1 + 1 + 2 * 3 + 7 * 1 * 1 + 7^2)} = 0.093 \quad [8]$$

$$Sim_{WPI_m}(I, R) = \frac{2 * 2}{(2 + 0 + 2 * 2 + 8 * 2 * 0 + 8^2)} = 0.055 \quad [9]$$

$$Sim_{WPI_m}(I, F) = \frac{2 * 4}{(4 + 0 + 2 * 4 + 6 * 4 * 0 + 6^2)} = 0.166 \quad [10]$$

d'où le noeud fils F participe plus que le noeud frère B , le noeud frère B participe plus que le noeud père R , ce qui répond à nos objectifs.

Dans nos travaux, cette mesure est utilisée pour calculer la participation de chaque noeud pertinent dans la représentation sémantique de l'image. Nous définissons la mesure de représentation sémantique comme suit :

$$Rep(I, N) = \frac{2 * N^3}{N1 + N2 + 2 * N3 + spec(I, N) + depth_b^2} * S_N \quad [11]$$

où I est le noeud image, N est le noeud qui participe à la représentation de l'image et S_N est le score du noeud N , calculé dans une première étape à l'aide d'un système de recherche d'information structurée classique.

Le score final de chaque image est calculé comme suit :

$$S(I) = \sum_{i=1} Rep(I, N_i) \quad [12]$$

avec N_i est un noeud de document contenant l'image.

Notre approche peut s'appliquer de deux façons, soit en utilisant uniquement les noeuds textuels comme noeuds participants à la représentation sémantique de l'image, soit en calculant un score pour chaque noeud interne et en utilisant ces noeuds internes pour la représentation sémantique de l'image (voir Figure 2).

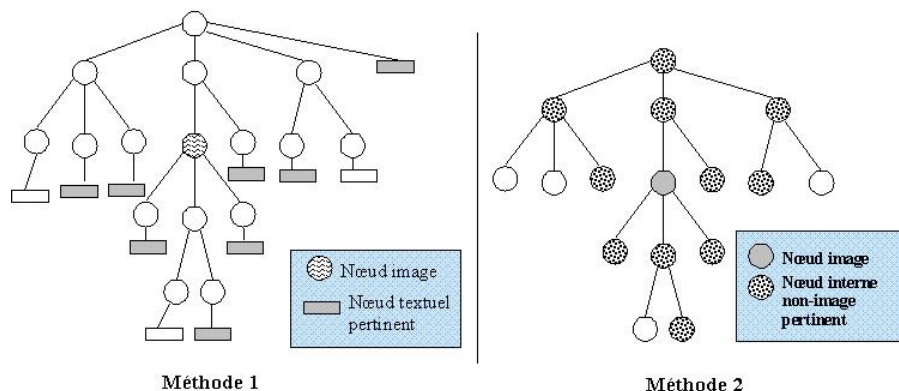


Figure 2. Représentation sémantique des images : par les noeuds textuels (Méthode 1), par les noeuds internes (Méthode 2)

Ces deux méthodes proposées peuvent être appliquées non seulement aux images mais aussi aux autres médias tels que le son et la vidéo. En effet elles nécessitent pas d'informations liées au contenu du média, mais uniquement le contexte textuel et structurel.

En examinant les deux méthodes théoriquement, nous estimons que la première méthode va amener à des résultats meilleurs que la deuxième pour les raisons suivantes : la méthode 1 propage le contenu des noeuds textuels vers l'image une seule fois et ceci en utilisant notre mesure proposée. Par contre, la méthode 2 propage dans un premier temps, le contenu des noeuds textuels vers les noeuds internes en utilisant une formule de propagation et d'agrégation décrite dans la section suivante, et dans un deuxième temps, notre mesure sera appliquée. Ceci peut induire à une utilisation multiple de la même information textuelle : cette information participe autant de fois qu'elle a d'ascendants, et ceci peut influencer négativement sur les scores des images.

Une autre limite de la deuxième méthode réside dans le fait que le nombre total de noeuds internes du document influence sur le score de l'image. En effet, une image contenue dans un document composé de 10 noeuds internes peut être pénalisée par rapport à une autre image moins pertinente contenue dans un document composé de 1000 noeuds.

4. Evaluation

Notre méthode est évaluée dans le cadre de la campagne d'évaluation INEX 2006. Dans les paragraphes suivants, nous présentons le système XFIRM, qui nous sert à calculer les scores des noeuds feuilles et internes des documents XML. Ensuite, nous présentons la campagne d'évaluation INEX ainsi que la collection utilisée dans nos

évaluations et les mesures de pertinence utilisées. Enfin, nous discutons les résultats obtenus par utilisation des noeuds textuels et des noeuds internes.

4.1. Le système XFIRM

Le système XFIRM est développé au sein de notre équipe par (Sauvagnat, 2005) pour la recherche textuelle dans des documents semi-structurés. Il est basé sur une méthode de propagation de la pertinence pour la recherche dans des documents XML. Le traitement se fait en deux temps. Tout d'abord, des scores de pertinence sont calculés au niveau des noeuds feuilles. Ces scores sont ensuite propagés vers les noeuds internes à travers l'arbre du document. A la fin du traitement, une liste triée de sous-arbres est retournée à l'utilisateur.

On trouvera ci-dessous la description du traitement des deux types de requêtes évaluées par XFIRM :

- Les requêtes portant sur le contenu seulement et composées de simples mots clés (requêtes CO).
- Les requêtes composées de contraintes structurelles et de contenu (requêtes CO+S).

4.1.1. Traitement des requêtes portant sur le contenu seulement (Content Only)

- Calcul du score des noeuds feuilles

Soit $q = t_1, \dots, t_n$ une requête composée de termes t_i . Les scores de pertinence des noeuds feuilles ln sont calculés suivant la fonction de similarité $RSV(q, ln)$ suivante :

$$RSV(q, ln) = \sum_{i=1}^n w_i^q * w_i^{ln}, \text{ où } w_i^q = tf_i^q \text{ et } w_i^{ln} = tf_i^{ln} * idf_i * ief_i \quad [13]$$

w_i^q et w_i^{ln} sont respectivement les scores du terme t_i dans la requête q et dans le noeud feuille ln . tf_i^q et tf_i^{ln} sont respectivement la fréquence du terme t_i dans q et dans ln , $idf_i = \log(|D|/(|di| + 1)) + 1$, avec $|D|$ le nombre total de documents dans la collection, et $|di|$ le nombre de documents contenant t_i , et ief_i est la fréquence inverse d'élément du terme t_i , avec $ief_i = \log(|N|/(|nfi| + 1)) + 1$, où $|nfi|$ est le nombre de noeuds feuilles contenant t_i et $|N|$ est le nombre total de noeuds feuilles de la collection.

- Calcul de score des noeuds internes par propagation

Après avoir calculé un score de pertinence pour chaque noeud feuille, nous évaluons pour chaque noeud interne un score de pertinence r_n fonction des scores de pertinence de ses noeuds feuilles et du score de pertinence r racine du document entier. Ce dernier paramètre est introduit pour la raison suivante : on se base sur l'idée que le concepteur d'un document suit une certaine unité dans ses idées, même si le contenu du document

est hétérogène, et qu'un noeud appartenant à un document fortement pertinent doit être mieux classé qu'un noeud se trouvant dans un document de pertinence moindre.

$$r_n = \rho * |L'_n| \cdot \sum_{ln_k \in L_n} \alpha^{dist(n,ln_k)-1} * RSV(q, ln_k) + (1 - \rho) * r_{racine} \quad [14]$$

$dist(n,ln_k)$ est la distance entre le noeud n et le noeud feuille ln_k dans l'arbre de document, c'est à dire le nombre d'arcs qui séparent n de ln_k , et $\alpha \in]0..1]$ permet d'adapter l'importance du paramètre $dist$. $|L'_n|$ est le nombre de noeuds feuilles descendants de n et ayant une valeur de pertinence différente de zéro (selon l'équation 13). $\rho \in]0..1]$, permet d'introduire la pertinence du document dans l'évaluation de pertinence des noeuds internes, et r_{racine} est le score de pertinence de l'élément *racine*, c'est à dire le score de pertinence de tout le document, évalué en utilisant l'équation 14 avec $\rho = 1$.

4.1.2. Traitement des requêtes orientées contenu et structure (CO+S)

Dans les requêtes CO+S, un type d'élément cible (c'est à dire devant être retourné) est précisé par l'utilisateur. Par exemple, dans la requête *article[France] // ec : section[culture française] // paragraphe[Paris]*, l'utilisateur souhaite obtenir une *section* parlant de *culture française* contenue dans un *article* sur la *France* et contenant un *paragraphe* à propos de *Paris*. L'évaluation des requêtes CO+S se fait de la façon suivante :

1) Les requêtes XFIRM sont décomposées en sous-requêtes élémentaires *ESQ*, qui sont de la forme : $ESQ = tg[q]$, où tg est un nom de balise, c'est-à-dire une condition structurelle, et $q = t_1, \dots, t_n$ est une condition de contenu composée de simples mots-clés.

2) Des scores de pertinence sont ensuite calculés entre les noeuds feuilles et les conditions de contenu des *ESQ*.

3) Ces scores de pertinence sont propagés dans l'arbre de document pour répondre aux conditions structurelles des *ESQ*.

4) Les requêtes originales sont évaluées grâce à une agrégation et propagation vers le haut et vers le bas des scores de pertinence vers les éléments cibles.

L'étape 2 est traitée en utilisant l'équation 13. Dans l'étape 3, la valeur de pertinence r_n d'un noeud n à une $ESQ = tg[q]$ est calculée avec la formule suivante :

$$r_n = \begin{cases} \sum_{ln_k \in L_n} \alpha^{dist(n,ln_k)-1} * RSV(q, ln_k) & \text{si } n \in \text{construct}(tg) \\ 0 & \text{sinon} \end{cases} \quad [15]$$

où la fonction $\text{construct}(tg)$ permet la création d'un ensemble composé de noeuds ayant tg comme nom de balise, et $RSV(q, ln_k)$ est évaluée durant l'étape 1 avec l'équation 13. La fonction $\text{construct}(tg)$ utilise un index dictionnaire qui fournit pour une balise tg donnée, les balises considérées comme équivalentes. Ce dictionnaire est construit manuellement.

Pour plus de détails sur le système XFIRM, voir (Sauvagnat, 2005).

4.2. INEX : Collection, tâche et mesure d'évaluation

INEX (*Initiative for the Evaluation of XML Retrieval*) est actuellement la seule campagne d'évaluation des différents systèmes de recherche d'information pour des documents XML. Le but principal d'INEX est de promouvoir l'évaluation de la recherche sur des documents XML en fournissant une collection de test, des procédures d'évaluation et un forum pour permettre aux différentes organisations participantes de comparer leurs résultats. La collection de test consiste en un ensemble de documents XML, requêtes et jugements de pertinence.

Le langage de requêtes utilisé dans XFIRM est NEXI (Trotman *et al.*, 2004). Nous nous intéressons ici à la tâche multimedia qui a eu lieu en 2007 pour la troisième fois, et plus particulièrement à la sous tâche Multimedia Fragments.

Une requête Multimedia Fragment contient des contraintes structurelles et peut contenir ou non des besoins images.

Le système doit retourner à l'utilisateur des fragment multimédia (texte et/ou image). Dans nos expérimentations, nous avons évalué deux types de requête : (1) les requêtes originales de la tâche Multimedia Fragment (type requête=fragment), où l'élément retourné par le système n'est pas forcément une image ; (2) des requêtes construites à partir des requêtes originales en enlevant toutes les contraintes structurelles et en exigeant que l'élément retourné par le système soit toujours une image (type requête=image).

Pour les requêtes de type *fragment*, nous divisons chaque requête en sous-requêtes. Le traitement de ces dernières est fait selon la contrainte structurelle, si elle est image, nous utilisons notre méthode, sinon, nous utilisons la méthode de traitement des requêtes *CO+S* du système *XFIRM*. A la fin, nous faisons la propagation vers l'élément cible de la requête grâce au processus de propagation de *XFIRM*. Dans cette tâche Multimedia Fragment, la collection utilisée en 2006 et 2007 est la collection Wikipedia d'INEX (Denoyer *et al.*, 2006). Cette collection contient 344.642 images et la profondeur moyenne de la structure XML est de 6.72.

Pour la sous tâche Multimedia Fragments, la mesure de pertinence utilisée est *MAeP* (Mean Average effort Precision) (Kazai *et al.*, 2005) : c'est la moyenne non-interpolée d'effort-précision.

L'effort-précision (ep) à un niveau donné de gain-rappel (gr) indique l'effort nécessaire à un utilisateur parcourant les réponses retournées par un système, relatif à l'effort nécessaire pour atténuer le même niveau de gain-rappel utilisant un classement de réponses idéal. *MAep* est utilisée pour moyenner les valeurs d'effort-précision pour chaque rang auquel une réponse pertinente est retournée.

4.3. Résultats avec représentation sémantique des images par les noeuds textuels

Le tableau 1 présente les résultats de l'évaluation de notre méthode en utilisant les noeuds feuilles pertinents :

Tableau 1. Résultats de la méthode de représentation sémantique des images avec les noeuds textuels

Run	Type requête	Mesure	MAeP
Run1	fragment	Rep_{WP}	0.01614
Run2	fragment	Rep_{WPSpec}	0.01576
Run3	fragment	Rep_{WPI_m}	0.01681
Run4	image	Rep_{WP}	0.01814
Run5	image	Rep_{WPSpec}	0.01751
Run6	image	Rep_{WPI_m}	0.01814

Rep_{WP} correspond aux résultats avec la mesure Wu-Palmer original (Equation 2). Rep_{WPSpec} correspond aux résultats avec la mesure Wu-Palmer après son adaptation avec la fonction *Spec* (Equation 7).

Rep_{WPI_m} correspond aux résultats avec la mesure Wu-Palmer après notre adaptation (Equation 11) avec $\beta = 2$.

La première remarque en comparant tous les runs est que ceux utilisant les requêtes de type *image* (*Run4*, *Run5*, *Run6*) donnent des résultats meilleurs que les runs utilisant les requêtes de type *fragment* (*Run1*, *Run2*, *Run3*). Ceci peut être expliqué par le fait que la tâche de recherche Multimedia Fragment est encore mal spécifiée et que les personnes jugeant les requêtes préfèrent avant tous des images.

Afin de confirmer ces résultats liés aux type de requêtes, nous devons faire d'autres évaluations. En effet, nous ne nous sommes pas basés sur la même méthode pour évaluer les différents contraintes structurelles (la méthode proposée dans cet article pour les contraintes structurelles de type image, et la méthode COS de XFIRM pour les autres contraintes structurelles.)

En comparant les trois premiers runs, nous constatons que les meilleurs résultats sont obtenus avec notre mesure (Run : MAeP=0.01681).

Dans les trois derniers runs (*Run4*, *Run5*, *Run6*), il n'y a pas de différence entre notre mesure proposée et la mesure de Wu-Palmer originale. Mais l'utilisation de la mesure Rep_{WPSpec} a dégradé les résultats.

4.4. Résultats avec représentation sémantique des images par les noeuds internes

Le tableau 2 montre les résultats obtenus en appliquant notre méthode avec les noeuds internes. Pour notre adaptation (*Run3* et *Run6*), nous avons utilisé $\beta = 2$.

Tableau 2. Résultats de la méthode de représentation sémantique des images avec les noeuds internes

Run	Type requête	Mesure	MAeP
Run1	fragment	Rep_{WP}	0.01623
Run2	fragment	Rep_{WPSpec}	0.01582
Run3	fragment	Rep_{WPI_m}	0.01322
Run4	image	Rep_{WP}	0.01733
Run5	image	Rep_{WPSpec}	0.01705
Run6	image	Rep_{WPI_m}	0.01690

De manière identique aux expérimentations précédentes, les runs utilisant le type *image* (*Run4*, *Run5*, *Run6*) donnent des résultats meilleurs que les runs utilisant le type *fragment* (*Run1*, *Run2*, *Run3*).

D'après les résultats de ce tableau, nous remarquons que notre mesure n'est pas efficace avec l'utilisation des noeuds internes, et les résultats sont dégradés (*Run3*, *Run6*). Ceci peut être expliqué par la raison suivante : le nombre de noeuds internes représentant les fils et les frères est négligeable devant le nombre des noeuds internes représentant les noeuds ascendants. Le score final de l'image est calculé par la somme des différents scores des noeuds internes participants, ce qui augmente la participation des noeuds ascendants par rapport aux noeuds fils et frères.

4.5. Discussion

En comparant globalement les deux évaluations de notre approche (une par l'utilisation des noeuds textuels et une par l'utilisation des noeuds internes), nous constatons que la première montre son intérêt par rapport à la deuxième. Nous pouvons conclure que l'utilisation directe du contenu informatif des noeuds est meilleure que l'utilisation indirecte de ce contenu (à travers les noeuds internes)

En comparant l'évaluation des deux types de requêtes avec les deux méthodes, les requêtes de type *image* donnent des résultats meilleurs que les requêtes de type *fragment*.

La comparaison des trois mesures de représentation sémantique des images en utilisant la structure montre que la meilleure mesure pour l'utilisation des noeuds internes est celle de Wu-Palmer original. Pour l'utilisation des noeuds textuels, notre mesure améliore un peu les résultats par rapport à la mesure Wu-Palmer originale.

Ces résultats sont cependant loin des meilleurs résultats obtenus par les participants d'INEX. D'autres facteurs influençant les résultats comme l'indexation, la pondération des termes et l'utilisation d'autres éléments de contexte (liens, noms d'images...) doivent encore être étudiés.

Ces expérimentations nous auront en tout cas permis de valider l'intérêt de notre méthode par rapport à de précédentes expérimentations (Hlaoua *et al.*, 2007).

5. Conclusion et perspectives

Nous avons proposé dans cet article une méthode permettant de représenter sémantiquement les images par d'autres éléments du document XML. Afin de quantifier la participation de chaque élément dans la représentation sémantique de l'image, nous avons proposé une mesure de représentation inspirée d'une mesure de similarité sémantique entre les concepts d'une ontologie, qui est basée sur le nombre d'arcs séparant les deux concepts.

Nous avons appliqué cette mesure selon deux manières différentes : une première fois avec les noeuds textuels, et une seconde fois avec les noeuds internes. En évaluant ces deux méthodes, la première a montré son intérêt par rapport à la deuxième.

Les perspectives envisageables à nos travaux sont nombreuses :

- tester d'autres formules de pondération de termes, en essayant de pondérer le texte entourant l'image d'une façon appropriée permettant de la qualifier,
- évaluer d'autres mesures de représentation sémantique basées sur les graphes,
- étudier la différence entre le type de l'élément retourné : image ou fragment,
- ajouter d'autres éléments pour la représentation sémantique des images tels que les liens et le nom de l'image,
- évaluer la méthode proposée dans INEX 2007.

6. Bibliographie

- Baziz M., Indexation conceptuelle guidée par ontologie pour la recherche d'information, Thèse de doctorat, Université Paul Sabatier, Toulouse, France, décembre, 2005.
- Bhogal J., Macfarlane A., Smith P., « A review of ontology based query expansion », *Information Processing & Management*, vol. 43, n° 4, p. 866-886, July, 2007.
- Denoyer L., Gallinari P., « The Wikipedia XML corpus », *SIGIR Forum*, vol. 40, n° 1, p. 64-69, 2006.
- Fuhr N., Lalmas M., Malik S., Kazai G., *Advances in XML Information Retrieval and Evaluation, 4th International Workshop of the Initiative for the Evaluation of XML Retrieval, INEX 2005, Dagstuhl Castle, Germany, November 28-30, 2005, Revised Selected Papers*, vol. 3977 of *Lecture Notes in Computer Science*, Springer, 2006.
- Hirst G., St-Onge D., « Lexical Chains as representation of context for the detection and correction malapropisms », 1997.
- Hlaoua L., Torjmen M., Pinel-Sauvagnat K., Boughanem M., « XFIRM at INEX 2006. Ad-hoc, Relevance Feedback and MultiMedia tracks », in , N. Fuhr, , M. Lalmas, , A. Trotman (eds), *International Workshop of the Initiative for the Evaluation of XML Retrieval (INEX), Dagstuhl, Allemagne, 18/12/06-20/12/06*, vol. LNCS 4518, Springer, <http://www.springerlink.com>, p. 373-386, mars, 2007.
- Kazai G., Lalmas M., « INEX 2005 Evaluation Measures », *INEX*, p. 16-29, 2005.
- Kong Z., Lalmas M., « Integrating Xlink and Xpath to retrieve structured multimedia documents in digital libraries », *RIAO 2004 Conference*, 2004.

- Kong Z., Lalmas M., « XML Multimedia Retrieval », *SPIRE*, p. 218-223, 2005.
- Lin D., « An Information-Theoretic Definition of Similarity », *Proceedings of 15th International Conference On Machine Learning*, 1998.
- Lu X. A., Keefer R. B., « Query Expansion/Reduction and its Impact on Retrieval Effectiveness », *TREC*, 1994.
- Rada R., Mili H., Bicknell E., Blettner M., « Development and application of a metric on semantic nets », *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 19, n° 1, p. 17-30, 1989.
- Sauvagnat K., Modèle flexible pour la recherche d'information dans des corpus de documents semi-structurés, Thèse de doctorat, Université Paul Sabatier, Toulouse, France, juin, 2005.
- Torjmen M., Pinel-Sauvagnat K., Boughanem M., « Utilisation du contexte textuel et structurel pour la recherche d'images dans des documents XML », *Congrès Informatique des Organisations et Systèmes d'Information et de Décision (INFORSID), Perros-Guirec, 22/05/07-25/05/07*, Université de Rennes 1- Laboratoire IRISA, p. 21-36, mai, 2007.
- Trotman A., Sigurbjornsson B., « Narrowed Extended XPath I (NEXI) », *INEX*, p. 16-40, 2004.
- Ventresque A., « Recherche d'Information efficace utilisant la sémantique : le focus », *CORIA, Session Jeunes Chercheurs*, p. 377-382, 2006a. 2-9520326-6-1.
- Ventresque A., « Une mesure de similarité sémantique utilisant des résultats de psychologie », *CORIA, Session Jeunes Chercheurs*, p. 371-376, 2006b. 2-9520326-6-1.
- Wu Z., Palmer M., « Verb semantics and lexical selection », *Proceedings of the 23rd Annual Meetings of the Associations for Computational Linguistics*, p. 133-138, 1994.
- Zargayouna H., « Contexte et sémantique pour une indexation de documents semi-structurés », *Conférence en Recherche d'Information et Applications*, p. 571-581, Mars, 2004.