Propositions pour la recherche contextuelle d'images dans des documents XML

Utilisation de la structure des documents XML pour la recherche d'images

Mouna Torjmen

IRIT SIG-RI
118 route de Narbonne
31 062 Toulouse Cedex4
torjmen@irit.fr

RÉSUMÉ. *Cet article s'inscrit dans le cadre de la recherche d'images dans des documents XML. Ce type de recherche peut utiliser des informations sémantiques en plus des informations visuelles de l'image. Nous nous proposons ici d'explorer des pistes pour la recherche de ces informations sémantiques au sein des documents XML, en supposant qu'une image peut être présentée par les autres éléments non images du document XML. Nous proposons d'une part une méthode pour choisir quels sont les éléments d'un document XML pouvant participer au mieux à la représentation de l'image, et d'autre part, une mesure qui évalue pour chaque élément non image sa participation dans la représentation de l'image.*

ABSTRACT. *This article focusses on images retrieval in XML documents. Images retrieval can use sementical information in addition to visual features.*
We propose here to explore some ideas to find this semantic information in XML document, by assuming that an image can be represented by the other non-image elements of the XML document. We propose a method to choose which are the elements of a XML document that can better take part to the representation of the image, and a measure which evaluates for each non-image element its participation to the representation of the image.

MOTS-CLÉS : *XML, image, multimedia, recherche d'information.*

KEYWORDS: *XML, image, multimedia, information retrieval.*

CORIA 2007, pp. 479–484
Saint-Étienne, 28–30 mars 2007

1. Introduction

Au fil des évolutions documentaires et numériques, la Recherche d'Information est devenue un champ transdisciplinaire. Elle doit aujourd'hui être capable de prendre en compte l'information structurelle et multimédia contenue au sein des documents (images, son, vidéo,...). La naissance du standard XML (eXtensible Markup Language) a révolutionné les documents électroniques. Sa particularité consiste à séparer le 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 nœuds représentant les éléments ou les attributs, et les nœuds feuilles sont les nœuds contenant les valeurs (texte,image,...). Dans un document XML, les informations sur un nœud multimedia peuvent être trouvées dans les nœuds textuels qui lui sont liés. Par conséquent, ces informations textuelles peuvent être employées pour représenter l'élément multimédia, et permettent d'assurer une recherche directe des données multimedia par des requêtes textuelles (Kong *et al.,* 2004).
Dans cet article, nous nous intéressons plus particulièrement à la recherche d'images dans les documents XML en utilisant le contexte textuel et structurel.
Nous proposons de calculer une représentation textuelle de chaque image à travers les nœuds non-images les plus proches et jugés pertinents par un système de recherche d'information classique. Dans la suite de cet article, nous assimilerons la pertinence d'un nœud à sa pertinence système. Pour calculer la pertinence des nœuds non-images, nous avons utilisé le modèle XFIRM qui évalue un score pour chaque nœud textuel et propage ensuite ce score vers le haut jusqu'au nœud racine (Sauvagnat, 2005). Par conséquent, les nœuds pertinents (internes et feuilles) auront un score de pertinence qui pourra servir pour le calcul du score du nœud image.
Nous organisons cet article comme suit : la section 2 décrit quels sont les nœuds non-images pouvant être choisis pour la description de l'image, et dans la section 3, nous proposons une mesure pour le calcul de la représentation de l'image en fonction des nœuds non-images déja choisis.

2. Définition de la *zone descriptive* d'une image

On appelle *zone descriptive* d'une image la zone constituée par les nœuds jugés pertinents les plus proches de cet objet multimédia. Ils seront employés pour la représentation de l'image. La figure 1 montre un exemple d'une *zone descriptive*.
L'utilisation des nœuds descendants de l'image est primordiale car ils contiennent généralement des informations spécifiques à l'image comme par exemple la *légende* ou le *titre*. Par conséquent, ils forment une partie fixe de la *zone descriptive*. Les nœuds ascendants forment aussi une partie fixe de la *zone descriptive* car ils permettent de prendre en compte tout le contexte du document. En effet, les scores des nœuds ascendants sont calculés par une propagation et agrégation des scores des nœuds textuels (Sauvagnat, 2005). L'utilisation des nœuds descendants et ascendants est cependant insuffisante pour calculer une représentation de l'image, car d'autres nœuds peuvent contenir des informations pertinentes. Pour cela, nous pouvons par exemple utiliser

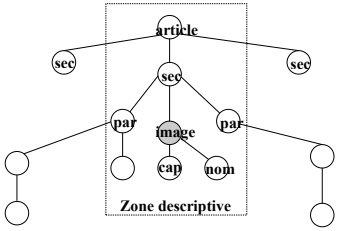

Figure 1. *Présentation d'une zone descriptive d'une image*

aussi les nœuds frères avec leurs nœuds descendants. Dans certains documents, la détermination de la *zone descriptive* d'une image par les nœuds descendants, les nœuds ascendants et les nœuds frères est suffisante (zone descriptive de niveau 1). Mais dans d'autres documents, la définition de la zone descriptive en utilisant seulement ces nœuds est insuffisante, comme le montre l'article de (Kong *et al.,* 2005).

Kong and Lalmas dans (Kong *et al.,* 2005) ont divisé tout le contenu textuel du document en plusieurs *Region Knowledge*[1] *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 déja utilisé ; *2nd ancestor level RK*,..., *Nth ancestor level RK*. Des expérimentations faites avec la collection INEX LonelyPlanet[2] basées sur le modèle vectoriel ont montré que l'utilisation du *Self level RK* seulement ainsi que du *Sibling level RK* seulement ne donne pas de bons résultats en fonction de la précision moyenne. Ceci peut être expliqué par deux raisons : (1) ces *RKs* contiennent généralement quelques mots, donc la probabilité de trouver des termes similaires aux termes de la requête est faible et (2) le vocabulaire est généralement très spécifique à la description de l'image alors que le vocabulaire de la requête peut être plus général. Par contre, l'utilisation des *RKs* de premier, deuxième et troisième ancêtre a conduit à de bons résultats, et ceci peut être expliqué par : (1) ces régions sont plus riches textuellement : il y a une probabilité plus élevée de trouver les termes de la requête dans ces régions et (2) le vocabulaire utilisé dans ces régions ressemble au vocabulaire utilisé dans la requête. Enfin, l'utilisation des nœuds descendants des plus hauts ancêtres n'a pas donné de bons résultats car il n'y a pas de nœuds images dans ces hauts niveaux.
Ces résultats sont certes liés à la collection utilisée mais nous constatons bien l'importance du choix des régions de connaissance.
En se basant sur ce qui précède, et sur les résultats des premières expériences faites avec la *zone descriptive de niveau 1* à INEX 2006 (Hlaoua *et al.,* 2006), nous proposons d'élargir la *zone descriptive* utilisée pour le calcul d'une représentation de l'image. Au lieu d'utiliser les descendants du premier ancêtre, nous remontons d'un

1. Le contenu textuel de l'objet multimédia et des élements l'entourant hiérarchiquement.
2. Initiative for the Evaluation of XML Retrieval. http ://inex.is.informatik.uni-duisburg.de/

niveau et nous utilisons les descendants du deuxième ancêtre (zone descriptive de niveau 2). Si cette zone est encore insuffisante, nous remontons d'un autre niveau (zone descriptive de niveau 3), etc. (cf. figure 2)

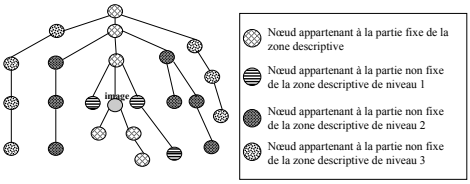

Figure 2. *Différentes parties des zones descriptives*

Nous ne pouvons pas décider d'une manière exacte et déterministe la *zone descriptive* d'une image car elle dépend fortement de la collection utilisée. Nous proposons alors d'utiliser un paramètre p_1 pour le niveau de la zone descriptive : *Zone descriptive de niveau (p_1)* avec $0 < p_1 < n$ et n le nombre de niveaux entre l'image et le nœud racine.

3. Proposition d'une mesure de représentation de l'image

Une fois que la *zone descriptive* est définie, il faut calculer une représentation de l'image en fonction des nœuds de la *zone descriptive* avec l'intuition que les nœuds descendants participent plus que les nœuds frères et les nœuds frères participent plus que les nœuds ascendants dans cette représentation. Le document XML peut être représenté par un arbre, donc nous pouvons le considérer comme une ontologie très simplifiée où les nœuds sont les concepts qui sont organisés hiérarchiquement avec la relation ***est partie de***. Par exemple, l'élément *section* ***est partie de*** *article*.
L'idée est alors de se servir d'une mesure de similarité sémantique entre les termes d'une ontologie pour calculer combien chaque nœud, appartenant à la *zone descriptive*, peut participer à la représentation de l'image. On considère le nœud *image* comme un concept *C1* et le nœud à utiliser comme un autre concept *C2*. Dans le domaine des ontologies, il existe plusieurs mesures pour calculer la similarité sémantique entre les concepts. Selon (Lin, 1998), la mesure la plus simple à implémenter et la plus performante est celle de Wu-Palmer (Wu *et al.,* 1994) :

$$Sim(C1, C2) = \frac{2*N_3}{N_1+N_2+2*N_3},$$

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'élement racine.
Dans la figure 3, la similarité entre *I* et son frère *B* est : $Sim(I, B) = \frac{2*3}{1+1+2*3} = 0.75$.
La similarité entre *I* et son fils *L* est : $Sim(I, L) = \frac{2*4}{0+3+2*4} = 0.72$.
(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 (dans notre exemple : *Sim (I,L)<Sim (I,B)*) alors qu'elle envisage ramener tous les fils d'un concept avant ses frères.
Pour cela (Zargayouna, 2004) a proposé de pénaliser les frères en ajoutant une fonction *spec (C1,C2)*, qui calcule la spécifité de deux concepts par rapport au concept le plus bas (*bottom*) (cf. figure3) :

$$spec(C1, C2) = depth_b(C) * distance(C, C_1) * distance(C, C_2),$$

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

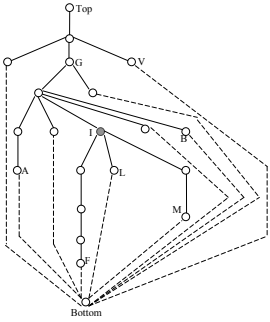

Figure 3. *Relations entre les concepts et Bottom*

Ainsi la mesure de similarité devient : $Sim(C1, C2) = \frac{2*N_3}{N_1+N_2+2*N_3+spec(C1,C2)}$.
Dans notre exemple, *Sim(I,L)* devient supérieur à *Sim(I,B)* :

$$Sim(I, L) = \frac{2*4}{0+3+2*4+4*1*0} = 0.72. ; Sim(I, B) = \frac{2*4}{0+3+2*4+5*1*1} = 0.61.$$

Cette nouvelle mesure ne convient pas encore à nos objectifs car nous pouvons avoir la similarité entre un concept et son ascendant supérieure à la similarité entre ce concept et son frère alors que nous souhaiterions que les concepts frères participent plus que les concepts ascendants dans la représentation. Par exemple dans la figure 3 :

$$Sim(I, G) = \frac{2*2}{2+0+2*2+6*2*0} = 0.66 ; Sim(I, B) = \frac{2*3}{1+1+2*3+5*1*1} = 0.46.$$

Nous proposons alors de pénaliser les concepts ascendants en intégrant un paramètre *(Niv)* : si les deux concepts ont la même profondeur alors *Niv*=0, sinon *Niv*=1.
Puisque ce nouveau paramètre va influencer aussi sur les fils (*Niv* entre un concept et son fils est 1), nous le multiplions par $depth_b(C)$ pour garantir que les concepts fils sont toujours les premiers retournés. La mesure devient donc :

$$Sim(C1,C2) = \frac{2*N_3}{N_1+N_2+2*N_3+spec(C1,C2)+Niv*depth_b(C)}.$$

Cette mesure sera nommée *Rep(I,N)* au lieu de *Sim(C1,C2)*, puisque nous travaillons dans le cadre de représentation d'un nœud image *I* par un nœud *N* appartenant à la *zone descriptive*. Nous la multiplions aussi par le score du nœud *N* qui est déja calculé par le système XFIRM (W_N) (Sauvagnat, 2005). Selon l'exemple de la figure 3 :

$$Rep(I,G) = \frac{2*2}{2+0+2*2+6*2*0+6*1} * W_G = 0.33 * W_G \text{ (nœud ascendant)},$$
$$Rep(I,B) = \frac{2*3}{1+1+2*3+5*1*1+5*0} * W_B = 0.46 * W_B \text{ (nœud frère)},$$
$$Rep(I,L) = \frac{2*4}{0+1+2*4+4*0*1+1*4} * W_L = 0.61 * W_L \text{ (nœud descendant)}.$$

Pour le score final de l'image, nous pouvons par exemple additionner toutes les représentations des nœuds.

4. Conclusion

Dans cet article, nous nous sommes interessés à la recherche contextuelle d'images dans les documents XML. Pour cela, nous avons proposé d'utiliser une *zone descriptive* (nœuds servant à représenter l'image) qui peut être paramétrable selon la collection. Une fois que cette *zone descriptive* est définie, une mesure qui calcule combien chaque nœud pertinent participe à la représentation de l'image est utilisée. Cette mesure est inspirée de la mesure de Wu-Palmer(Wu *et al.,* 1994) pour la similarité des termes dans une ontologie, adaptée par (Zargayouna, 2004) à la mesure de similarité entre documents XML. Nous envisageons maintenant d'évaluer ces propositions sur la collection de test INEX.

5. Bibliographie

Hlaoua L., Torjmen M., Pinel-Sauvagnat K., Boughanem M., « XFIRM at INEX 2006 - Preliminary work. Ad-hoc, Relevance Feedback and MultiMedia tracks », *INEX, Dagstuhl, Allemagne*, 2006.

Kong Z., Lalmas M., « Integrating Xlink and Xpath to retrieve structured multimedia documents in digital librairies », *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.

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.

Wu Z., Palmer M., « Verb semantics and lexical selection », *Proceedings of the 23nd Annual Meetings of the Associations for Computational Linguistics*, p. 133-138, 1994.

Zargayouna H., « Contexte et sémantique pour une indexation de documents semi-structrés », *COnference en Recherche d'Information et Applications*, p. 571-581, Mars, 2004.