

Formalisation de règles d'indexation MeSH pour un usage automatique

A. Névéal^{a, b}, F. Florea^{a, b}, B. Thirion^a et SJ. Darmoni^{a, b}

^a CISMéF & CGSIS, CHU de Rouen-1, rue de Germont - 76031 Rouen cedex.

^b Laboratoire PSI, FRE 2645, CNRS - INSA & Université de Rouen

Introduction

Afin de faciliter l'accès des professionnels de santé, des étudiants en médecine et des patients à une information de qualité en santé, le Catalogue et Index des Sites Médicaux Francophones (CISMéF[1]) sélectionne, recense et décrit les principaux sites institutionnels dans le domaine de la santé. A l'heure actuelle, ces tâches sont effectuées manuellement. L'objectif de nos travaux est l'indexation automatique de ressources de santé à l'aide de termes du vocabulaire contrôlé Medical Subject Headings (MeSH) - mots clés et qualificatifs - et de types de ressources CISMéF [2] afin d'étendre la couverture du catalogue. Ainsi, nous avons constitué un dictionnaire électronique pouvant être utilisé par un système automatique. Nous poursuivons ce travail par la constitution d'une bibliothèque de graphes (transducteurs) traduisant des règles d'indexation portant (1) sur l'utilisation de paires mot clé/qualificatif MeSH et (2) sur l'utilisation de triplets mot clé/qualificatif \ type de ressource. Ce dernier type de règle intervient notamment dans l'indexation de ressources contenant à la fois du texte et des images: les types de ressource concernés ont trait à l'imagerie médicale, comme par exemple <radiographie> ou <échographie>.

Constitution des graphes

La recherche des règles d'indexation s'effectue par l'intermédiaire d'une étude de corpus indexé en compagnie d'un expert en documentation médicale. L'expert a pour tâche d'expliquer les raisons qui ont conduit au choix des descripteurs (paires ou triplets), en pointant les éléments textuels de la ressource qui les soutiennent. A partir d'exemples successifs, le cas échéant, il est possible d'extraire une règle d'indexation suffisamment générale pour donner lieu à une formalisation, puis à la construction d'un graphe INTEX/NOOJ. Ce travail a été entrepris pour décrire l'emploi des 83 qualificatifs du MeSH. La Figure 1 présente le graphe obtenu pour l'emploi du qualificatif <utilisation>. Nous souhaitons également l'étendre à la description de l'emploi de certains types de ressources de l'imagerie médicale. Ciblée sur les légendes accompagnant les images, l'extraction de triplets aura le double avantage de conforter le diagnostic de

reconnaissance de la modalité des images (ou type de ressource) établi à l'aide de la méthode introduite dans [3] et de proposer une indexation fine de la ressource. Ainsi, en combinant l'analyse du texte à l'analyse des attributs d'image (texture et statistiques) [3], nous progressons vers une véritable indexation combinée texte/image. La Figure 2 présente un graphe préliminaire, décrivant l'emploi de types de ressources affiliés pour l'indexation.

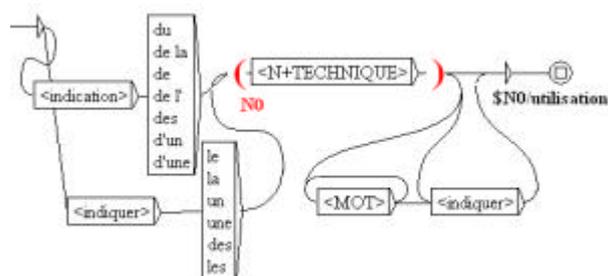


Figure 1: Graphe décrivant l'emploi du qualificatif <utilisation>



Figure 2: Graphe préliminaire décrivant certains emplois des types de ressource affiliés

References

- [1] Darmoni SJ, Leroy JP, Thirion B, Baudic F, Douyère M and Piot J. CISMéF: a structured Health resource guide. *Meth Inf Med* 2000; 39(1): 30-5.
- [2] Douyère M, Soualmia LF., Névéal A., Rogozan A., Dahamna B., Leroy JP., Thirion B., Darmoni SJ. (2004) Enhancing the MeSH thesaurus to retrieve French online health resources in a quality-controlled gateway. *Health Info Libr J.* 2004 Dec;21(4):253-6.
- [3] Florea FI, Rogozan A, Bensrhair A and Darmoni SJ. Medical image retrieval by content and keyword in a on-line health-catalogue context, *Proc. Mirage* 2005 : 229-36