

Extraction d'informations médicales au LIMSI

Cyril Grouin¹, Louise Deléger¹, Anne-Lyse Minard¹,
Anne-Laure Ligozat¹, Asma Ben Abacha¹, Delphine Bernhard¹,
Bruno Cartoni², Brigitte Grau¹, Sophie Rosset¹, Pierre Zweigenbaum¹
(1) LIMSI-CNRS, BP133, 91403 Orsay Cedex, France
(2) Département de Linguistique, Université de Genève, Suisse
{prenom.nom}@limsi.fr, bruno.cartoni@unige.ch, louise.deleger@cchmc.org

Les textes présents dans les dossiers de patients dans les hôpitaux contiennent des informations précieuses pour diverses tâches médicales comme par exemple la prise de décision diagnostique et thérapeutique. Il est de ce fait utile de chercher à extraire automatiquement ces informations. Récemment, des évaluations internationales annuelles de systèmes d'extraction d'informations à partir de documents cliniques en anglais ont été organisées par l'institut i2b2 aux États-Unis (Uzuner *et al.*, 2010). Cette démonstration présente de façon concrète les outils mis en place au LIMSI pour participer aux évaluations i2b2 2009 et i2b2/VA 2010. Ces outils prennent en entrée les phrases d'un texte en anglais, et utilisent au besoin les annotations réalisées par des outils précédemment appliqués, pour détecter les informations requises, fournies sous la forme d'annotations déportées. Résultats en 2009 : F=0.773, 8^e/20 ; en 2010 : F=0.773, 12^e/22 (piste 1), F=0.931, 5^e/21 (piste 2), F=0.709, 3^e/16 (piste 3).

La tâche 2009 consistait à extraire les prescriptions médicamenteuses (médicament, dosage, fréquence, quantité, mode d'administration, durée, et raison de la prescription). Notre méthode repose sur des lexiques et l'observation des régularités d'expression de ces prescriptions, implémentées sous la forme de patrons fondés sur les classes sémantiques correspondant aux types d'informations recherchées (Deléger *et al.*, 2010; Grouin *et al.*, 2011).

Trois tâches étaient proposées en 2010 : (i) détection de trois types de concepts médicaux : problèmes (maladies, symptômes), traitements (y compris médicamenteux) et examens ; (ii) détermination du statut d'un problème médical (présent, absent, hypothétique, etc.) ; et (iii) détection de sept types de relations entre problèmes, traitements et examens. Nous avons mis en place des méthodes à base de connaissances expertes et de TAL (lexiques, patrons, étiquetage morphosyntaxique, analyse syntaxique) ainsi que des méthodes à base d'apprentissage supervisé (SVM, CRF) entraînés sur les corpus annotés fournis par les organisateurs de la campagne (Bernhard & Ligozat, 2011; Grouin *et al.*, 2011; Minard *et al.*, 2011a,b), et cherché à combiner ces deux types de méthodes.

Ce travail a été partiellement financé par les projets Akenaton (ANR-07-TecSan-001), InterSTIS (ANR-07-TecSan-010) et Quæro (financement Oseo, agence française pour l'innovation et la recherche).

Références

- BERNHARD D. & LIGOZAT A.-L. (2011). Analyse automatique de la modalité et du niveau de certitude : application au domaine médical. In *TALN 2011*, Montpellier. À paraître.
- DELÉGER L., GROUIN C. & ZWEIGENBAUM P. (2010). Extracting medical information from narrative patient records : the case of medication-related information. *J Am Med Inform Assoc*, **17**, 555–558.
- GROUIN C., DELÉGER L., CARTONI B., ROSSET S. & ZWEIGENBAUM P. (2011). Accès au contenu sémantique en langue de spécialité : extraction des prescriptions et concepts médicaux. In *TALN 2011*. À paraître.
- MINARD A.-L., LIGOZAT A.-L., BEN ABACHA A., BERNHARD D., CARTONI B., DELÉGER L., GRAU B., ROSSET S., ZWEIGENBAUM P. & GROUIN C. (2011a). Hybrid methods for improving information access in clinical documents : Concept, assertion, and relation identification. *J Am Med Inform Assoc*. À paraître.
- MINARD A.-L., LIGOZAT A.-L. & GRAU B. (2011b). Extraction de relations dans des comptes rendus hospitaliers. In *Actes des 22emes Journées francophones d'Ingénierie des Connaissances*.
- UZUNER O., SOLT I. & CADAG E. (2010). Extracting medication information from clinical text. *J Am Med Inform Assoc*, **17**(5), 514–518.