

Regroupement des Définitions de Sigles Biomédicaux

Ousmane Djanga, Hanine Hamzioui, Mickaël Hatchi, Isabelle Mougenot, Mathieu Roche

LIRMM, Université Montpellier 2 – CNRS UMR5506

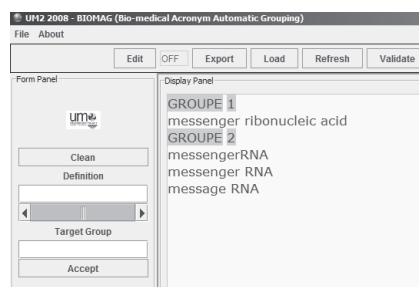
Résumé. L'application présentée permet de regrouper les définitions de sigles issues des sciences du vivant par des mesures de proximité lexicale (approche automatique) et une intervention de l'expert (approche manuelle).

Le logiciel présenté constitue un premier effort d'exploitation des sigles et de leurs définitions dans les sciences du vivant. Il s'agit de regrouper les définitions des sigles sémantiquement proches (Okazaki et Ananiadou (2006)). Ainsi, notre logiciel applique une mesure de proximité lexicale (formule (1)) décrite par Maedche et Staab (2002) qui s'appuie sur une distance d'édition (notée E). Cette dernière calcule la somme minimale des opérations (suppression, insertion et remplacement) pour transformer l'une des deux chaînes de caractères en l'autre.

$$Str(ch1, ch2) = \max\left\{0, \frac{\min\{|ch1|, |ch2|\} - E(ch1, ch2)}{\min\{|ch1|, |ch2|\}}\right\} \in [0, 1] \quad (1)$$

Par exemple, les définitions "zona occludens" et "zonula occludens" propres au sigle ZO ont une distance d'édition égale à 2 (deux insertions) permettant d'obtenir une mesure de proximité égale à $\max\left\{0, \frac{14-2}{14}\right\} = 0.85$. Cette valeur élevée suggère que nous pouvons regrouper automatiquement ces deux définitions.

L'apport de notre logiciel développé en Java porte ensuite sur la capacité donnée à l'utilisateur d'effectuer des changements de groupes déterminés automatiquement (déplacement d'une définition vers un autre groupe ou création de nouveaux groupes). La capture d'écran ci-contre montre une partie de l'interface graphique proposée aux utilisateurs.



L'objectif à court terme est donc de proposer un logiciel offrant des facilités d'exploitation des sigles couramment manipulés par les biologistes. À plus long terme, les vocabulaires et les ontologies des sciences du vivant pourraient être mis à contribution afin d'ajouter des fonctionnalités de partage et de diffusion de l'information.

Références

- Maedche, A. et S. Staab (2002). Measuring similarity between ontologies. In *Proc. of the European Conference on Knowledge Acquisition and Management - EKAW*, pp. 251–263.
- Okazaki, N. et S. Ananiadou (2006). A Term Recognition Approach to Acronym Recognition. In *Proceedings of ACL*, pp. 643–650.