

HMIN 315 - mini projet s de TP - 2 sujets au choix - en prendre un.

Extracteur de relations sémantiques

Concevoir un outil qui va chercher des pages textuelles (genre wikipédia) et extrait les relations sémantiques présentes dans ces pages.

Exemple, soit le texte suivant :

La **rate** est un organe profond, situé dans l'hypocondre gauche en regard de la 10^e côte (*côte splénique*), accolé à la face latérale de l'estomac, la grande courbure. Elle est donc en position *thoraco-abdominale*. Elle joue un rôle dans l'immunité et dans le renouvellement des cellules sanguines.

=>

rate r_isa organe

rate r_isa organe profond

rate r_lieu hypocondre gauche

rate r_associated immunité

rate r_associated renouvellement des cellules sanguines

* utilisation de patrons lexicaux

\$s est un type de \$y => \$x r_isa \$y

* avec contraintes :

\$x possède un \$y / \$x r_isa être vivant & \$y r_isa organe => r_has_part

\$x possède un \$y / \$x r_isa être humain \$y r_isa artéfact => r_own

Liste de patrons : <http://www.jeuxdemots.org/rezo-ask.php?text=1&templates=1>

Liste des relations de JDM : <http://www.jeuxdemots.org/jdm-about-detail-relations.php>

Se baser sur les K de JDM. Extraire les sous-parties de la KD avec :

<http://www.jeuxdemots.org/rezo-dump.php>

* extraire aussi sans schéma explicite : la belle maison => maison r_carac belle

* identifier statistiquement de nouveaux schémas en se basant sur des chaînes de caractères fréquentes et la connaissance de la KB (JDM)

* l'outil doit être rapide

* gérer la coréférences (pronom, adjectif possessifs)

Le tigre a des dents. Elles sont acérées.

=> dent r_carac acérée / dents r_carac acérées / dents de tigre r_carac acérées

mais aussi : tigre r_has_part dent/dents

Pièges divers :

la chatte de la voisine est dans le jardin

* => voisine r_lieu jardin

il faut bien identifier le \$x et le \$y

le cambrioleur utilise un pied de biche pour rentrer

=> cambrioleur r_make_use_of pied

il faut bien identifier les termes composés

Winograd

A partir d'un (petit) texte, pouvoir répondre à des questions :

Lancelot se rue sur le dragon

> Qui se rue sur le dragon ?

=> Lancelot

> Avec quoi Lancelot fait cela ?

=> probablement avec une épée

=> probablement avec un kalasnikov (mauvais)

=> probablement avec un objet (trop général)

=> probablement avec une arme (ok)

> Pourquoi ?

Lancelot est un chevalier, un chevalier a une épée (probablement), une épée est une arme, on peut attaquer avec une arme

Analyser le texte

Analyser les questions

Complémenter l'analyse de texte si nécessaire

Fournir les réponses sous forme textuelle

En interne, l'analyse extrait des relations. Une question est traduit en une relation à variable(s). Correspondance avec le graphe de relations. Augmenter les relations par inférence en fonction de la question.

Se baser sur les K de JDM. Extraire les sous-parties de la KD avec :

<http://www.jeuxdemots.org/rezo-dump.php>

Pour trouver des termes composés à partir du terme x => rezo dump avec x et relation type 11