

Outil d'extraction de phrases à mots polysémiques

sujet de TER L3 informatique 2014-2015

Encadrant : Mathieu Lafourcade, LIRMM, lafourca@lirmm.fr

Résumé : L'objet de ce projet est de modéliser et implémenter un outil d'extraction de phrases à partir de textes du Web. L'utilisateur pourra utiliser cet outil en indiquant un terme, que l'outil considèrera comme élément de recherche sur des sites d'information (Le Monde, Google news, etc.). Un certain nombre de phrases en français contenant le mot demandé seront retournées. On tiendra compte des **différentes formes que le mot peut prendre** (pluriel, féminin, formes conjuguées).

Le formulaire devra pouvoir être utilisé par programme et le résultat retourné décodable automatiquement. Il y a donc deux usages à cet outil, un manuel et un pouvant être automatisé.

La liste des sites à fouiller devra être configurable.

Programmation en php, mysql, javascript,...

Prérequis : PHP, très bonne maîtrise du français

Ressources utilisable : dictionnaire ABU <http://abu.cnam.fr/> et lexique/réseau de JeuxDeMots : <http://www.jeuxdemots.org/jdm-about.php>