

Corpus et méthode expérimentale : projet

Florian Boudin

8 septembre 2014

Méthodologie à suivre

1. Lecture d'articles
 - (a) Kim et al., *Baldwin Semeval-2010 task 5 : Automatic keyphrase extraction from scientific articles*
<http://www.aclweb.org/anthology/S10-1004>
 - (b) Kim et al., *Evaluating N-gram based Evaluation Metrics for Automatic Keyphrase Extraction*
<http://www.aclweb.org/anthology/C10-1065>
 - (c) Hasan and Ng, *Automatic Keyphrase Extraction : A Survey of the State of the Art*
<http://www.aclweb.org/anthology/P14-1119>
 - (d) Bougouin et al., *TopicRank : Graph-Based Topic Ranking for Keyphrase Extraction*
<http://www.aclweb.org/anthology/I13-1062>
2. Analyse du corpus WikinewsKeyphraseCorpus
 - <https://github.com/adrien-bougouin/WikinewsKeyphraseCorpus>
 - Caractéristiques du corpus : nature des documents, droits de redistribution, taille, etc.
 - Annotation du corpus : *guidelines*, points faibles
3. Modification du corpus
 - Proposition de nouvelles *guidelines*
 - Modification du corpus existant et ajout de nouveaux documents
4. Valorisation du corpus créé
 - e.g. article soumis à LREC, RECITAL