

# Création de corpus : contextes de citations

Florian Boudin

Corpus et méthode expérimentale - 2013

## Méthodologie à suivre

1. Lecture d'articles
  - Radev et Abu-Jbara, *Rediscovering ACL discoveries through the lens of ACL anthology network citing sentences*, <http://www.aclweb.org/anthology-new/W/W12/W12-3201.pdf>.
  - Teufel, Siddharthan et Tidhar, *Automatic classification of citation function*, <http://www.aclweb.org/anthology-new/W/W06/W06-1613.pdf>.
  - Qazvinian et Radev, *Identifying Non-Explicit Citing Sentences for Citation-Based Summarization*, <http://www.aclweb.org/anthology-new/P/P10/P10-1057.pdf>.
  - Elkiss, Shen, Fader, Erkan, States et Radev. *Blind men and elephants : What do citation summaries tell us about a research article ?*, [http://homes.cs.washington.edu/~afader/bib\\_pdf/jasist08.pdf](http://homes.cs.washington.edu/~afader/bib_pdf/jasist08.pdf)
2. Etat de l'art des corpus existants
  - e.g. *ACL Anthology Network*
3. Définir les caractéristiques du corpus
  - Nature des documents, droits de redistribution, taille, etc.
4. Annotation du corpus
  - Définition des *guidelines* puis test et modification
5. Valorisation du corpus créé
  - e.g. article soumis à LREC, RECITAL