

Presentation de rssint

Anas Alaoui Romain Bressan

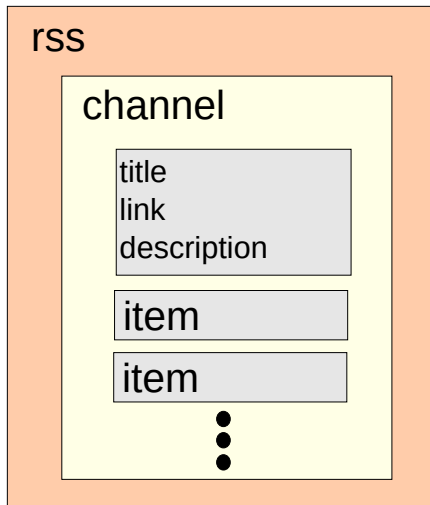
7 janvier 2016

Objectifs

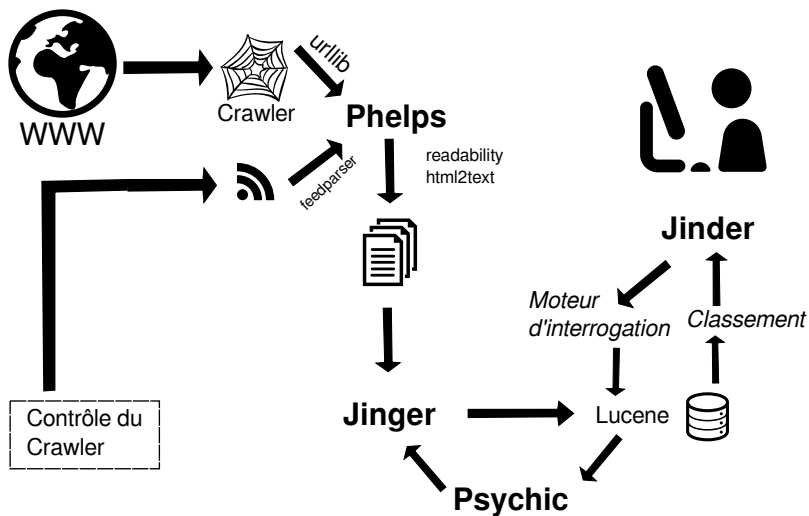
Réaliser un système de veille sur Internet qui exploite des flux RSS.

- Lecteur de flux RSS et ATOM
- Collecteur : conversion (extraction de texte), encodage, langue, stemmer, boilerplate
- Indexeur : Lucene
- Requeteur : recherche dans l'index
- Classifieur : classer un document à partir des documents déjà classés avec KNN

Flux RSS



Architecture



Dépendances

- feedparser, Python, BSD 2-Clause
- html2text, Python, GNU GPL 3.0
- readability, Python, Apache License 2.0
- lucene, Java, Apache License 2.0

Travail rendu

- Lecteur de flux **RSS** et **ATOM**
- **Collecteur** : conversion (extraction de texte), encodage, langue, stemmer, boilerplate
- **Indexeur** : Lucene
- **Requeteur** : recherche dans l'index
- **Classifieur** : classer un document à partir des documents déjà classés avec KNN

Démonstration