

Unitex vs. Outilex

Matthieu Constant
Université Paris-Est

Le projet Outilex

- Financé par le Ministère de l'Industrie
- Durée : 2002-2006
- Coordinateur : IGM
- 10 partenaires :
 - 4 académiques
 - 3 PME
 - 3 grands groupes industriels
- But : construction d' une plateforme logicielle de traitement automatique des langues naturelles

Le projet Outilex à l'IGM

- Développement du cœur de la plateforme
 - automates finis et réseaux récurrents de transitions
 - ressources linguistiques
 - formats XML
- Développeur : Olivier Blanc
- Langage : C++
- Logiciels existants : Intex, Unitex, NLTK, Gate, FSM, ...

Outilex vs. Unitex

Caractéristiques communes

- Même chaîne de traitement
 - tokenisation
 - application de ressources lexicales
 - application de grammaires
- Applications :
 - Annotation de textes
 - Concordancier

Outilex vs. Unitex

Différences

- Ressources linguistiques :
 - Format XML pour Outilex
- Objet en entrée des grammaires :
 - Unitex : texte
 - Outilex : automate du texte
- Application des grammaires :
 - Unitex : recherche et annotation de formes
 - Outilex : vrai parseur (Earley)

Perspectives

- Maintenance de deux logiciels en parallèle difficile !
- Vers la fusion des fonctionnalités d'Outilex dans Unitex