

Interrogations en langue naturelle

Projet M1

Ludovic Bonnefoy Romain Deveaud

Tutoré par Marc El-Bèze et encadré par Eric Charton

Jeudi 18 juin 2009

Introduction

La recherche d'information, le langage naturel et NLGbAse

Moteurs de recherche intégrant la sémantique

Présentation de NLGbAse

Algorithmes déployés

Catégorisation d'une question

Extraction de mots-clés

Conclusion

Introduction

► Trululu

Moteurs de recherche intégrant la sémantique

- Google, Powerset, Hakia...

Moteurs de recherche intégrant la sémantique

- ▶ Google, Powerset, Hakia...
- ▶ Algorithmes ayant recours à des sources extérieures.

Moteurs de recherche intégrant la sémantique

- ▶ Google, Powerset, Hakia...
- ▶ Algorithmes ayant recours à des sources extérieures.
- ▶ Enrichissement et activité communautaire indispensables pour la validité et la récence des informations.

Moteurs de recherche intégrant la sémantique

- ▶ Google, Powerset, Hakia...
- ▶ Algorithmes ayant recours à des sources extérieures.
- ▶ Enrichissement et activité communautaire indispensables pour la validité et la récence des informations.
- ▶ NLGbAse : base de données classifiée (ontologie) issue de Wikipédia.

Présentation de NLGbAse

- Trois outils de recherche d'informations.

Présentation de NLGbAse

- ▶ Trois outils de recherche d'informations.
- ▶ Un moteur "classique", prenant en entrée des mots-clés et appliquant un algorithme de compacité.

Présentation de NLGbAse

- ▶ Trois outils de recherche d'informations.
- ▶ Un moteur "classique", prenant en entrée des mots-clés et appliquant un algorithme de compacité.
- ▶ Un moteur "sémantique", reprenant le même algorithme que le précédent, mais permettant de sélectionner les résultats appartenant à une catégorie sémantique précise.

Présentation de NLGbAse

- ▶ Trois outils de recherche d'informations.
- ▶ Un moteur "classique", prenant en entrée des mots-clés et appliquant un algorithme de compacité.
- ▶ Un moteur "sémantique", reprenant le même algorithme que le précédent, mais permettant de sélectionner les résultats appartenant à une catégorie sémantique précise.
- ▶ Un moteur "extracteur d'informations", basé sur un algorithme de compacité, permettant d'obtenir une information précise éventuellement contenu dans un document.

Conclusion

► Trululu