

TER

MultiWiki : une approche pour la création automatique de deux taxonomies intégrées aux p

February 4, 2017

1 Vocabulaire

- taxonomie : de décrire les organismes vivants et de les regrouper en entités appelées taxons afin de les identifier puis les nommer et enfin les classer et de les reconnaître via des clés de détermination dichotomiques.
- hypernomie : relation sémantique hiérarchique d'une unité Lexicale à une autre selon laquelle l'extension du premier terme, plus général, englobe l'extension du second, plus spécifique.

2 Objectif

créer un ensemble de taxonomies dans un langage arbitraire. On construit en anglais puis on projette vers d'autres langues. Le procédé utilise un nouvel algorithme qui exploite les informations disponibles dans des taxonomies pour renforcer la création de nouvelles taxonomie. L'information taxonomique dans MultiWiki est caractérisé par une qualité plus haute et une plus grande couverture que les ressources comme DBpedia, Yago, Menta, Wikinet, LHD et wiki-Taxonomie ou d'autres langages. Multiwiki est disponible en ligne. (Création d'un outil d'extraction dans les pages wikipedia)

3 Introduction

(nécessité du sens des mots non ambigu)

3.1 des logiciels comme wikipédia

permettent de créer un réseau de catégories wikipédia

3.2 DBpedia

cartographie manuellement les infoBoxes de wikipedia à travers les langages pour créer des ontologies

3.3 Menta

combine des informations taxonomiques du monde d'internet avec wikipédia (infoBoxes et catégories)

3.4 LHD

crée des liens basiques quoique généraux basé sur des règles.

3.5 Cependant

le type de savoir extrait par ces approches est partiel, incomplet et hétérogène.

3.6 Les plus de MultiWiki

- utiliser les infoBoxes et les pages wikipédia
- plusieurs langages possibles
- méthode qui dépasse la barrière langagière
- (peut même répertorier des concepts spécifiques à une culture)

4 arrière-plan et contributions

4.1 l'idée clef de la bitaxonomie de wikipédia

permet d'utiliser les outils fournis par wikipédia comme les pages wikipédia, les catégories, les liens vers d'autres pages, les inventaires des sens (révèle un problème de DBpédia)

4.2 une bitaxonomie

est une paire $B = (T_P, T_C)$ de taxonomie avec

- T_P : la taxonomie de la page wikipédia
- T_C : la taxonomie de catégories de wikipédia

On peut donc créer un graphe $T_P(T_C)$: l'ensemble des arêtes hypernomiques découvert par cet algorithme.

4.3 les différentes étapes :

pour se débarrasser de la barrière linguistique, il faut acquérir une bitaxonomie en anglais et ensuite projeter l'information dans d'autres langages.

4.4 cet article parle de

- nouvel algo d'extraction
- utilisation de la bitaxonomie
- nouvelle approche de traduction
- nouvelle approche au-delà du langage
- meilleures fouilles de données

5 création de la bitaxonomie de wikipédia en anglais

étapes de la bitaxonomie :

- création d'une page initiale de taxonomie
- création de la bitaxonomie
- raffinement de la bitaxonomie

6 phase 1 : création de la taxonomie de la page

6.1 Objectif

introduire la taxonomie simple de pages wikipédia. Soit P : ensemble de toutes les pages wikipédia. Soit $T_P = (P, E)$: le graphe orienté avec

- noeuds=pages
- arêtes=vide

Pour chaque $p \in P$ (pages), on identifie la génération $p_h \in P$. Pour créer une arête (p, p_h) et l'ajouter à E .

6.2 Exemple :

$E = E \cup \text{apple, fruit}$. Pour le faire, on procède en 2 étapes.

- extraction d'hypernomie
- rendre l'hypernomie non ambiguë

7 évaluation taxonomique des pages

7.1 évaluation des mesures (pouvoir mesurer la qualité d'une taxonomie)

7.2 données des pages : 1000 pages wikipédia

7.3 ordre des liens hypernomiques

7.4 résultats

8 phase2 : introduire la bitaxonomie p78 algo trop cool

8.1 algorithme bitaxonomie (voir détails)

8.2 initialisation $B = (T_P, T_C)$

8.3 les 4 étapes

8.4 paramètre modifié et condition d'arrêt

8.5 paramètre accordé

9 phase3 : raffinement de la bitaxonomie

9.1 les problèmes

- nécessité d'une catégorie associée à une page wikipédia
- fournir un lemme hypernomique

Donc cet algorithme ne peut pas être appliqué partout. Donc le raffinement sert à prendre en charge les "oubliés" de l'algorithme.

9.2 évaluation de la bitaxonomie anglaise

9.3 travail lié

9.4 évaluation comparative

(pourquoi MultiWiki est le meilleur)

9.5 projeter la bitaxonomie

règle de projection : $X_E \text{is-a } Y_E \wedge X_E \| X_F \wedge Y_E \| Y_F \Rightarrow X_F \text{is-a } Y_F$
 $\forall X_E, Y_E \in T_E,$
 $X_F, Y_F \in T_F$

9.6 évaluation multilingues

9.7 conclusion