

Couplage d'espaces sémantiques et de graphes pour le Deft 2011 : une approche automatique non supervisée

V. Hoareau, M. Ahat, C. Peterman & M. Bui

Laboratoire CHArt – Université de Paris 8, Ecole Pratique des Hautes Etudes

1er juillet 2011

1 Introduction

- L'objectif de notre contribution
- Les espaces sémantiques
- Le couplage espace sémantique et graphe
- Un exemple d'application

2 Méthode

- Principe générale
- Du graphe complet à la relation bi-univoque
- La chaine de traitement et l'apprentissage

3 Résultats

4 Discussion et Conclusion

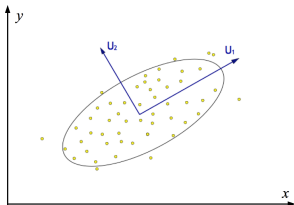
References20

L'objectif de notre contribution

- Représenter un espace sémantique par un graphe disposant des mêmes propriétés sémantiques
- Proposer des algorithmes reposant sur les graphes (et non plus seulement sur les espaces sémantiques)
- Proposer un algorithme non-supervisé pour attribuer un résumé à un article

Les espaces sémantiques

- 1 Des objets mathématiques qui représentent des mots, des documents, des concepts, des utilisateurs dans un espace multi-dimensionnel.
- 2 Des *implémentations* de l'hypothèse distributionnelle :
 - "Des termes qui ont un sens similaire, apparaissent dans un contexte similaire"
 - Des termes qui ont un sens similaires, sont représentés par des vecteurs similaires ou dans les même régions de l'espace (appelées *aires sémantiques*).
- 3 *Les méthodes* LSA, HAL, RI, BEAGLE, PLSA, LDA en sont des exemples



Le couplage espace sémantique et graphe

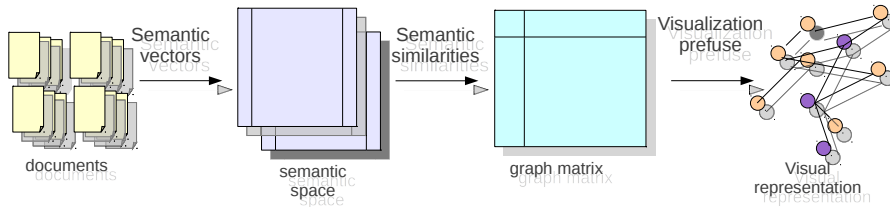
Principe général

- 1 Construire un espace sémantique à partir d'un ensemble de document
- 2 Construire une matrice de connexité à partir des distances mesurées entre chaque documents
- 3 Appliquer différents algorithmes sur la matrice/le graphe

Visualisation d'une base documentaire (Hoareau, Ahat, Medernach, & Bui, 2011)

Navigation entre les documents et les mots par

- relation doc – doc, doc – mot, mot – mot;
- recherche par mot clé;
- visualisation – **Force directed** et **Radial** graphe.



Force directed graphe : vue globale et locale

Corpus Deft 07 – jeux video

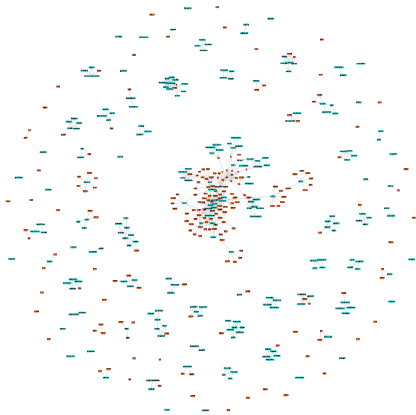


Figure: vue globale

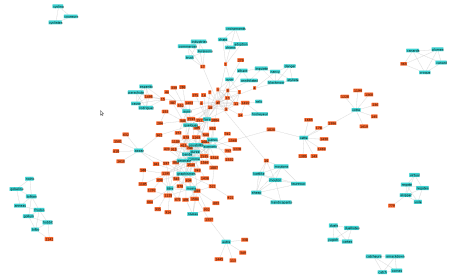


Figure: vue locale sur des clusters

Radial graphe : structure locale et inter-relations

Corpus Inria

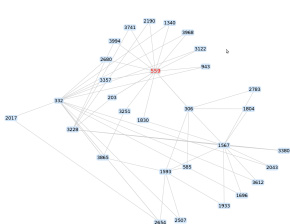


Figure: 30 docs

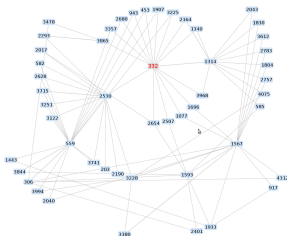


Figure: 50 docs

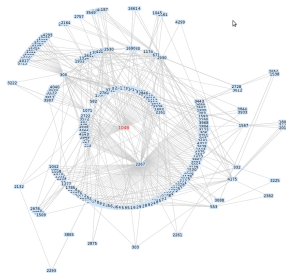


Figure: 200 docs

1 Introduction

- L'objectif de notre contribution
- Les espaces sémantiques
- Le couplage espace sémantique et graphe
- Un exemple d'application

2 Méthode

- Principe générale
- Du graphe complet à la relation bi-univoque
- La chaine de traitement et l'apprentissage

3 Résultats

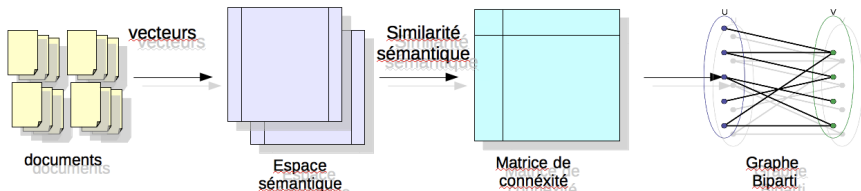
4 Discussion et Conclusion

References20

Couplage espace sémantique et graphe pour le Deft 2011

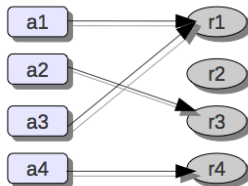
Principe générale

- 1 Construire un espace sémantique à partir de l'ensemble des documents (articles + résumés)
- 2 Construire une matrice de connexité à partir des distances mesurées entre chaque documents
- 3 Construire un graphe complet et le transformer en graphe biparti où chaque résumé n'est associé qu'à un et un seul article



La méthode de réduction du graphe complet

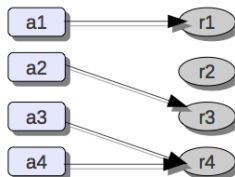
	r1	r2	r3	r4
a1	10	9	8	7
a2	6	8	10	9
a3	9	7	5	8
a4	4	8	6	10



Étape 1 : graphe avec ambiguïté

La méthode de réduction du graphe complet

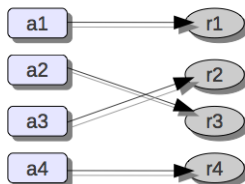
	r1	r2	r3	r4
a1	10	9	8	7
a2	6	8	10	9
a3	0	7	5	8
a4	4	8	6	10



Étape 2 : graphe avec ambiguïté

La méthode de réduction du graphe complet

	r1	r2	r3	r4
a1	10	9	8	7
a2	6	8	10	4
a3	0	7	5	0
a4	4	8	6	10



Étape 3 : graphe sans ambiguïté

La chaîne de traitement et l'apprentissage

La chaîne de traitement

- 1 xml2mysql \Rightarrow script (php)
- 2 mysql2lucene \Rightarrow LuSql^a (java)
- 3 espace sémantique \Rightarrow SemanticVectors^b (Widdows & Cohen, 2010) (java)
- 4 matrices et graphes \Rightarrow java

^a<http://lab.cisti-icist.nrc-cnrc.gc.ca/cistilabswiki/index.php/LuSql>

^b<http://code.google.com/p/semanticvectors/>

L'apprentissage

- 1 Unité d'apprentissage : l'article entier ; le résumé
- 2 Aucun pré-traitement linguistique

1 Introduction

- L'objectif de notre contribution
- Les espaces sémantiques
- Le couplage espace sémantique et graphe
- Un exemple d'application

2 Méthode

- Principe générale
- Du graphe complet à la relation bi-univoque
- La chaine de traitement et l'apprentissage

3 Résultats

4 Discussion et Conclusion

References20

Résultats

- Score pour la piste 1 : 1
- Score pour la piste 2 : 0,995

1 Introduction

- L'objectif de notre contribution
- Les espaces sémantiques
- Le couplage espace sémantique et graphe
- Un exemple d'application

2 Méthode

- Principe générale
- Du graphe complet à la relation bi-univoque
- La chaine de traitement et l'apprentissage

3 Résultats

4 Discussion et Conclusion

References20

Je vous remercie.

- Hoareau, Y. V., Ahat, M., Medernach, D., & Bui, M. (2011). Un outil de navigation dans un espace sémantique. In A. Khenchaf & P. Poncelet (Eds.), *Extraction et gestion des connaissances (egc'2011)* (Vol. RNTI-E-20, p. 275-278). Hermann-Éditions.
- Levorato, V., Van Le, T., Lamure, M., & Bui, M. (2009). Classification prétopologique basée sur la complexité de Kolmogorov. *Studia informatica universalis*, 7, 199–222.
- Widdows, D., & Cohen, T. (2010). The semantic vectors package: New algorithms and public tools for distributional semantics. *International Conference on Semantic Computing*, 0, 9-15.