



Création et validation de signatures sémantiques :
application à la mesure de similarité sémantique et à la
substitution lexicale

Mokhtar Boumedyen BILLAMI, Núria GALA

LIF-CNRS UMR 7279, Aix-Marseille Université

mokhtar.billami@lif.univ-mrs.fr, nuria.gala@univ-amu.fr

Vendredi 23 Juin 2017

Sommaire

- 1 Introduction : représentations sémantiques de mots
- 2 Création de signatures sémantiques
- 3 Évaluation intrinsèque : mesures de similarité sémantique
- 4 Évaluation extrinsèque : substitution lexicale
- 5 Conclusion et perspectives

Représentations sémantiques (1/2)

– Deux types de source d'information :

- ➊ À base de corpus (modèles distributionnels)
 - + *Word embeddings (plongements lexicaux)*
 - LSA [Turney et Pantel, 2010]
 - Word2Vec (CBOW et SKIP-GRAM [Mikolov et *al.*, 2013])
 - GLOVE [Pennington et *al.*, 2014]
- ➋ À base de ressources lexicales et encyclopédiques
 - NASARI utilise BabelNet [Camacho-Collados et *al.*, 2016]
 - SASA utilise Wikipédia [Wu et Giles, 2015]
 - SRW utilise Wiktionnaire [Zesch et *al.*, 2008]

Représentations sémantiques (2/2)

Deux types de similarité sémantiques

Sur les mots eux-mêmes : la similarité est entre les vecteurs de mots tenant compte de l'ensemble des sens possibles

Sur les sens des mots : la similarité entre deux mots est celle de leurs sens les plus proches [Budanitsky et Hirst, 2006]

Représentations sémantiques (2/2)

Deux types de similarité sémantiques

Sur les mots eux-mêmes : la similarité est entre les vecteurs de mots tenant compte de l'ensemble des sens possibles

Sur les sens des mots : la similarité entre deux mots est celle de leurs sens les plus proches [Budanitsky et Hirst, 2006]

+ Nos représentations ?

Représentations sémantiques (2/2)

Deux types de similarité sémantiques

Sur les mots eux-mêmes : la similarité est entre les vecteurs de mots tenant compte de l'ensemble des sens possibles

Sur les sens des mots : la similarité entre deux mots est celle de leurs sens les plus proches [Budanitsky et Hirst, 2006]

+ Nos représentations ?

- À base de la ressource (réseau lexical) [JeuxDeMots](#) [Lafourcade, 2007]
- Utilisation du premier type de similarité sémantique

Modèle de représentation sémantique à base du réseau lexical JeuxDeMots

Signature sémantique d'un élément linguistique

- Forme spéciale du modèle d'espace vectoriel (VSM, *Vector Space Model* [Turney et Pantel, 2010])
- Comme le modèle VSM, le poids associé à une dimension indique l'importance de la dimension pour l'élément linguistique à représenter
- Les poids sont estimés sur la base des propriétés structurelles de la ressource utilisée

Modèle de représentation sémantique à base du réseau lexical JeuxDeMots

Signature sémantique d'un élément linguistique

- Forme spéciale du modèle d'espace vectoriel (VSM, *Vector Space Model* [Turney et Pantel, 2010])
- Comme le modèle VSM, le poids associé à une dimension indique l'importance de la dimension pour l'élément linguistique à représenter
- Les poids sont estimés sur la base des propriétés structurelles de la ressource utilisée

Exemples d'application

- Est-ce que la similarité entre FOURNAISE et FOUR est plus forte que celle entre FOURNAISE et INSTRUMENT ?
- Est-ce que PRIX représente un substitut pertinent pour INTÉRÊT par rapport à AVANTAGE dans un contexte décrivant le sens *finance* ?

Relations dans JeuxDeMots

- Une signature est typée et pondérée et peut dépendre d'un ou de plusieurs types de relations
- Les instances de relations sont pondérées dans [JeuxDeMots](#)

Relation	Type de relation	Nombre d'instances	Exemple
Synonyme	Lexical	746 128	<i>outil</i> → <i>instrument</i>
Acception	Associatif	93 899	<i>outil</i> → <i>dispositif</i>
Domaine	Sémantique	670 055	<i>outil</i> → <i>mécanique</i>
Agent	Prédicatif	1 055 271	<i>outil</i> ← <i>fonctionner</i>
Patient	Prédicatif	65 397	<i>outil</i> ← <i>fabriquer</i>
Hyperonyme	Sémantique	3 155 763	<i>outil</i> → <i>matériel</i>
Hyponyme	Sémantique	712 125	<i>outil</i> → <i>marteau</i>
Idée associée	Associatif	17 768 142	<i>outil</i> → <i>bricolage</i>

Table 1: Relations utilisées pour la création de signatures sémantiques¹

¹La base lexicale utilisée est datée de Janvier 2017
<http://www.jeuxdemots.org/JDM-LEXICALNET-FR/?C=M;O=D>

Production de signatures sémantiques

- Signature pour chaque relation
- Signatures à base de combinaison de plusieurs relations
 - ① Combinaison de *Synonyme* et *Acception*
 - ② Combinaison d'*Hyperonyme* et d'*Hyponyme*
 - ③ Combinaison d'*Agent* avec *Patient*
 - ④ Toutes les relations sont prises avec un même coefficient
 - ⑤ Toutes les relations sont prises avec une importance supérieure à certaines relations
 - {“*Domaine*” : 6, “*Acception*” : 5, “*Hyperonyme*” : 4, “*Hyponyme*” : 4, “*Synonyme*” : 3, “*Agent*” : 2, “*Patient*” : 2, “*Idée associée*” : 1}
- + Normalisation des signatures avec la norme infinie

$$\|\vec{X}\|_{\infty} = \max(|x_1|, |x_2|, \dots, |x_n|)$$

- Un seuil est fixé à 0,01 pour la validation des dimensions des signatures

Similarité sémantique entre mots (A et B)

- Deux techniques qui tiennent compte du niveau distributionnel :

– **Cosinus**

$$Sim_{Cos}(S_A, S_B) = \frac{S_A \cdot S_B}{\|S_A\| \|S_B\|}$$

– **Weighted Overlap (WO)**, source :
[Pilehvar et al., 2013]

$$Sim_{WO}(S_A, S_B) = \frac{\sum_{d \in D} (r_d(S_A) + r_d(S_B))^{-1}}{\sum_{i=1}^{|D|} (2i)^{-1}}$$

- Quatre fonctions d'activation (inspiration du type de fonction proposé par [Lafourcade, 2011, 21–22]) :

$$Act_j : \mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{N}^+ \rightarrow \mathbb{R}^+$$

$$- \text{Act}_1(\mathbf{A}, \mathbf{B}, \mathbf{R}) = \begin{cases} 1 & \text{si } A \in S_B \text{ ou } B \in S_A \\ Sim_{Cos}(S_A, S_B) & \text{sinon} \end{cases}$$

$$- \text{Act}_3(\mathbf{A}, \mathbf{B}, \mathbf{R}) = \max(A[B], B[A], Sim_{Cos}(S_A, S_B))$$

$$- \text{Act}_2(\mathbf{A}, \mathbf{B}, \mathbf{R}) = \begin{cases} 1 & \text{si } A \in S_B \text{ ou } B \in S_A \\ Sim_{WO}(S_A, S_B) & \text{sinon} \end{cases}$$

$$- \text{Act}_4(\mathbf{A}, \mathbf{B}, \mathbf{R}) = \max(A[B], B[A], Sim_{WO}(S_A, S_B))$$

Évaluation intrinsèque : mesures de similarité sémantique

- Qualité de nos représentations évaluée par rapport à un jugement humain
- Utilisation de la liste de référence RG-65 pour le français [Joubarne et Inkpen, 2011]

But de la création de la liste RG-65

Étudier la similarité sémantique et contextuelle pour un ensemble de 65 paires de noms communs évaluées sur une échelle de 0 (non liés) à 4 (complètement liés).

- 18 évaluateurs humains dont le français est considéré comme langue maternelle
- La liste proposée par Joubarne et Inkpen présente une traduction (avec nouveau jugement) de la liste d'origine RG-65 pour l'anglais [Rubenstein et Goodenough, 1965]

Jeux de données RG-65 [Joubarne et Inkpen, 2011]



Figure 1: Les 48 mots du vocabulaire correspondant au jeu de données RG-65²

²Deux paires sont ignorées : (**cock, rooster**) traduite en (**coq, coq**) et (**cemetery, graveyard**) traduite en (**cimetièrè, cimetièrè**).

Évaluation de la similarité sémantique (1/2)

Types de signature	Couverture (%)	Act ₁	Act ₂	Act ₃	Act ₄
<i>r_synonyme</i>	95.83 ³	0.88	0.85	0.89	0.85
<i>r_acception</i>	77.08	0.80	0.80	0.79	0.80
<i>r_synonyme_acception</i>	95.83	0.88	0.88	0.87	0.87
<i>r_domaine</i>	81.25	0.43	0.39	0.40	0.32
<i>r_agent</i>	58.33	0.23	0.35	0.23	0.35
<i>r_patient</i>	60.42	0.51	0.47	0.51	0.47
<i>r_agent_patient</i>	72.92	0.25	0.33	0.25	0.33
<i>r_hyperonyme</i>	85.42	0.46	0.51	0.43	0.48
<i>r_hyponyme</i>	87.5	0.45	0.45	0.44	0.44
<i>r_hyperonyme_hyponyme</i>	89.58	0.44	0.48	0.43	0.47
<i>r_idée_associée</i>	95.83	0.88	0.88	0.82	0.81
<i>r_traits_avec_coeff</i>	95.83	0.86	0.86	0.81	0.81
<i>r_traits_égaux</i>	95.83	0.87	0.87	0.85	0.85

Table 2: Corrélations de Pearson obtenues selon différentes signatures avec différentes configurations

³ *asylum* peut se traduire par *refuge* ou *asile* et *goblet* s'écrit *gobelet* en français.

Évaluation de la similarité sémantique (2/2)

Système	Corrélation de Pearson (r)	Corrélation de Spearman (ρ)
<i>r_traits_avec_coeff</i>	0.86	0.85
<i>r_traits_égaux</i>	0.87	0.85
<i>r_idée_associée</i>	0.88	0.83
<i>r_synonyme</i>	0.88	0.75
DEPGLOVE	0.48	0.50
NASARI _{COS}	0.80	0.77
NASARI ⁴ _{WO}	0.82	0.78

Table 3: Corrélations de Pearson et Spearman obtenues selon différentes signatures avec utilisation de la première configuration, comparaison avec les résultats obtenus par NASARI et DEPGLOVE sur un ensemble de 60 paires couvertes par tous les systèmes

+ $r = 0.29$ pour PMI et $r = 0.17$ pour SOC-PMI (Joubarne et Inkpen)⁵.

⁴Utilisation de l'algorithme proposé par [Camacho-Collados et al., 2015, 570]

⁵Pointwise Mutual Information; Second Order Co-occurrence PMI 

Évaluation extrinsèque : substitution lexicale

Intérêt pour l'évaluation de la similarité sémantique ?

- 1 Évaluation extrinsèque pour laquelle la similarité sémantique à un rôle prépondérant pour que des différences la concernant puissent être observées vis-à-vis de la tâche de substitution
- 2 Le niveau contextuel est pris en compte

Évaluation extrinsèque : substitution lexicale

Intérêt pour l'évaluation de la similarité sémantique ?

- 1 Évaluation extrinsèque pour laquelle la similarité sémantique à un rôle prépondérant pour que des différences la concernant puissent être observées vis-à-vis de la tâche de substitution
- 2 Le niveau contextuel est pris en compte

Tâche de substitution lexicale ?

Sous-tâche 1 : génération de candidats substitués pour le mot-cible à remplacer^a

Sous-tâche 2 : choix de l'un des candidats en fonction du contexte

^aUtilisation des signatures à base de *Synonyme*. Une présélection des candidats en tenant compte seulement des synonymes ayant un poids d'importance ≥ 0.8

Jeu d'évaluation fourni dans SemDis [Fabre et al., 2014]

- 30 mots-cibles à substituer
- À chaque mot-cible, 10 phrases sont proposées (300 phrases au total⁶)



Figure 2: 10 noms cibles

Figure 3: 10 verbes cibles

Figure 4: 10 adjectifs cibles

⁶Utilisation de l'outil Talismane [Urieli, 2013] pour une analyse morpho-syntaxique.

Évaluation de la substitution lexicale (1/3)

Algorithme de substitution lexicale ?

Mesurer la similarité entre chaque candidat substitut et l'ensemble de mots pleins de la phrase contenant le mot-cible à remplacer, hors ce dernier (modèle déjà proposé par [Ferret, 2014] que nous appellerons **Sub_Lex**).

- **Mesures d'évaluation (SemEval 2007, SemDis 2014)**

[McCarthy et Navigli, 2009, Fabre et *al.*, 2014]

- 1– **best** : le système est évalué par rapport à la première substitution proposée. Le meilleur score renvoie le substitut choisi majoritairement par les annotateurs.
- 2– **oot** (*out of ten*) : le système est évalué par rapport à tous les substituts proposés (dans la limite de 10). Le meilleur score obtainable correspond au nombre maximum de réponses couvertes par les annotateurs.

Évaluation de la substitution lexicale (2/3)

Systeme	best	oot
<i>JDM_TraitsÉgaux_FctAct3</i>	.079	.308
<i>JDM_TraitsÉgaux_FctAct2</i>	.077	.308
<i>JDM_TraitsÉgaux_FctAct4</i>	.077	.308
<i>JDM_TraitsIdéeAssociée_FctAct2</i>	.076	.313
<i>JDM_TraitsÉgaux_FctAct1</i>	.076	.307
<i>JDM_TraitsAvecCoeff_FctAct1</i>	.076	.309
<i>JDM_TraitsAvecCoeff_FctAct3</i>	.074	.307
<i>JDM_TraitsIdéeAssociée_FctAct3</i>	.073	.304
<i>JDM_TraitsAvecCoeff_FctAct2</i>	.071	.307
<i>JDM_TraitsIdéeAssociée_FctAct4</i>	.069	.315
<i>JDM_TraitsIdéeAssociée_FctAct1</i>	.068	.304
<i>JDM_TraitsAvecCoeff_FctAct4</i>	.061	.307
Baseline (<i>premiers synonymes, sans contexte</i>)	.029	.269

Table 4: Résultats pour la tâche de substitution lexicale selon différentes signatures avec différentes configurations

Évaluation de la substitution lexicale (3/3)

Systeme	best	oot
<i>JDM_TraitsÉgaux_FctAct3</i>	.079	.308
<i>JDM_TraitsAvecCoeff_FctAct1</i>	.076	.309
<i>JDM_TraitsIdéeAssociée_FctAct4</i>	.069	.315
Baseline (<i>premiers synonymes, sans contexte</i>)	.029	.269
<i>DEPGLOVE</i>	.034	.284
<i>CEA_list-word_cos_sent</i> (Sub_Lex)	.075	.236
<i>CEA_list-fredist_cos_sent</i> (Sub_Lex)	.040	.236
<i>CEA_list-isc_cos_sent</i> (Sub_Lex)	.033	.287
<i>CEA_list-isc_l2_sent</i> (Sub_Lex)	.010	.231

Table 5: Résultats pour la tâche de substitution lexicale, comparaison avec les résultats obtenus en utilisant DEPGLOVE et les systèmes de SemDis à base de **Sub_Lex**

Conclusion et perspectives

Conclusion

- Description d'un nouveau modèle de représentation sémantique de mots à base du réseau lexical [JeuxDeMots](#)
- Évaluation intrinsèque (mesures de similarité sémantique) et extrinsèque (substitution lexicale)
- Utilisation de différentes fonctions d'activation pour mesurer la similarité sémantique
- Nos systèmes sont plus performants que certains systèmes de l'état de l'art comme DEPGLOVE ou NASARI

Conclusion et perspectives

Perspectives

- Évaluation intrinsèque sur une autre liste de référence plus large que celle de RG-65
- Ajout d'un module de désambiguïsation sémantique avant la substitution lexicale et comparaison des résultats avec et sans désambiguïsation
- Des mots au texte : construction de signatures sémantiques pour les textes et comparaison entre les éléments linguistiques possédant un niveau de granularité différent

En vous remerciant pour votre attention

Évaluation de la substitution lexicale

Système	best				oot			
	Nom	Adj.	Verbe	Total	Nom	Adj.	Verbe	Total
<i>JDM.TraitsÉgaux.FctAct3</i>	.078	.100	.059	.079	.261	.323	.339	.308
<i>JDM.TraitsÉgaux.FctAct2</i>	.068	.111	.053	.077	.269	.322	.332	.308
<i>JDM.TraitsÉgaux.FctAct4</i>	.071	.098	.063	.077	.269	.325	.331	.308
<i>JDM.TraitsIdéeAssociée.FctAct2</i>	.081	.092	.054	.076	.277	.346	.317	.313
<i>JDM.TraitsÉgaux.FctAct1</i>	.075	.099	.053	.076	.258	.322	.340	.307
<i>JDM.TraitsAvecCoeff.FctAct1</i>	.077	.094	.056	.076	.259	.326	.341	.309
<i>JDM.TraitsAvecCoeff.FctAct3</i>	.078	.075	.070	.074	.260	.322	.340	.307
<i>JDM.TraitsIdéeAssociée.FctAct3</i>	.060	.095	.063	.073	.270	.323	.318	.304
<i>JDM.TraitsAvecCoeff.FctAct2</i>	.066	.096	.052	.071	.264	.324	.334	.307
<i>JDM.TraitsIdéeAssociée.FctAct4</i>	.067	.093	.046	.069	.283	.344	.317	.315
<i>JDM.TraitsIdéeAssociée.FctAct1</i>	.060	.097	.047	.068	.268	.326	.318	.304
<i>JDM.TraitsAvecCoeff.FctAct4</i>	.058	.066	.058	.061	.261	.327	.334	.307
<i>baseline (premiers synonymes, sans contexte)</i>	.029	.051	.006	.029	.247	.303	.258	.269
<i>DEPGLOVE</i>	.017	.033	.053	.034	.242	.280	.331	.284
<i>CEA_list-word_cos_sent (Sub_Lex)</i>	.075	.074	.076	.075	.195	.245	.268	.236
<i>CEA_list-fredist_cos_sent (Sub_Lex)</i>	.032	.028	.060	.040	.181	.225	.303	.236
<i>CEA_list-isc_cos_sent (Sub_Lex)</i>	.025	.034	.040	.033	.233	.287	.340	.287
<i>CEA_list-isc_l2_sent (Sub_Lex)</i>	.004	.012	.015	.010	.163	.230	.300	.231

Table 6: Résultats pour la tâche de substitution lexicale selon différentes signatures avec différentes configurations, comparaison avec les résultats obtenus en utilisant DEPGLOVE et les systèmes de SemDis à base de **Sub_Lex**

Références



Baroni M., Bernardini S., Ferraresi A. et Zanchetta E. (2009)

The WaCky wide web: a collection of very large linguistically processed web-crawled corpora.

Language Resources and Evaluation, 43(3), 209–226.



Baroni M., Dinu G. et Kruszewski G. (2014)

Don't count, predict ! A systematic comparison of context-counting vs. context-predicting semantic vectors.

52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference, 1, 238–247.



Biran O., Brody S. et Elhadad N. (2011)

Putting It Simply: A Context-aware Approach to Lexical Simplification.

In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics : Human Language Technologies : Short Papers, volume 2 of HLT '11, p. 496–501, Stroudsburg, PA, USA.



Budanitsky A. et Hirst G. (2006)

Evaluating WordNet-based Measures of Lexical Semantic Relatedness.

Comput. Linguist., 32(1), 13–47.

Références



Camacho-Collados J., Pilehvar M. T. et Navigli R. (2015)
NASARI : a Novel Approach to a Semantically-Aware Representation of Items.
In Mihalcea R., Chai J. Y. et Sarkar A., Eds., HLT-NAACL, p. 567–577 : The Association for Computational Linguistics.



Camacho-Collados J., Pilehvar M. T. et Navigli R. (2016)
NASARI : Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities.
Artif. Intell., 240, 36–64.



Corley C. et Mihalcea R. (2005)
Measuring the Semantic Similarity of Texts.
In Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment, EMSEE '05, p. 13–18, Stroudsburg, PA, USA : Association for Computational Linguistics.



Deerwester S. C., Dumais S. T., Landauer T. K., Furnas G. W. et Harshman R. A. (1990)
Indexing by Latent Semantic Analysis.
JASIS, 41(6), 391–407.

Références



Evert S. (2005)

The statistics of word cooccurrences : word pairs and collocations.
Ph.D. thesis, Université de Stuttgart.



Fabre C., Hathout N., Ho-Dac L.-M., Morlane-Hondère F., Muller P., Sajous F., Tanguy L. et Van De Cruys T. (2014)

Présentation de l'atelier SemDis 2014 : sémantique distributionnelle pour la substitution lexicale et l'exploration de corpus spécialisés.

In Actes de l'atelier SemDis 2014, 21e Conférence sur le Traitement Automatique des Langues Naturelles, p. 196–205, Marseille, France.



Ferret O. (2014)

Utiliser un modèle neuronal générique pour la substitution lexicale.

In Actes de l'atelier SemDis 2014, 21e Conférence sur le Traitement Automatique des Langues Naturelles, p. 218–227, Marseille, France.



Harris Z. (1954)

Distributional structure.

Word, 10(23), 146–162.

Références



Joubarne C. et Inkpen D. (2011)

Comparison of Semantic Similarity for Different Languages Using the Google n-gram Corpus and Second-Order Co-occurrence Measures.

In Advances in Artificial Intelligence - 24th Canadian Conference on Artificial Intelligence, Canadian AI 2011, St. John's, Canada, May 25-27, 2011. Proceedings, p. 216–221.



Lafourcade M. (2007)

Making people play for Lexical Acquisition with the JeuxDeMots prototype.

In SNLP'07 : 7th International Symposium on NLP, Pattaya, Chonburi, Thaïlande.



Lafourcade M. (2011)

Lexique et analyse sémantique de textes - structures, acquisitions, calculs, et jeux de mots.

Mémoire d'habilitation à diriger les recherches, Université Montpellier 2, LIRMM.



Lavie A. et Denkowski M. J. (2009)

The METEOR Metric for Automatic Evaluation of Machine Translation.

Machine Translation, 23(2-3), 105–115.

Références



Lin D. (1998)

Automatic Retrieval and Clustering of Similar Words.

In Proceedings of the 17th International Conference on Computational Linguistics, volume 2 of COLING '98, p. 768–774, Stroudsburg, PA, USA : Association for Computational Linguistics.



Matuschek M. et Gurevych I. (2013)

Dijkstra-WSA : A Graph-Based Approach to Word Sense Alignment.

Transactions of the Association for Computational Linguistics, 1, 151–164.



McCarthy D. et Navigli R. (2009)

The English Lexical Substitution Task.

Language Resources and Evaluation, 43(2), 139–159.



Mikolov T., Chen K., Corrado G. et Dean J. (2013)

Efficient Estimation of Word Representations in Vector Space.

CoRR, abs/1301.3781.

Références



Navigli R. (2009)

Word Sense Disambiguation : A Survey.
ACM Computing Surveys, 41(2), 1–69.



Navigli R. et Ponzetto S. P. (2012)

BabelNet : The Automatic Construction, Evaluation and Application of a Wide-coverage Multilingual Semantic Network.
Artif. Intell., 193, 217–250.



Otegi A., Arregi X., Ansa O. et Agirre E. (2015)

Using knowledge-based relatedness for information retrieval.
Knowl. Inf. Syst., 44(3), 689–718.



Pennington J., Socher R. et Manning C. D. (2014)

Glove : Global Vectors for Word Representation.
In EMNLP, volume 14, p. 1532–1543.

Références



Pilehvar M. T., Jurgens D. et Navigli R. (2013)

Align, Disambiguate and Walk : A Unified Approach for Measuring Semantic Similarity.

In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, volume 1, p. 1341–1351.



Ploux S. et Victorri B. (1998)

Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes.

Traitement Automatique des Langues, 39(1), 161–182.



Rubenstein H. et Goodenough J. B. (1965)

Contextual Correlates of Synonymy.

Commun. ACM, 8(10), 627–633.



Salton G., Wong A. et Yang C. S. (1975)

A Vector Space Model for Automatic Indexing.

Commun. ACM, 18(11), 613–620.

Références



Turney P. D. et Pantel P. (2010)

From Frequency to Meaning : Vector Space Models of Semantics.

J. Artif. Int. Res., 37(1), 141–188.



Urieli A. (2013)

Robust French syntax analysis : reconciling statistical methods and linguistic knowledge in the Talismane toolkit.

Ph.D. thesis, Université de Toulouse II le Mirail.



Voorhees E. M. (1994)

Query Expansion Using Lexical-semantic Relations.

In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '94, p. 61–69, New York, NY, USA.



Wu Z. et Giles C. L. (2015)

Sense-aware Semantic Analysis : A Multi-prototype Word Representation Model Using Wikipedia.

In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI'15, p. 2188–2194 : AAAI Press.



Zesch T., Muller C. et Gurevych I. (2008)

Using Wiktionary for Computing Semantic Relatedness.

In Proceedings of the 23rd National Conference on Artificial Intelligence, volume 2 of AAAI'08, p. 861–866 : AAAI Press.



Jones K. S. (1972)

A statistical interpretation of term specificity and its application in retrieval.

Journal of Documentation, 28, 11–21.