

**A HYBRID SEMANTIC SIMILARITY FEATURE-BASED TO SUPPORT
MULTIPLE ONTOLOGIES**

NURUL ASWA BINTI OMAR

A thesis submitted in
fulfillment of the requirement for the award of the
Doctor of Philosophy

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia

AUGUST, 2017

In the name of Allah, The Most Beneficent, The Most Merciful.

Thank you Allah for giving me such wonderful people.

Deep appreciation to my beloved husband,
Beni Widarman Bin Yus Kelana

My children,
Muhammad Faris Aiman
Nur Fatihah Aleeya

My parents and father-in-law,
Omar Othman; Shakinah Yaakob; Yus Kelana; Noraini

and my siblings (Nurul Atiqah, Muhammad Syafiq, Nurul Athirah, Muhammad Syamil, Muhammad Syarafi).

Thank you for your prayers, understanding, caring, compromising and everything.

ACKNOWLEDGEMENT

Firstly, I would like to express my thanks to Assoc. Prof. Dr. Shahreen Binti Kasim, and Assoc. Prof. Dr. Mohd Farhan Bin Md Fudzee for their willingness to accept me as a Ph.D. student. Their guidance, support, determination, encouragement, understanding and patience along this journey is greatly appreciated.

I am also grateful to the Ministry of Higher Education (MOHE) for sponsoring me under the Scheme of Academic Training for Public Higher Education Institution (SLAI) during my studies.

I am also greatly indebted to the Faculty of Computer Science and Information Technology (FSKTM) and the Centre for Graduate Studies (CGS) of Universiti Tun Hussein Onn Malaysia (UTHM) for providing good facilities and an inspiring environment for me to complete this study comfortably.

I am dedicating special thanks to my friends, especially the postgraduate members (Dr. Isredza, Dr. Norhanifah, Norhanani, Norlida, Mahfuzah, Yana Mazwin and Dr. Yusliza) for their help, motivation, encouragement and knowledge sharing throughout this journey. Many thanks to Dr. Mohd Arfian Ismail, Dr. Sazali Khalid, Dr. Riswan Efendi, Dr. Adila Firdaus for helping in completing my thesis journey.

Last but not least, I would like to thank my husband, my son, my parents, my father-in-law and my siblings for their prayers, support, understanding, compromise and everything.

ABSTRACT

Semantic similarity between concepts, words, and terms is of great importance in many applications dealing with textual data, such as Natural Language Processing (NLP). Semantic similarity is defined as the closeness of two concepts, based on the likeliness of their meaning. It is also more ontology-based, due to their efficiency, scalability, lack of constraints and the availability of large ontologies. However, ontology-based semantic similarity is hampered by the fact that it depends on the overall scope and detail of the background ontology. Coupled with the fact that only one ontology is exploited, this leads to insufficient knowledge, missing terms and inaccuracy. This limitation can be overcome by exploiting multiple ontologies. Semantic similarity with multiple ontologies potentially leads to better accuracy because it is able to calculate the similarity of these missing terms from the combination of multiple knowledge sources. This research was conducted for developing the taxonomy of semantic similarity that contributes to understanding the current approaches, issues and data involved. This research aims to propose and evaluate ontological features for semantic similarity with multiple ontologies. Additionally, this research aims to develop and evaluate a feature-based mechanism (Hyb-TvX) to measure semantic similarity with multiple ontologies which can improve the accuracy of the similarity. This research used two benchmark datasets of biomedical concepts from Perdesen and Hliaoutakis. Similarity value, correlation and *p*-value were also used in the evaluation of the relationship between the concept pair of multiple ontologies. The findings indicate that the use of a semantic relationship of concepts (hypernym, hyponym, sister term and meronym) can improve the baseline method up to 75%. Besides that, the Hyb-TvX mechanism produces the highest correlation value compared to the other two methods, that is 0.759 and the result correlation is significant. Finally, the ability to discover similarity concepts with multiple ontologies could be also exploited in other domains besides biomedicine as future research.

ABSTRAK

Persamaan semantik antara konsep, perkataan dan terma adalah penting dalam pelbagai aplikasi yang berkaitan dengan data teks seperti pemprosesan bahasa tabii. Persamaan semantik merupakan satu pendekatan bagi mengenalpasti persamaan konsep melalui perbandingan makna. Persamaan semantik lebih cenderung menggunakan ontologi berdasarkan kecekapan, berskala, kurang kekangan serta mempunyai ontologi yang besar. Walaubagaimanapun, penggunaan ontologi di dalam persamaan semantik masih dibatasi oleh kebergantungan ontologi yang terperinci dan hanya satu ontologi diterokai yang menyebabkan ketidakcukupan pengetahuan, kehilangan terma dan ketidaksetepatan persamaan. Permasalahan ini boleh diatasi dengan mengeksplotasi kepelbagaiian ontologi. Persamaan semantik dari pelbagai ontologi berpotensi meningkatkan ketepatan persamaan dengan pengiraan persamaan dalam situasi kehilangan terma serta gabungan pelbagai sumber ontologi. Kajian ini dijalankan untuk membangunkan taksonomi persamaan semantik dalam memahami pendekatan semasa, isu dan data yang terlibat. Kajian ini bertujuan mencadangkan dan menilai ciri-ciri ontologi untuk persamaan semantik dari pelbagai ontologi. Disamping itu, kajian ini juga bertujuan untuk membangunkan dan menilai mekanisme berdasarkan ciri-ciri (Hyb-TvX) bagi pengukuran persamaan semantik pelbagai ontologi dalam meningkatkan nilai ketepatan persamaan. Kajian ini telah menggunakan dua penanda aras set data bioperubatan konsep yang terdiri daripada Perdesen dan Hliaoutakis. Nilai persamaan, kolerasi dan nilai p juga digunakan bagi melihat perhubungan dan kepentingan hubungan antara konsep dari pelbagai ontologi. Dapatkan kajian menunjukkan penggunaan perhubungan semantik konsep (*hypernym*, *hyponym*, *sister term* dan *meronym*) telah meningkatkan kolerasi sebanyak 75%. Selain itu, kaedah pengukuran Hyb-TvX menghasilkan nilai kolerasi yang tinggi berbanding dua kaedah sebelum ini iaitu 0.759 dengan keputusan kolerasi signifikan. Akhir sekali, penyelidikan persamaan pelbagai ontologi boleh dieksplotasi dalam bidang selain bioperubatan di masa depan.

TABLE OF CONTENTS

DECLARATION	ii
DEDICATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
ABSTRAK	vi
TABLE OF CONTENTS	vii
LIST OF TABLES	xii
LIST OF FIGURES	xvi
LIST OF ALGORITHMS	xix
LIST OF SYMBOLS AND ABBREVIATIONS	xx
LIST OF APPENDICES	xxv
LIST OF PUBLICATIONS	xxvi
CHAPTER 1 INTRODUCTION	1
1.1 Overview	1
1.2 Problem statement	3
1.3 Research objectives	5
1.4 Research questions	5
1.5 Scope and research significance	6
1.6 Organization of the thesis	9

CHAPTER 2 LITERATURE REVIEW	10
2.1 Introduction	10
2.2 Similarity	11
2.3 Words similarity	12
2.4 Semantic similarity	16
2.5 Knowledge-based	19
2.5.1 Ontology-based	19
2.5.2 General purpose ontologies used with semantic similarity approaches	24
2.5.3 Domain specific ontologies used with semantic similarity approaches	27
2.6 Classification of semantic similarity approaches according to ontology	33
2.7 Single ontology	37
2.7.1 Structure-based approach	37
2.7.2 Information content-based approach	40
2.7.3 Feature-based approach	42
2.7.4 Hybrid-based approach	43
2.8 Multiple ontologies	43
2.8.1 Matching method	44
2.8.2 Ontological feature matching for multiple ontologies	45
2.8.3 Semantic similarity measure for multiple ontologies	47
2.9 Biomedical domain as a case study	56
2.10 Discussion of similarity	59
2.11 Chapter summary	61
CHAPTER 3 RESEARCH METHODOLOGY	63
3.1 Introduction	63
3.2 Structure of semantic similarity	63
3.3 Research framework	64
3.4 Data sources and preparation	68
3.4.1 WordNet	71

3.4.2	MeSH	72
3.5	Instrumentation and result analysis	72
3.5.1	Hardware and software requirements	74
3.5.2	Testing and analysis	74
3.6	Parameter	78
3.6.1	The proposed parameter	79
3.6.2	The classical probabilistic theory	92
3.7	Chapter summary	93

CHAPTER 4 THE ONTOLOGICAL FEATURES

ALGORITHM IN SEMANTIC SIMILARITY FOR MULTIPLE ONTOLOGIES		95
4.1	Introduction	95
4.2	Ontological features algorithm	97
4.2.1	Matching matrix semantic relationship	97
4.2.2	Semantic overlapping between subsumers	99
4.2.3	Selecting the LCS	101
4.3	Result and discussion	103
4.3.1	Ontological features evaluation result	103
4.3.2	Matching method results	106
4.3.3	Correlation between similarity measurement and method for multiple ontologies	109
4.4	Chapter summary	112

CHAPTER 5 Hyb-TvX: A HYBRID SEMANTIC SIMILARITY

FEATURE-BASED MEASUREMENT FOR MULTIPLE ONTOLOGIES		113
5.1	Introduction	113
5.2	Hyb-TvX: A hybrid semantic similarity feature- based measurement	116
5.2.1	TvX-1: similarity measurement level 1	122
5.2.2	TvX-2: similarity measurement level 2	125
5.3	Result and discussion	128

5.3.1 Experimental results	128
5.4 Chapter summary	133
CHAPTER 6 CONCLUSION AND FUTURE WORKS	135
6.1 Introduction	135
6.2 Research summary	135
6.3 To develop the taxonomy of semantic similarity that contributes understanding towards current approaches, issues and data related to the topic	137
6.3.1 What is the appropriate approach for semantic similarity with multiple ontologies in order to understand the issues and data involved?	137
6.4 To propose and evaluate an ontological features algorithm in semantic similarity for multiple ontologies in terms of features matching accuracy	138
6.4.1 Does the feature-based measurement used is appropriate in measuring the ontological features algorithm proposed?	138
6.4.2 Does the proposed ontological features algorithm increase the accuracy of feature matching?	139
6.4.3 Is there any relationship between dependent variable (human scored) and independent variables (similarity values for each method)?	140
6.5 To develop and evaluate feature-based mechanism to measure semantic similarity of multiple ontologies to increase similarity accuracy	141

6.5.1	How does the proposed parameters support in improving the accuracy of similarity for multiple ontologies?	141
6.5.2	Does it a significant relationship between the Hyb-TvX methods with human scored?	143
6.6	Contribution of the research	143
6.7	Future research	144
REFERENCES		145
APPENDIX		157



PTTA UTHM
PERPUSTAKAAN TUNKU TUN AMINAH

LIST OF TABLES

2.1	Symbol and meaning in relation to similarity	12
2.2	Words similarity approach	15
2.3	Fields and applications that use the semantic similarity approach	17
2.4	Advantages and disadvantages based on the background information in the semantic similarity approach	19
2.5	Classification approach according to the categories of ontology	38
2.6	Advantages and disadvantages of the Tversky method	42
2.7	Types of ontological features in each semantic similarity method	45
2.8	Descriptions of two concepts from WordNet and MeSH	49
2.9	Method feature-based approach for multiple ontologies	56
2.10	Summary of datasets for single ontology used in previous work	58
2.11	Summary of datasets for multiple ontologies that used in previous work	58
3.1	Set of 36 medical term pairs with averaged experts similarity scores (Hliaoutakis <i>et al.</i> , 2006)	68
3.2	Set of 30 medical term pairs with averaged physician and coder ratings scores (Pedersen <i>et al.</i> , 2007)	70

3.3	The level of strength correlation coefficient	77
3.4	Example of results human similarity and proposed method	77
3.5	Calculate the correlation coefficient (r) table	77
3.6	Result implementation parameters for situation 1	80
3.7	Result implementation parameters for situation 2	81
3.8	Result implementation parameters for situation 3	82
3.9	Result implementation parameters for situation 4	83
3.10	Result implementation parameters for situation 5	84
3.11	Result implementation parameters for situation 6	85
3.12	Result implementation parameters for situation 7	86
3.13	Result implementation parameters for situation 8	87
4.1	Comparison previous method and OntF	102
4.2	Similarity result of the proposed method (OntF) using four similarity feature-based measurements in Pedersen benchmark datasets based on physician ratings	104
4.3	Similarity result of the proposed method (OntF) using four similarity feature-based measurements in Pedersen benchmark datasets based on coder ratings	105
4.4	Similarity result of the proposed method (OntF) using four similarity feature-based measurements in Hlioutakis benchmark datasets based on expert ratings	106

4.5	Similarity results of multiple ontologies. The bold column represents the proposed method (OntF)	108
4.6	Correlation values based on four similarity feature-based measurements using different methods. The row in bold shows the results of the proposed method (OntF)	110
4.7	Improvements that the OntF achieved over the average of the three other methods using the Sánchez measure based on Physician, Coder and Expert ratings	111
4.8	Improvements that the OntF achieved over the average of the three other methods using the Jaccard measure based on Physician, Coder and Expert ratings	111
4.9	Improvements that the OntF achieved over the average of the three other methods using the Dice measure based on Physician, Coder and Expert ratings	112
4.10	Improvements that the OntF achieved over the average of the three other methods using the Ochiai measure based on Physician, Coder and Expert ratings	112
5.1	Example for concepts extracting a set of token	118
5.2	Example for concepts with tokenization of words	118
5.3	Example for concepts and synonym	119
5.4	Examples of concepts	122
5.5	Example for concepts and synonym	124
5.6	Example of concepts and features	126
5.7	Comparison similarity of the proposed method with the X-similarity and Rodriguez	

	and the Egenhofer method using Pedersen benchmarks datasets	129
5.8	Comparison similarity of the proposed method with the X-similarity and Rodriguez and Egenhofer method using Hlioutakis benchmarks datasets	130
5.9	Comparison similarity of proposed method with physician, coder and experts ratings (averaged)	131
5.10	Correlation of similarity method on feature- based approach for multiple ontologies according to the WordNet and MeSH dataset	133
5.11	<i>p</i> -value of similarity method on feature- based approach for multiple ontologies	133
6.1	Investigated the proposed parameters (α) in eight situations involve condition $ comp\ B $, $ comp\ A $, differentiation complement and intersection value	143

LIST OF FIGURES

1.1	Example of terminological matching between subsumer of antibiotic (WordNet) and antibacterial agent (MeSH)	4
1.2	Comparison of the number of articles published every year and the number of citations per year (http://apps.webofknowledge.com)	8
2.1	Content structure	10
2.2	The computer ontology	20
2.3:	Semantic similarity in single ontology for male and female WordNet ontology	22
2.4	Semantic similarity in multiple ontologies for intracranial hemorrhage (in Snomed-CT) and brain neoplasms (in MeSH)	23
2.5	The snapshot of lexical database for english (WordNet)	25
2.6	The snapshot of WordNet content for renal failure	25
2.7	The snapshot of the Cyc web page.	26
2.8	Snapshot of the UMLS web page.	28
2.9	The snapshot of Medical Subject Headings (MeSH)	29
2.10	The snapshot of MeSH content for kidney disease	29
2.11	The snapshot of Snomed-CT web pages	30
2.12	An example of a biological process tree for biological_process (GO: 0008150). It is the	

parent while biological regulation (GO:0065007), localisation (GO:00051179), phosphorus utilisation (GO:0006794), and signaling (GO:00023052) are its children.	32
2.13 The snapshot web page of SDTS	33
2.14 Semantic similarity for single ontology	34
2.15 Semantic similarity for multiple ontologies	35
2.16 Single ontology	37
2.17 The concept similarity between C_1 and C_2	40
2.18 WordNet taxonomy, showing most-specific subsumers of nickel and dime and of nickel and credit card. Solid lines represent <i>is-a</i> links; dashed lines indicate that some intervening nodes have been omitted	41
2.19 Ontologies O_1 and O_3	54
2.20 The process to develop hybrid semantic similarity feature-based for multiple ontologies	62
3.1 Structure of semantic similarity method	64
3.2 Research framework	65
3.3 Research structure	66
3.4 Example of a taxonomical tree in WordNet	73
3.5 Example of a taxonomical tree MeSH	74
3.6 $ comp\ B $ is larger than $ comp\ A $	87
3.7 $ comp\ B $ is smaller than $ comp\ A $	88
3.8 Results for condition 1 in situations 1, 2, 3 and 4	90
3.9 Results for condition 2 in situations 5, 6, 7 and 8	91
4.1 Ontological modelling of the concept lupus in MeSH	98
4.2 MeSH ontology for <i>myocardium</i> concept	100
5.1 The flow process of Hyb-TvX	115

5.2	Venn diagram for intersection	116
5.3	Venn diagram for union	117
5.4	Venn diagram for complement	117
5.5	Intersection of (C_1) and (C_2)	118
5.6	Maximum tokenization of (C_1) and (C_2)	119
5.7	Intersection synonym of (C_1) and (C_2)	120
5.8	Union synonym of (C_1) and (C_2)	120
5.9	Venn diagram for complement A	121
5.10	Venn diagram for complement B	121



PTT AUTHM
PERPUSTAKAAN TUNKU TUN AMNAH

LIST OF ALGORITHMS

4.1	The proposed ontological features algorithm (OntF)	101
4.2	Features algorithm from Sánchez <i>et al.</i> , (2012)	102
4.3	Features algorithm from Batet <i>et al.</i> , (2013)	102
5.1	The process Hyb-TvX algorithm	128

LIST OF SYMBOLS AND ABBREVIATIONS

\forall	-	For all, any, each, every of element.
\in	-	Is element of
σ	-	Selection of
\geq	-	Greater than or equal to
A	-	Set A
B	-	Set B
o	-	Objects
\times	-	Cross
R	-	Real number
$=$	-	Similar to
x	-	Concept x
y	-	Concept y
\cap	-	Intersection
$*$	-	Multiplication
C_1	-	First concept
C_2	-	Second concept
C_3	-	Third concept
N_1	-	N_1 are the number of <i>is-a</i> links from A
N_2	-	N_2 are the number of <i>is-a</i> links from B
N_3	-	N_3 are the number of <i>is-a</i> links from C
LCS	-	Lower common subsume
$W(c)$	-	$W(c)$ is the set of words (nouns) in the corpus
c/C	-	Concept (c) or concept (C)
Σ	-	Total

N	-	The total number of word (noun) tokens in the corpus
$S(a, b)$	-	Similarity between concept a and b
$Sim(C_1, C_2)$	-	Similarity concept 1 and concept 2
$Sim(a, b)$	-	Similarity concept a and concept b
O	-	Ontology
O_1	-	First ontology
O_2	-	Second ontology
R_i	-	Type (i) of semantic relationship for first ontology
R_j	-	Type (j) of semantic relationship for second ontology
X	-	X correspondences to sets of a
Y	-	Y correspondences to sets of b
$X \cap Y / A \cap B$	-	Set X union set Y or Set A union set B
$ X - Y $	-	The relative complement of Y in X
$ Y - X $	-	The relative complement of X in Y
$A \cup B$	-	Set A union set B
$1 - \alpha > 0$	-	$1 - \alpha$ must be more the zero
α	-	Parameter for complement
$depth(a^p)$	-	Depth of ontology for concept a
$depth(b^q)$	-	Depth of ontology for concept b
$\alpha(a^p, b^q)$	-	Parameter for complement a and b
$w_w, w_u, w_n \geq 0$	-	Weighting parameter must be same or more than zero
(S_w)	-	Similarity word matching
(S_u)	-	Similarity feature matching
$(S_n) / S_{neighborhood}$	-	Similarity neighborhood
$S_p(a^p, b^q)$	-	Similarity parts for concept a and b
$S_f(a^p, b^q)$	-	Similarity function for concept a and b
$S_u(a^p, b^q)$	-	Similarity attribute or concept a and b
S_{synset}	-	Similarity synset or synonym
$S_{description}$	-	Similarity description
$Total_{hypo}$	-	All hyponym

$\text{total}_{\text{hypo}}O_i(S_i)$	- All hyponym in ontology i for subsumer i
(r)	- Pearson correlation coefficient
n	- Number of word pairs
$(\sum x_i y_i)$	- Total multiplication of human judgments (x_i) and y_i is the corresponding i th element in the list of similarity value
$(\sum x_i)$	- Total value human judgment or physician judgment
$(\sum y_i)$	- Total value similarity based similarity measurement method
M_R	- Matrix relationship
M_S	- Matrix subsume
$R_i(S_i)$	- The row represent the subsumer of relationship
$R_j(S_j)$	- The column represents the subsumer of relationship
Hyb-TvX	- Proposed method (A Hybrid Semantic Similarity Feature-based Measurement)
TvX-1	- Similarity Measurement level 1
TvX-2	- Similarity Measurement level 2
(Int)	- Intersection
(Un)/ U	- Union
(comp)	- Complement
(max)	- Take the maximum value of the words
$x \in A$	- x element of set A
$x \in B$	- x element of set B
$\text{comp } A/A'$	- Complement for set A
$\text{comp } B/B'$	- Complement for set B
$x \in U$	- x element of set Union
$x \notin A$	- x not element of set A
($\text{Int} C_1, C_2 $)	- Intersection for concept 1 and concept 2
($\text{max} C_1, C_2 $)	- Maximum for concept 1 and concept 2
($\text{Int} A, B $)	- Intersection set A and set B
($\text{Un} A, B $)	- Union set A and set B

$S_c(C_1, C_2)$	-	Similarity concept for concept 1 and concept 2
$S_s(C_1, C_2)$	-	Similarity synonym for concept 1 and concept 2
$S_f(C_1, C_2)$	-	Similarity features for concept 1 and concept 2
(S_c)	-	Similarity concept
(S_s)	-	Similarity synonym
(S_f)	-	Similarity features
$\lfloor \max(S_c, S_s, S_f) \rfloor$	-	Identify value maximum between S_c, S_s, S_f
(w_a)	-	The proposed parameter for $ comp A $
(w_b)	-	The proposed parameter for $ comp B $
GB	-	Gigabyte
RAM	-	Random Access Memory
Ghz	-	Gigahertz
PHP	-	Hypertext preprocessor
GIS	-	Geographic information systems
STS	-	Semantic textual similarity
WordNet	-	Ontology WordNet
Snomed-CT	-	Systemized Nomenclature of Medicine Clinical Term
MeSH	-	Medical Subject Heading
GO	-	Gene ontology
IC	-	Information content
UMLS	-	Unified Medical Language System
ICD	-	International Classification Disease
S	-	Synonym
SENSUS	-	Ontology SENSUS
Cyc KB	-	Cyc knowledge base
KB	-	Knowledge base
CPT	-	Current procedural terminology
ICD-10-CM	-	International Classification of Diseases, Tenth Revision, Clinical Modification
LOINC	-	Logical Observation Identifiers Names and Codes

NLM	- National Library of Medicine
MH	- Mesh Heading
STDS	- Spatial Data Transfer Standard
LCA	- Lower common ancestor
STS	- Semantic textual similarity
RM	- Root Matching
TM	- Terminological Matching
TM	- Terminological Subsumption
SS	- Semantic Subsumption
LCS	- Lower common subsume



PTTA UTHM
PERPUSTAKAAN TUNKU TUN AMINAH

LIST OF APPENDICES

A1	List of total concept datasets WordNet for Hliaoutakis <i>et al.</i> , (2006) benchmark	157
A2	List of total concept datasets MeSH for Hliaoutakis <i>et al.</i> , (2006) benchmark	159
A3	List of total concept datasets WordNet for Pedersen <i>et al.</i> , (2007) benchmark	161
A4	List of total concept datasets MeSH for Pedersen <i>et al.</i> , (2007) benchmark	162

REFERENCES

- Abdelrahman, A. M. B., & Kayed, A. (2015). A Survey on Semantic Similarity Measures between Concepts in Health Domain. *American Journal of Computational Mathematics*, 5(June), 204–214. <https://doi.org/10.4236/ajcm.2015.52017>.
- Aggarwal, N. (2012). Cross lingual semantic search by improving semantic similarity and relatedness measures. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7650 LNCS, 375–382. <https://doi.org/10.1007/978-3-642-35173-0-26>.
- Al-Mubaid, H., & Nguyen, H. A (2009). Measuring semantic similarity between biomedical concepts within multiple ontologies. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 39(4), 389–398. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5061528
- Al-Mubaid, H., & Nguyen, H. A. (2006). A cluster-based approach for semantic similarity in the biomedical domain. In *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings* (pp. 2713–2717). <https://doi.org/10.1109/IEMBS.2006.259235>.
- Alasoud, A. K. (2009). A Multi-Matching Technique for Combining Similarity Measures in Ontology Integration. Concordia University School of Graduate Studies. Concordia University. Thesis-PhD
- Allison, L., & Dix, T. I. (1986). A bit-string longest-common-subsequence algorithm. *Information Processing Letters*. [https://doi.org/10.1016/0020-0190\(86\)90091-8](https://doi.org/10.1016/0020-0190(86)90091-8).
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., & Al., A. P. D. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1), 25–29. <https://doi.org/10.1038/75556>.
- Banerjee, S., & Pedersen, T. (2002). An Adapted Lesk Algorithm for Word Sense

- Disambiguation Using WordNet. *International Conference on Intelligent Text Processing and Computational Linguistics*, 2276, 136–145. https://doi.org/10.1007/3-540-45715-1_11.
- Batet, M., Harispe, S., Ranwez, S., Sánchez, D., & Ranwez, V. (2014). An information theoretic approach to improve semantic similarity assessments across multiple ontologies. *Information Sciences*, 283, 197–210. <https://doi.org/10.1016/j.ins.2014.06.039>.
- Batet, M., & Sánchez, D. (2014). A Semantic Approach for Ontology Evaluation. *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, 138–145. <https://doi.org/10.1109/ICTAI.2014.30>.
- Batet, M., Sánchez, D., & Valls, A. (2011). An ontology-based measure to compute semantic similarity in biomedicine. *Journal of Biomedical Informatics*, 44(1), 118–125. <https://doi.org/10.1016/j.jbi.2010.09.002>.
- Batet, M., Sánchez, D., Valls, A., & Gibert, K. (2010). Exploiting taxonomical knowledge to compute semantic similarity: An evaluation in the biomedical domain. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6096 LNAI, 274–283. https://doi.org/10.1007/978-3-642-13022-9_28.
- Batet, M., Sánchez, D., Valls, A., & Gibert, K. (2013). Semantic similarity estimation from multiple ontologies. *Applied Intelligence*, 38(1), 29–44. <https://doi.org/10.1007/s10489-012-0355-y>.
- Bernstein, P. A., & Rahm, E. (2001). Generic Schema Matching , Ten Years Later, 695–701.
- Blank, A. (2003). Words and concepts in time: Towards diachronic cognitive onomasiology. *Trends In Linguistics Studies And Monographs*, 6–25. Retrieved from <http://www.metaphorik.de/01/blank.htm>.
- Bodenreider, O. (2004). The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 32(suppl 1), D267–D270. <https://doi.org/10.1093/nar/gkh061>.
- Budanitsky, A., & Hirst, G. (2006a). Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics*. <https://doi.org/10.1162/2006.32.1.13>.
- Budanitsky, A., & Hirst, G. (2006b). Evaluating WordNet-based Measures of Semantic Distance. *Computational Linguistics*, 32, 13–47.

- [https://doi.org/10.1162/coli.2006.32.1.13.](https://doi.org/10.1162/coli.2006.32.1.13)
- Bulskov, H., Knappe, R., & Andreasen, T. (2002). On measuring similarity for conceptual querying. *Flexible Query Answering Systems*, 100–111. Retrieved from http://link.springer.com/chapter/10.1007/3-540-36109-X_8.
- Choi, N., Song, I.-Y., & Han, H. (2006). A Survey on Ontology Mapping, *College of Information Science and Technology Drexel University, Philadelphia, PA 19014* 35(3): 34–41.
- Cohen, W. (2000). Data integration using similarity joins and a word-based information representation language. *ACM Transactions on Information Systems*. <https://doi.org/10.1145/352595.352598>.
- Couto, F. M., Silva, M. J., & Coutinho, P. M. (2007). Measuring semantic similarity between Gene Ontology terms. In *Data and Knowledge Engineering*, 61, 137–152. <https://doi.org/10.1016/j.datak.2006.05.003>
- Cross, V., Yu, X., & Hu, X. (2013). Unifying ontological similarity measures: A theoretical and empirical investigation. In *International Journal of Approximate Reasoning* (Vol. 54, pp. 861–875). <https://doi.org/10.1016/j.ijar.2013.03.003>
- Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*. <https://doi.org/10.1145/363958.363994>.
- Deselaers, T., & Ferrari, V. (2011). Visual and semantic similarity in ImageNet. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 1777–1784). <https://doi.org/10.1109/CVPR.2011.5995474>.
- Dice, L. R. (1945). Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26, 297–302. <https://doi.org/10.2307/1932409>.
- Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Peng, Y., Reddivari, P., Doshi, V., Sachs, J. (2004). Swoogle : A Semantic Web Search and Metadata Engine. In *Proceedings of the thirteenth ACM international conference on Information and knowledge management* (pp. 652–659).
- <https://doi.org/10.1145/1031171.1031289>.
- Doan, A., Madhavan, J., Domingos, P., & Halevy, A. (2004). Ontology matching: A machine learning approach. In *Handbook on Ontologies*, 1–20. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-24750-0_19.
- Ehrig, M., & Staab, S. (2004). QOM—quick ontology mapping. *The Semantic Web—ISWC 2004*, 1–28. Retrieved from <http://link.springer.com/chapter/10.1007/978->

3-540-30475-3_47.

- Elavarasi, S., Akilandeswari, J., & Menaga, K. (2014). A Survey on Semantic Similarity Measure. *Ijrat.org*, 2(3), 389–398. Retrieved from <http://www.ijrat.org/downloads/march-2014/paper id-232014114.pdf>.
- Erkan, G., & Radev, D. R. (2004). LexRank: Graph-based lexical centrality as salience in text summarization. *Journal of Artificial Intelligence Research*, 22, 457–479. <https://doi.org/10.1613/jair.1523>.
- Euzenat, J., & Shvaiko, P. (2007). Ontology Mapping. Springer Berlin Heidelberg.
- Fellbaum, C. (1998). WordNet: An Electronic Lexical Database. *British Journal Of Hospital Medicine London England 2005* (Vol. 71). <https://doi.org/10.1139/h11-025>.
- Gabrilovich, E., & Markovitch, S. (2007). Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis. *Proceedings of the 20th International Joint Conference on Artificial Intelligence* (pp. 1–6). <https://doi.org/10.1145/2063576.2063865>.
- Garla, V. N., & Brandt, C. (2012). Semantic similarity in the biomedical domain: an evaluation across knowledge sources. *BMC Bioinformatics*, 13, 261. <https://doi.org/10.1186/1471-2105-13-261>.
- Gomaa, W., & Fahmy, A. (2013). A survey of text similarity approaches. *International Journal of Computer Applications*, 68(13), 13–18.
- Grossman, J., Grossman, M., & Katz, R. (1980). The first system of weighted differential and integral calculus, ISBN 0-9771170-1-4
- Harper, D. (2016). Online Etymology Dictionary. Access at June 25, 2016, from <http://www.etymonline.com/index.php?l=s&p=42>
- Hirst, G., & St-Onge, D. (1998). Lexical chains as representations of context for the detection and correction of malapropisms. *WordNet - An Electronic Lexical Database*, 305–332. <https://doi.org/citeulike-article-id:4893262>.
- Hliaoutakis, A. (2005). Semantic Similarity Measures in MeSH Ontology and their application to Information Retrieval on Medline. Technical University of Crete, Greek, Thesis PhD.
- Hliaoutakis, A., Varelas, G., Voutsakis, E., Petrakis, E. G. M., & Milius, E. (2006). Information Retrieval by Semantic Similarity. *International Journal on Semantic Web and Information Systems*, 2, 55–73. <https://doi.org/10.4018/jswis.2006070104>.

- Iannone, L., Palmisano, I., & Fanizzi, N. (2007). An algorithm based on counterfactuals for concept learning in the Semantic Web. *Applied Intelligence*, 26, 139–159. <https://doi.org/10.1007/s10489-006-0011-5>.
- Islam, A., & Inkpen, D. (2008). Semantic text similarity using corpus-based word similarity and string similarity. *ACM Transactions on Knowledge Discovery from Data*. <https://doi.org/10.1145/1376815.1376819>.
- Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de La Société Vaudoise Des Sciences Naturelles*, 37, 547–579. <https://doi.org/citeulike-article-id:4118068>.
- Jaro, M. A. (1995). Probabilistic linkage of large public health data files. *Statistics in Medicine*, 14, 491–498. <https://doi.org/10.1002/sim.4780140510>.
- Jean-Mary, Y. R., Shironoshita, E. P., & Kabuka, M. R. (2009). Ontology Matching with Semantic Verification. *Journal of Web Semantics*, 7(3), 235–251. <https://doi.org/10.1016/j.websem.2009.04.001>.
- Jiang, J., & Conrath, D. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. *arXiv Preprint Cmp-lg/9709008*. Retrieved from <http://arxiv.org/abs/cmp-lg/9709008>.
- Jiang, R., Gan, M., & Doy, X., (2013). From ontology to semantic similarity: calculation of ontology-based semantic similarity. *The Scientific World Journal*, 2013, 793091. <https://doi.org/10.1155/2013/793091>.
- Jiang, Y., Zhang, X., Tang, Y., & Nie, R. (2015). Feature-based approaches to semantic similarity assessment of concepts using Wikipedia. *Information Processing and Management*, 51(3), 215–234. <https://doi.org/10.1016/j.ipm.2015.01.001>.
- Kasim, S. (2011). Fuzzy C-Means Clustering By Incorporating Biological Knowledge And Multi-Stage Filtering To Improve Gene Function Prediction. University Teknologi Malaysia.Thesis-PhD
- Kleinsorge, R., Tilley, C., & Willis, J. (2002). Unified Medical Language System (UMLS). *Encyclopedia of Library and Information Science*, 369–378. <https://doi.org/10.1002/9781118479612.ch16>.
- Ko, Y., Park, J., & Seo, J. (2004). Improving text categorization using the importance of sentences. *Information Processing and Management*, 40, 65–79. [https://doi.org/10.1016/S0306-4573\(02\)00056-0](https://doi.org/10.1016/S0306-4573(02)00056-0).
- Kolb, P. (2009). Experiments on the difference between semantic similarity and

- relatedness. *Proceedings of Nodalida 2009* (pp. 81–88).
- Kondrak, G. (2005). N -Gram Similarity and Distance. *Lecture Notes in Computer Science*, 3772, 115–126. <https://doi.org/10.1.1.69.2441>.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*. <https://doi.org/10.1037/0033-295X.104.2.211>.
- Lapata, M., & Barzilay, R. (2005). Automatic evaluation of text coherence: Models and representations. *IJCAI International Joint Conference on Artificial Intelligence* (pp. 1085–1090).
- Leacock, C., & Chodorow, M. (1998). Combining Local Context and WordNet Similarity for Word Sense Identification. *WordNet: An electronic lexical database*. (pp. 265–283). <https://doi.org/citeulike-article-id:1259480>.
- Lesk, M. (1986). Automatic sense disambiguation using machine readable dictionaries. *Proceedings of the 5th annual international conference on Systems documentation - SIGDOC '86* (pp. 24–26). <https://doi.org/10.1145/318723.318728>.
- Li, H., Tian, Y., & Cai, Q. (2011). Improvement of semantic similarity algorithm based on WordNet. *Proceedings of the 2011 6th IEEE Conference on Industrial Electronics and Applications, ICIEA 2011* (pp. 564–567). <https://doi.org/10.1109/ICIEA.2011.5975649>.
- Li, Y., McLean, D., Bandar, Z. A., O'Shea, J. D., & Crockett, K. (2006). Sentence similarity based on semantic nets and corpus statistics. *IEEE Transactions on Knowledge and Data Engineering*, 18(8), 1138–1150. <https://doi.org/10.1109/TKDE.2006.130>.
- Lin, C.-Y., & Hovy, E. (2003). Automatic evaluation of summaries using N-gram co-occurrence statistics. *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - NAACL '03, 2003*, 71–78. <https://doi.org/10.3115/1073445.1073465>.
- Lin, D. (1998a). An Information-Theoretic Definition of Similarity. *Proceedings of ICML*, 296–304.
- Lin, D. (1998b). Extracting Collocations from Text Corpora. *First Workshop on Computational Terminology* (pp. 57–63). <https://doi.org/10.1.1.56.1687>.

- Liu, Y., & Zong, C. (2004). Example-based Chinese-english MT. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics* (Vol. 7, pp. 6093–6096). <https://doi.org/10.1109/ICSMC.2004.1401354>.
- Liu, H., Bao, H., & Xu, D. (2012). Concept vector for semantic similarity and relatedness based on WordNet structure. *Journal of Systems and Software* (Vol. 85, pp. 370–381). <https://doi.org/10.1016/j.jss.2011.08.029>
- Lund, K., Burgess, C., & Atchley, R. A. (1995). Semantic and Associative Priming in High-Dimensional Semantic Space. In *Cognitive Science Proceedings, LEA* (pp. 660–665). <https://doi.org/10.1016/j.jconhyd.2010.08.009>.
- Madhavan, J., Bernstein, P. A., Doan, A., & Halevy, A. (2005). Corpus-based schema matching. In *Proceedings - International Conference on Data Engineering* (pp. 57–68). <https://doi.org/10.1109/ICDE.2005.39>.
- Maind, A., Deorankar, A., & Chatur, P. (2012). Measurement Of Semantic Similarity Between Words: A Survey. *International Journal of Computer Science, Engineering & Informatics*, 2(6), p51.
- Martinez-Gil, J., & Aldana-Montes, J. F. (2013). Semantic similarity measurement using historical google search patterns. *Information Systems Frontiers*, 15(3), 399–410. <https://doi.org/10.1007/s10796-012-9404-7>.
- Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38, 39–41. <https://doi.org/10.1145/219717.219748>.
- Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*. <https://doi.org/10.1080/01690969108406936>.
- Mohameth-François., Ranwez, S., Montmain, J., Regnault, A., Crampes, M., & Ranwez, V. (2012). User centered and ontology based information retrieval system for life sciences. *BMC Bioinformatics*, 13 Suppl 1, S4. <https://doi.org/10.1186/1471-2105-13-S1-S4>.
- Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48, 443–453. [https://doi.org/10.1016/0022-2836\(70\)900574](https://doi.org/10.1016/0022-2836(70)900574).
- Neter, J., Kutner, M.H., Nachtsheim,CJ., Wasserman, W., (1996). Applied Linear Statistical Modelslrwin, Chicago.
- Noy, N., & McGuinness, D. (2001). Ontology development 101: A guide to creating

- your first ontology. *Development*, 32, 1–25.
<https://doi.org/10.1016/j.artmed.2004.01.014>.
- Ochiai A. (1957). Zoogeographical studies on the soleoid fishes found Japan and its neighboring regions. *Japan Society Fish Sciences*, 22:526–30.
- Palmer, M., & Wu, Z. (1994). Verb Semantics And Lexical. *Proceeding ACL '94 Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics*, 133–138.
- Papineni, K., Roukos, S., Ward, T., & Zhu, W. (2002). BLEU: a method for automatic evaluation of machine translation. *Proceedings of the 40th annual meeting on association for computational linguistics*, 311–318.
<https://doi.org/10.3115/1073083.1073135>
- Patwardhan, S., & Pedersen, T. (2006). Using WordNet-based Context Vectors to Estimate the Semantic Relatedness of Concepts. *11th Conference of the European Chapter of the Association for Computational Linguistics*, 1501, 1–8.
<https://doi.org/citeulike-article-id:1574418>.
- Pedersen, T., Pakhomov, S. V. S., Patwardhan, S., & Chute, C. G. (2007). Measures of semantic similarity and relatedness in the biomedical domain. *Journal of Biomedical Informatics*, 40, 288–299. <https://doi.org/10.1016/j.jbi.2006.06.004>.
- Pesquita, C., Faria, D., Falcão, A. O., Lord, P., & Couto, F. M. (2009). Semantic similarity in biomedical ontologies. *PLoS Computational Biology*.
<https://doi.org/10.1371/journal.pcbi.1000443>.
- Petrakis, E., Varelas, G., Hliaoutakis, A., & Raftopoulou, P. (2006). X-Similarity: Computing Semantic Similarity between Concepts from Different Ontologies. *Journal of Digital Information Management*, 4(4), 233. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.114.3247>.
- Pirró, G. (2009). A semantic similarity metric combining features and intrinsic information content. *Data and Knowledge Engineering*, 68, 1289–1308.
<https://doi.org/10.1016/j.datak.2009.06.008>.
- Pirró, G., & Euzenat, J. (2010). A feature and information theoretic framework for semantic similarity and relatedness. *The Semantic Web-ISWC 2010*. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-17746-0_39.
- Pirrò, G., Ruffolo, M., & Talia, D. (2009). SECCO: On building semantic links in peer-to-peer networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*

- (Vol. 5480 LNCS, pp. 1–36). https://doi.org/10.1007/978-3-642-00685-2_1.
- Pirró, G., & Seco, N. (2008). Design , Implementation and Evaluation of a New Semantic Similarity Metric Combining Features and Intrinsic Information Content. *Lecture Notes in Computer Science Volume 5332* (Vol. 5332, pp. 1271–1288). <https://doi.org/10.1007/978-3-540-88873-4>.
- Quine, W. V. O. (1969). Ontological Relativity & Other Essays. *Ontological Relativity & Other Essays*, 1–25. <https://doi.org/10.2307/2024305>.
- Rada, R., Mili, H., Bicknell, E., & Blettner, M. (1989). Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(1), 17–30. <https://doi.org/10.1109/21.24528>.
- Resnik, P. (1995). Using information content to evaluate seantic similarity in a taxonomy. *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*.
- Richardson, R., Smeaton, A., & Murphy, J. (1994). Using WordNet as a knowledge base for measuring semantic similarity between words. Retrieved from <http://ducra.dcu.ie/wpapers/1994/1294.pdf>.
- Rodríguez, M. A., & Egenhofer, M. J. (2003a). Comparing Geospatial Entity Classes : An Asymmetric and Context-Dependent Similarity Measure. *International Journal of Geographical Information Science*, vol. 18, no. 3, pp. 229–256, 2004
- Rodríguez, M. A., & Egenhofer, M. J. (2003b). Determining semantic similarity among entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering*, 15, 442–456. <https://doi.org/10.1109/TKDE.2003.1185844>.
- Rubenstein, H., & Goodenough, J. B. (1965). Contextual correlates of synonymy. *Communications of the ACM*. <https://doi.org/10.1145/365628.365657>.
- Saleena, B., & Srivatsa, S. K. (2014). Using concept similarity in cross ontology for adaptive e-Learning systems. *Journal of King Saud University - Computer and Information Sciences*, 27(1), 1–12. <https://doi.org/10.1016/j.jksuci.2014.03.007>.
- Sánchez, D., & Batet, M. (2012). A new model to compute the information content of concepts from taxonomic knowledge. *International Journal on Semantic Web and Information Systems*, 8, 34–50. <https://doi.org/10.4018/jswis.2012040102>.
- Sánchez, D., & Batet, M. (2011). Semantic similarity estimation in the biomedical domain: An ontology-based information-theoretic perspective. *Journal of*

- Biomedical Informatics*, 44, 749–759. <https://doi.org/10.1016/j.jbi.2011.03.013>.
- Sánchez, D., & Batet, M. (2013). A semantic similarity method based on information content exploiting multiple ontologies. *Expert Systems with Applications*, 40(4), 1393–1399. <https://doi.org/10.1016/j.eswa.2012.08.049>.
- Sánchez, D., Batet, M., Isern, D., & Valls, A. (2012). Ontology-based semantic similarity: A new feature-based approach. *Expert Systems with Applications*, 39(9), 7718–7728. <https://doi.org/10.1016/j.eswa.2012.01.082>.
- Sánchez, D., Batet, M., Valls, A., & Gibert, K. (2010). Ontology-driven web-based semantic similarity. *Journal of Intelligent Information Systems*, 35, 383–413. <https://doi.org/10.1007/s10844-009-0103-x>.
- Sánchez, D., & Isern, D. (2011). Automatic extraction of acronym definitions from the Web. *Applied Intelligence*, 34, 311–327. <https://doi.org/10.1007/s10489-009-0197-4>.
- Sánchez, D., Isern, D., & Millan, M. (2011). Content annotation for the semantic web: An automatic web-based approach. *Knowledge and Information Systems*, 27, 393–418. <https://doi.org/10.1007/s10115-010-0302-3>.
- Sánchez, D., Solé-Ribalta, A., Batet, M., & Serratosa, F. (2012). Enabling semantic similarity estimation across multiple ontologies: an evaluation in the biomedical domain. *Journal of Biomedical Informatics*, 45(1), 141–55. <https://doi.org/10.1016/j.jbi.2011.10.005>.
- Saruladha, K., Aghila, G., & Bhuvaneswary, A. (2011a). COSS: Cross Ontology Semantic Similarity measure-An information content based approach. *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*, 485–490. <https://doi.org/10.1109/ICRTIT.2011.5972360>.
- Saruladha, K., Aghila, G., & Bhuvaneswary, A. (2011b). Information content based semantic similarity approaches for multiple biomedical ontologies. *Communications in Computer and Information Science* (Vol. 191 CCIS, pp. 327–336). https://doi.org/10.1007/978-3-642-22714-1_34.
- Schallehn, E., Sattler, K. U., & Saake, G. (2004). Efficient similarity-based operations for data integration. *Data and Knowledge Engineering*, 48, 361–387. <https://doi.org/10.1016/j.datak.2003.08.004>.
- Schickel-Zuber, V., & Faltings, B. (2007). OSS: A semantic similarity function based on hierarchical ontologies. *IJCAI International Joint Conference on Artificial Intelligence* (pp. 551–556).

- Schütze, H. (1998). Automatic Word Sense Discrimination. *Computational Linguistics*, 24, 97–123.
- Schwartzman, S. (1996). The Words of Mathematics: An Etymological Dictionary of Mathematical Terms Used in English (Spectrum). *The Mathematical Association of America* (September 1996).
- Schwering, A. (2008). Approaches to semantic similarity measurement for geo-spatial data: A survey. *Transactions in GIS*. <https://doi.org/10.1111/j.1467-9671.2008.01084.x>.
- Singh, P. (2004). Web Ontology to Facilitate Semantic Web. *2nd International CALIBER-2004, New Delhi, 11-13 February, 2004* (pp. 11–13).
- Slimani, T. (2013). Description and Evaluation of Semantic similarity Measures Approaches. *Journal of Computer Applications of Computer Applications*, 80-no 10, 1–10.
- Smith, T. F., & Waterman, M. S. (1981). Identification of common molecular subsequences. *Journal of Molecular Biology*, 147, 195–197. [https://doi.org/10.1016/0022-2836\(81\)90087-5](https://doi.org/10.1016/0022-2836(81)90087-5).
- Solé-Ribalta, A., Sánchez, D., Batet, M., & Serratosa, F. (2014). Towards the estimation of feature-based semantic similarity using multiple ontologies. *Knowledge-Based Systems*, 55, 101–113. <https://doi.org/10.1016/j.knosys.2013.10.015>.
- Spasic, I., Ananiadou, S., McNaught, J., & Kumar, A. (2005). Text mining and ontologies in biomedicine: Making sense of raw text. *Briefings in Bioinformatics*, 6, 239–251. <https://doi.org/10.1093/bib/6.3.239>.
- Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, 25(1–2), 161–197. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6).
- Sussna, M. (1993). Word sense disambiguation using a massive of computer for free-text semantic indexing network. *CIKM '93 Proceedings of the Second International Conference on Information and Knowledge Management*, 67–74. <https://doi.org/10.1145/170088.170106>.
- Thiagarajan, R., Manjunath, G., & Stumptner, M. (2008). Computing Semantic Similarity Using Ontologies. *Computing*, Germany. Retrieved from <http://www.hpl.hp.com/techreports/2008/HPL-2008-87.pdf>

- Tversky, A. (1977). Features of similarity. *Psychological Review*. <https://doi.org/10.1037/0033-295X.84.4.327>.
- Vicent, C., Sánchez, D., & Moreno, A. (2013). An automatic approach for ontology-based feature extraction from heterogeneous textual resources. *Engineering Applications of Artificial Intelligence*, 26, 1092–1106. <https://doi.org/10.1016/j.engappai.2012.08.002>.
- Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., & Hübner, S. (2001). Ontology-Based Integration of Information - A Survey of Existing Approaches. *IJCAI-01 Workshop: Ontologies and Information Sharing* (pp. 108–117). <https://doi.org/10.1016/j.ijhcs.2008.07.007>.
- Wang, L., & Chen, M. (2013). A New Model to Compute Semantic Similarity from Multi-ontology. *China Semantic Web Symposium and Web Science Conference. Communications in Computer and Information Science book series (CCIS, vol: 406)* 1–10.
- Wang, J. Z., Du, Z., Payattakool, R., Yu, P. S., & Chen, C.-F. (2007). A new method to measure the semantic similarity of GO terms. *Bioinformatics (Oxford, England)*, 23, 1274–1281. <https://doi.org/10.1093/bioinformatics/btm087>.
- Wegrzyn-Wolska, K., & Szczepaniak, P. S. (2005). Classification of RSS-Formatted Documents Using Full Text Similarity Measures. *Lecture Notes in Computer Science*, 3579, 400–405.
- Winkler, W. E. (1990). String Comparator Metrics and Enhanced Decision Rules in the Fellegi-Sunter Model of Record Linkage. *Proceedings of the Section on Survey Research* (pp. 354–359).
- Yatskevich, M., & Giunchiglia, F. (2007). Element level semantic matching using WordNet. *CiteSeerX*.
- Yuhua Li, Zuhair A. Bandar, and D. M. (2003). An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on Knowledge and Data Engineering*, 15(4), 871–882. <https://doi.org/10.1109/TKDE.2003.1209005>.
- Zhang, S., Zheng, X., & Hu, C. (2015). A Survey of Semantic Similarity and its Application to Social Network Analysis. *IEEE International Conference on Big Data (Big Data)* (pp. 1–3).