SemanticRank: Ranking Keywords and Sentences Using Semantic Graphs

George Tsatsaronis
Web: http://www.idi.ntnu.no/~gbt/
e-mail: gbt@idi.ntnu.no

NTNU
Norwegian University of Science and Technology

Iraklis Varlamis
Web: http://www.dit.hua.gr/~varlamis/
e-mail: varlamis@hua.gr

HAROKOPION UNIVERSITY

Kjetil Nørvåg
Web: http://www.idi.ntnu.no/~noervaag/
e-mail: Kjetil.Norvag@idi.ntnu.no

NTNU
Norwegian University of Science and Technology

Presentation Layout

- Introduction and Motivation
- Contributions
- SemanticRank
  - Semantic Graph Construction
  - Ranking nodes
- Experimental Evaluation
  - Keyword Extraction
  - Text Summarization
- Conclusions and Future Directions

Introduction

- Selection of the most important text elements
  - Keyword Selection
    - Top ranked terms as the most representative (TF-IDF, variations)
    - Co-occurrence frequency for recognizing composite terms
    - Applications: summary keywords, tags extraction, classification, retrieval
  - Sentence Selection
    - More difficult: Rank sentences of a given text passage
    - Observation: Most important sentences occur in the beginning
    - Cohesion-based methods
    - Applications: text summarization, title extraction, snippet extraction, search engines

Motivation

- Graph based ranking algorithms
  - Very helpful to rank Web pages, members in a social network, authors in bibliographic databases
  - Implies the existence of explicit links between nodes (pages, members, authors)

- Use of graph based ranking algorithms for selecting the most important keywords and sentences
  - Problems:
    - No explicit links - How to construct/weigh edges?
    - What algorithms are suitable for ranking in this case?
  - Proposed Solution:
    - Create graphs from text using implicit links based on semantic similarity
    - Apply modified/weighted versions of ranking algorithms to the semantic graphs

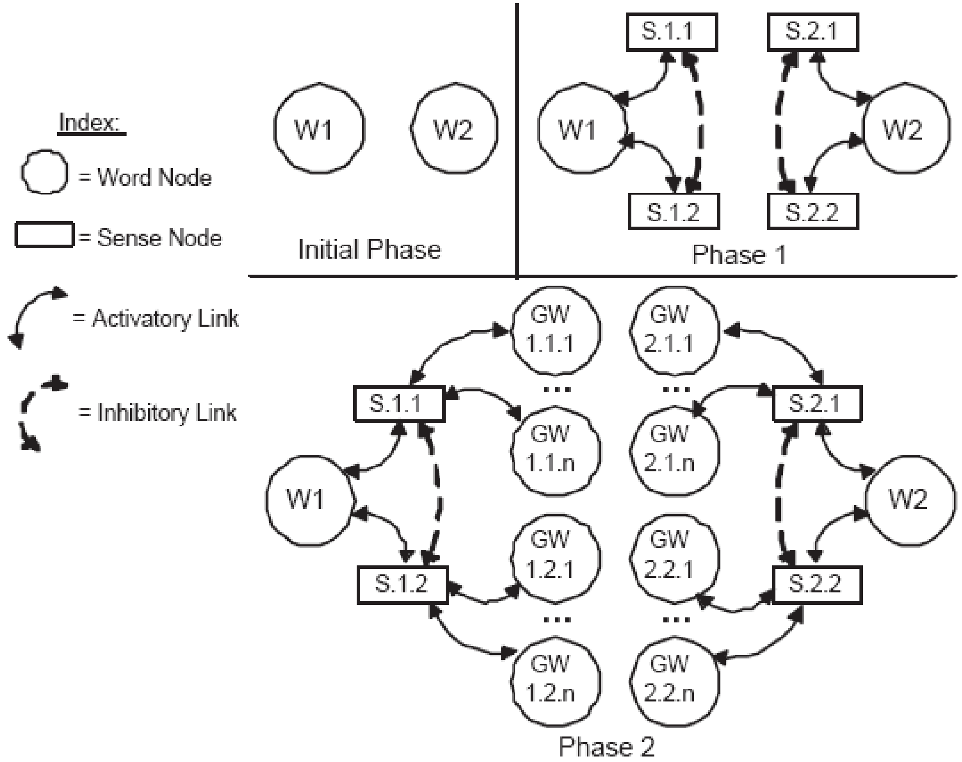
Contributions

- SemanticRank
  - A novel algorithm for constructing semantic graphs from text and rank keywords/sentences
  - Combination of knowledge based measures of semantic relatedness employed
- Modular Design
  - Allows any graph-based ranking algorithm to be employed
- Experimental evaluation using a variety of graph-based ranking algorithms
  - Keyword selection
  - Text summarization

Earlier use of Semantic Graphs in NLP (Word Sense Disambiguation)

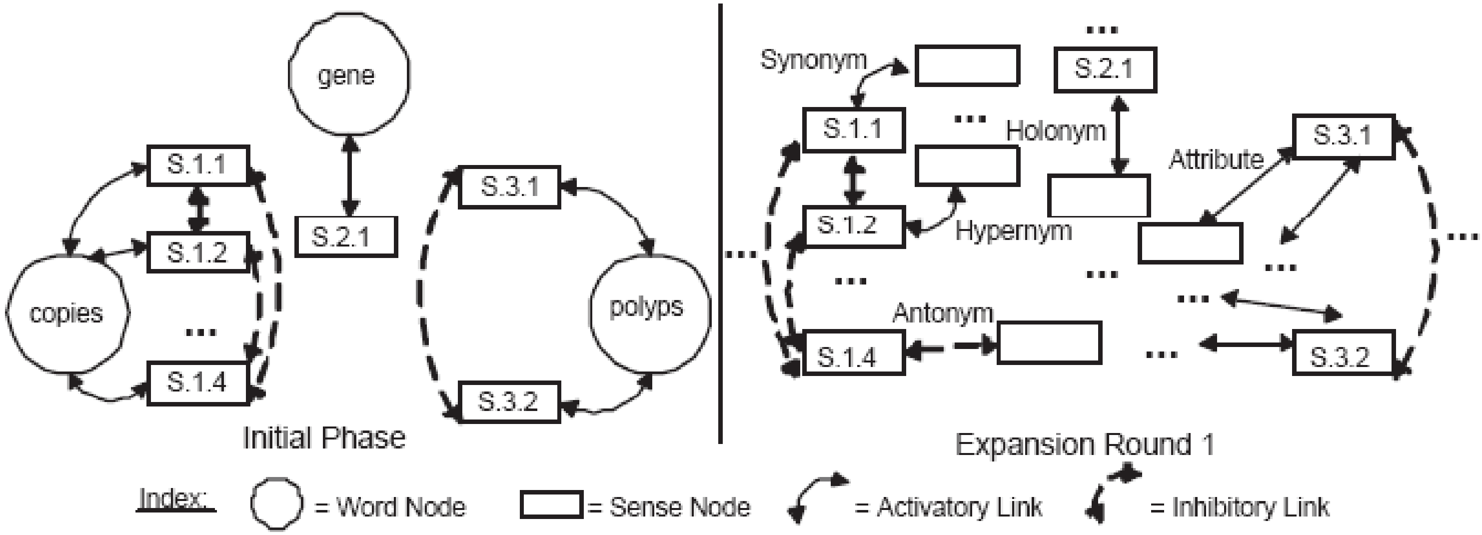
- Graph-based methods demonstrate SoA results amongst unsupervised WSD methods [Sinha and Mihalcea 2007].
- An example of an earlier approach: [Veronis and Ide 1990]

- Spreading of activation (social network analysis) to process the network.



Semantic Networks Creation

- Tsatsaronis et. al. [IJCAI 2007] proposed a new method for constructing semantic networks
  - Use all of the available semantic information from WN
  - Use edges weighting scheme
  - Example: “If both copies of a certain gene were knocked out, benign polyps would develop”



Applications of Semantic Networks

- Semantic similarity/relatedness [Budanitsky and Hirst 2006]

- *Omiotis* measure, Tsatsaronis et al. [JAIR 2010]
  - Relatedness computation between:
    - Term pairs
    - Sentence pairs

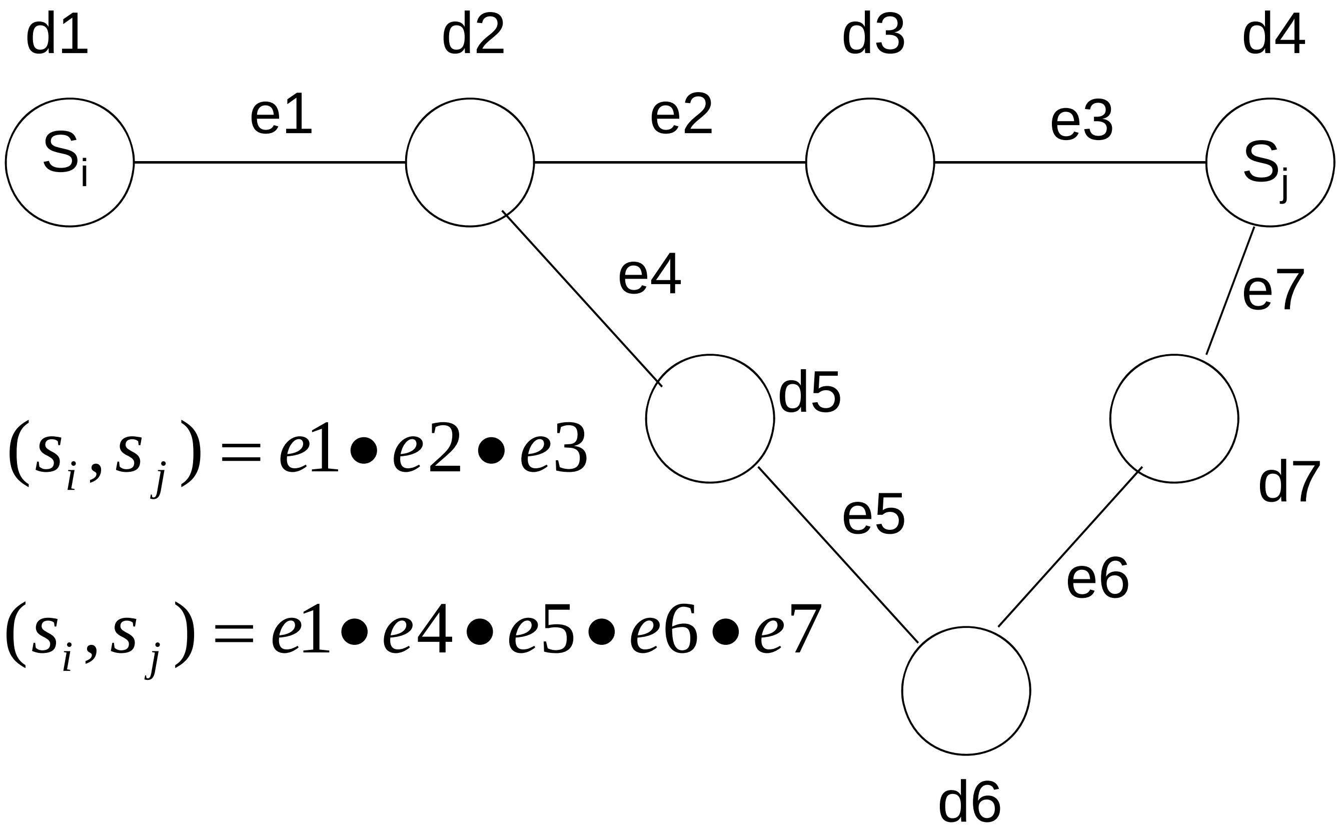
- Publicly available: http://omiotis.hua.gr

- Very good performance in word-to-word and sentence-to-sentence similarity

SemanticRank: Semantic Graphs Creation

- Keyword Graphs
  - Pre-processing
    - stopword removal
    - Composite term recognition
  - Terms as nodes
  - Compute all pair wise term semantic relatedness values as edges
    - 2 measures used: WordNet based: Omiotis, and Wiki-based: Milne and Witten (2008)
- Sentence Graphs
  - Pre-processing
    - Sentence splitting
  - Sentences as nodes
  - Compute all pair wise sentence semantic relatedness values as edges
    - Omiotis used

Omiotis: Semantic Relatedness Between Terms Tsatsaronis et al. (2010)



$$SCM(s_i, s_j) = e1 \bullet e2 \bullet e3$$

$$SCM(s_i, s_j) = e1 \bullet e4 \bullet e5 \bullet e6 \bullet e7$$

$$SPE(s_i, s_j) = \frac{1}{d\max}\frac{2d1d2}{d1+d2} \bullet \frac{1}{d\max}\frac{2d2d3}{d2+d3} \bullet \frac{1}{d\max}\frac{2d3d4}{d3+d4}$$

Omiotis: Semantic Relatedness Between Sentences

A

t_{11}

t_{12}

….

t_i

….

t_{1n}

$SR(t_i,t_{21})$

$SR(t_i,t_{22})$

……..

$SR(t_i,t_{2m})$

B

t_{21}

t_{22}

….

t_j

….

t_{2m}

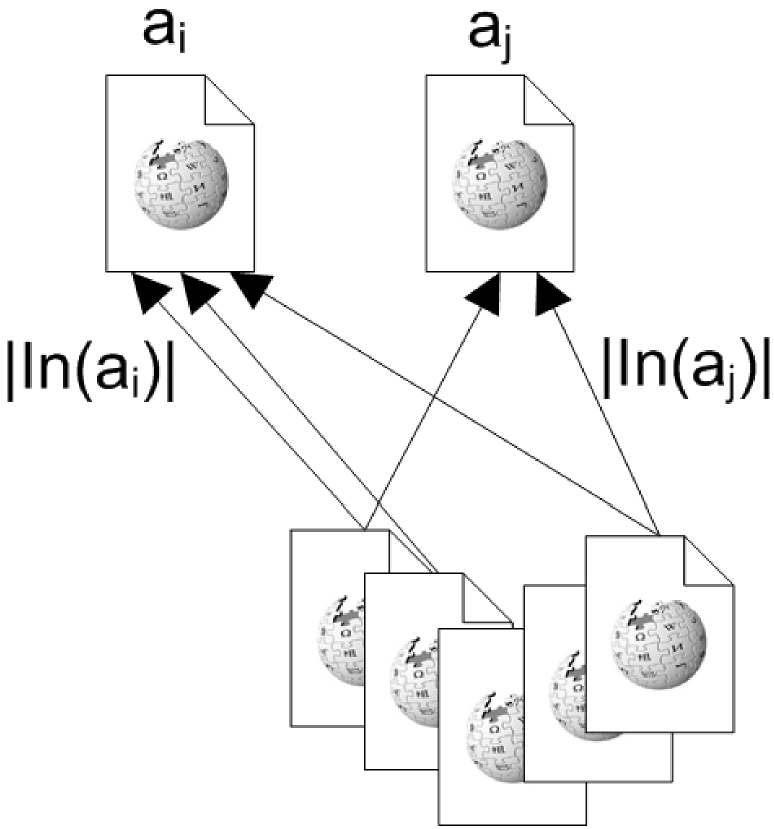
$$x(i) = \arg\max_{j \in [1,|B|]} (\lambda_{i,j} \cdot SR(t_i, t_j))$$

$$\lambda_{i,j} = \frac{2 \cdot TFIDF(t_i, A) \cdot TFIDF(t_j, B)}{TFIDF(t_i, A) + TFIDF(t_j, B)}$$

$$\zeta_1(A,B) = \frac{1}{|A|} \cdot (\sum_{i=1}^{|A|} \lambda_{i,x(i)} \cdot SR(t_i, t_{x(i)}))$$

$$\boxed{OMIOTIS(A,B) = \frac{1}{2}[\zeta_1(A,B) + \zeta_2(A,B)]}$$

Milne and Witten (2008)



$$SR_{Wiki}(t_i, t_j) = \frac{\log(\max\{|In(a_i)|, |In(a_j)|\}) - \log(|In(a_i) \cap In(a_j)|)}{\log(|W|) - \log(\min\{|In(a_i)|, |In(a_j)|\})}$$

SemanticRank: Measure of Semantic Relatedness

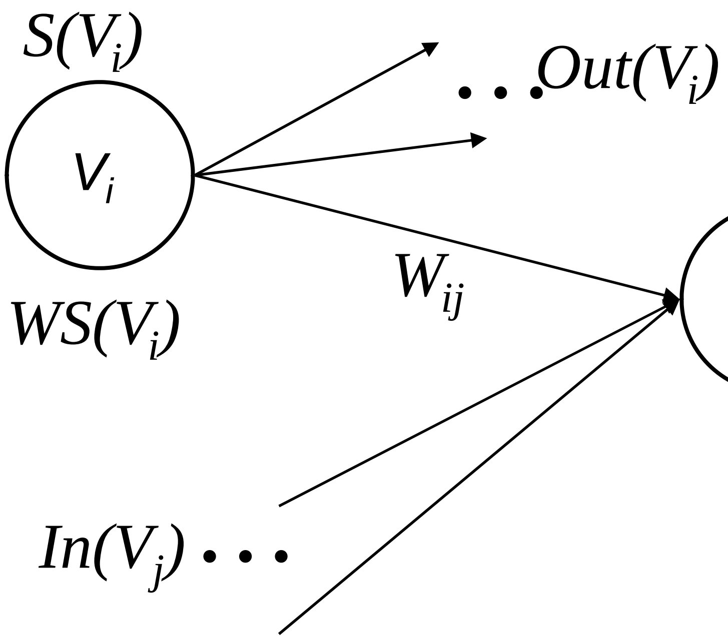
- Keyword Graphs' edges
  - Combination of Omiotis and Milne and Witten

$$SRT(t_i, t_j) = \begin{cases} 1, & t_i = t_j \\ SR_{WN}(t_i, t_j), & \text{if } t_i, t_j \in WordNet \\ SR_{Wiki}(t_i, t_j), & \text{if } t_i, t_j \in Wikipedia \\ 0, & otherwise \end{cases}$$

- Sentence Graphs' edges

$$SRT(s_i, s_j) = Omiotis(s_i, s_j)$$

PageRank

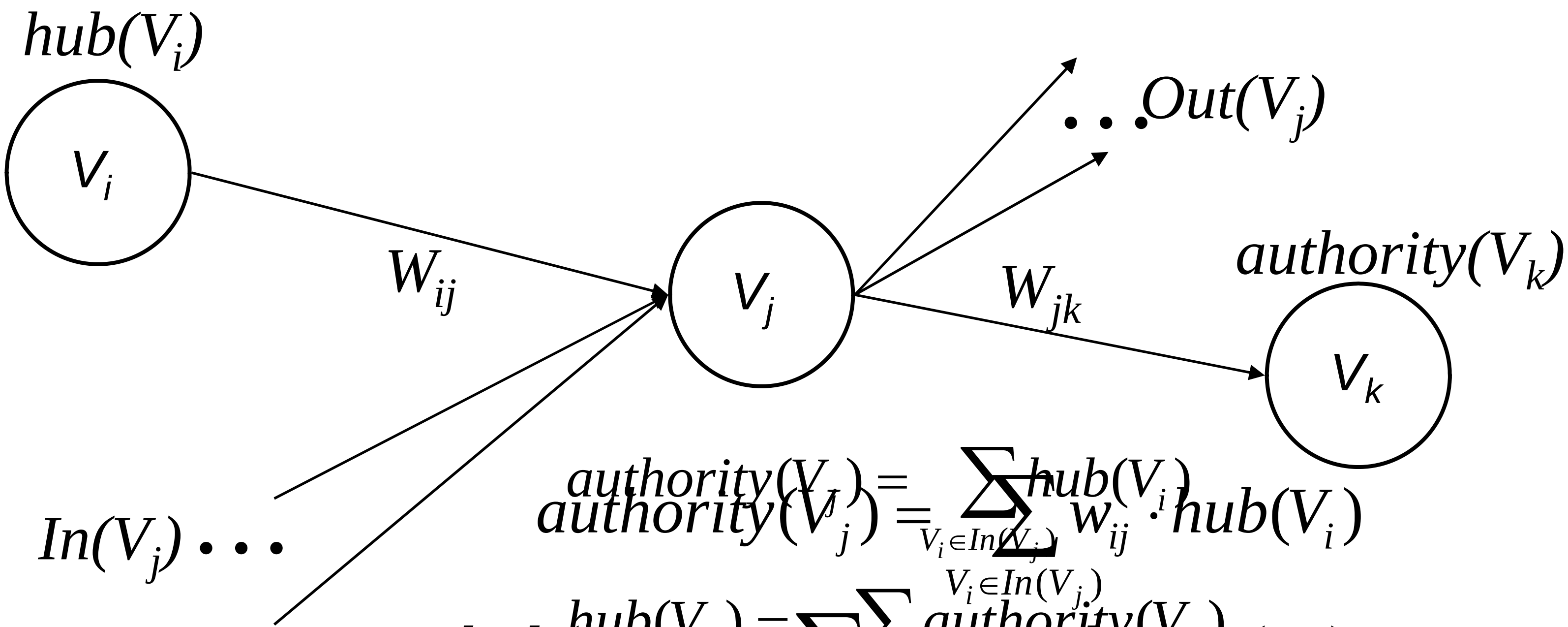


$$S(V_j) = (1-d) + d \cdot \sum_{i \in In(V_j)} \frac{S(V_i)}{|Out(V_i)|}$$

$$WS(V_j) = (1-d) + d \cdot \sum_{V_i \in In(V_j)} \frac{w_{ij} \cdot WS(V_i)}{\sum_{V_k \in Out(V_i)} w_{ik}}$$

- [Brin and Page 1998]
- [Mihalcea et al. 2004]

HITS



$$authority(V_j) = \sum_{V_i \in In(V_j)} hub(V_i)$$

$$authority(V_j) = \sum_{V_i \in In(V_j)} w_{ij} \cdot hub(V_i)$$

$$hub(V_j) = \sum_{V_k \in Out(V_j)} authority(V_k)$$

$$hub(V_j) = \sum_{V_k \in Out(V_j)} w_{jk} \cdot authority(V_k)$$

- [Kleinberg 1999]

Two Modifications of PageRank for Keyword Graphs

- Averaged PageRank Weighting

$$APW(t_i) = \frac{1}{2}\left(\frac{WPR(t_i)}{WPR_{\max}} + \frac{TF-IDF(t_i, d_j)}{TF-IDF_{\max}}\right)$$

- Priors Biased PageRank

$$P-PR(t_i) = (1-\beta_{t_i}) + \beta_{t_i} \cdot \sum_{j \in IN(t_i)} \frac{w_{ij} \cdot P-PR(t_j)}{\sum_{k \in OUT(t_j)} w_{jk}}$$

Experimental Evaluation: Keyword Extraction

- Inspec Database
  - 500 abstracts, mean number of assigned terms: 7.63
- SemanticRank performs better by almost 10 p.p. in F-Measure
- APW and P-PR seem to be more appropriate for ranking keywords

Method		P	R	F
Sem (k=5)	WPR	0.396	0.121	0.1853
	WHITS	0.348	0.088	0.14
	APW	0.556	0.185	0.278
	P-PR	**0.659**	0.226	0.337
Sem (k=10)	WPR	0.368	0.2463	0.296
	WHITS	0.335	0.138	0.195
	APW	0.498	0.331	0.398
	P-PR	0.524	0.352	0.422
Sem (k=15)	WPR	0.371	0.364	0.368
	WHITS	0.355	0.241	0.287
	APW	0.449	0.442	0.446
	P-PR	0.451	0.441	0.446
Sem (k=20)	WPR	0.376	0.466	0.417
	WHITS	0.374	0.312	0.34
	APW	0.421	**0.532**	**0.47**
	P-PR	0.418	0.514	0.46
USem (k=5)	UPR	0.057	0.046	0.048
	UHITS	0.061	0.053	0.055
USem (k=10)	UPR	0.06	0.102	0.07
	UHITS	0.06	0.108	0.072
USem (k=15)	UPR	0.052	0.116	0.069
	UHITS	0.054	0.123	0.072
USem (k=20)	UPR	0.052	0.14	0.074
	UHITS	0.053	0.151	0.076
Michalcea (2004)		0.312	0.431	0.362
Hulth (2003)		0.252	0.517	0.339

Experimental Evaluation: Text Summarization (1/2)

- DUC Data
  - DUC 2001, and 2002 single document summarization task

System		F-Measure
Sem	WPR	0.40996(0.39067 – 0.4292)
	WHITS	0.3651(0.3435 – 0.38609)
USem	UPR	0.2951(0.2727 – 0.3195)
	UHITS	0.3132(0.2901 – 0.3375)
T		0.4131(0.3922 – 0.434)
P		0.4039(0.3843 – 0.4226)
O		0.3905(0.3663 – 0.4132)
V		0.3885(0.368 – 0.4085)
Q		0.3857(0.3616 – 0.4089)
Baseline		0.3549(0.3329 – 0.3756)

System		F-Measure
Sem	WPR	0.4971(0.4799 – 0.5164)
	WHITS	0.3836(0.3815 – 0.4047)
USem	UPR	0.3086(0.297-0.32084)
	UHITS	0.2851(0.2735-0.297)
TextRank		0.4904
S27		0.5011
S31		0.4914
S28		0.489
S21		0.4869
S29		0.4681
Baseline		0.4779

- SemanticRank performs comparably to the top-5 participating systems for both data sets
- Weighted versions perform always much better

Experimental Evaluation: Text Summarization (2/2)

- DUC Data
  - DUC 2007, multi-document summarization (update task)

System		F (R-2)	F (R-SU4)
Sem	WPR	0.093	0.133
	WHITS	0.078	0.115
USem	UPR	0.031	0.069
	UHITS	0.028	0.062
S40		0.111	0.143
S55		0.098	0.135
S45		0.096	0.132
S44		0.093	0.136
S47		0.093	0.130
Baseline		0.085	0.122

- Same observations as in the single-document summarization
- Baseline method in this case is different

Conclusions

- A novel algorithm for ranking keywords and sentences has been introduced
- Semantic graphs are created based on the semantic similarity between keywords and sentences
- Experimental evaluation shows that SemanticRank can select very successfully the most central keywords and sentences
- Weighted versions of graph ranking algorithms perform always better in all examined cases compared to the unweighted

Future Directions

- Examine more graph-based ranking algorithms (e.g., personalized PageRank, or other variations)

- Embed the algorithm into more concrete applications
  - Sentiment analysis
  - Opinion mining

- Semantic Social Networks

Questions

Thank you very much for your attention!

Questions/Comments?

References (1/4)

- [Budanitsky et al. 2006] Budanitsky, A. and Hirst, G, *Evaluating WordNet-based Measures of Lexical Semantic Relatedness*, Computational Linguistics, 32(1), pp. 13-47,, 2006.
- [Leacock et al. 1998] Leacock, C., and Chodorow, M. (1998). Combining local context and WordNet similarity for word sense identification. In C. Fellbaum (ed.), *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press, pp. 265-283
- [Mavroeidis et. al. 2005] D. Mavroeidis, G. Tsatsaronis, M. Vazirgiannis, M. Theobald, G. Weikum, *Word Sense Disambiguation for Exploiting Hierarchical Thesauri in Text Classification*, PKDD 2005.
- [Resnik 1995] Resnik, P. *Using information content to evaluate semantic similarity*. IJCAI 1995.
- [Tsatsaronis et al. 2007] Tsatsaronis, G., Androutsopoulos, I. and Vazirgiannis, M., *Word sense disambiguation with spreading activation networks generated from thesauri. In: Proc. of the 20th IJCAI. (2007) 1725–1730.*
- [Finkelstein et al. 2002] Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., Ruppin, E.: Placing search in context: The concept revisited. ACM TOIS 20(1) (2002) 116–131
- [Bloehdorn et al. 2006] S. Bloehdorn, R. Basili, M. Cammisa, and A. Moschitti. *Semantic kernels for text classification based on topological measures of feature similarity*. In Proc. of the ICDM 2006, pages 18–22, 2006.

References (2/4)

- [Landauer et al. 1998] Landauer, T., Foltz, P., Laham, D.: Introduction to latent semantc analysis. Discourse Processes 25 (1998) 259–284
- [Sanderson 1994] Sanderson, M.: Word sense disambiguation and information retrieval. In: Proc. of the 17th SIGIR, Ireland, ACM (1994) 142–151
- [Jiang et al. 1997] Jiang, J., Conrath, D.: Semantic similarity based on corpus statistics and lexical taxonomy. In: Proc. of ROCLING X. (1997) 19–33
- [Hirst et al. 1998] Hirst, G., St-Onge, D.: Lexical chains as representations of context for the detection and correction of malapropisms. In: WordNet: An Electronic Lexical Database, chapter 13, Cambridge, The MIT Press (1998) 305–332
- [Leacock et al. 1998] Leacock, C., Miller, G., Chodorow, M.: Using corpus statistics and wordnet relations for sense identification. Computational Linguistics 24(1) (March 1998) 147–165
- [Lin 1998] Lin, D.: An information-theoretic definition of similarity. In: Proc. of the 15th International Conference on Machine Learning. (1998) 296–304
- [Turney 2006] Turney, P.D., Similarity of Semantic Relations, Computational Linguistics, 32(3), pp. 379-316.

References (3/4)

- [Trieschnigg et al. 2008] Trieschnigg, D., Meij, E., Rijke, M. and Kraaij W., : Measuring Concept Relatedness Using Language Models. SIGIR (2008), pp. 823-824.
- [Jarmasz et al. 2003] Jarmasz, M., Szpakowicz, S.: Roget’s thesaurus and semantic similarity. In: Proc. of Conference on Recent Advances in Natural Language Processing. (2003) 212–219.
- [Song et al. 1999] Song, F., Croft, B.,: A General Language Model for Information Retrieval, CIKM (1999), 316-321.
- [Lesk 1986] Lesk, M. *Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone*. SIGDOC 1986.
- [Montoyo et. Al. 2005] A. Montoyo, A. Suarez, G. Rigau, M. Palomar, *Combining Knowledge-based and Corpus-based Word-Sense-Disambiguation Methods.* JAIR 2005.
- [Veronis and Ide 1990] J. Veronis and N. Ide, *Word Sense Disambiguation with very large neural networks extracted from machine readable dictionaries*, COLING 1990.
- [Crestani 1997] F. Crestani, *Application of spreading activation techniques in information retrieval.* Artificial Intelligence Review, 11:453-482, 1997.

References (4/4)

- [Haliday and Hasan 1976] Halliday, M. and Hasan, R. Cohesion in English. Longman 1976.
- [Cowie et. Al. 1992] J. Cowie, J. Guthrie and L. Guthrie, Lexical disambiguation using simulated annealing, COLING 1992.
- [Ide and Veronis 1998] N. Ide, and J. Veronis. *Word Sense Disambiguation: the state of the art. Computational Linbguistics, 24(1): 1-40, 1998.*
- [Mihalcea and Tarau 2004] Mihalcea, R., and Tarau, P.,: "TextRank: Bringing Order into Texts", EMNLP 2004.
- [Mihalcea and Csomai] Mihalcea, R. and Csomai, A.,: "Senselearner: Word Sense Disambiguation for all words in unrestricted text", ACL 2005.
- [Kohomban and Lee 2005] Kohomban, U.S., and Lee, W.S.: "Learning Semantic Classes for Word Sense Disambiguation", ACL 2005.
- [Navigli 2006] Navigli, R.: "Online word sense disambiguation with structural semantic interconnections", EACL 2006.
- [Hoste et al. 2002] Hoste, V. and Daelemans, W. and Hendrickx, I. and van den Bosch, A., "Evaluating the results of a memory-based word-expert approach to unrestricted word sense disambiguation"