Word Sense Disambiguation and Named-Entity Disambiguation using graph-based algorithms

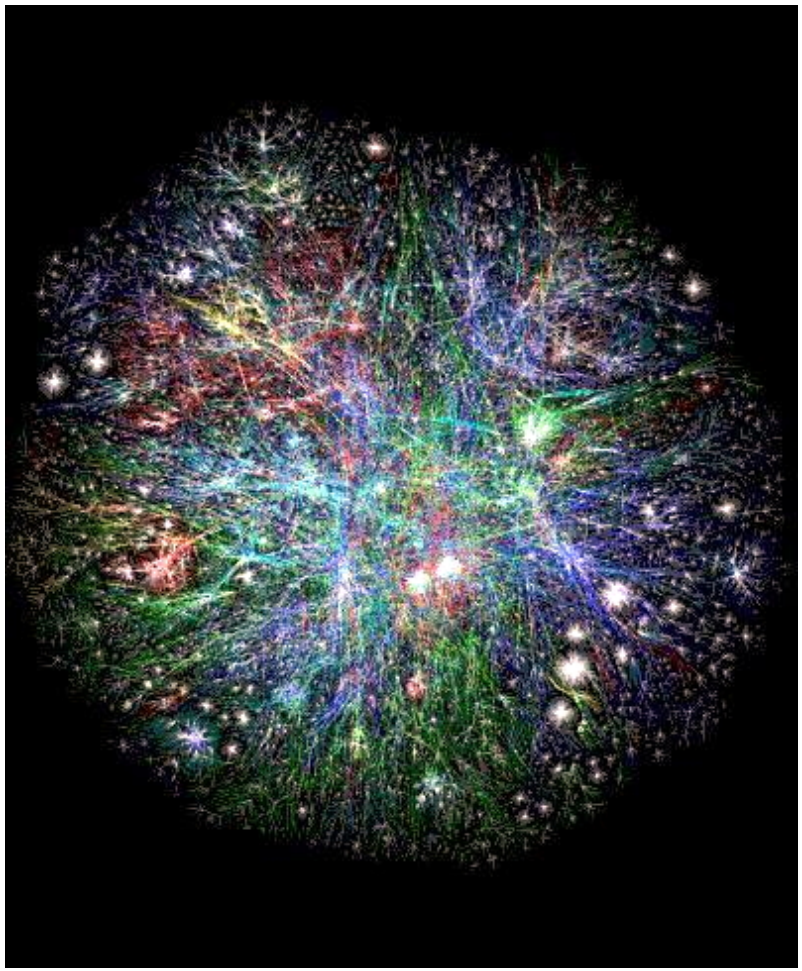
Eneko Agirre
ixa2.si.ehu.es/eneko

IXA NLP Group
University of the Basque Country

WSAP in Copenhaguen, 2014

Algorithms on Large Graphs

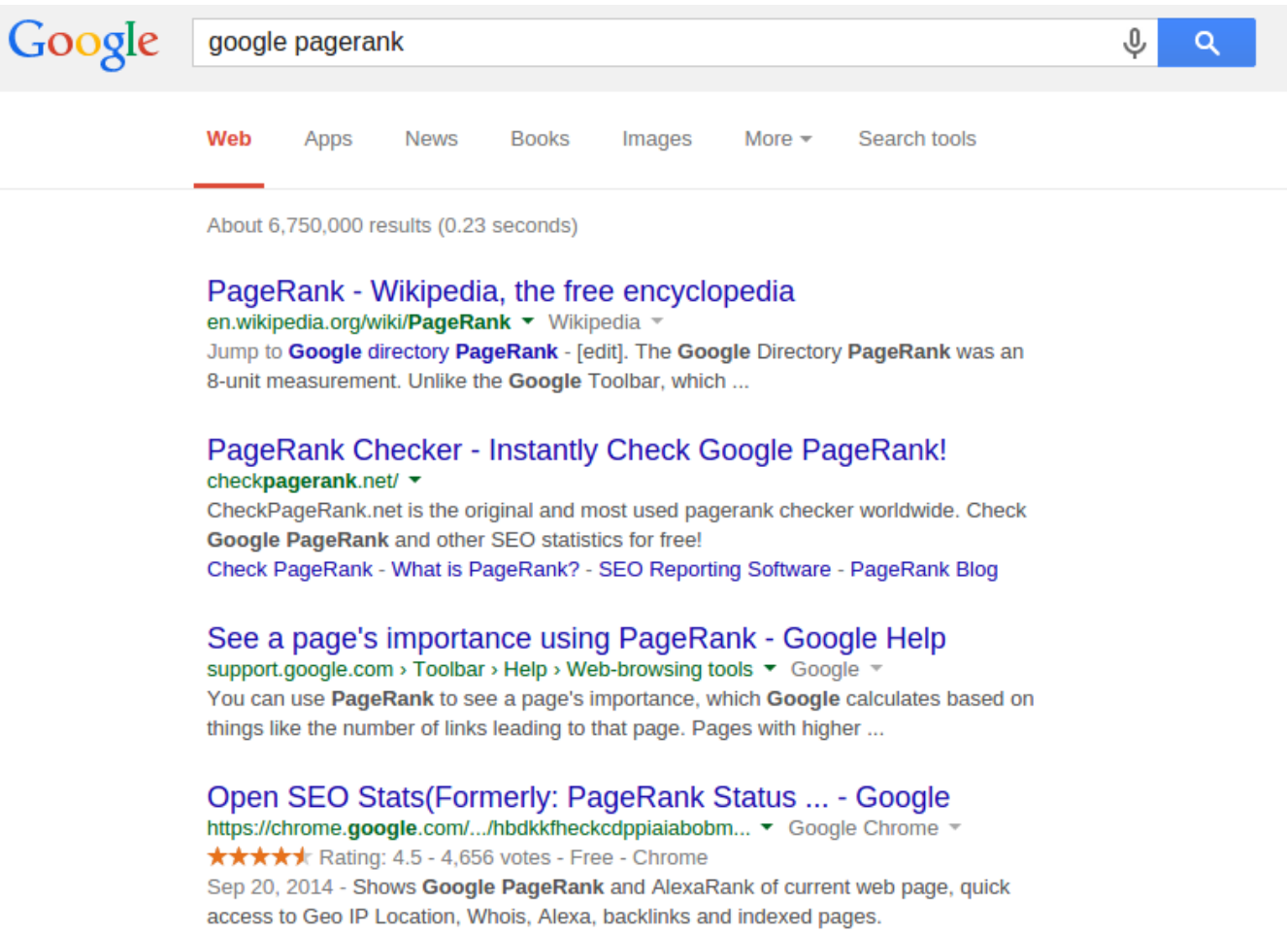
WWW, Random walks, PageRank and Google



source: http://opte.org

Algorithms on Large Graphs

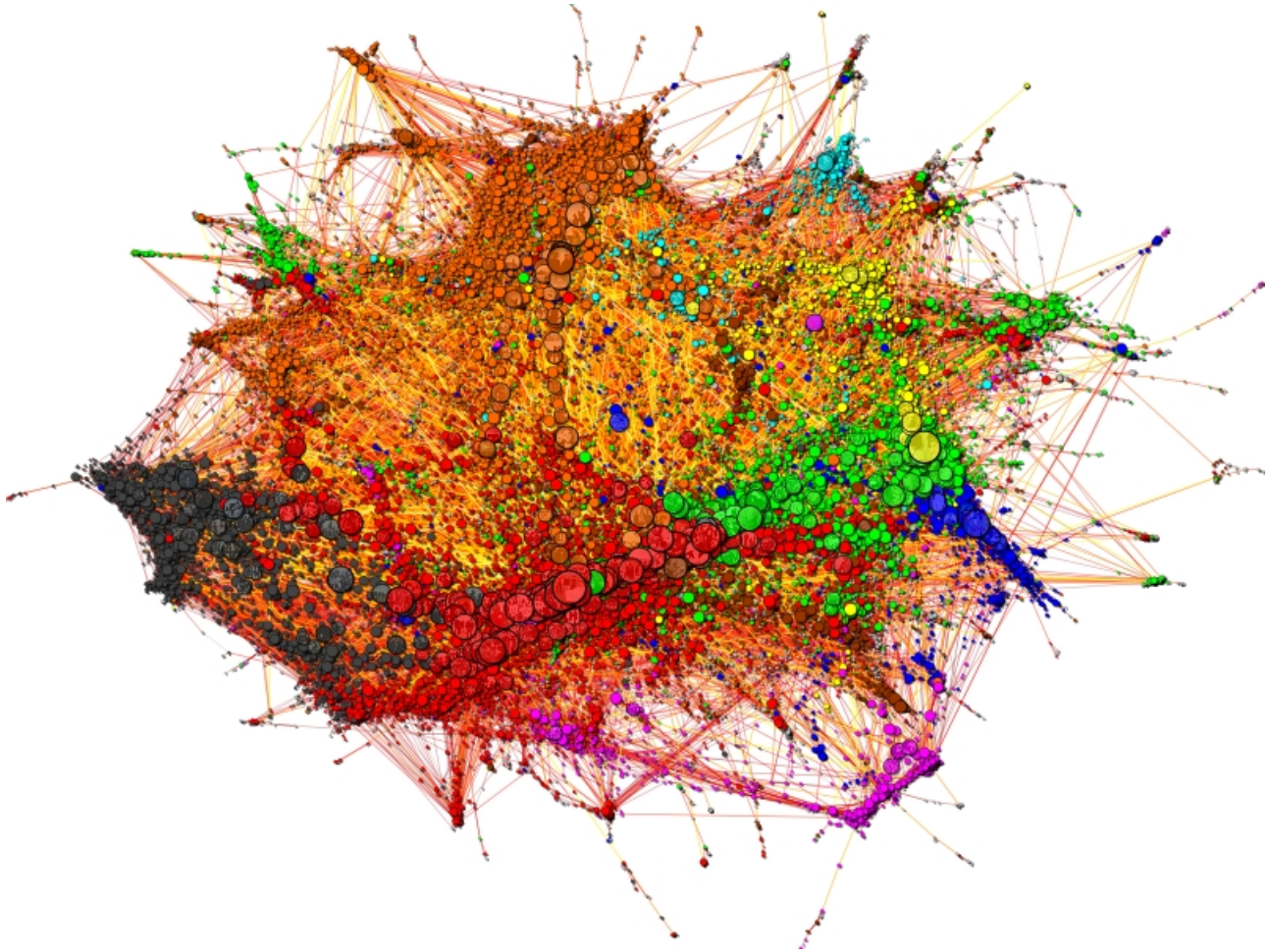
WWW, Random walks, PageRank and Google



source: http://opte.org

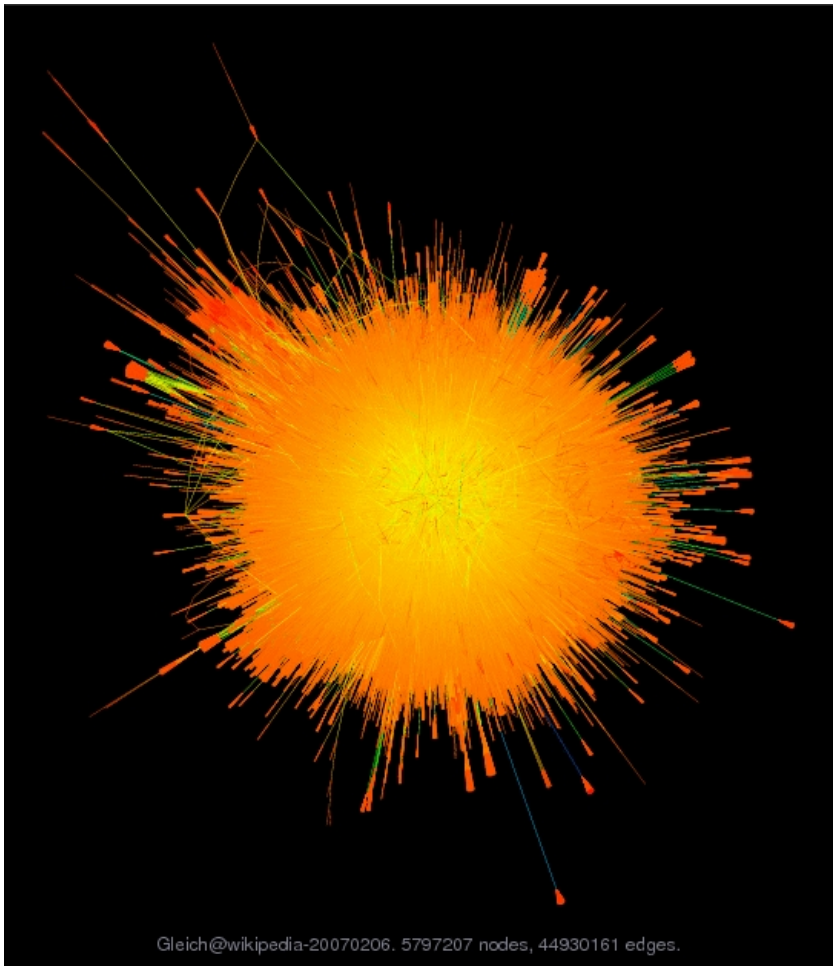
Algorithms on Large Graphs

Linked Data



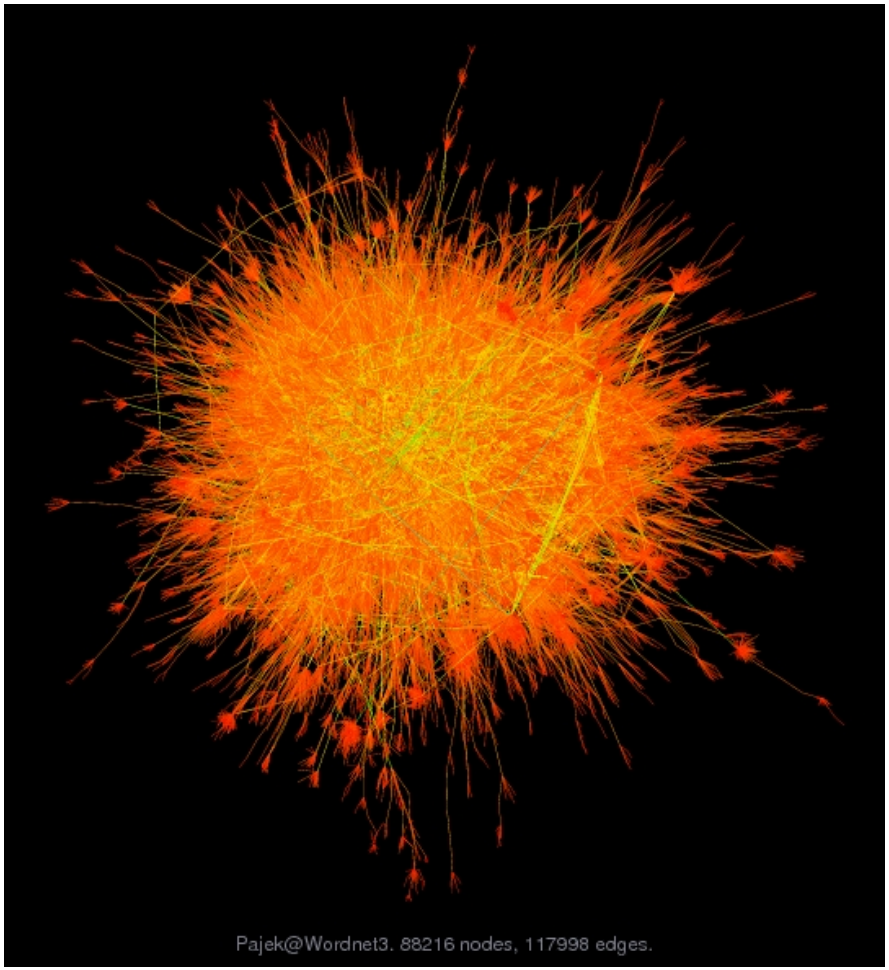
Algorithms on Large Graphs

Wikipedia (DBpedia)



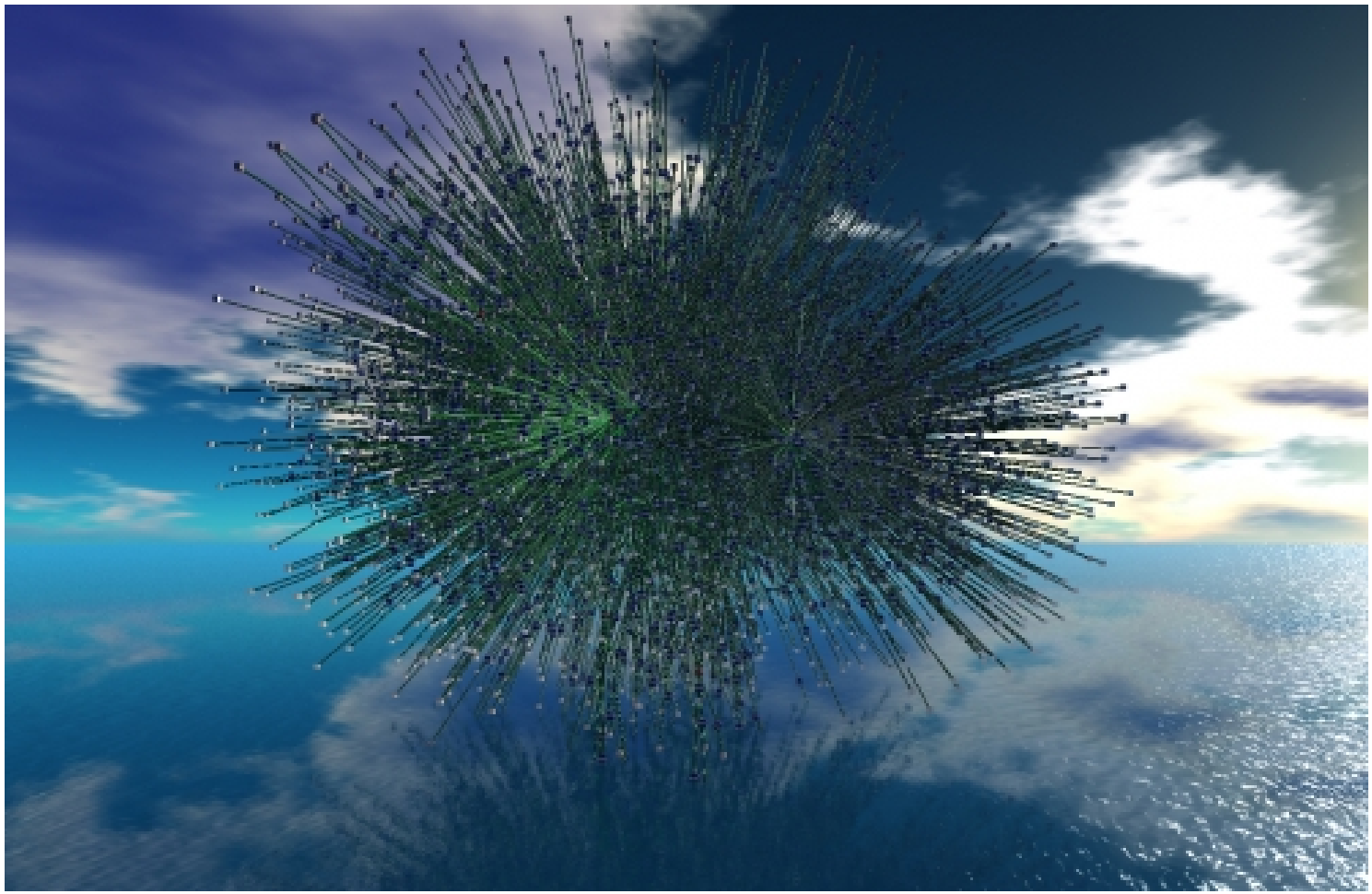
Algorithms on Large Graphs

WordNet

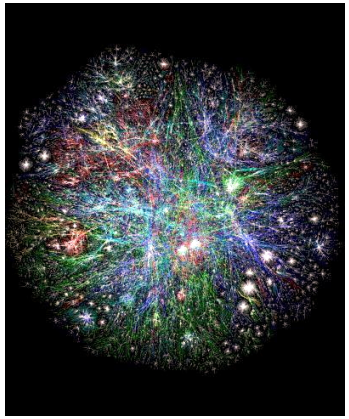


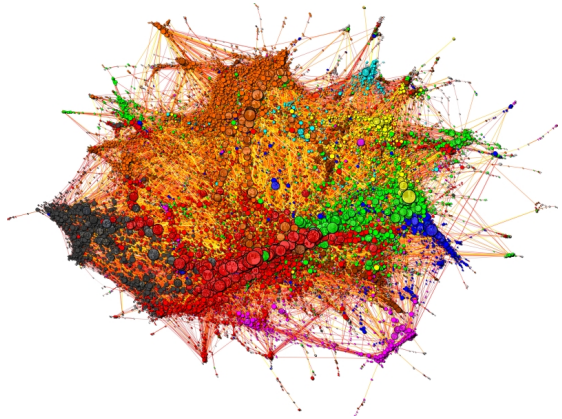
Algorithms on Large Graphs

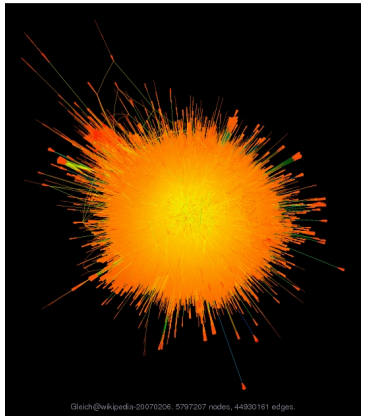
Unified Medical Language System

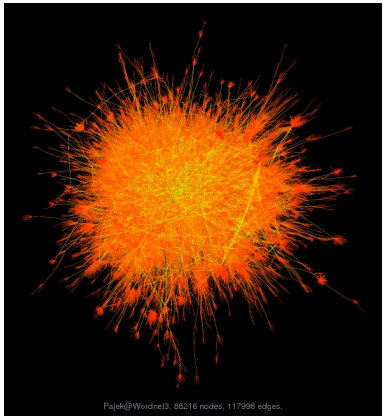


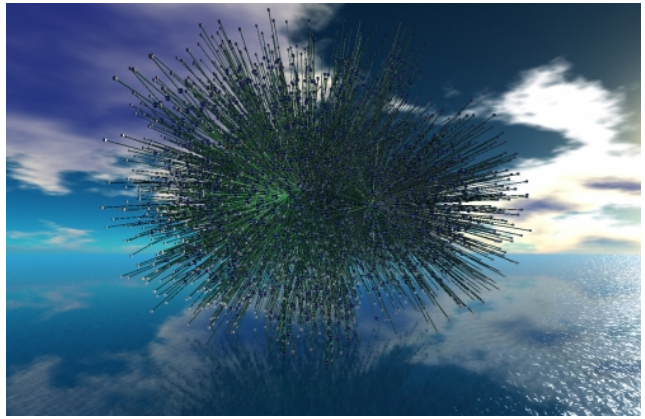
Algorithms on Large Graphs











sources: http://sixdegrees.hu/ http://www2.research.att.com/~yifanhu/
http://www.cise.ufl.edu/research/sparse/matrices/Gleich/ http://www.ebremer.com/

Text Understanding

- Understanding of broad language, what's behind the surface strings

  Barcelona *boss says that Jose Mourinho is*
  'the best ***coach*** *in the world'*

Text Understanding

- Understanding of broad language, what's behind the surface strings

  Barcelona *boss says that Jose Mourinho is*
  'the best ***coach*** *in the world'*

Text Understanding

- Understanding of broad language, what's behind the surface strings

  __Barcelona__ boss says that Jose Mourinho is 'the best __coach__ in the world'

Text Understanding: Knowledge Bases and Graph algorithms

How far can we go with current KBs and graph-based algorithms?

- Ground words in context to KB concepts and instances
  Word Sense Disambiguation
  Named Entity Disambiguation, Entity Linking, Wikification
- Similarity between concepts, instances and words
- Improve ad-hoc information retrieval
- Applied to WordNet(s), UMLS, Wikipedia
- Excellent results
- Open source software and data: `http://ixa2.si.ehu.es/ukb/`

Text Understanding: Knowledge Bases and Graph algorithms

How far can we go with current KBs and graph-based algorithms?

- Ground words in context to KB concepts and instances
  Word Sense Disambiguation
  Named Entity Disambiguation, Entity Linking, Wikification
- Similarity between concepts, instances and words
- Improve ad-hoc information retrieval
- Applied to WordNet(s), UMLS, Wikipedia
- Excellent results
- Open source software and data: `http://ixa2.si.ehu.es/ukb/`

Text Understanding: Knowledge Bases and Graph algorithms

How far can we go with current KBs and graph-based algorithms?

- Ground words in context to KB concepts and instances
  Word Sense Disambiguation
  Named Entity Disambiguation, Entity Linking, Wikification
- Similarity between concepts, instances and words
- Improve ad-hoc information retrieval
- Applied to WordNet(s), UMLS, Wikipedia
- Excellent results
- Open source software and data: `http://ixa2.si.ehu.es/ukb/`

Outline

Outline

Wordnet, Pagerank and Personalized PageRank (*with Aitor Soroa*)

- **WordNet** is the most widely used hierarchically organized lexical database for English (Fellbaum, 1998)
- Broad coverage of nouns, verbs, adjectives, adverbs
- Main unit: *synset* (concept)
  - **coach#1, manager#3, handler#2**
    someone in charge of training an athlete or a team.

- Relations between concepts:
  synonymy (built-in), hyperonymy, antonymy, meronymy, entailment, derivation, gloss
- Closely linked versions in several languages

Wordnet

Representing WordNet as a graph:

- Nodes represent concepts
- Edges represent relations (undirected)
- In addition, directed edges from words to corresponding concepts (senses)

Wordnet

handle#v6
managership#n3
derivation
derivation
trainer#n1
sport#n1
hyperonym
coach#n1
domain
teacher#n1
hyperonym
coach
coach#n2
derivation
tutorial#n1
coach#n5
holonym
hyperonym
holonym
fleet#n2
public_transport#n1
seat#n1

Random Walks: PageRank

- Given a graph, ranks nodes according to their relative structural importance
- If an edge from n_i to n_j exists, a vote from n_i to n_j is produced
  - Strength depends on the rank of n_i
  - The more important n_i is, the more strength its votes will have.
- PageRank is more commonly viewed as the result of a random walk process
  - Rank of n_i represents the probability of a random walk over the graph ending on n_i, at a sufficiently large time.

Random Walks: PageRank

- G: graph with N nodes $n_1, \ldots, n_N$
- d_i: outdegree of node i
- M: $N \times N$ matrix

$$M_{ji} = \begin{cases} \frac{1}{d_i} & \text{an edge from } i \text{ to } j \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

PageRank equation:

$$\mathbf{Pr} = cM\mathbf{Pr} + (1-c)\mathbf{v}$$

- surfer follows edges
- surfer randomly jumps to any node (teleport)

c: damping factor: the way in which these two terms are combined

Random Walks: PageRank

- G: graph with N nodes $n_1, \ldots, n_N$
- d_i: outdegree of node i
- M: $N \times N$ matrix

$$M_{ji} = \begin{cases} \frac{1}{d_i} & \text{an edge from } i \text{ to } j \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

PageRank equation:

$$\mathbf{Pr} = cM\mathbf{Pr} + (1-c)\mathbf{v}$$

- surfer follows edges
- surfer randomly jumps to any node (teleport)

c: damping factor: the way in which these two terms are combined

Random Walks: PageRank

- G: graph with N nodes $n_1, \ldots, n_N$
- d_i: outdegree of node i
- M: $N \times N$ matrix

$$M_{ji} = \begin{cases} \frac{1}{d_i} & \text{an edge from } i \text{ to } j \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

PageRank equation:

$$\mathbf{Pr} = cM\mathbf{Pr} + (1-c)\mathbf{v}$$

- surfer follows edges
- surfer randomly jumps to any node (teleport)

c: damping factor: the way in which these two terms are combined

Random Walks: PageRank

- G: graph with N nodes $n_1, \ldots, n_N$
- d_i: outdegree of node i
- M: $N \times N$ matrix

$$M_{ji} = \begin{cases} \frac{1}{d_i} & \text{an edge from } i \text{ to } j \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

PageRank equation:

$$\mathbf{Pr} = cM\mathbf{Pr} + (1-c)\mathbf{v}$$

- surfer follows edges
- surfer randomly jumps to any node (teleport)

c: damping factor: the way in which these two terms are combined

Random Walks: PageRank

- G: graph with N nodes $n_1, \ldots, n_N$
- d_i: outdegree of node i
- M: $N \times N$ matrix

$$M_{ji} = \begin{cases} \frac{1}{d_i} & \text{an edge from } i \text{ to } j \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

PageRank equation:

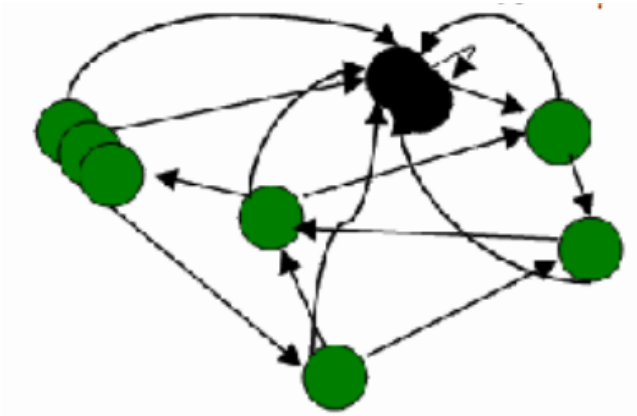
$$\mathbf{Pr} = cM\mathbf{Pr} + (1-c)\mathbf{v}$$

- surfer follows edges
- surfer randomly jumps to any node (teleport)

c: damping factor: the way in which these two terms are combined

Random Walks: Personalized PageRank

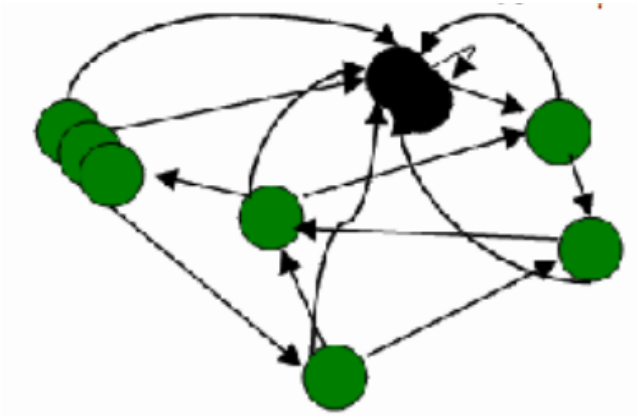
$$\mathbf{Pr} = cM\mathbf{Pr} + (1-c)\mathbf{v}$$



- PageRank: $\mathbf{v}$ is a stochastic normalized vector, with elements $\frac{1}{N}$
  - Equal probabilities to all nodes in case of random jumps
- **Personalized PageRank**, non-uniform $\mathbf{v}$ (Haveliwala 2002)
  - Assign stronger probabilities to certain kinds of nodes
  - Bias PageRank to prefer these nodes
- For ex. if we concentrate all mass on node i
  - All random jumps return to n_i
  - Rank of i will be high
  - High rank of i will make all the nodes in its vicinity also receive a high rank
  - Importance of node i given by the initial $\mathbf{v}$ spreads along the graph

Random Walks: Personalized PageRank

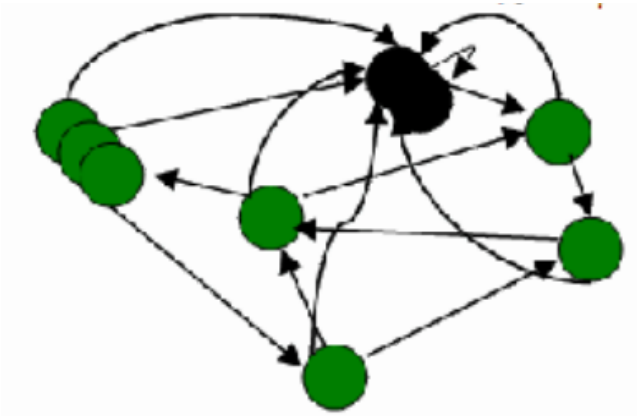
$$\mathbf{Pr} = cM\mathbf{Pr} + (1-c)\mathbf{v}$$



- PageRank: $\mathbf{v}$ is a stochastic normalized vector, with elements $\frac{1}{N}$
  - Equal probabilities to all nodes in case of random jumps
- **Personalized PageRank**, non-uniform $\mathbf{v}$ (Haveliwala 2002)
  - Assign stronger probabilities to certain kinds of nodes
  - Bias PageRank to prefer these nodes
- For ex. if we concentrate all mass on node i
  - All random jumps return to n_i
  - Rank of i will be high
  - High rank of i will make all the nodes in its vicinity also receive a high rank
  - Importance of node i given by the initial $\mathbf{v}$ spreads along the graph

Random Walks: Personalized PageRank

$$\mathbf{Pr} = cM\mathbf{Pr} + (1-c)\mathbf{v}$$



- PageRank: $\mathbf{v}$ is a stochastic normalized vector, with elements $\frac{1}{N}$
  - Equal probabilities to all nodes in case of random jumps
- **Personalized PageRank**, non-uniform $\mathbf{v}$ (Haveliwala 2002)
  - Assign stronger probabilities to certain kinds of nodes
  - Bias PageRank to prefer these nodes
- For ex. if we concentrate all mass on node i
  - All random jumps return to n_i
  - Rank of i will be high
  - High rank of i will make all the nodes in its vicinity also receive a high rank
  - Importance of node i given by the initial $\mathbf{v}$ spreads along the graph

Outline

Word Sense Disambiguation (WSD)

- Goal: determine senses of the open-class words in a text.
  - **"Nadal is sharing a house with his uncle and coach, Toni."**
  - **"Our fleet comprises coaches from 35 to 58 seats."**



- Knowledge Base (e.g. WordNet):
  - **coach#1** someone in charge of training an athlete or a team.
  - coach#2 a person who gives private instruction (as in singing, acting, etc.).
  - ...
  - **coach#5** a vehicle carrying many passengers; used for public transport.

Word Sense Disambiguation (WSD)

- Goal: determine senses of the open-class words in a text.
  - **"Nadal is sharing a house with his uncle and coach, Toni."**
  - **"Our fleet comprises coaches from 35 to 58 seats."**



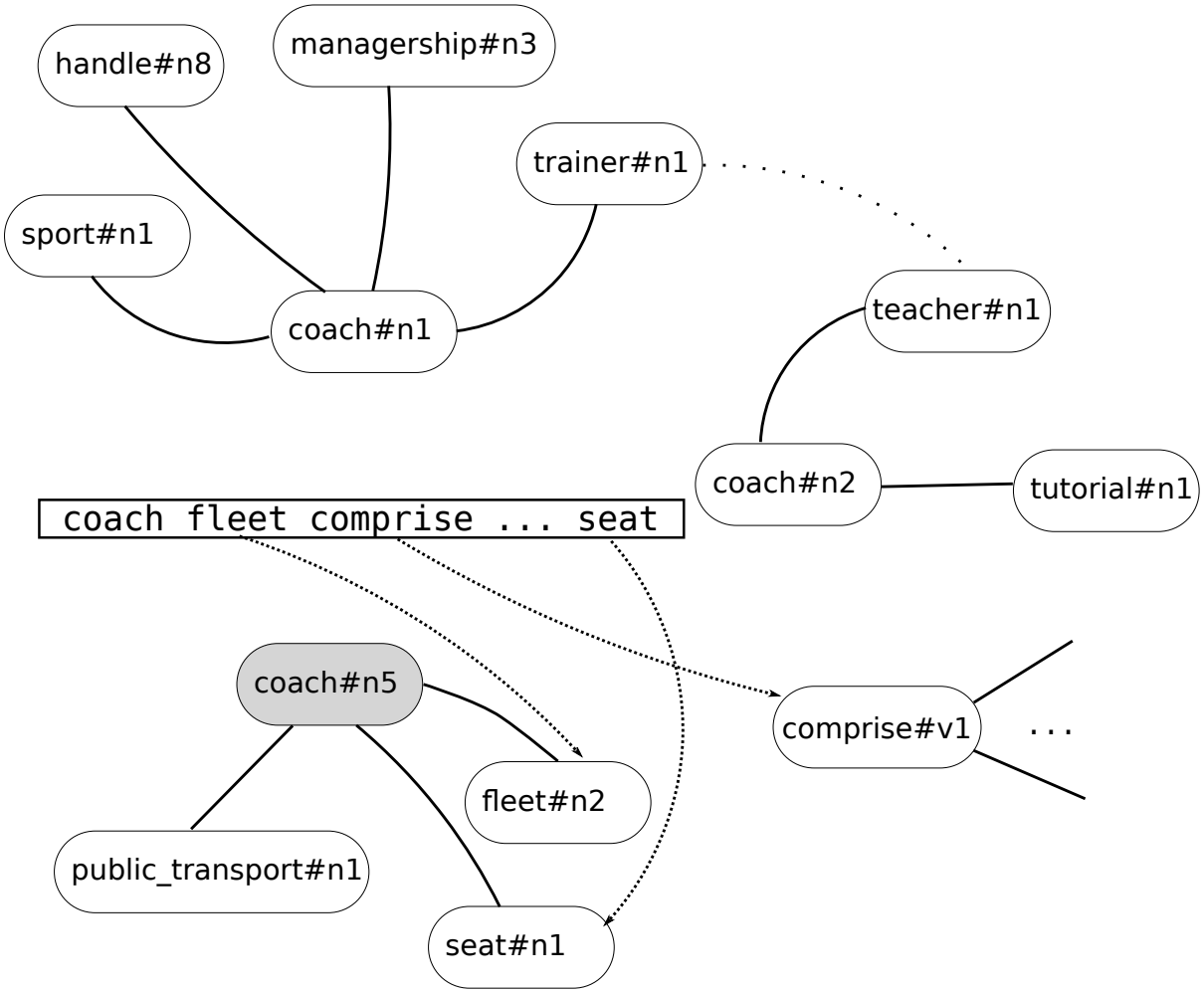
- Knowledge Base (e.g. WordNet):
  - **coach#1** someone in charge of training an athlete or a team.
  - coach#2 a person who gives private instruction (as in singing, acting, etc.).
  - ...
  - **coach#5** a vehicle carrying many passengers; used for public transport.

Using Personalized PageRank for WSD
(with Aitor Soroa, Oier Lopez de Lacalle)

For each word W_i, $i = 1 \ldots m$ in the context

- Initialize $\mathbf{v}$ with uniform probabilities over words W_i
  Context words act as source nodes
  injecting probability mass into the concept graph
- Run Personalized PageRank
- Choose highest ranking sense for target word

Using Personalized PageRank (PPR)



Results according to relations

relation	#	F1	ablation
Antonimy	8K	19.1	**59.9**
Meronymy (part-of)	21K	23.4	59.6
Derivation	32K	35.4	59.6
Taxonomy	89K	37.4	**59.9**
Disambiguated gloss	550K	**59.9**	47.1
All relations		59.7	

Results and comparison to related work

System	S2AW	S3AW	S07CG	(N)
(Agirre et al. 2008)		56.8		
(Tsatsaronis 2010)	58.8	57.4		
(Ponzetto and Navigli, 2010)				(79.4)
(Moro and Navigli, 2014)				(**84.6**)
PPR$_{w2w}$	**59.7**	**57.9**	**80.1**	(83.6)
MFS	60.1	62.3	78.9	(77.4)
(Ponzetto and Navigli, 2010)			81.7	(85.5)
(Zhong et al. 2010)	68.2	67.6	82.6	(82.3)

Outline

UMLS and biomedical text
(with Aitor Soroa and Mark Stevenson)

- Ambiguity believed not to occur on specific domains
  - **On the Use of Cold Water as a Powerful Remedial Agent in Chronic Disease.**
  - **Intranasal ipratropium bromide for the common cold.**
- 11.7% of the phrases in abstracts added to MEDLINE in 1998 were ambiguous (Weeber et al. 2011)

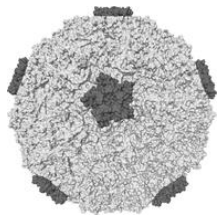
- Unified Medical Language System (UMLS) Metathesaurus
- Concept Unique Identifiers (CUIs)
  - C0234192: Cold (Cold Sensation) [Physiologic Function]
  - C0009264: Cold (cold temperature) [Natural Phenomenon or Process]
  - C0009443: Cold (Common Cold) [Disease or Syndrome]

UMLS and biomedical text
(with Aitor Soroa and Mark Stevenson)

- Ambiguity believed not to occur on specific domains
  - **On the Use of Cold Water as a Powerful Remedial Agent in Chronic Disease.**
  - **Intranasal ipratropium bromide for the common cold.**
- 11.7% of the phrases in abstracts added to MEDLINE in 1998 were ambiguous (Weeber et al. 2011)

- Unified Medical Language System (UMLS) Metathesaurus
- Concept Unique Identifiers (CUIs)
  - C0234192: Cold (Cold Sensation) [Physiologic Function]
  - C0009264: Cold (cold temperature) [Natural Phenomenon or Process]
  - C0009443: Cold (Common Cold) [Disease or Syndrome]





WSD and biomedical text

- Thesaurus in Metathesaurus: ($\sim$1M CUIs)
  Alcohol and other drugs, Medical Subject Headings, Crisp Thesaurus, SNOMED Clinical Terms, etc.
- Relations in the Metathesaurus between CUIs ($\sim$5M):
  parent, can be qualified by, related possibly sinonymous, related other
- We applied Personalized PageRank.
- Evaluated on NLM-WSD, 50 ambiguous terms (100 instances each)

KB	#CUIs	#relations	Acc.	Terms
AOD	15,901	58,998	51.5	4
MSH	278,297	1,098,547	44.7	9
CSP	16,703	73,200	60.2	3
SNOMEDCT	304,443	1,237,571	62.5	29
all above	572,105	2,433,324	64.4	48
all relations	-	5,352,190	70.4	50
(Jimeno and Aronson, 2011)	-	-	68.4	50

WSD and biomedical text

- Thesaurus in Metathesaurus: ($\sim$1M CUIs)
  Alcohol and other drugs, Medical Subject Headings, Crisp Thesaurus, SNOMED Clinical Terms, etc.
- Relations in the Metathesaurus between CUIs ($\sim$5M):
  parent, can be qualified by, related possibly sinonymous, related other
- We applied Personalized PageRank.
- Evaluated on NLM-WSD, 50 ambiguous terms (100 instances each)

KB	#CUIs	#relations	Acc.	Terms
AOD	15,901	58,998	51.5	4
MSH	278,297	1,098,547	44.7	9
CSP	16,703	73,200	60.2	3
SNOMEDCT	304,443	1,237,571	62.5	29
all above	572,105	2,433,324	64.4	48
all relations	-	5,352,190	70.4	50
(Jimeno and Aronson, 2011)	-	-	68.4	50

Outline

Named Entity Disambiguation
(with Aitor Soroa, Ander Barrena)

- Goal: given a Named Entity mention, determince instance in KB (aka Entity Linking, Wikification)
- Represent Wikipedia (DBpedia) as graph:
  ∼5M articles
  ∼90M hyperlinks

Named Entity Disambiguation
(with Aitor Soroa, Ander Barrena)

- Goal: given a Named Entity mention, determince instance in KB (aka Entity Linking, Wikification)
- Represent Wikipedia (DBpedia) as graph:
  ∼5M articles
  ∼90M hyperlinks

Highveld Lions cricket team

From Wikipedia, the free encyclopedia

The **Highveld Lions** is the name used by the combined Gauteng and North West first class cricket teams in South Africa.

The home venues are the New Wanderers Stadium in Johannesburg and Senwes Park in Potchefstroom. The combined team plays in the Sunfoil Series first class cricket competition as well as in the Momentum 1 Day Cup and Ram Slam T20 Challenge limited over competitions.

Honours [edit]

- **Sunfoil Series (0) -** ; **shared () -**
- **Momentum 1 Day Cup (0) -** ; **shared (1) -** 2012-13 shared with Nashua Cape Cobras
- **Ram Slam T20 Challenge (2) -** 2006–07, 2012-13
- **Champions League Twenty20 (0) -** ; **Runners up (1) -** 2011-2012

Squad [edit]

- No. denotes the player's squad number, as worn on the back of their shirt.

Named Entity Disambiguation

- **Alan Kourie, CEO of the Lions franchise, had discussions with Fletcher in Cape Town.**

  - Brisbane Lions, an Australian rules football team
  - BC Lions, a Canadian football team
  - Chandigarh Lions, a team from the Indian Cricket League
  - Detroit Lions, an American football team
  - Finland men's national ice hockey team or the Lions
  - Highveld Lions cricket team, a South African cricket team
  - Huonville Football Club, Australian rules football club in Tasmania
  - Leicester Lions, a British speedway team
  - New Yorker Lions, an American football team from Braunschweig, Germany
  - LHC Les Lions, an ice hockey team in Lyon, France

Named Entity Disambiguation

Main steps:

- Named Entity Recognition in text (NER)
- Candidate generation: use titles, redirects, text in anchors
- Disambiguation: Personalized PageRank
- NIL detection and clustering: no corresponding instance in the KB

- Evaluation: accuracy (we don't do NILs or NIL clustering)

  TAC-KBP 2009 **78.8** vs. 76.5 (Best system)
  TAC-KBP 2010 **83.6** vs. 80.6 (Best system)
  TAC-KBP 2013 **81.7** vs. 77.7 (Best system)

Outline

Random walks for similarity
(with Aitor Soroa, Montse Cuadros, German Rigau)

Given two words estimate how similar they are.

gem jewel

Given a pair of words (w_1, w_2): (Hughes and Ramage, 2007)

- Initialize teleport probability mass on w_1
- Run Personalized Pagerank, obtaining $\vec{w}_1$
- Initialize w_2 and obtain $\vec{w}_2$
- Measure similarity between $\vec{w}_1$ and $\vec{w}_2$ (e.g. cosine)

Similarity datasets

RG dataset			WordSim353 dataset		
cord	smile	0.02	king	cabbage	0.23
rooster	voyage	0.04	professor	cucumber	0.31
	...			...	
glass	jewel	1.78	investigation	effort	4.59
magician	oracle	1.82	movie	star	7.38
	...			...	
cemetery	graveyard	3.88	journey	voyage	9.29
automobile	car	3.92	midday	noon	9.29
midday	noon	3.94	tiger	tiger	10.00

80 pairs, 51 subjects
Similarity

353 pairs, 16 subjects
Similarity and relatedness

Results

Method	Source	WS353	RG
(Hughes and Ramage, 2007)	WordNet	0.55	-
(Finkelstein et al. 2007)	Corpora (LSA)	0.56	-
(Agirre et al. 2009)	Corpora	0.66	0.88
PPR	WordNet	0.69	0.87
(Huang et al. 2012)	Corpora (NN)	0.71	-
(Baroni et al., 2014)	Corpora (NN)	0.71	0.84
PPR	Wikipedia	0.73	0.86
(Gabrilovich and Markovitch, 2007)	Wikipedia	0.75	0.82
(Reisinger and Mooney, 2010)	Corpora	0.77	-
(Pihlevar et al. 2013)	BabelNet	-	0.87
PPR	Wiki + WNet	0.79	0.91
(Radinsky et al. 2011)	Corpora (Time)	0.80	-

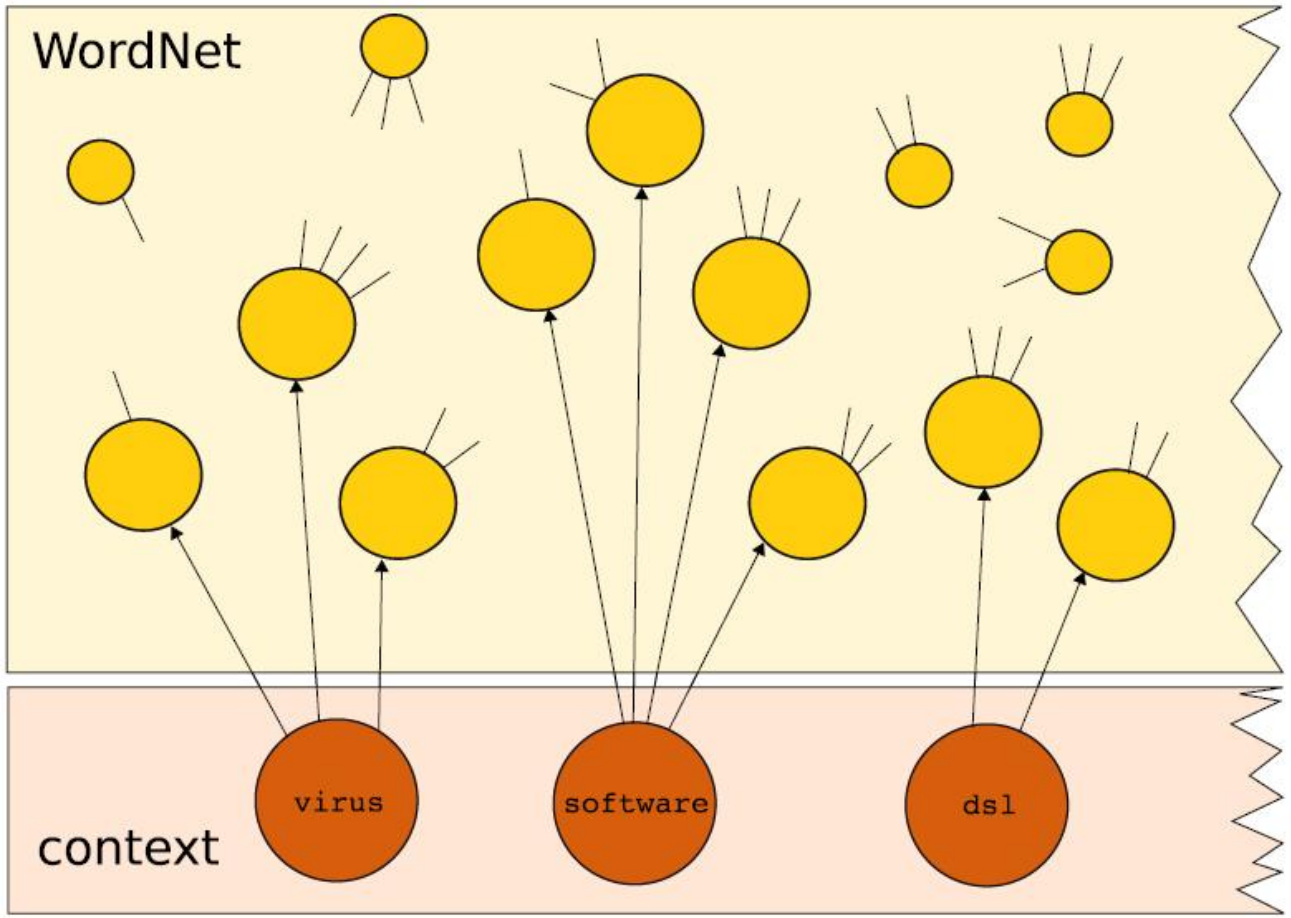
Outline

Similarity and Information Retrieval *(with Arantxa Otegi and Xabier Arregi)*

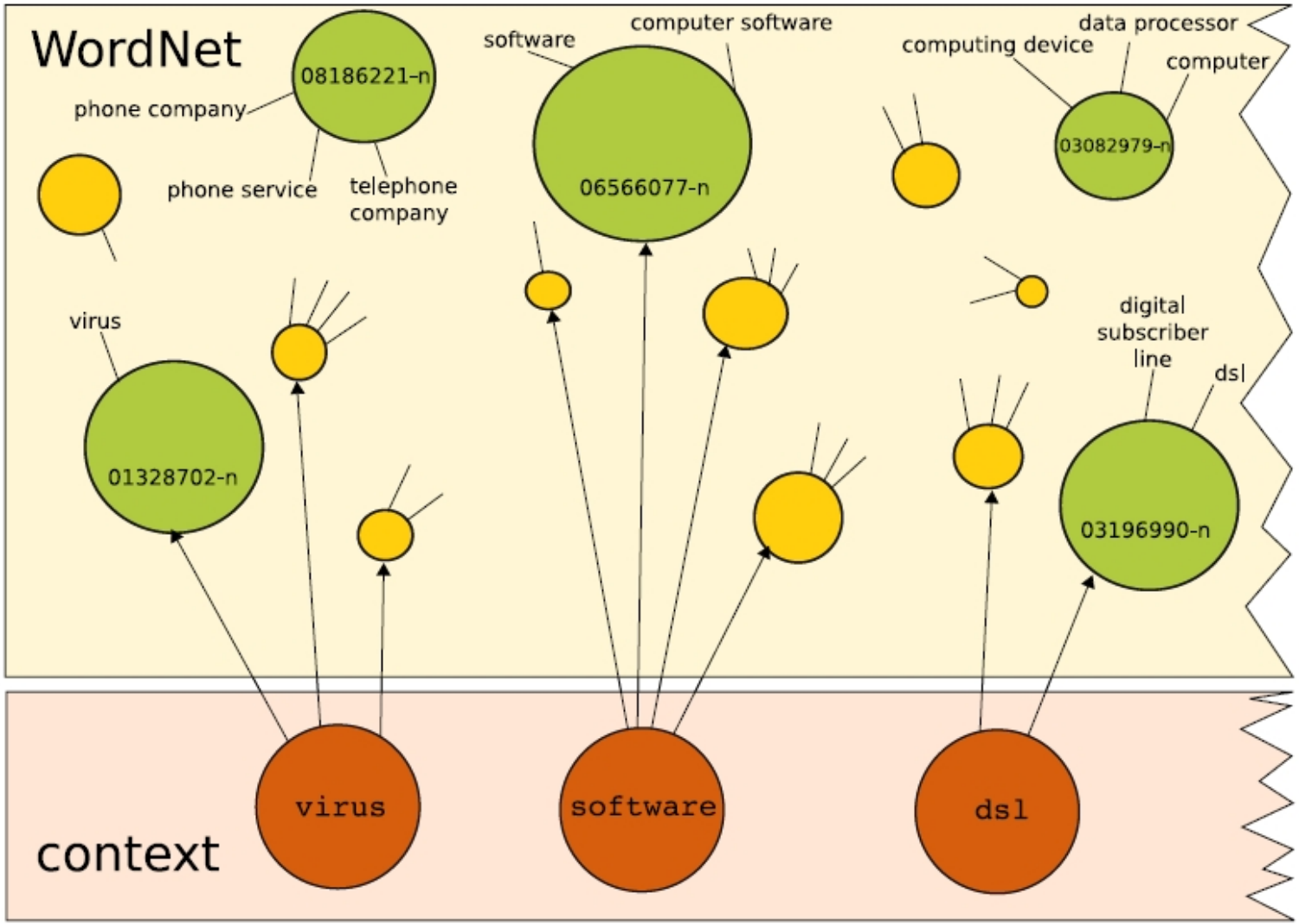
- Document expansion (aka clustering and smoothing) has been shown to be successful in ad-hoc IR
- **Use WordNet and similarity to expand documents**
- Example:
  - I can't **install DSL** because of the **antivirus program**, any hints?
  - You should turn off **virus** and anti-spy software. And thats done within each of the **softwares** themselves. Then turn them back on later after **setting up** any **DSL softwares**.
- Method:
  - Initialize random walk with document words
  - Retrieve top k synsets
  - Introduce words on those k synsets in a secondary index
  - When retrieving, use both primary and secondary indexes

Example

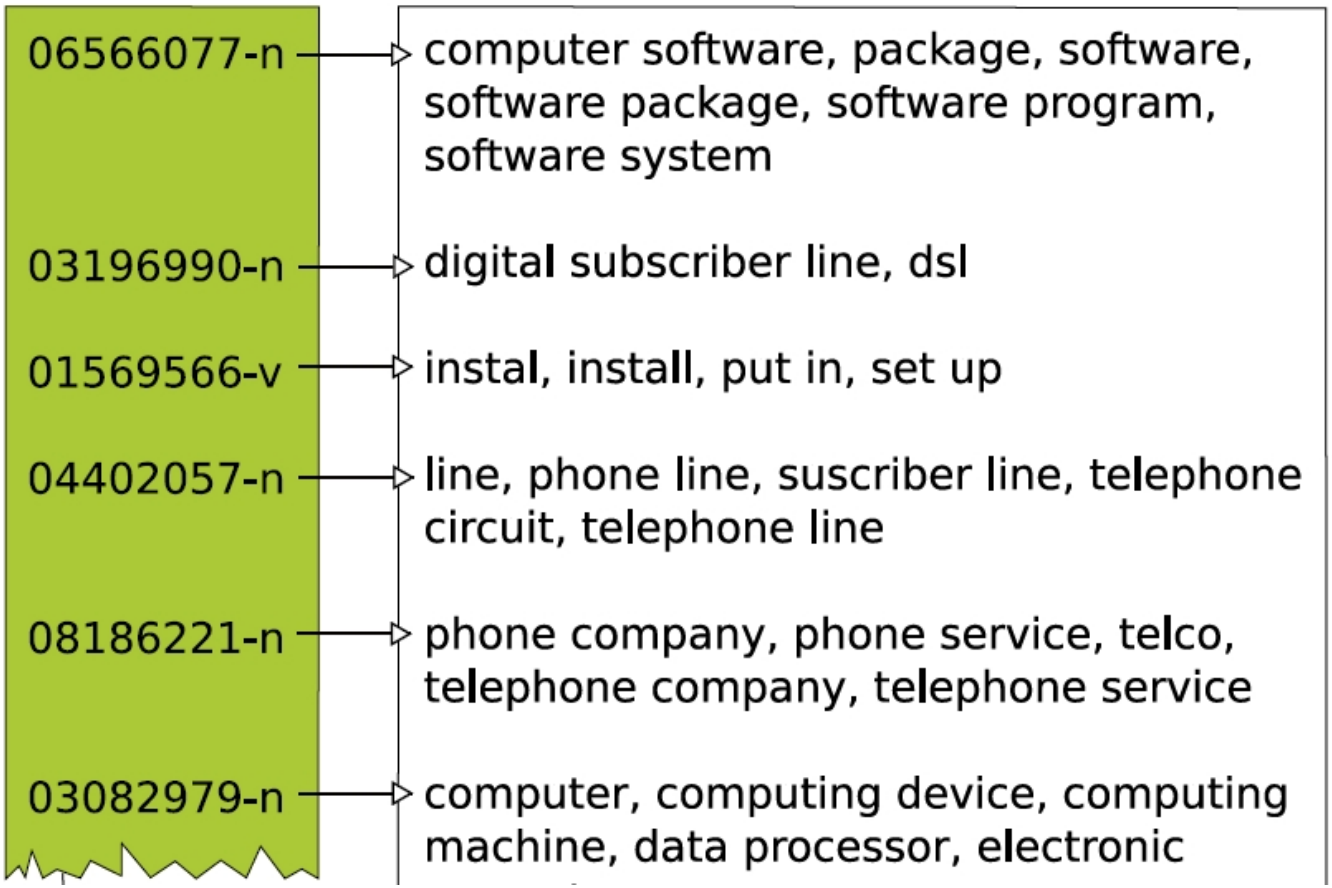
You should turn off virus and anti-spy software. And thats done within each of the softwares themselves. Then turn them back on later after setting up any DSL softwares.



Example


WordNet
08186221-n
phone company
phone service
telephone company
software
computer software
06566077-n
computing device
data processor
computer
03082979-n
virus
01328702-n
digital subscriber line
dsl
03196990-n
context
virus
software
dsl

Example



Query: I can't install DSL because of the antivirus program, any hints?

Experiments

- BM25 ranking function
- Combine 2 indexes: original words and expansion terms
- Parameters: k_1, b (BM25) λ (indices) k (concepts in expansion)
- Three collections:
  - Robust at CLEF 2009
  - Yahoo Answer!
  - RespubliQA (IR for QA)
- Summary of results:
  - Default parameters: 1.43% - 4.90% improvement in all 3 datasets
  - Optimized parameters: 0.98% - 2.20% improvement in 2 datasets
- Robustness on suboptimal parametrizations: 5.77% - 19.77% improvement in 4 out of 6
- Particularly on short documents

Experiments

- BM25 ranking function
- Combine 2 indexes: original words and expansion terms
- Parameters: k_1, b (BM25) λ (indices) k (concepts in expansion)
- Three collections:
  - Robust at CLEF 2009
  - Yahoo Answer!
  - RespubliQA (IR for QA)
- Summary of results:
  - Default parameters: 1.43% - 4.90% improvement in all 3 datasets
  - Optimized parameters: 0.98% - 2.20% improvement in 2 datasets
- Robustness on suboptimal parametrizations: 5.77% - 19.77% improvement in 4 out of 6
- Particularly on short documents

Experiments

- BM25 ranking function
- Combine 2 indexes: original words and expansion terms
- Parameters: k_1, b (BM25) λ (indices) k (concepts in expansion)
- Three collections:
  - Robust at CLEF 2009
  - Yahoo Answer!
  - RespubliQA (IR for QA)
- Summary of results:
  - Default parameters: 1.43% - 4.90% improvement in all 3 datasets
  - Optimized parameters: 0.98% - 2.20% improvement in 2 datasets
- Robustness on suboptimal parametrizations: 5.77% - 19.77% improvement in 4 out of 6
- Particularly on short documents

Outline

Conclusions

- Knowledge-based method for WSD, NED and similarity
- State-of-the-art results in similarity and NED
- Best graph-based results in all tasks
  - Specific experiments: link overlap (NGD), subgraphs
  - Exploits whole structure of very large KB, simple, few knobs
  - Key for performance: selection of relations in the graph
- Publicly available at `http://ixa2.si.ehu.es/ukb`
  - Both programs and data (WordNet, UMLS, Wikipedia to come soon)
  - Including program to construct graphs from KBs
  - GPL license, open source, free

Conclusions

- Knowledge-based method for WSD, NED and similarity
- State-of-the-art results in similarity and NED
- Best graph-based results in all tasks
  - Specific experiments: link overlap (NGD), subgraphs
  - Exploits whole structure of very large KB, simple, few knobs
  - Key for performance: selection of relations in the graph
- Publicly available at `http://ixa2.si.ehu.es/ukb`
  - Both programs and data (WordNet, UMLS, Wikipedia to come soon)
  - Including program to construct graphs from KBs
  - GPL license, open source, free

Future

- Beyond terms (Semeval 2015 task2 Sematic Text Similarity)
- Explore other sources of links: co-occurrence graphs
- Multi-linguality and cross-linguality
- Beyond bag of words: incorporate syntactic structure
- Include supervision

Future

- Beyond terms (Semeval 2015 task2 Sematic Text Similarity)
- Explore other sources of links: co-occurrence graphs
- Multi-linguality and cross-linguality
- Beyond bag of words: incorporate syntactic structure
- Include supervision

Word Sense Disambiguation and Named-Entity Disambiguation using graph-based algorithms

Eneko Agirre
ixa2.si.ehu.es/eneko

IXA NLP Group
University of the Basque Country

WSAP in Copenhaguen, 2014

References I


Agirre, E., Arregi, X. and Otegi, A. (2010).
Document Expansion Based on WordNet for Robust IR.
In Proceedings of the 23rd International Conference on Computational Linguistics (Coling) pp. 9–17,.


Agirre, E., de Lacalle, O. L. and Soroa, A. (2009).
Knowledge-Based WSD on Specific Domains: Performing better than Generic Supervised WSD.
In Proceedings of IJCAI, Pasadena, USA.


Agirre, E., Lacalle, d. O. L. and Soroa, A. (2014).
Random Walks for Knowledge-Based Word Sense Disambiguation.
Computational Linguistics *40*.


Agirre, E. and Soroa, A. (2009).
Personalizing PageRank for Word Sense Disambiguation.
In Proceedings of EACL-09, Athens, Greece.

References II


Agirre, E., Soroa, A., Alfonseca, E., Hall, K., Kravalova, J. and Pasca, M. (2009).
A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches.
In Proceedings of annual meeting of the North American Chapter of the Association of Computational Linguistics (NAAC), Boulder, USA.


Agirre, E., Soroa, A. and Stevenson, M. (2010).
Graph-based Word Sense Disambiguation of Biomedical Documents.
Bioinformatics *26*, 2889–2896.


Eneko Agirre, Montse Cuadros, G. R. and Soroa, A. (2010).
Exploring Knowledge Bases for Similarity.
In Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), (Calzolari, N., ed.), pp. 373–377, European Language Resources Association (ELRA), Valletta, Malta.

References III


Otegi, A., Arregi, X. and Agirre, E. (2011).
Query Expansion for IR using Knowledge-Based Relatedness.
In Proceedings of the International Joint Conference on Natural Language Processing.


Otegi, A., Arregi, X., Ansa, O. and Agirre, E. (2014).
Using knowledge-based relatedness for information retrieval.
Knowledge and Information Systems *In press*, 1–30.


Stevenson, M., Agirre, E. and Soroa, A. (2011).
Exploiting Domain Information for Word Sense Disambiguation of Medical Documents.
Journal of the American Medical Informatics Association ,, 1–6.


Yeh, E., Ramage, D., Manning, C., Agirre, E. and Soroa, A. (2009).
WikiWalk: Random walks on Wikipedia for Semantic Relatedness.
In ACL workshop "TextGraphs-4: Graph-based Methods for Natural Language Processing.