



Science Information Applications

Ulrich Schäfer

DFKI Language Technology Lab

Paper/bibliographic search

Numbers from one year ago

<http://academic.research.microsoft.com/>

- for many research areas; graphical browsers (Windows only...)
- "explore ~~37,472,555~~ 48,774,763 publications and ~~19,327,188~~ 21,932,046 authors": people, organizations, citation network, CfP calendar, research trends

<http://scholar.google.com>

- textual paper content search, author search

DBLP (<http://www.informatik.uni-trier.de/~ley/db/>): ~~1.8~~ 2.1 million entries, mainly computer science and related field; only bibl. metadata with links to open or closed access papers

Bielefeld Academic Search (<http://www.base-search.net/>): ~~32,663,572~~ 40,879,686 papers from ~~2,085~~ 2,428 sources: metadata with links to open or closed access papers

CiteceerX (<http://citeseerx.ist.psu.edu/index>): digital library, search engine and citation statistics for computer and information science papers, also a software infrastructure

Open Access Portals:

Scientific Commons (<http://en.scientificcommons.org>): ~~38,245,864~~ 38,354,162 documents from 1269 sources

ArXiv (<http://lanl.arxiv.org>): Open access to ~~728,365~~ 812,535 e-prints in Physics, Mathematics, Computer Science, Quantitative Biology, Quantitative Finance and Statistics

Publisher's Portals

Springer

Elsevier

Thomson-Reuters Web of Science

Universities, e.g. SciDok (SULB)

Thousands of other indexes and portals...

Citation Analysis

Pioneer: Eugene Garfield (1955), see references
founder of ISI (Information Sciences Institute, USC,
Marina del Rey, CA)

Related Research fields:

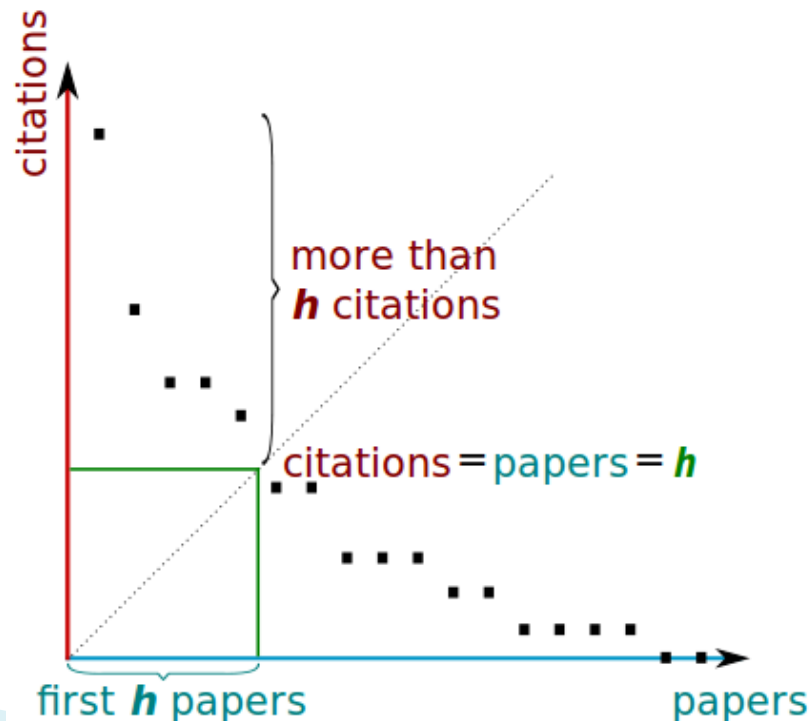
- **Scientometrics**
- **Bibliometrics**
- **Library Science**
- **Information Science**

Citation Analysis

Citation Index

h-index (or Hirsch index, after Jorge E. Hirsch)

A scientist has index h if h of his/her N papers have at least h citations each, and the other $(N - h)$ papers have no more than h citations each.



The key to (almost) everything in citation analysis and search: String distance metrics...

1. **Levenshtein distance**: number of edits from s_1 to s_2

2. **Jaro distance**:
$$d_j = \frac{1}{3} \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m - t}{m} \right)$$

(i.e., normalized metric: 0=no, 1=full match; m =# of matches, $t=1/2$ # of transpositions)

3. **Jaro-Winkler**: Jaro with weight for prefix changes

There are **many more**...

→ Exercise python + external Levenshtein module (src from <http://pypi.python.org/pypi/python-Levenshtein/>)

Exercise: python-levenshtein library

Ubuntu/Debian:

```
sudo apt-get install python-levenshtein
```

```
python
```

```
from Levenshtein import distance, hamming, jaro, jaro_winkler
```

```
>>> distance("scientometrics", "bibliometrics")
```

```
5
```

```
>>> hamming("bibliometrics", "scientometric")
```

```
13
```

```
>>> jaro("scientometrics", "bibliometrics")
```

```
0.6672771672771672
```

```
>>> jaro_winkler("scientometrics", "bibliometrics")
```

```
0.6672771672771672
```

```
>>> jaro("scientometrics", "scientomanics")
```

```
0.8772893772893773
```

```
>>> jaro_winkler("scientometrics", "scientomanics")
```

```
0.9754578754578754
```

Java variant (different library): Simmetrics

<http://sourceforge.net/projects/simmetrics/>

<http://web.archive.org/web/20081224234350/>

<http://www.dcs.shef.ac.uk/~sam/stringmetrics.html>

The case of Medical Science

Elaborated Ontologies:

- **MeSH** (Medical Subject Headlines, <http://www.nlm.nih.gov/mesh/>)
- **UMLS** (Unified Medical Language System, <http://www.nlm.nih.gov/research/umls/>)

Huge text databases: PubMed/Medline (publication metadata and abstracts only...):

<http://www.ncbi.nlm.nih.gov/pubmed/>

There are many more...

Related research field: Literature analysis/text mining as subfield of Bioinformatics

Computational Linguistics

LT World (<http://www.lt-world.org>)

- Underlying ontology and data: people, organisations, projects, conferences, news, links, resources, tools, etc.
- Largely hand-crafted content, limited terminology resources, no publication metadata nor publication content

ACL Anthology (<http://www.aclweb.org/anthology>)

- Open access digital library of more than 25,000 CL papers from 1967 until today, including the complete CL Journal.
- Content search via Google custom search and DFKI's **Searchbench**
- Incomplete publication metadata (will be improved)
- Citation Network: <http://clair.si.umich.edu/clair/anthology/>

Using more NLP for Science Information Application

Motivation: go beyond citation graphs and indexes, text retrieval/fulltext and metadata search

Users want to see original, full content of papers, not just bibliographic metadata, abstracts and references

Interesting areas for NLP:

- improve search → semantic search ("find what I mean")
 - search for complex propositions, synonyms, in context
 - preprocess textual content: parsing, coreferences, etc.
- automatic terminology, taxonomy & ontology extraction from text
- qualitative citation analysis
- automatic summarization
- question answering, learning by reading, expert systems, ...

Parsing Science with NLP (more or less...)

MEDIE is a semantic search engine to retrieve biomedical correlations from MEDLINE articles (Sætre et al., 2008)

SciBorg: UK-based research project on parsing and named entity recognition of chemistry papers from a publisher

Wolfram Alpha: Question answering, specialized tools and database: <http://www.wolframalpha.com/>

NLP pipeline: Text extraction

Preprocessing 1: Text extraction from digital and scanned documents

commercial (O)CR:

- Omnipage, Abbyy

Open source (O)CR:

- Tesseract (<http://code.google.com/p/tesseract-ocr/>)

Open source layout recognition on top of Tesseract:

- Ocropus (<http://code.google.com/p/ocropus/>)

Alternatives for native (not scanned) PDF:

- Apache PDFbox: <http://pdfbox.apache.org/>
- Poppler/Xpdf: <http://poppler.freedesktop.org/>

Text and metadata extraction from office file formats etc.:

- Apache POI (<http://projects.apache.org/projects/poi.html>),
- Aperture (<http://aperture.sourceforge.net/>)

NLP Pipeline

Preprocessing 2:

- text filtering (remove non-text character sequences)
- de-hyphenation
- XML Markup (optional, e.g. **TEI P5**, **Docbook**,...), containing information on section headings, footnotes, tables, character styles such as *Italics*, page numbers, figures and tables, captions, ... Potentially useful for detecting argumentative zones, citation classification, emphasized tokens marked for parsing, etc.
- Example: XML file: paper.xml

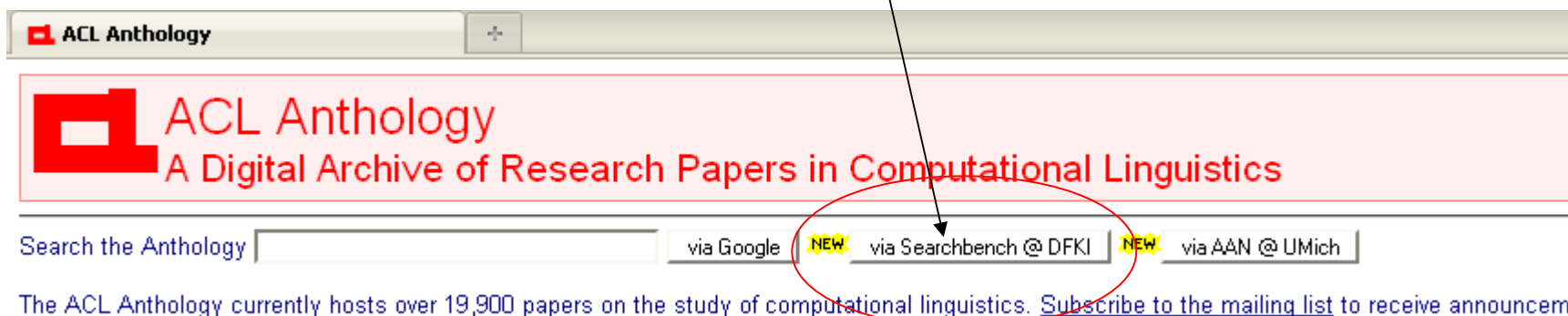
NLP Pipeline

Preprocessing 3:

- Sentence boundary recognition
- Tokenization
- PoS tagging (for unknown word guessing, term extraction, ...)
- Named entity recognition
- Parsing
- Semantics extraction
- Index preparation
- (Structured) indexing with Apache **Lucene/Solr**

ACL Anthology Searchbench

- <http://aclasb.dfki.de>
- Released at ACL-2011
- Combines semantic, full-text and bibliographic search in 25,000 papers of the ACL Anthology from the past 46 years, incl. CL journal
- ACL Anthology start page links to it!



ACL Anthology
A Digital Archive of Research Papers in Computational Linguistics

Search the Anthology

via Google via Searchbench @ DFKI via AAN @ UMich

The ACL Anthology currently hosts over 19,900 papers on the study of computational linguistics. [Subscribe to the mailing list](#) to receive announcements.

ACL Anthology Searchbench - Startpage



Add and remove filters for the papers you are interested in (or [remove all currently set filters](#) ☒).

- Statements ☒
- Plain Text ☒
- Extracted Topics ☒
- Publication ☒
- Authors ☒
- Year ☒
- Title ☒
- Affiliations ☒
- Affiliation Sites ☒

Welcome to the ACL Anthology Searchbench

The ACL Anthology Searchbench provides semantic, full text and bibliographic search in papers of the [ACL Anthology](#) (or actually a nearly complete subset of all existing anthology papers without ROCLING, which currently corresponds to more than 22,000 papers).

Getting Started

By clicking the buttons in the sidebar on the left you can add filters for the papers you are interested in. This view will then show you a list of all papers matching the set filters. Selecting a document from this list will show it in the view below.

Quick Start Examples

- ▶ Maybe you'd like to start off with a filter on papers of two years ago? Simply click this year link to use it as a new publication year filter: [2010](#) ☒
- ▶ Or how about looking at [papers containing statements on "using discriminative models" having the term "parsing" in their title?](#)
- ▶ Or how about looking at [papers containing statements on "computing semantic similarity" which appeared between 2006 and 2009?](#)
- ▶ Or how about [ACL papers by Manning which talk about "combining classifiers"?](#)
- ▶ Or have a look at [papers that Mirella Lapata co-authored with people from Saarland University.](#)

Troubleshooting

If highlighting sentences in the PDF View shouldn't work for you, then please consult [our Frequently Asked Questions section](#).

Simple Glossary

The Searchbench can also be used as a simple glossary of terms in Computational Linguistics. Just create a statements filter with the term to define as the semantic subject and "is" as the predicate.

Examples:

- ▶ [s:HPSG p:is](#) (to look up "HPSG")
- ▶ [s:mutual information p:is](#) (to look up "mutual information")

Citation Browsing

The Searchbench features CiBro, a citation browser for papers in the ACL Anthology. CiBro visually displays citation relations between papers in a citation graph. You can find CiBro on the "Citations" tab of each paper.

Here is [an example what the browser looks like](#).

ACL Anthology Searchbench

The screenshot shows the ACL Anthology Searchbench interface. On the left, there is a sidebar with several sections:

- Results list:** Points to the main list of search results.
- Search filters:** Includes buttons for 'add', 'remove', and 'edit', and a section for 'Statements' with a 'delete reading' button.
- Document view:** Points to the 'Plain Text' view option.
- Sentences view:** Points to the 'Extracted Topics' view option.
- PDF view:** Points to the 'PDF' view option.
- Citation browser:** Points to the 'Citations' view option.
- Online help:** Points to the 'Help' button at the bottom.

The main content area displays a list of search results. The first result is:

- An Improved Redundancy Elimination Algorithm for Underspecified Representations (2006)** by Koller, Alexander and Thater, Stefan. The abstract states: "This algorithm successively **deletes** eliminable splits from the chart, which reduces the set of described readings while making sure that at least one representative of each original equivalence class remains."
- Linguistic Theory in Statistical Language Learning (1998)** by Christer Samuelsson. The abstract states: "Then the rules of the grammar take turn **discarding** morphological readings based on their syntactic context. ... To limit the desolation brought about by the grammar, no rule is allowed to **remove** a word's last reading. ... If it does well, i.e., if it **removes** a lot of incorrect readings, but few correct ones, it is considered good and retained."
- The Evolution of Dominance Constraint Solvers (2005)** by Koller, Alexander and Thater, Stefan. The abstract states: "The individual readings can be enumerated from the description if they are needed, and this enumeration process should be efficient; but it is also possible to **eliminate** readings that are infelicitous given knowledge about the world or

The second result is a detailed view of the paper "An Improved Redundancy Elimination Algorithm for Underspecified Representations" (P06-1052) by Koller, Alexander and Thater, Stefan. It includes a 'Content' tab, a 'PDF' viewer, and a 'Citations' list. The 'Conclusion' section reads: "We presented an algorithm for redundancy elimination on underspecified chart representations. This algorithm successively **deletes** eliminable splits from the chart, which reduces the set of described readings while making sure that at least one representative of each original equivalence class remains. Equivalence is defined with respect to a certain class of rewriting systems; this definition approximates semantic equivalence of the described formulas and fits well with the underspecification". The 'Citations' list includes works by R. P. Chaves (2003), A. Copestake et al. (2004), M. Egg et al. (2001), and D. Flickinger (2002).

Research Fields in TAKE

Coreference resolution

Deep parsing and semantic tuple extraction

Combined semantic search



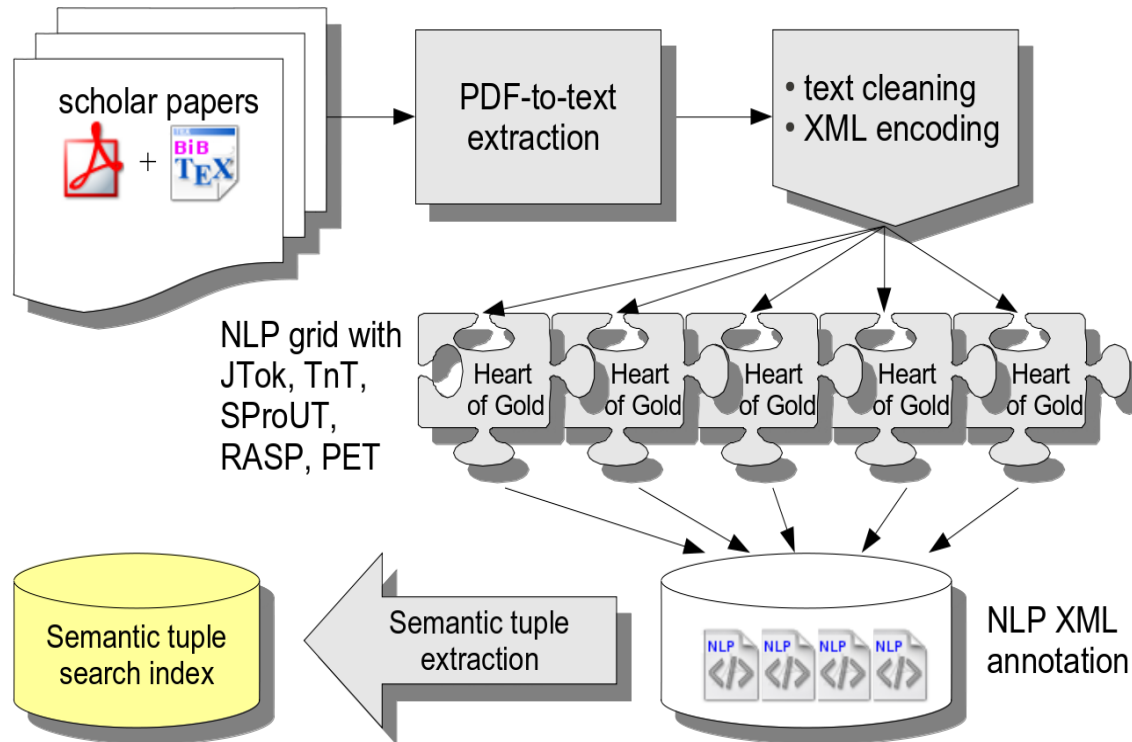
Citation Analysis

Unsupervised multi-word domain term extraction

Taxonomy extraction

Glossary extraction

Paper Parsing Architecture



Common NL Pre-Processing

Boost in Deep Parsing Coverage and Efficiency

ACL Anthology Parsing: breakthrough by combining

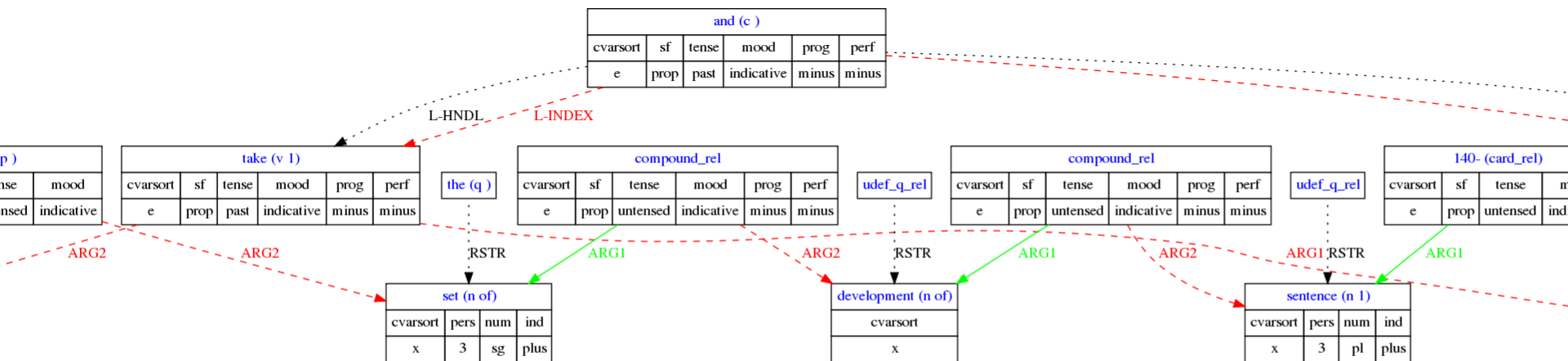
- **chart pruning**: directed search during parsing to increase performance, and also coverage for longer sentences (Cramer & Zhang, 2010)
- **chart mapping**, a novel method for integrating preprocessing information (Adolphs et al, 2008)
- new grammar (ERG) with better handling of open word classes
- fine-grained named entity recognition, including citation patterns (SProUT)
- new parse ranking model (WeScience; Oepen '09)

→ Improvement of overall coverage from 63% to now **>83%** full parses (now 4.9 million sentences)

DMRS to Semantic Tuple Conversion

From W07-1209, section 3

“We took the raw strings from the 140-sentence development set and parsed them with each of the state-of-the-art probabilistic parsers.”



SEMAITIC SUBJECT	SEMAITIC PREDICATE	SEMAITIC FIRST OBJECT	SEMAITIC SECOND OBJECT	ADJUNCTS
We	took	the raw strings from the 140-sentence development set		
We	parsed	them		with each of the state-of-the-art probabilistic parsers.

Asking Solr Index (simplified)

Query:

"method improve baseline"

is translated into Apache Solr query:

subj:method +pred:(improve OR ameliorate OR better OR meliorate) +(rest:baseline)

result (1 of 72) →

could also be used for question answering...

```
<doc> <!-- each doc is a single quirple sentence here -->
  <float name="score">1.2502118</float>
  <date name="timestamp">2009-01-27T10:46:38.452Z
</date>
  <str name="aclaid">W05-0814</str>
  <int name="offset">198</int>
  <int name="sentno">87</int>
  <int name="page">4</int>
  <str name="prefix">W05-0814-s87-p4</str>
  <str name="qgen">PET</str>
  <str name="sentence">Our model and training
    method improve upon a strong baseline for
    producing 1-to-many alignments.
</str>
  <str name="subj">Our model training method</str>
  <int name="subj_start">0</int>
  <int name="subj_end">28</int>
  <str name="pred">improve</str>
  <int name="pred_start">30</int>
  <int name="pred_end">36</int>
  <str name="rest">upon a strong baseline for
    producing 1-to-many alignments
</str>
  <int name="rest_start">38</int>
  <int name="rest_end">94</int>
</doc>
```

Searchbench: Statement Search Options

strict

only find strictly affirmative statements with a predicate matching only the entered one.

default

find generally affirmative or neutral statements with a predicate matching either the entered one or a synonym of it.

lax

as before, but additionally find statements with negated or neutral predicates matching antonyms of the entered predicate.

maximal

find statements with the entered predicate or a synonym/antonym thereof, irrespective of whether the predicate is negated or not

Multiword Domain Term Extraction

Based on an extended implementation of the Frantzi & Ananiadou 2000 approach (C-Value/NC-Value)

Example in Searchbench: „data structure + speech recognition + partial results + ...

Also basis for taxonomy and glossary extraction

THE "WHITEBOARD" ARCHITECTURE: A WAY TO INTEGRATE HETEROGENEOUS COMPONENTS OF NLP SYSTEMS

Christian Boitet (CNRS, France)

Mark Seligman (ATR Interpreting Telecommunications Research Laboratories, Kyoto Japan)

Content

PDF

Citations

THE "WHITEBOARD" ARCHITECTURE: A WAY TO INTEGRATE HETEROGENEOUS COMPONENTS OF NLP SYSTEMS

Abstract

We present a new software architecture for NLP systems made of heterogeneous components, and demonstrate an architectural prototype we have built at ATR in the context of Speech Translation.

KEYWORDS: Distributed NLP systems, Software architectures, Whiteboard.

INTRODUCTION Speech translation systems must integrate components handling speech recognition, machine translation and speech synthesis.

Speech recognition often uses special hardware.

word lattice
time span
time interval
speech translation
first layer
chart parser

Automatic Taxonomy Extraction - Evaluation with OntoGWAP

ONTO GAMES

[Home](#) | [Instructions](#) | [Invaders](#) | [Tetris](#) | [Quiz](#) | [My scores](#) | [Hall of fame](#) | [Feedback](#) | [Logout\(ulrich\)](#)

ONTO Quiz

Question

'pause' is-a 'disfluency' ?

POINTS

23

- "disfluency" is not a valid domain term
- "pause" is not a valid domain term
- yes
- no
- I don't know

High

local collocational clue domain model duration model

Shoot concepts that are NOT "knowledge source"

Time left 01:43

Show preview

Score 00000

Time 00:59

cyc

concept ontology

Tetris

Examples of extracted hypernym-hyponym pairs (including invalid pairs)

Hyper-/Hyponym Extraction: Evaluation

The competition lasted 10 days.

61 players participated,

32 Tetris players

10 Invaders players

26 Quiz participants

2940 pairs presented to the players (31% of the entire set;
pooling)

3-way agreements: 639 (490 is-a, 149 is-not-a)

5-way agreements: 298 (239 is-a, 59 is-not-a)

Citation Classification & Navigation

Intentions, Implicatures and Processing of Complex Questions

[Sanda Harabagiu, Steven Maiorano, Alessandro Moschitti, Cosmin Bejan] (ID: W04-2505)

Cited papers:

[W03-1006](#) [John Chen, Owen Rambow]
Use of Deep Linguistic Features for the Recognition and Labeling of Semantic Arguments

[P03-1003](#) [Abdessamad Echihabi, Daniel Marcu]
A Noisy-Channel Approach to Question Answering

[#02-32230](#) [Chin-Yew Lin E H Hovy, U Hermjakob, Deepak Ravichandran]
Using knowledge to facilitate pinpointing of factoid answers

[W03-1008](#) [Daniel Gildea, Julia Hockenmaier]
Identifying Semantic Roles Using Combinatory Categorical Grammar

[#02-0835](#) [D Gildea, D Jurafsky]
Automatic labeling of semantic roles

[P02-1031](#) [Daniel Gildea, Martha Palmer]
The Necessity of Parsing for Predicate Argument Recognition

[#01-31805](#) [King-Shy Goh, Edward Chang, Kwang-Ting Cheng]
SVM binary classifier ensembles for image classification

[#90-0391](#) [H Grice]
Logic and Conversation

[#75-32237](#) [Grice]
Syntax and Semantics Vol. 3: Speech Acts

[N04-1030](#) [Sameer S Pradhan, Wayne H Ward, Kadri Hacioglu, James H Martin, Dan Jurafsky]
Shallow Semantic Parsing using Support Vector Machines

[#00-4201](#) [S Harabagiu, M Paçca, S Maiorano]
Experiments with open-domain textual question answering

Navigation **Zoom** **Citation depth**

Back Center to PaperID

Forward + Overview

Calculate Endings Calculate Fanout

1 2 3 4 5 6 7 8 9 10

Sentiments

Negative Neutral

Agree Undefined

Recycle

Typed (Qualified) Citation Classification

Classify citation sentences into categories such as use, refutation, neutral, confirmative, ...

Possibly several categorized citations contribute to an overall classification of the reference from one paper to another (colored edge in the graphical user interface)

Rule-based approaches with PoS-, lexical, syntactical patterns: not robust, low overall recall and precision

→ Novel approach with semi-supervised learning on citation classification addresses two problems:

- expensive manual annotation
- unbalanced class distribution

New Citation Browser for ACL Searchbench

#02-5302
Jeh
et al. 5

#05-23494
Wan
et al.

#05-23469
Malin
et al.

#05-23425
Bekker
et al. 7

#06-23511
Yang
et al.

#06-23488
Minkov
et al. 5

P10-1006
Han
et al.

P11-1095
Han
et al. 16

using

similar

has investigated

disambiguated

similar

similar

was also

P11-1095 → D07-1074
are critically, proposed

Structural Semantic Relatedness: A Knowledge-Based Method to Named Entity Disambiguation

Xianpei Han, Jun Zhao

Cited Papers:

- #98-0614: Christiane Fellbaum (1998) *WordNet: An Electronic Lexical Database*
- #98-0143: Dekang Lin (1998) *An information-theoretic definition of similarity*
- P98-1012: Amit Bagga, Breck Baldwin (1998) *Entity-Based Cross-Document Coreferencing Using the Vector Space Model*
- #99-1464: Ricardo A Baeza-Yates, Berthier Ribeiro-Neto (1999) *Modern Information Retrieval*
- #02-5302: Glen Jeh, Jennifer Widom (2002) *Simrank: A measure of structural-context similarity*
- W03-0405: Gideon Mann, David Yarowsky (2003) *Unsupervised Personal Name Disambiguation*
- P04-1076: Cheng Niu, Wei Li, Rohini K Srihari (2004) *Weakly Supervised Learning for Cross-document Person Name Disambiguation Supported by Information Extraction*

Go to Paper...

Calculate Fanout

Depth:

Year Range:

Center

Show Preview-Nodes

1

1998

2011

Overview



ACL

Exercise 2

- Try to find the paper „Steven Abney; Steven Bird: The Human Language Project: Building a Universal Corpus of the World’s Languages“ from the ACL 2010 main conference on the various systems (the links on slide 2) plus ACL Anthology Network and ACL Anthology Searchbench.
- Try to find a part of that paper (sentence, keywords, statement) using these systems,
- Report on your findings

Literature

- Lutz Bornmann and Hans-Dieter Daniel. 2008. What do citation counts measure? A review of studies on 13 citing behavior. *Journal of Documentation*, 64(1):45–80. DOI 10.1108/00220410810844150.
- Steven Bird, Robert Dale, Bonnie Dorr, Bryan Gibson, Mark Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir Radev, and Yee Fan Tan. 2008. The ACL anthology reference corpus: A reference dataset for bibliographic research. In *Proceedings of the Language Resources and Evaluation Conference (LREC-2008)*, Marrakesh, Morocco.
- K. Frantzi, S. Ananiadou, and H. Mima. 2000. Automatic recognition of multi-word terms: the Cvalue/NC-value method. *International Journal on Digital Libraries*, 3:115–130.
- M. Hearst. 1992. Automatic Acquisition of Hyponyms from Large Text Corpora. In *Proc. of the 14th Coling Conference*, pages 539–545.
- Z. Kozareva, E. Riloff, and E. Hovy. 2008. Semantic Class Learning from the Web with Hyponym Pattern Linkage Graphs. In *Proc. of ACL*, pages 1048–1056.
- Ann Copestake and Dan Flickinger. 2000. An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the 2nd Conference on Language Resources and Evaluation (LREC-2000)*, pages 591–598, Athens, Greece.
- Ann Copestake, Dan Flickinger, Ivan A. Sag, and Carl Pollard. 2005. Minimal recursion semantics: an introduction. *Journal of Research on Language and Computation*, 3(2–3):281–332.
- CJ Rupp, Ann Copestake, Peter Corbett, and Ben Waldron. 2007. Integrating general-purpose and domain-specific components in the analysis of scientific text. In *Proceedings of the UK e-Science Programme All Hands Meeting 2007 (AHM2007)*, Nottingham, UK.
- Rune Sætre, Sagae Kenji, and Jun'ichi Tsujii. 2008. Syntactic features for protein-protein interaction extraction. In Christopher J.O. Baker and Su Jian, editors, *Short Paper Proceedings of the 2nd International Symposium on Languages in Biology and Medicine (LBM 2007)*, pages 6.1–6.14, Singapore, 1. ISSN 1613-0073319.

Literature

Eugene Garfield. 1955. Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, 123:108–111.

Eugene Garfield. 1965. Can citation indexing be automated? In Mary Elizabeth Stevens, Vincent E. Giuliano, and Laurence B. Heilprin, editors, *Statistical Association Methods for Mechanical Documentation*. National Bureau of Standards, Washington, DC. NBS Misc. Pub. 269.

Vladimir I. Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8):707–710.

David A. Pendlebury. 2009. The use and misuse of journal metrics and other citation indicators. *Archivum Immunologiae et Therapiae Experimentalis*, 57(1):1–11. DOI 10.1007/s00005-009-0008-y.

Dragomir R. Radev, Pradeep Muthukrishnan, and Vahed Qazvinian. 2009. The ACL anthology network corpus. In *Proceedings of the ACL Workshop on Natural Language Processing and Information Retrieval for Digital Libraries*, Singapore.

Simone Teufel, Advaith Siddharthan, and Dan Tidhar. 2006. Automatic classification of citation function. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 103–110, Sydney, Australia.

Ulrich Schäfer, Bernd Kiefer, Christian Spurk, Jörg Steffen, Rui Wang: The ACL Anthology Searchbench. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL HLT 2011), System Demonstrations*, pages 7-13, 2011. ISBN 978-1-932432-90-9. Portland, OR, USA.

Magdalena Wolska, Ulrich Schäfer, The Nghia Pham: Bootstrapping a Domain-specific Terminological Taxonomy from Scientific Text. *9th International Conference on Terminology and Artificial Intelligence (TIA)*, pages 17-23, Paris, France, 2011.

Literature

- Adolphs, P., Oepen, S., Callmeier, U., Crysmann, B., Flickinger, D., Kiefer, B.: Some fine points of hybrid natural language parsing. In: Proc. of LREC. pp. 1380-1387. Marrakesh, Morocco (2008).
- Callmeier, U.: PET – A platform for experimentation with efficient HPSG processing techniques. Natural Language Engineering 6(1), 99-108 (2000).
- Cramer, B., Zhang, Y.: Constraining robust constructions for broad-coverage parsing with precision grammars. In: Proc. of COLING. pp. 223-231. Beijing, China (2010).
- Flickinger, D., Oepen, S., Ytrestøl, G.: WikiWoods: Syntacto-semantic annotation for English Wikipedia. In: Proc. of LREC. pp. 1665-1671. Valletta, Malta (2010).