
VleWs4DFKI

Topic Extraction from Scientific Literature for Competency Management

Paul Buitelaar, Thomas Eigner
Competence Center Semantic Web &
Language Technology Lab
DFKI GmbH
Saarbrücken, Germany

Views4DFKI - the Basic Idea

- Competency (Expertise) management at DFKI through automatic topic extraction
 - DFKI as a research institute is an example of a knowledge-based organisation in which a large variety of competencies (expertise topics) are available, being developed, disappearing, etc.
- Visualize expertise topics in the context of name occurrences in DFKI web pages
 - Expertise visualization will help to increase inter-researcher and inter-research-group understanding of competencies
- Network analysis of related topics, researchers and documents
 - Network analysis can help to provide expert finding services, automated task group formation, etc.

Expertise Topic Extraction

We presented an ontology-based approach to information extraction in the soccer domain that aims at the automatic generation of a knowledge base from match reports and the subsequent visualization of the extracted information through automatic hyperlinking. We argue that such an approach is innovative and enhances the user experience.

Context Pattern

approach to

Topic Pattern

semantic tagging , using various corpora to derive relevant underspecified lexical

JJ NN

, VBG JJ NN TO VB JJ JJ JJ

semantic tagging

Expertise Topic Management

DFKI

Home | Profil | Sitemap | FAQ | Kontakt

User: paul00

Aktuelles Forschung Persönliches Organisation Services

Top-News

- 14.10.2008, [Grippeerschuldung 2008](#)
- 02.10.2008, [Neue elektronische Briefvorlagen](#)
- 29.09.2008, [***GI-Mitteilung #111](#)
- 16.09.2008, [Nächste Betriebsversammlung am 18.11.08 in KL](#)
- 15.09.2008, [\[DFKI-SB\] Fassadensanierung](#)

Information

VieWs4DFKI allows DFKI researchers to define a set of scientific or technical topics that are of relevance to their research and which they would like to be associated with. To support this process, VieWs4DFKI automatically extracts such topics from scientific articles as provided by [Google Scholar](#). The topics that we were able to extract for you are listed below. Please select those that you are comfortable with and rank them according to your preferences. Additionally you are able to define 0 to 5 topics manually. The 5 highest ranked extracted topics and the manually defined topics will be stored and displayed with your name as it occurs in the DFKI web pages. A [screenshot](#) may explain more clearly how this will look like. We hope you will like this new service! Please send any feedback to paulb@dfki.de

Please note that this service will be available for now only from within the DFKI network, expert topics linked to your name will be visible only to those with proper access rights from within DFKI.

Expert Topic Management of Paul Buitelaar

Expert Topics defined by user

Extracted Expert Topics

Please select and rank up to 5 extracted expert topics.

Select	Expert Topic	Rank
<input checked="" type="checkbox"/>	Semantic Web	↑ ↓
<input checked="" type="checkbox"/>	LEXICAL SEMANTICS	↑ ↓
<input checked="" type="checkbox"/>	ontology learning	↑ ↓
<input type="checkbox"/>	semantic annotation	↑ ↓
<input checked="" type="checkbox"/>	information extraction	↑ ↓
<input type="checkbox"/>	cross-language information retrieval	↑ ↓
<input type="checkbox"/>	lexical semantic encoding	↑ ↓
<input type="checkbox"/>	bilingual dictionary look-up	↑ ↓
<input type="checkbox"/>	sense disambiguation	↑ ↓
<input type="checkbox"/>	domain specific sense assignment	↑ ↓

Mitarbeiter-Schnellsuche

Nachname:

Vorname:

Suchen [erweiterte Suche](#)

Monatsübersicht

<< Oktober >>

Mo	Di	Mi	Do	Fr	Sa	So
29	30	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3	4	5	6	7	8	9

Visualization of Expertise Topics

The image shows a screenshot of the DFKI Semantics Web website. The website layout includes a header with the DFKI logo and navigation links for 'Deutsch' and 'English'. A left sidebar contains a menu with items like 'STARTSEITE', 'AKTUELLES', 'ÜBER UNS', 'FORSCHUNG', 'LIVING LABS', 'KOMPETENZENTRENEN', 'e-Learning', 'Sprachtechnologie', 'Semantisches Web', 'Projekte', 'Virtuelles Büro der Zukunft', 'Computational Culture', 'Human-Centered Visualization', 'TRANSFERZENTRUM', 'PRESSE', 'KONTAKT', 'DFKI VIDEO', and 'IMPRESSUM'. The main content area is titled 'Kompetenzzentrum Semantisches Web' and contains several paragraphs of text describing the center's mission and activities. A right sidebar features a 'Aktuelles' section with news items and an 'IFA 2008' logo. An overlay visualization is positioned over the 'Kontakt' section, showing a list of expertise topics: 'Sprachtechnologie', 'Semantisches Web', 'Visitenkarte', 'Publikationen', and 'Expertise'. A box labeled 'Semantic Web' lists specific topics: 'LEXICAL SEMANTICS', 'ontology learning', 'information extraction', and 'large-scale semantic tagging'.

DFKI Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

Kompetenzzentrum Semantisches Web

Deutsch English

Suche

Startseite

AKTUELLES

ÜBER UNS

FORSCHUNG

LIVING LABS

KOMPETENZENTRENEN

e-Learning

Sprachtechnologie

Semantisches Web

Projekte

Virtuelles Büro der Zukunft

Computational Culture

Human-Centered Visualization

TRANSFERZENTRUM

PRESSE

KONTAKT

DFKI VIDEO

IMPRESSUM

Kompetenzzentrum Semantisches Web

Das CCSW ist ein virtuelles Forschungszentrum innerhalb des DFKI, dessen Ziel es ist, interne Synergien zwischen Projekten und Forschungsbereichen zu ermöglichen, die sich mit der Entwicklung und dem Einsatz von Semantic Web und Web 3.0 Technologien befassen, insbes. XML, RDF, RDF Schema, OWL, Regelsprachen und (semantische) Web Services.

Das Semantic Web ist eine zentrale Aktivität des World Wide Web Consortium (W3C) mit dem Ziel, Web-Informationen in maschinenlesbarer Form zu repräsentieren, so dass diese nicht nur dargestellt werden können, sondern um Automatisierung, Integration und Wiederverwendung zwischen Applikationen zu ermöglichen. Das DFKI hat eine lange Tradition auf den Forschungsgebieten, die für das Semantic Web relevant sind, z.B. Ontologien, semantische Metadaten, Regelsysteme und Informationsextraktion.

Ein Schwerpunkt des CCSW ist verteiltes Informations-Wissensmanagement auf der Basis von standardisierten Semantic Web Wissensrepräsentationssprachen wie RDF/S und OWL. Das CCSW fasst das Semantic Web als eine übergeordnete Struktur auf, die einer Vielfalt von Teilwebs entspringt, von denen jedes die Bedürfnisse einer spezifischen Web-Community befriedigt, z.B.:

- auf Ontologien basierende eScience-Anwendungen auf dem Gebiet der Biomedizin und auf anderen Gebieten
- semantische Analyse für Business Intelligence im Finanzsektor und öffentlichen Verwaltungssektor
- Community-basierte intelligente Web-Anwendungen in der Medienindustrie

Unser Slogan "Semantic Web von unten" betont unseren von den Daten ausgehenden Ansatz für das Semantic Web. Hierbei werden z.B. Methoden für die semiautomatische Extraktion von semantischen Metadaten und Ontologien unter Ausnutzung von Natural Language Processing und Verfahren des maschinellen Lernens betrachtet. Ein weiteres wichtiges Gebiet ist die semantische Informationsintegration, d.h. Abgleich, Verschmelzung und Transformation von heterogenen Datenquellen - Datenbanken, semistrukturierte Wissensbasen, unstrukturierter Text und Multimediadokumente - unter Verwendung semantischer Metadaten, Ontologien und Regeln.

Das CCSW bietet Training und Consulting-Dienstleistungen für die Industrie, öffentliche Verwaltung und private Organisationen zu allen Aspekten des Semantic Web an, z.B.:

- Ontologieaufbau, -lernen, -implementierung und -wartung
- Versehen von konventionellen Daten (z.B. textuelle und Multimediadaten) mit Wissensmarkup
- Inferenzen und Reasoning

Leitung: Dipl.-Inform. Michael Sirenk, Dr. Paul Buitelaar

Kontakt

Deutsches Forschungszentrum für Künstliche Intelligenz
Trippstädter Straße 122
D-67663 Kaiserslautern
Deutschland

Telefon: +49 (0)631 / 205-75 130 | +49 (0)681 / 302-5325
Fax: +49 (0)631 / 205-75 103
Homepage: <http://semanticsweb.dfi.de>

Aktuelles

- DFKI präsentiert auf der REHACARE Sicherheitskomponenten für teilautonome Rollstühle
Pressemitteilung
- DFKI - 20 Jahre Innovation pur
Sonstige Seiten
- Künstliche Intelligenz 2008 in Kaiserslautern
Pressemitteilung
- DFKI präsentiert sich am 16.09.2008 auf dem Parlamentarischen Abend in der Landesvertretung der Freien Hansestadt Bremen in Berlin
Pressemitteilung
- IFA 2008
Besuchen Sie uns in TWF, Halle 2.03, Stand 14
- DFKI im Bundeskanzleramt beim Tag der offenen Tür
Pressemitteilung
- Advanced Tangible Interface Lab
Neue Interaktionen mit dem Wi Balance Board

Deutschland Land der Ideen

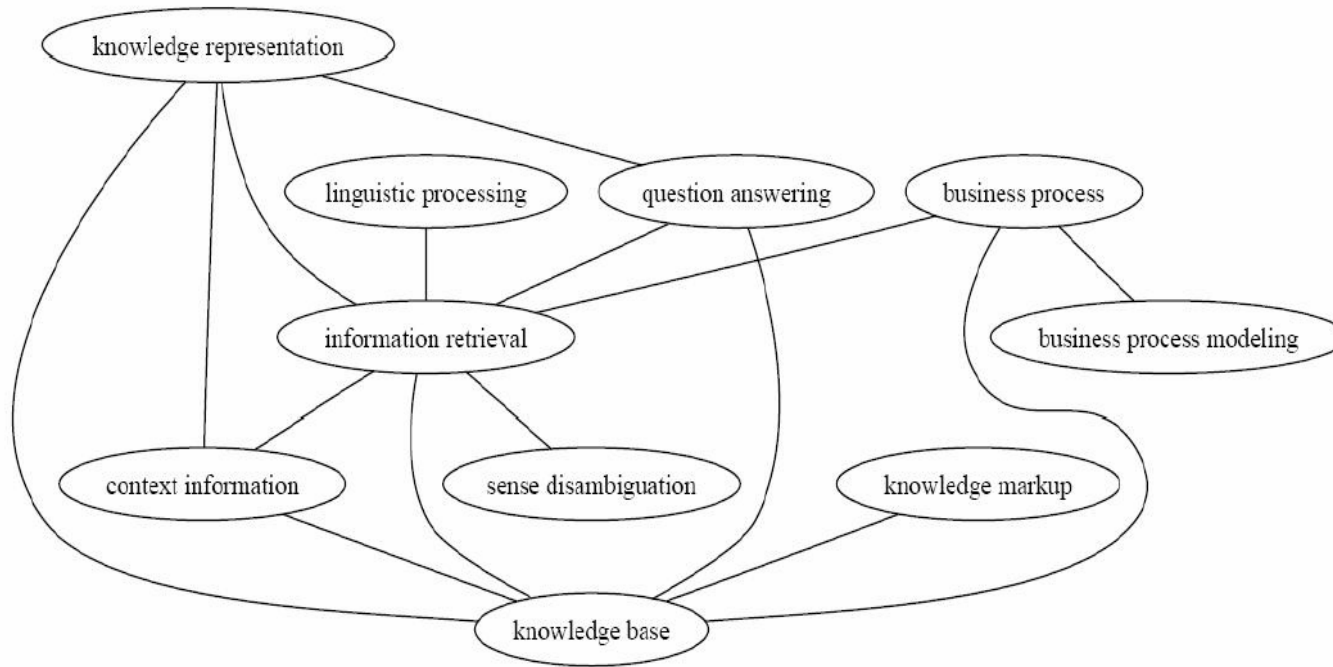
Paul Buitelaar

- Sprachtechnologie
- Semantisches Web
- Visitenkarte
- Publikationen
- Expertise

Semantic Web

- LEXICAL SEMANTICS
- ontology learning
- information extraction
- large-scale semantic tagging

Expertise Topic Network Analysis



- Network analysis based on co-occurrence frequency of topics within documents

Experiment – Data Set

- 199 DFKI researchers
 - Almost all researchers - from all departments: natural language processing, information retrieval, knowledge management, business informatics, image processing, robotics, agent systems
 - Download scientific publications
- 3253 downloaded scientific publications
 - Using the Google API - Google Scholar
 - In HTML, generated automatically from PDF, PS, ...
 - Apply context patterns
- 7946 extracted topic text segments
 - Topic text segments up to 10 words long
 - Part-of-speech tagged with TnT
 - Apply topic patterns – one topic per topic segment

Expertise Topic Extraction

Context Patterns

Topic Patterns

approach to

~~semantic tagging~~ , using various corpora to derive relevant underspecified lexical
JJ NN , VBG JJ NN TO VB JJ JJ JJ
semantic tagging

tools for

~~ontology adaptation~~ and for mapping different ontologies should be an
NN NN CC IN VBG JJ NNS MD VB DT
ontology adaptation

approach for

~~modeling similarity measures~~ which tries to avoid the mentioned problems
JJ NN NNS WDT VBZ TO VB DT VBN NNS
modeling similarity measures

methods for

~~domain specific semantic lexicon construction~~ that builds on the reuse
NN JJ JJ NN NN WDT VBZ IN DT NN
domain specific semantic lexicon construction

Evaluation – Data Set

- Randomly selected subset of 85 researchers
 - Includes researchers without assigned topics, e.g. due to sparse data in their case
 - 48 submitted evaluation results

Evaluation - Interface

Evaluierung

Topics für Paul Buitelaar:

- mining large text corpora
- other ongoing relevant research
- systematic polysemous classes
- infer knowledge structures
- large text corpora
- large-scale semantic tagging
- lexical knowledge acquisition
- lexical semantic analysis
- new relation instances
- ongoing relevant research
- seminar location Ciravegna
- text-based ontology extraction
- verb semantic classes
- Word Sense Disambiguation
- systematic polysemy
- concept tagging
- sense disambiguation
- knowledge markup
- Language Comprehension
- polymorphic type

Evaluation - Results

Level of Correctness	Number of Researchers
0-10%	7
11-20%	1
21-30%	3
31-40%	9
41-50%	6
51-60%	9
61-70%	10
71-80%	3
81-100%	0
	48

Current Work

- Extend context patterns
- Extend topic patterns



Current Work - extend context patterns

#	Context Pattern	Extracted Topics (examples)
80	approach to	quality evaluation, document structure analysis, ontology learning, ...
29	approach for	lip motion analysis, ontology learning, ontology evaluation, ...
25	approaches to	automated sales dialogs, document structure analysis, ...
10	approaches for	song similarity computation, dialog act classification, ...
10	based on	case retrieval, order-sorted algebra, systematic polysemous classes, ...
9	task of	dialogue modeling, plan recognition, multimodal fusion, ...
8	framework for	web service discovery, contextual reasoning, multimodal dialog systems, ...
7	aspects of	concept tagging, linguistic processing, systematic polysemy, ...
7	area of	dialogue modeling, information retrieval, gesture synthesis, ...
7	evaluation of	page dewarping, knowledge base, reference models, business component, ...

Current Work - extend topic patterns

129	NNP NNP	Language Text,
67	NNP NNP NNP	Introduction Shallow NLG, Web Service Coordination, ...
55	NN NN	dialogue component, specification method, proof search, ...
40	NNP NNP NNP NNP	Discourse Modelling Context Representation, ...
34	NNP NNP NNP NNP NNP	Lexicalized Head-Driven Syntactic Generation Tilman, ...
29	JJ JJ NN	second essential feature, flexible shallow approach, ...
27	NNP NN	Shallow generation, Behavioral Modeling, ...
21	JJ NN NN	low resolution image, semantic web approach, ...
19	DT NN NN	the translation equivalent, an evaluation method, ...
17	JJ NN TO JJ NN	spectrometry-based approach to quantitative analysis, ...

Conclusions

- Pros

- Real-life application, experimentation

- Cons

- No real semantics yet