

TechWatch

Technology and Market Observation powered by SMILA

PD Dr. Günter Neumann
DFKI, Deutsches Forschungszentrum für Künstliche Intelligenz
GmbH, Juni 2011

Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages



» Technologies:

- » Combine methods from Bibliometrics, Information Wrapping, Text Mining, Information Extraction, Semantic Search and Knowledge Management.

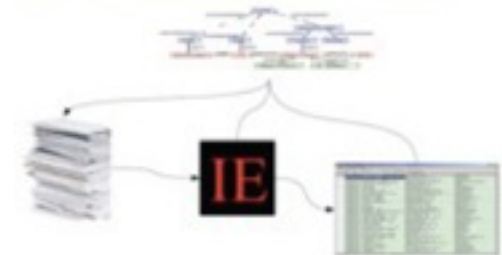
» Applications:

- » Search for publications and patents
- » The analysis of extracted relationships between themes, authors, time behavior and organizations.
- » Search for Key-Players of innovations.
- » Web-based search for trend statements.
- » Presentation of a knowledge domain as ontology, which can be used in further functionalities that improve the search and analysis

- » Cooperation between DFKI and TKS since 2006
- » Ontology-based search for key players in the area of welding and joining technologies
 - » Authors, inventors, companies, institutes, technologies
 - » Advanced patent search (EPO)
 - » Publications (Google Scholar)
 - » Visualization of the created innovation network and of new trends
- » Software development „TechWatchTool“
- » In Theseus-Ordo
 - » SMILA-fication of the system architecture
 - » Ontology-based information extraction
 - » Trend-Monitoring and analysis (new technologies, innovation pushers)



Ontology-based Information Extraction



- **Extraction of relations through linguistic analysis**
 - e.g., dependency analysis
- **Information that change over time**
 - e.g., „Increase of publications about lithium batteries“
- **Near-Synonyms**
 - „Approach“; „Method“
- **Collocations**
 - „come to a decision“, „make a picture“
- **Co-reference**
 - „dependency analysis“, „this method“
- **Hyponyms, Part-of Relationship**
 - „is a“, „consists of“
- **Meta data**
 - „is author of“, „is owner of the patent“

» Relevant Key Phrases

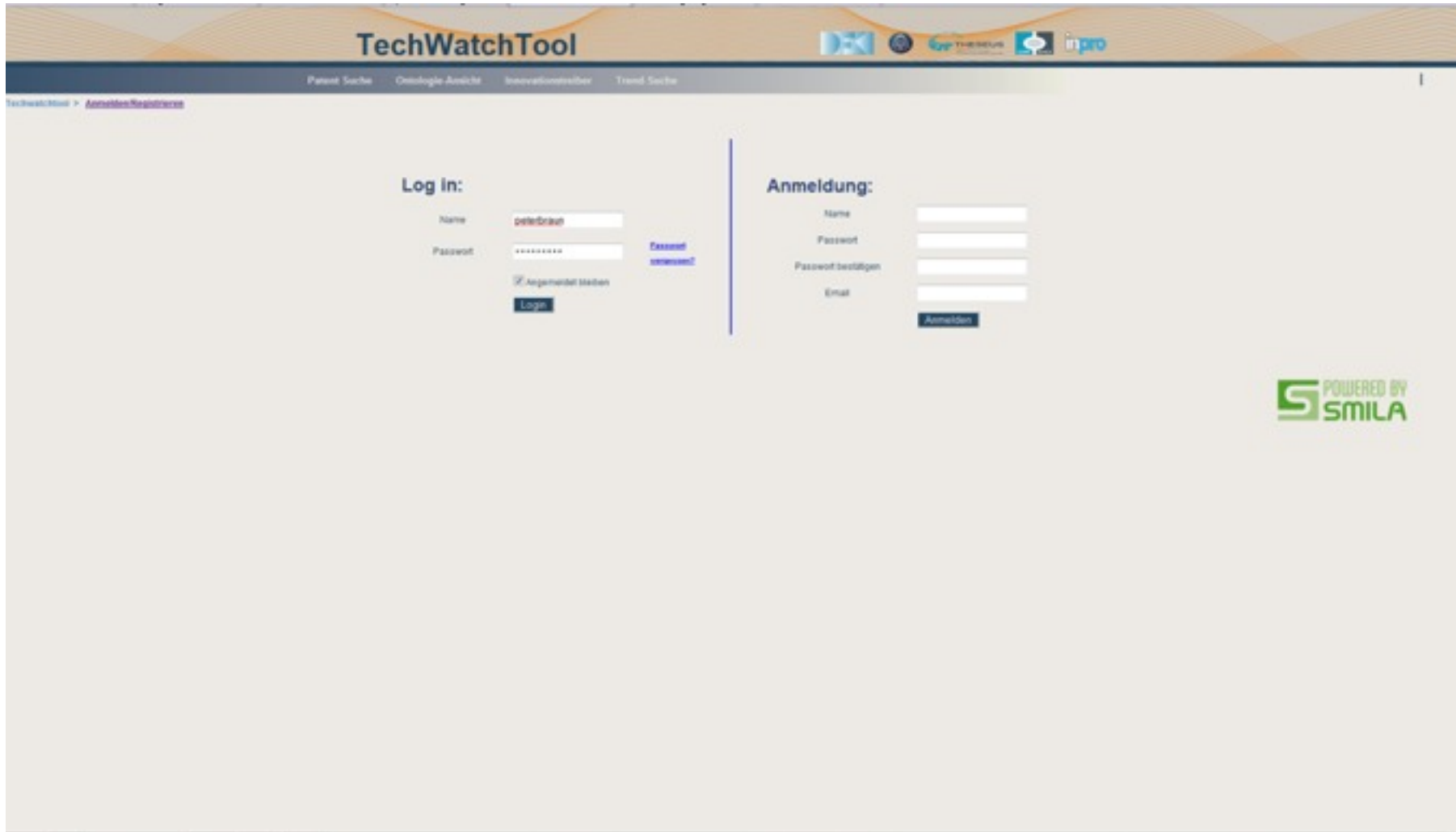
» „Technology of tomorrow“, „Future“, „Development“

» Sentence patterns

» „Lithium-ion batteries will power cars in the future.“

» „Lithium-ion technology is the future for hybrid and electric cars.“

» „The laser will always be good for a surprise also in the future“, predicts Peter Leibinger.



TechWatchTool

Patent Suche Ontologie Ansicht Innovationstreiber Trend Suche

TechWatchTool > [Anmelden/Registrieren](#)

Log in:

Name: peterbraut

Passwort: *****

Angemeldet bleiben

[Passwort vergessen?](#)

Login

Anmeldung:

Name: _____

Passwort: _____

Passwort bestätigen: _____

Email: _____

Anmelden

POWERED BY SMILA

TechWatchTool



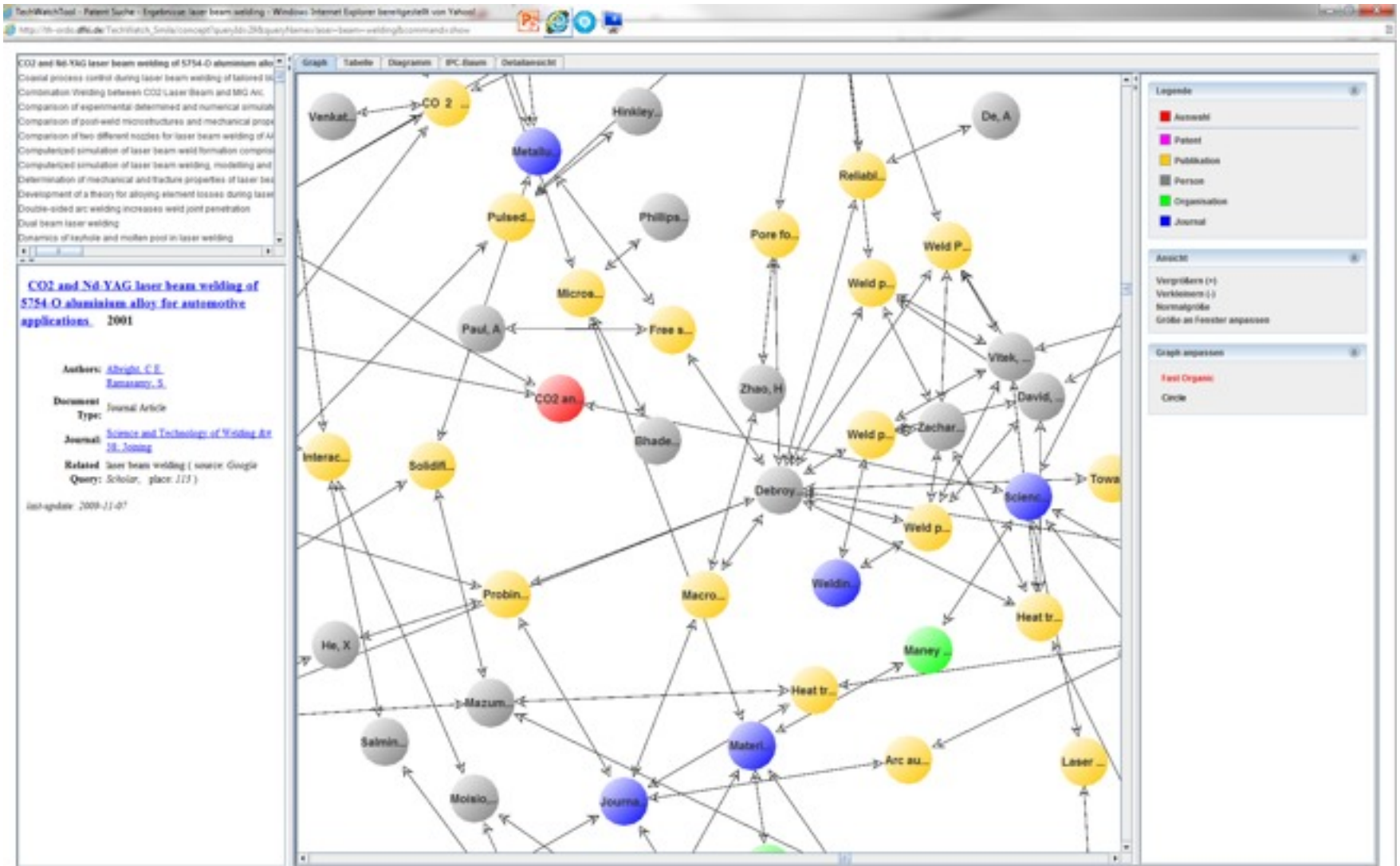
[Patent Suche](#) [Ontologie-Ansicht](#) [Innovationen-Viewer](#) [Trend-Suche](#)

TechWatchTool > [Patent-Suche](#)

TechWatch Suche

[Suche](#) [Erweiterte Suche](#)





Technische Patente: Ergebnisse einer Patent-Suche - Windows Internet Explorer bereitgestellt von Yahoo!

http://www.dfkf.de/Technische_Patente/Concept.asp?ID=205&queryName=beam+beam+welding&command=show

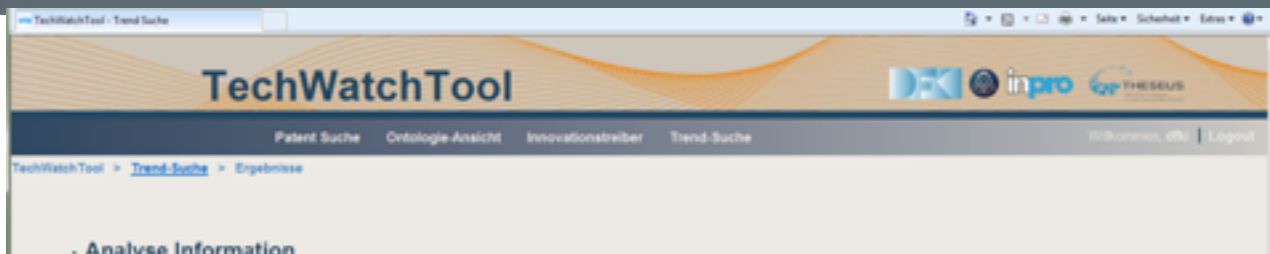
GEN ELECTRIC
GEN MOTORS CORP
GENERAL ELECTRIC CO
GULIN RES INST OF GEOLOGY FOR
HAEBLER LASER UND FEINWERKTEC
HAERBIN-INDUSTRIY UNIVERSITY
HARBIN INST OF TECHNOLOGY
HITACHI CONSTRUCTION MACHINERY
HITACHI LTD
HITACHI-PLANT TECHNOLOGIES LTD
HITACHI ZOSSEN CORP
HOCHSCHULE MITTELDAFH
HONDA MOTOR CO LTD

GEN ELECTRIC

All Patents in Database:

- [System and method of dual laser beam welding using first and second filler metals \(2012-04-01\)](#)
- [System and method of dual laser beam welding using first and second filler metals \(2011-03-08\)](#)
- [SYSTEM AND METHOD OF DUAL LASER BEAM WELDING OF FIRST AND SECOND FILLER METALS \(2011-03-24\)](#)
- [A method of high-powered laser beam welding of articles using a metallic shim protruding from the surfaces of the articles... Assembly therefore \(2010-11-01\)](#)
- [HIGH-POWERED LASER BEAM WELDING AND ASSEMBLY THEREFOR \(2010-09-30\)](#)
- [A method of high-powered laser beam welding of articles using a metallic shim protruding from the surfaces of the articles... \(2010-05-11\)](#)
- [DEVICE FOR ENHANCING EFFICIENCY OF AN ENERGY EXTRACTION SYSTEM \(2010-02-11\)](#)
- [PREHEATING USING A LASER BEAM \(2009-11-09\)](#)
- [Superconducting homopolar inductor alternator for power applications \(2009-10-07\)](#)
- [Thermal environmental barrier coating system for silicon-containing materials \(2009-06-01\)](#)

Patent	Publikation	Person	Journal	Organisation	Relation	Patentanzahl	Anzahl der Veröffentlichungen	Anmerkung
GEN ELECTRIC		applicant				5		0,75
FRAUNHOFER GES FORSCHUNG		applicant				7		0,97
ABEL, PETER		applicant				2		0,5
HONDA MOTOR CO LTD		applicant				2		0,49
COMAU INC		applicant				1		0,4
NEPSON STEEL CORP		applicant				5		0,35
ThyssenKrupp Steel AG		applicant				5		0,32
TOYOTA CAR CORP		applicant				19		0,29
POSCO		applicant				3		0,27
BOSCH GMBH ROBERT		applicant				7		0,26
CHISON INC		applicant				1		0,2
LASAS AG		applicant				1		0,2
ThyssenKrupp Patented Blanks GmbH		applicant				1		0,2
LOGENORA S P A		applicant				1		0,2
RES INST INC SCIENCE & TECH		applicant				4		0,19
MITSUBISHI ELECTRIC CORP		applicant				3		0,16
KOBE STEEL LTD		applicant				3		0,15
HOCHSCHULE MITTELDAFH		applicant				1		0,13
INST POLYTECHN TRBAYV UAA		applicant				1		0,13
INDIA MOTOR		applicant				3		0,119
CALSONIC KAWASO CORP		applicant				1		0,1
CHENGDU HOLE MANUFACTURE TECH		applicant				1		0,1
SUMITOMO HEAVY INDUSTRIES		applicant				2		0,09
NISSAN MOTOR		applicant				15		0,088
DOGGUANG HUST MFG ENGINEERING		applicant				2		0,087
WUHAN MOTOR CO LTD		applicant				1		0,084
HITACHI LTD		applicant				1		0,08
INST OF MECHANICS CHINESE ACAD		applicant				1		0,06
KUKA-ROBOTE GmbH		applicant				1		0,08
MITSUBISHI HEAVY IND LTD		applicant				5		0,06
Linde AG		applicant				2		0,07
DENSO CORP		applicant				3		0,04
Volkswagen AG		applicant				5		0,04
TOYOTA MOTOR CO		applicant				1		0,047
TOYOTA MOTOR CORP		applicant				4		0,042
WU CORP		applicant				3		0,038
HR LIQUIDE		applicant				1		0,038
NANLING NANO CO LTD		applicant				1		0,037
TRUMPF LASER UND SYSTEMTECHNIK		applicant				5		0,037
RENAULT SA		applicant				2		0,035
OTROEN SA		applicant				1		0,035
FHEBERLE LASER UND FEINWERKTEC		applicant				1		0,035
PEUGEOT		applicant				1		0,035
STI SHIPBUILDING CO LTD		applicant				1		0,035
ORIENT CHEMICAL IND		applicant				1		0,045
SHIKHARA REEF COY		applicant				1		0,044
JFE STEEL CORP		applicant				1		0,044
MURKATA KUKAI SEISAKUSHOKU		applicant				1		0,044
OSAMA UNIV		applicant				1		0,044
SUMITOMO WIRING SYSTEMS		applicant				1		0,044
SUNG WOO HTECH CO LTD		applicant				1		0,044
ITO YOSHIMIKO		applicant				2		0,044
DGCO CORP		applicant				1		0,04
HITACHI ZOSSEN CORP		applicant				1		0,04
SHENZHEN WANH E LASER TECHNOLOG		applicant				1		0,04
FTS GAMBH		applicant				1		0,038
GENERAL ELECTRIC CO		applicant				1		0,036
SHANGHAI WELDING BACKING FACT		applicant				1		0,036
Rockwell International Energy Systems Group, Co		publisher					1	0,033
SHENKHA		applicant				1		0,033
YUBA TECH CO LTD		applicant				1		0,033
HUANZHONG UNIVERSITY OF SCIENCE		applicant				2		0,029
THCO ELECTRONICS WAP GMBH		applicant				1		0,028
WENZHOU GREAT ELECTRICAL CO LT		applicant				1		0,029
American Society for Metals, Metals Park, Oh		publisher					1	0,027



TechWatchTool - Trend Suche

TechWatchTool

Patent Suche Ontologie-Ansicht Innovationstreiber Trend-Suche

Willkommen, ifb | Logout

TechWatchTool > [Trend-Suche](#) > Ergebnisse

Analyse Information

Suchanfrage: lithium-ion
 Innovationstreiber: Allgemein
 Gespeicherte Dateien: 05
 Personen-Suchergebnisse: 604
 Organisationen-Suchergebnisse: 1376

Sprache:
 Muster: /
 Trends in
 Abfrage n
 Abfrage n

DOWNLOAD

[results.zip](#)

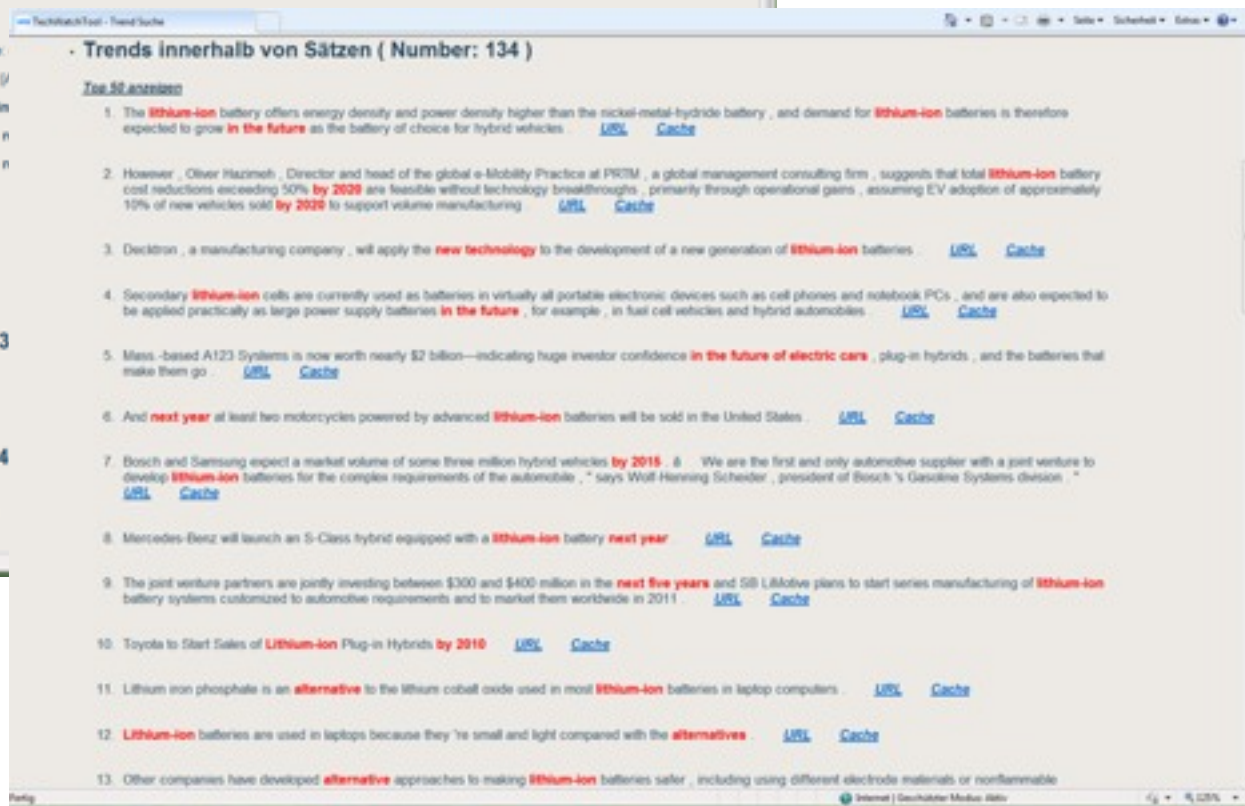
Trends innerhalb von Sätzen (Number: 13

[Top 50 anzeigen](#)
[alle anzeigen](#)

Abfrage relevanter Personen (Number: 14

[Top 14 anzeigen](#)
[alle Personen anzeigen \(Sortierung nach vorkommender Häufigkeit\)](#)

Fertig



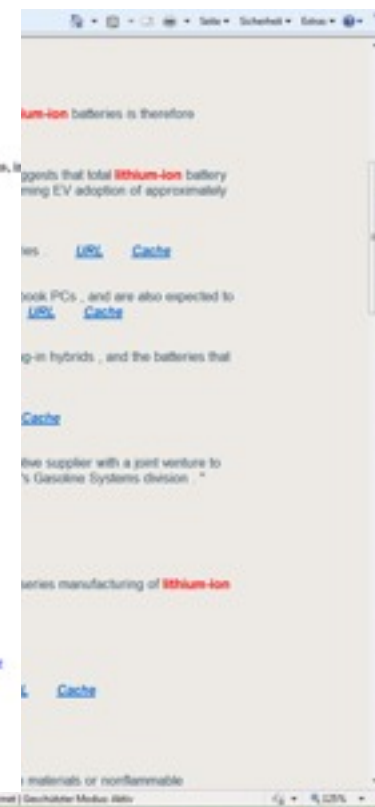
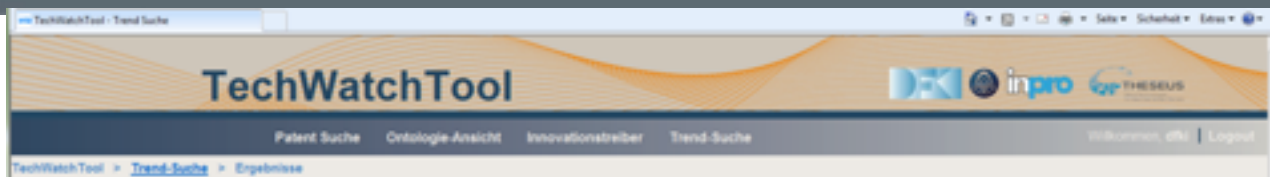
TechWatchTool - Trend Suche

Trends innerhalb von Sätzen (Number: 134)

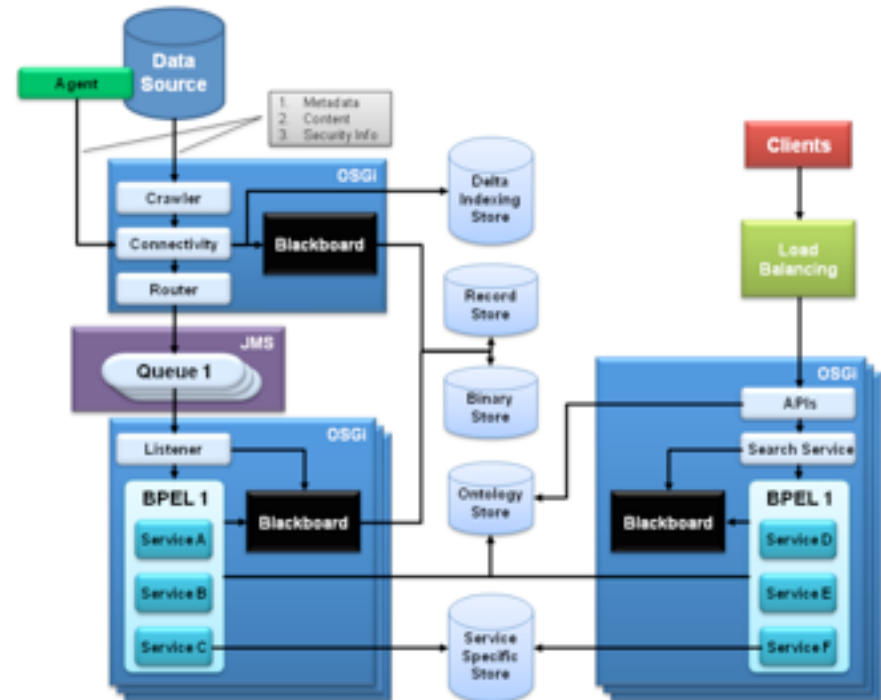
[Top 50 anzeigen](#)

- The **lithium-ion** battery offers energy density and power density higher than the nickel-metal-hydride battery , and demand for **lithium-ion** batteries is therefore expected to grow **in the future** as the battery of choice for hybrid vehicles . [URL](#) [Cache](#)
- However , Oliver Hazimeh , Director and head of the global e-Mobility Practice at PRCM , a global management consulting firm , suggests that total **lithium-ion** battery cost reductions exceeding 50% **by 2020** are feasible without technology breakthroughs , primarily through operational gains , assuming EV adoption of approximately 10% of new vehicles sold **by 2020** to support volume manufacturing . [URL](#) [Cache](#)
- Decatron , a manufacturing company , will apply the **new technology** to the development of a new generation of **lithium-ion** batteries . [URL](#) [Cache](#)
- Secondary **lithium-ion** cells are currently used as batteries in virtually all portable electronic devices such as cell phones and notebook PCs , and are also expected to be applied practically as large power supply batteries **in the future** , for example , in fuel cell vehicles and hybrid automobiles . [URL](#) [Cache](#)
- Mass-based A123 Systems is now worth nearly \$2 billion—indicating huge investor confidence **in the future of electric cars** , plug-in hybrids , and the batteries that make them go . [URL](#) [Cache](#)
- And **next year** at least two motorcycles powered by advanced **lithium-ion** batteries will be sold in the United States . [URL](#) [Cache](#)
- Bosch and Samsung expect a market volume of some three million hybrid vehicles **by 2015** . " We are the first and only automotive supplier with a joint venture to develop **lithium-ion** batteries for the complex requirements of the automobile , " says Wolf Henning Scheider , president of Bosch 's Gasoline Systems division . [URL](#) [Cache](#)
- Mercedes-Benz will launch an S-Class hybrid equipped with a **lithium-ion** battery **next year** . [URL](#) [Cache](#)
- The joint venture partners are jointly investing between \$300 and \$400 million in the **next five years** and SB LiMotive plans to start series manufacturing of **lithium-ion** battery systems customized to automotive requirements and to market them worldwide in 2011 . [URL](#) [Cache](#)
- Toyota to Start Sales of **Lithium-ion** Plug-in Hybrids **by 2010** . [URL](#) [Cache](#)
- Lithium iron phosphate is an **alternative** to the lithium cobalt oxide used in most **lithium-ion** batteries in laptop computers . [URL](#) [Cache](#)
- Lithium-ion** batteries are used in laptops because they 're small and light compared with the **alternatives** . [URL](#) [Cache](#)
- Other companies have developed **alternative** approaches to making **lithium-ion** batteries safer , including using different electrode materials or nonflammable

Fertig



- » The core components are integrated via wrapper methods
 - » Patent search
 - » Google Scholar
 - » DepatisNet
 - » Persistence
 - » Trend search
 - » Google Custom Search
 - » Document processing
 - » Trend patterns
 - » Relation extraction
 - » Named Entity recognition
 - » Persistence



LanguageID

- Automatic language identification
- 23 languages
- Speed: ~6ms
- Accuracy: 99,8%
- Implementation: Java, OSGi

MDParser

- Extraction of the dependency relations of sentences
- Languages: DE, EN & more!
- Speed: 40 Sent./Sec.
- Accuracy: 86% - 88%
- Implementation: Java, OSGi

NE-Hub

- Extraction of Named Entities: Persons, Organizations, Locations, etc.
- Integrates the result of different NE-recognizers (Sprout, OpenNLP, Stanford, LingPipe, etc.)
- Languages: DE, EN
- Implementation: Java, OSGi

DARE

- Semi-supervised extraction of relations
- Languages: DE, EN
- Implementation: Java, OSGi

Selection and orchestration of the components

Voting mechanism



Options

HoverMe for a hint.

	paragraph	sentence	token	name
Processors	dummy opennlp	dummy lingpipe stanford opennlp_sentence_en	dummy lingpipe stanford opennlp_token_en	dummy lingpipe stanford opennlp_sentence_en
Meta Processors				dummy voting

Selection and orchestration of the components

Voting mechanism

Sending



- A fine-grained extraction and search for entities and relations demands for a structural analysis of free text.
- Recently, dependency relations between words are getting more and more important for semantic applications
The dependency relationships uncover a flat semantic predicate argument structure.
- Our current R&D focus
 - Automatic learning of dependency grammars and parsers from corpora
 - Multilingual dependency parsing
 - Solution: MDParse



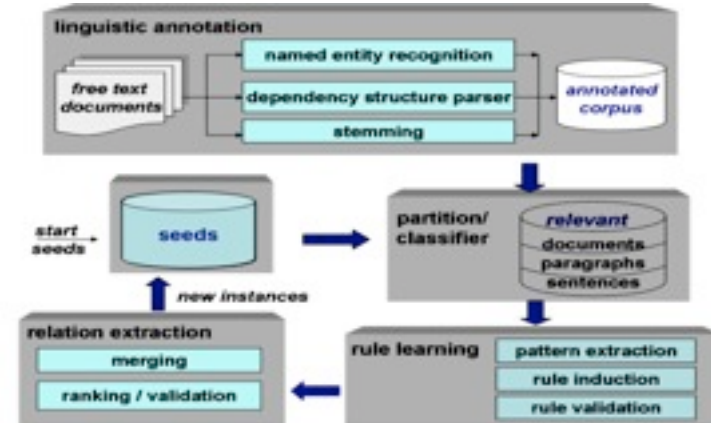
	Parsing Time	Sentences per Second	Tokens per Second
MDParser	73.188s	46.128	1015.55
MaltParser	1954.684s	1.73	38.02

RTE-6: 3376/74326 Sent./Words

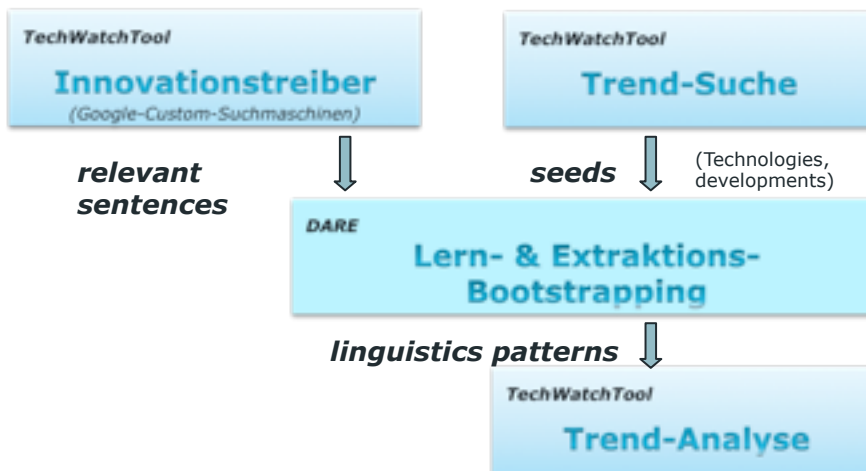
- A fine-grained extraction and search for entities and relations demands for a structural analysis of free text.
- Recently, dependency relations between words are getting more and more important for semantic applications
The dependency relationships uncover a flat semantic predicate argument structure.
- Our current R&D focus
 - Automatic learning of dependency grammars and parsers from corpora
 - Multilingual dependency parsing
 - Solution: MDParse



- Extraction of relations of different complexities.
- Automatic learning of linguistic patterns for the extraction of new instances and their projections.
- **Bootstrapping** between pattern learning and relation extraction which is initialized by means of a small set of seed instances



Integration into *TechWatchTool*



- Learning of linguistic patterns for the automatic recognition of new technologies and developments for a future trend analysis.
- Use of the found results (technologies & developments) of the trend analysis as new seeds
- Search of relevant sentences by means of innovation indicators.

uKeyWe

- Unsupervised Machine Learning algorithm for the extraction of key phrases
- Languages: language independent
- Implementation: Java, OSGi

ConExt

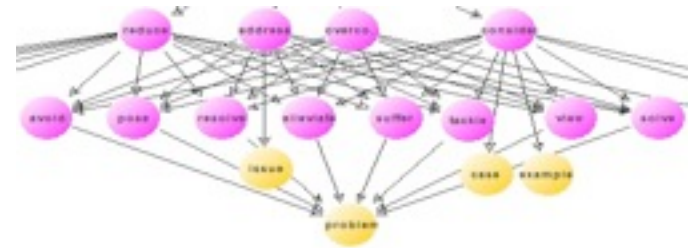
- Extraction of concept hierarchies from free texts
- Languages: DE, EN
- Implementation: Java, OSGi

TERA (Textual Entailment)

- Decides for two text fragments (phrases, sentences, paragraphs), whether the one semantically entails the other
- Languages: DE, EN
- Accuracy: best results at international competitions (cf. Text Analysis Conference - TAC)
- Implementation: Java, OSGi

- **Extraction of concept hierarchies** (based on Cimiano, Hotho, Staab; 2005)
- a method for extracting domain specific concept hierarchies from free text which has been syntactically analysed
- **Core idea:** words, which occur with same syntactic functions and context are semantically related.
- Can also be used to determine the terminology of a domain.
- **semantically related** words can be easier identified.
- **Information extraction** can benefit (e.g., by means of **Query Expansion, Clustering**)
- The strategy has been tested with several corpora:
 - business reports (3.964.925 sentences)
 - scientific publications (884.904 sentences)
 - Wikipedia, technical articles (628.560 Sentences)

Query expansion



Clustering



- **Extraction of concept hierarchies** (based on Cimiano, Hotho, Staab; 2005)

- a method for extracting hierarchies from free text syntactically analysed

- **Core idea:** words, which syntactic functions and are related.

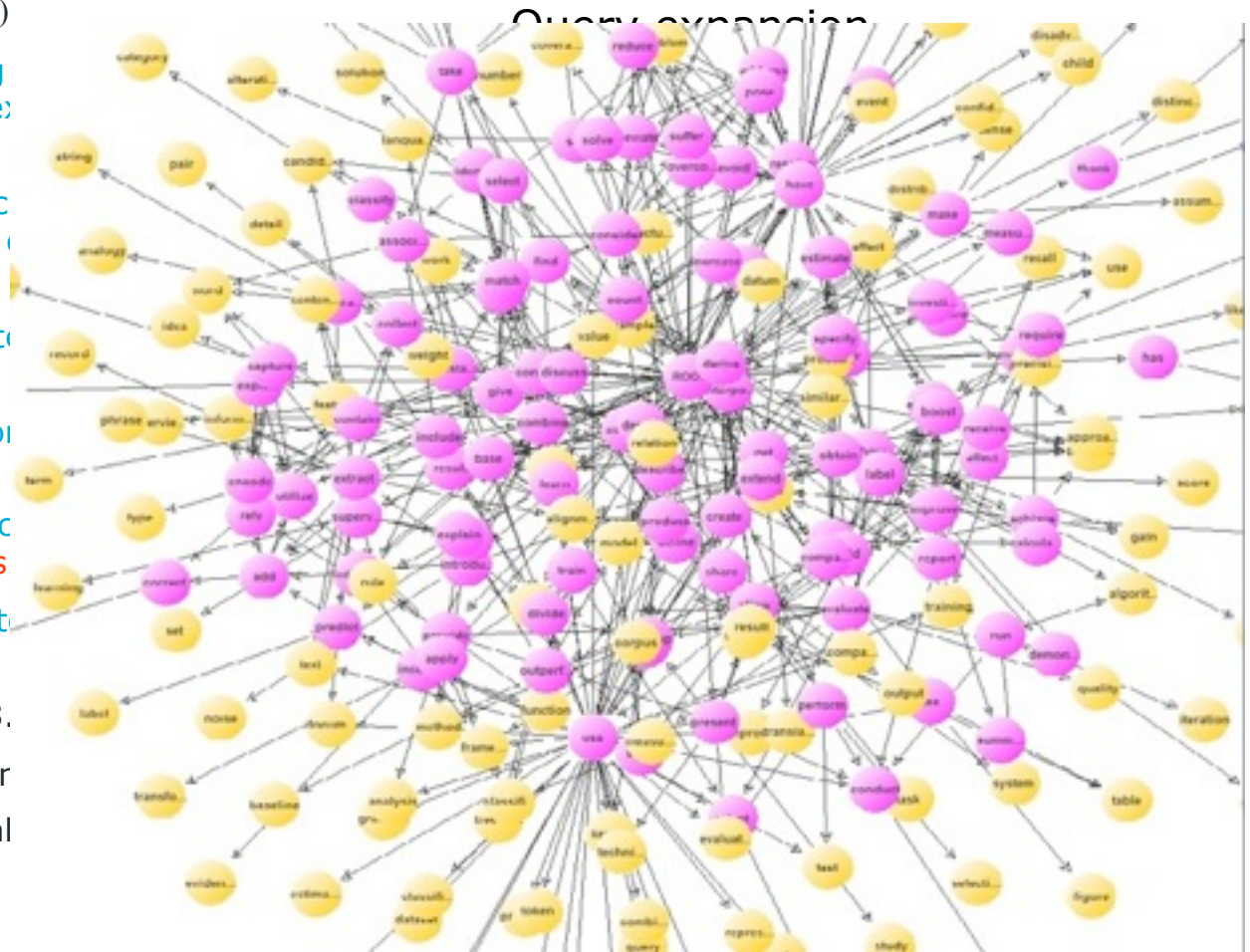
- Can also be used to detect the core of a domain.

- **semantically related words** identified.

- **Information extraction** can be done by means of **Query Expansion**

- The strategy has been tested on corpora:


- business reports (3.000 documents)
- scientific publications (100.000 documents)
- Wikipedia, technical documents (100.000 Sentences)



- Textual Inference
 - Determine semantic inference relationships between texts (Recognition of text variants/paraphrases)
 - Entailment: Does Text T entails Text H ?
 - Needed in many text analytics applications (e.g., information extraction, question answering, text summarization)
 - Q: „Wo acquired Overture?“
A: „Googles buying of Overture supports ...“
⇒: „X buys Y“ ≈ „X acquire Y“

- Textual Inference
 - Determine semantic inference relationships between texts (Recognition of text variants/paraphrases)
 - Entailment: Does Text T entails Text H ?
 - Needed in many text analytics applications (e.g., information extraction, question answering, text summarization)
 - Q: „Wo acquired Overture?“
A: „Googles buying of Overture supports ...“
⇒: „X buys Y“ ≈ „X acquire Y“

RTE-6 Main Task Example



Topic 918: Betty Friedan

H380: Betty Friedan is the author of "The Feminine Mystique"

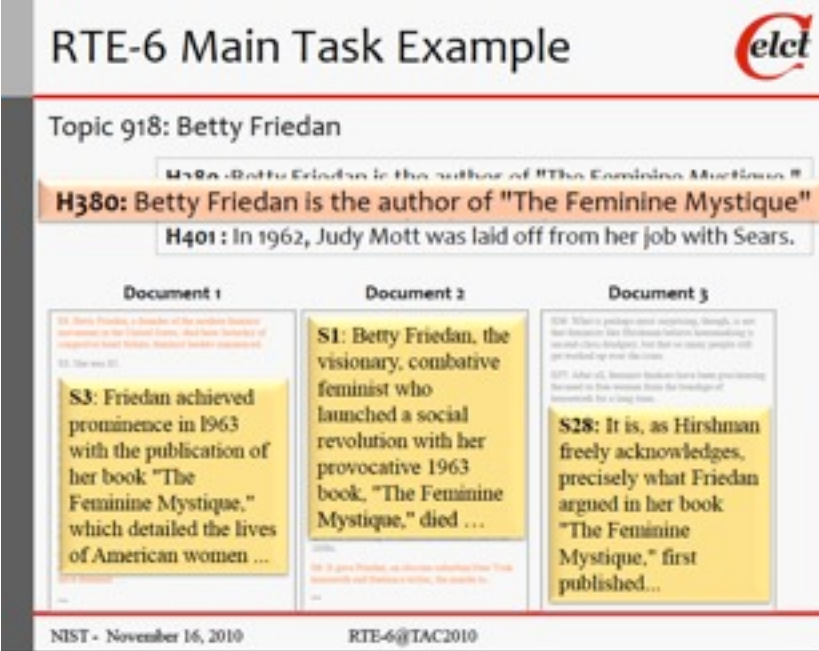
H380: Betty Friedan is the author of "The Feminine Mystique"


H401: In 1962, Judy Mott was laid off from her job with Sears.

Document 1	Document 2	Document 3
<p style="font-size: x-small; color: gray;">S3: Friedan achieved prominence in 1963 with the publication of her book "The Feminine Mystique," which detailed the lives of American women ...</p>	<p style="font-size: x-small; color: gray;">S1: Betty Friedan, the visionary, combative feminist who launched a social revolution with her provocative 1963 book, "The Feminine Mystique," died ...</p>	<p style="font-size: x-small; color: gray;">S28: It is, as Hirshman freely acknowledges, precisely what Friedan argued in her book "The Feminine Mystique," first published...</p>

NIST - November 16, 2010
RTE-6@TAC2010

- Textual Inference
 - Determine semantic inference relationships between texts (Recognition of text variants/paraphrases)
 - Entailment: Does Text T entails Text H ?
 - Needed in many text analytics applications (e.g., information extraction, question answering, text summarization)
 - Q: „Wo acquired Overture?“
A: „Googles buying of Overture supports ...“
⇒: „X buys Y“ ≈ „X acquire Y“
- Textual inference in TechWatch Ordo
 - semantic search
 - concept extraction/ontology learning
- Textual inference (Entailment) is still a very young and dynamic R&D area !
 - Progress through international competitions (NIST, USA)
 - DFKI with very good results
 - EU project Excitement -> prob. from 2012 -> ~ 3.4 million EUR funding (DFKI is one of the 4 core scientific partners, 3 industrial partners; analysis of customer interaction data)



RTE-6 Main Task Example 

Topic 918: Betty Friedan

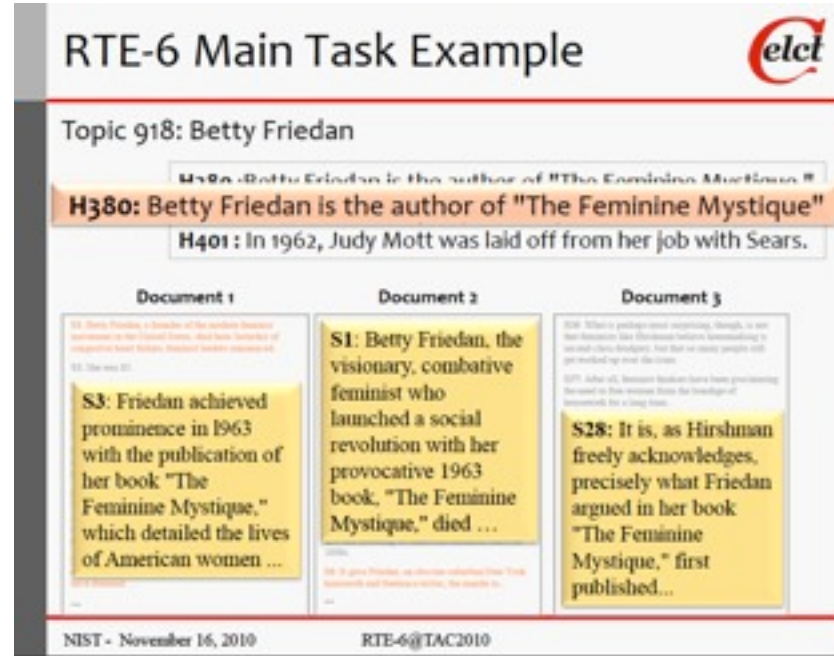
H380: Betty Friedan is the author of "The Feminine Mystique"


H401: In 1962, Judy Mott was laid off from her job with Sears.

Document 1	Document 2	Document 3
<p>S3: Friedan achieved prominence in 1963 with the publication of her book "The Feminine Mystique," which detailed the lives of American women ...</p>	<p>S1: Betty Friedan, the visionary, combative feminist who launched a social revolution with her provocative 1963 book, "The Feminine Mystique," died ...</p>	<p>S28: It is, as Hirshman freely acknowledges, precisely what Friedan argued in her book "The Feminine Mystique," first published...</p>

NIST - November 16, 2010 RTE-6@TAC2010

- Textual Inference
 - Determine semantic inference relationships between texts (Recognition of text variants/paraphrases)
 - Entailment: Does Text T entails Text H ?
 - Needed in many text analytics applications (e.g., information extraction, question answering, text summarization)
 - Q: „Wo acquired Overture?“
A: „Googles buying of Overture supports ...“
⇒: „X buys Y“ ≈ „X acquire Y“
- Textual inference in TechWatch Ordo
 - semantic search
 - concept extraction/ontology learning
- Textual inference (Entailment) is still a very young and dynamic R&D area !
 - Progress through international competitions (NIST, USA)
 - DFKI with very good results
 - EU project Excitement -> prob. from 2012 -> ~ 3.4 million EUR funding (DFKI is one of the 4 core scientific partners, 3 industrial partners; analysis of customer interaction data)



RTE-6 Main Task Example 

Topic 918: Betty Friedan

H380: Betty Friedan is the author of "The Feminine Mystique"

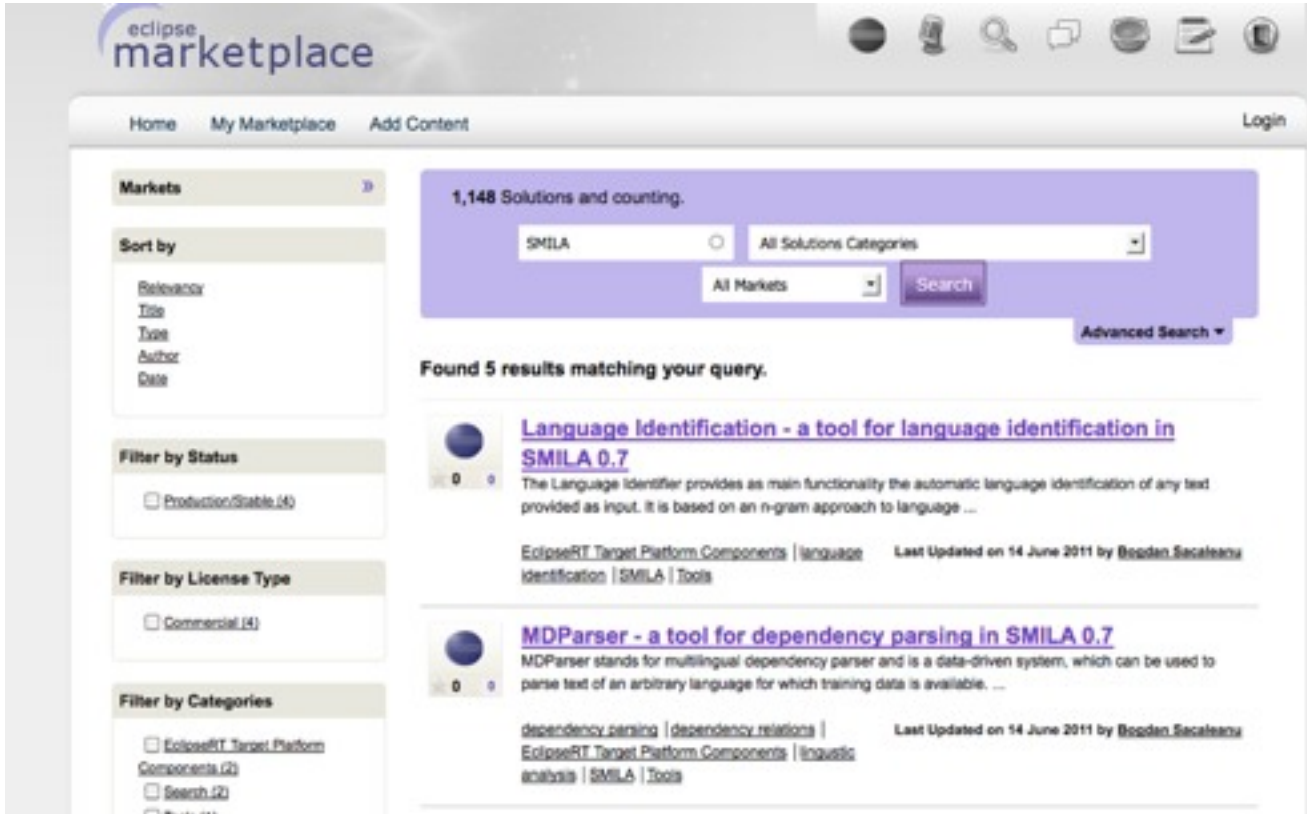
H401: In 1962, Judy Mott was laid off from her job with Sears.

Document 1	Document 2	Document 3
<p>S3: Friedan achieved prominence in 1963 with the publication of her book "The Feminine Mystique," which detailed the lives of American women ...</p>	<p>S1: Betty Friedan, the visionary, combative feminist who launched a social revolution with her provocative 1963 book, "The Feminine Mystique," died ...</p>	<p>S28: It is, as Hirshman freely acknowledges, precisely what Friedan argued in her book "The Feminine Mystique," first published...</p>

NIST - November 16, 2010 RTE-6@TAC2010

Focus:

- robustness, scalability
- organized by NIST, USA
- 2010: 19 Unis/Institute (world wide), DFKI, 4. best result in Novelty Subtask (with new team) (2009: 2. best result in Main Task.)



The screenshot shows the Eclipse Marketplace search interface. At the top, there are navigation links for Home, My Marketplace, Add Content, and Login. The main search area displays '1,148 Solutions and counting.' with a search bar containing 'SMILA' and 'All Solutions Categories'. Below the search bar, it says 'Found 5 results matching your query.' and lists two results:

- Language Identification - a tool for language identification in SMILA 0.7**
 The Language Identifier provides as main functionality the automatic language identification of any text provided as input. It is based on an n-gram approach to language ...
 EclipseRT Target Platform Components | language identification | SMILA | Tools
 Last Updated on 14 June 2011 by Bogdan Secaleanu
- MDParser - a tool for dependency parsing in SMILA 0.7**
 MDParser stands for multilingual dependency parser and is a data-driven system, which can be used to parse text of an arbitrary language for which training data is available. ...
 dependency parsing | dependency relations | EclipseRT Target Platform Components | linguistic analysis | SMILA | Tools
 Last Updated on 14 June 2011 by Bogdan Secaleanu

On the left side, there are filters for Sort by (Relevance, Title, Text, Author, Date), Filter by Status (Production/Stable (4)), Filter by License Type (Commercial (4)), and Filter by Categories (EclipseRT Target Platform Components (2), Search (2), Tools (1)).

Current

- Implementation: SMILA Piplets (Java-OSGi web service)
- Direct integration into SMILA-process
- License: Research & commercial

Planned

- NE-Hub – Multilingual architecture for NE recognition

Experience of our SMILA developers

Pros

- Rapid entry point for End-User by means of:
 - detailed documentation
 - many good examples
 - Out-of-the-box runnable system
- Parallelism of workflow through BPEL
- Fast composition of new apps by means of existing components and BPEL
- Rapid and competent support by means of the forum

Cons

- Complicated data structures
Version 0.7 – solved in Version 0.8
- Restricted extendability of the index algorithms in order to support several dynamically changing indices per user (a very specific requirement)