TechWatch
Technology and Market Observation
powered by SMILA

PD Dr. Günter Neumann
DFKI, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Juni 2011
Goal - Observation of Innovations and Trends

» 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)
• **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.
Patent Search
### Patent Search

#### Patent Database

<table>
<thead>
<tr>
<th>Patent</th>
<th>Publication</th>
<th>Person</th>
<th>Journal</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEN ELECTRIC</td>
<td>FRAUNHOFER GES FORSCHUNG</td>
<td>applicant</td>
<td>7</td>
<td>5.537</td>
</tr>
<tr>
<td>ABEL-PETER</td>
<td>HONDA MOTOR CO LTD</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>HONDA MOTOR CO LTD</td>
<td>COMAIR INC</td>
<td>2</td>
<td>3.25</td>
<td></td>
</tr>
<tr>
<td>NIPPON STEEL CORP</td>
<td>THOMAS KINGSLEIGH AG</td>
<td>2</td>
<td>2.727</td>
<td></td>
</tr>
<tr>
<td>TOSHIBA CORP</td>
<td>TỦNHUR PHARMA</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PANASONIC</td>
<td>CORKIN INC</td>
<td>3</td>
<td>3.208</td>
<td></td>
</tr>
<tr>
<td>LG CORP</td>
<td>THOMAS KINGSLEIGH AG</td>
<td>3</td>
<td>2.98</td>
<td></td>
</tr>
<tr>
<td>LGB CORP</td>
<td>DзВзА®ВВВг</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LG CORP</td>
<td>HONDA MOTOR CO LTD</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Graph

- **Title**: GEN ELECTRIC
- **Database**: Patent Search

#### Patent Details

- **Total Patents**: 30
- **Companies**: GEN ELECTRIC
- **Inventions**:
  - System and method of dual laser beam welding using first and second filler metals (2011-04-30)
  - System and method of dual laser beam welding using first and second filler metals (2011-09-08)
  - SYSTEM AND METHOD OF DUAL LASER BEAM WELDING OF FIRST AND SECOND FILLER METALS (2011-02-24)
  - A method of high-powered laser beam welding of articles using a metallic clad produced from the surfaces of the articles (Assembly thereof) (2010-11-08)
  - HIGH POWERED LASER BEAM WELDING AND ASSEMBLY THEREOF (2006-05-30)
  - A method of high-powered laser beam welding of articles using a metallic clad produced from the surfaces of the articles (2009-09-30)

#### Additional Information

- **Date**: Donnerstag, 21. Juli 2011
- **Publisher**: DFKI, Günter Neumann, Juni 2011
Trend analysis

- Analyse Information
  - Suchanfrage: lithium-ion
  - Innovationstreiber: Allgemein
  - Gespeicherte Dateien: 85
  - Personen-Suchergebnisse: 604
  - Organisationen-Suchergebnisse: 1576

- DOWNLOAD
  - results.zip

- Trends innerhalb von Sätzen (Number: 13)
  - Top 50 passagen
    - alle passagen

- Abfrage relevanter Personen (Number: 14)
  - Top 14 passagen
    - alle Personen anzeigen (Sortierung nach vorkommender Häufigkeit)

---

1. The lithium-ion battery offers energy density and power density higher than the nickel-metal-hyride battery, and demand for lithium-ion batteries is therefore expected to grow in the future as the battery of choice for hybrid vehicles.

2. However, Oliver Hazimeh, Director and head of the global e-Mobility Practice at PRTM, 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.

3. Declotron, a manufacturing company, will apply the new technology to the development of a new generation of lithium-ion batteries.

4. Secondary lithium-ion cells are currently used in 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.

5. 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.

6. And next year at least two motorcycles powered by advanced lithium-ion batteries will be sold in the United States.

7. 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 Schneider, President of Bosch's Gasoline Systems division.

8. Mercedes-Benz will launch an S-Class hybrid equipped with a lithium-ion battery next year.

9. 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.

10. Toyota to Start Sales of Lithium-ion Plug-in Hybrids by 2010.

11. Lithium iron phosphate is an alternative to the lithium cobalt oxide used in most lithium-ion batteries in laptop computers.

12. Lithium-ion batteries are used in laptops because they’re small and light compared with the alternatives.

13. Other companies have developed alternative approaches to making lithium-ion batteries safer, including using different electrode materials or nonflammable.
Trend analysis
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
TechWatch - DFKI Language Technology software

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
Example - NE-Hub

Selection and orchestration of the components

Voting mechanism
Example - NE-Hub

Selection and orchestration of the components

Voting mechanism
• 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: MDPARSER
• 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: MDPARSER
• 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

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

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

© DFKI, Günter Neumann, Juni 2011
- **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)
• **Extraction of concept hierarchies** (based on Cimiano, Hotho, Staab; 2005)
• a method for extracting hierarchies from free text: 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)
• 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)
Example
- Textual Inference

- 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“
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)
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: "Who acquired Overture?"
  A: "Google's buying of Overture supports ..."
  \( \Rightarrow \) "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)

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.)
**Eclipse Marketplace**

---

**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)