Language Technology Panel

Unlocking the Data Treasure Chest

A Network of Excellence

language technology
software

Standards and best practices for the Multilingual Web

The Forum for Europe’s Language Technology Industry
Big = Unstructured = Text = Multilingual

CONTENT IS KING

Content Control for Customers/Citizens

Engines on-page SEO consists only as good as its content will take it. The root of current trends is content is at the heart of a search engine. Our dedicated search team and content writers start the ranking. They write content that represents you, not only effective on one page, the best results for your business. The experience and talents to implement in a cohesively manner. SEO is only good if page SEO is an all-encompassing term that covers the root of current trends in search engine optimization. Popularity, the search campaign entails off-page profile, etc. Article optimization, discussion forum, a few. The off-page SEO is critical to support incredible services from others. With a local, localized feel, getting listed locally will search engine survival. All search engines spend significant time to put your business to rank on the search engines on local searches. Just as in regional or national. At incredible marketing, we specialize in getting exposure on the web as possible - including national, regional, or local. Your mobile search is taking over. Did you know that? Did you also know that for mobile search and standard search are different? Probably not, but many people do not understand that mobile search is its own beast. And becoming important to focus on. Currently, about 40% of all searches world come from a mobile device, and that number is growing exponentially. To dominate these rankings, it is not enough to merely have a mobile-friendly site. You must also optimize your website for mobile search criteria. We can help.
Interoperability thru Semantics

- screen
- monitor
- écran
- Bildschirm
- Monitor
- Display
- monitor

Terminology

Synonymy
The Panelists

Hans-Ulrich von Freyberg
CEO Cocomore AG

Philippe Wacker
Secretary General LT-Innovate

Hans Uszkoreit
Scientific Director DFKI GmbH

Moderated by:
Guðrún Magnúsdóttir
CSO ESTeam AB
Language Technology

Key-enabling technology for unlocking the Data Treasure Chest

Hans Uszkoreit, DFKI
Big Data – From Hype to Boom

- It will happen!
- Maybe under a different name.
What about Language Technology

- Have you read about the LT Hype
What about Language Technology

- Have you read about the LT Hype 🧡

http://www.meta-net.eu
What about Language Technology

- Have you read about the LT Hype

- Me neither!
What about Language Technology

- Have you read about the LT Hype

- Me neither!

- This may be our fault.
...because the boom has started

- Google semantic search
- Google, Apple, Nuance voice search
- Web Translation: Google, Bing, etc.
- Machine (aided) translation at EC, EP, EPO, most IT blue-chip enterprises
- IBM Watson
- Siri Voice Assistance

http://www.meta-net.eu
A boom without a hype?

- Do we understand that the two trends belong together?
- Can we bring them together before everybody else is earning money with the combination?
META-NET is a Network of Excellence Technologies for Multilingual Europe with 60 research centers in 34 countries

META-NET forges the Multilingual Europe Technology Alliance, which already has 725 Members in 60 countries

META-NET has published 31 White Papers each describing the state of one European Language in the Digital World

META-NET has issued the Strategic Research Agenda “Technologies for Multilingual Europe 2020”
Language among other Data

- Text
- Images
- Video
- Numbers
- Sound
Language among other Data

Text

Sound
Speech
Index/Metadata

Images
Legends/Tags

Numbers

Video
Speech
Language among other Data

- Text
- Sound
- Images
- Numbers
- Video
Language among other Data

Text

Sound

Images

Numbers

Video
Language among other Data

- Text
- Images
- Video
- Numbers
- Sound

Metadata/Semantics
Data Value Chain

Text

Sound

Images

Metadata/Semantics

Numbers

Video

Acquisition/Production/Collection

Storage

Link/Associate/Aggregate

Monitor/Analyse/Interpret

Deliver/Present/Visualize

Use/Share/Validate/Revise/
Language Technology in the Value Chain

- Text
- Sound
- Images
- Numbers
- Video

- Acquisition/Production/Collection
- Storage
- Link/Associate/Aggregate
- Monitor/Analyze/Interpret
- Deliver/Present/Visualize
- Use/Share/Validate/Revise
Language technologies are inevitable for adding value to big data for extracting information and knowledge from textual documents and social media, and for making them accessible across language borders for searching, linking and documenting non-linguistic, e.g. numerical or multimedia data, through effective metadata for semantic interfaces for the human consumer of the end products, e.g., through question answering, report generation and summarization.
Some Directions

- aggregate across language boundaries
- combine data of different type (e.g. numerical data with textual data)
Questions?
Comments?
Praise?
Complaints?
Thank You!
Research Priority Themes

- Translingual Cloud – Understanding everything, everywhere, everytime

- Social Intelligence – Technologies for e-participation and better decisions

- Socially aware interactive assistants
<table>
<thead>
<tr>
<th>Name</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michaela Bartelt</td>
<td>Electronic Arts, Germany</td>
</tr>
<tr>
<td>Joseph Mariani</td>
<td>CNRS/LIMSI, France</td>
</tr>
<tr>
<td>Will Burgett</td>
<td>Intel Corporation, United States</td>
</tr>
<tr>
<td>Roger Moore</td>
<td>University of Sheffield, UK</td>
</tr>
<tr>
<td>Johannes Bursch</td>
<td>Daimler AG, Germany</td>
</tr>
<tr>
<td>Jaap van der Meer</td>
<td>TAUS, The Netherlands</td>
</tr>
<tr>
<td>Nicoletta Calzolari</td>
<td>Istituto di Linguistica Computazionale “Antonio Zampolli”, CNR, Italy</td>
</tr>
<tr>
<td>Stelios Piperidis</td>
<td>Institute for Language and Speech Processing, Greece</td>
</tr>
<tr>
<td>Bill Dolan</td>
<td>Microsoft Research, USA</td>
</tr>
<tr>
<td>László Podhorányi</td>
<td>Vodafone Hungary, Hungary</td>
</tr>
<tr>
<td>Josef van Genabith</td>
<td>DCU, CNGL, Ireland</td>
</tr>
<tr>
<td>Gabor Proszeky</td>
<td>Morphologic, Hungary</td>
</tr>
<tr>
<td>Yota Georgakopolou</td>
<td>European Captioning Institute, UK</td>
</tr>
<tr>
<td>Georg Rehm</td>
<td>DFKI, Germany</td>
</tr>
<tr>
<td>Gregory Grefenstette</td>
<td>Exalead, France</td>
</tr>
<tr>
<td>Mirko Silvestrini</td>
<td>European Union of Associations of Translation Companies (EUATC), Italy</td>
</tr>
<tr>
<td>Jan Hajic</td>
<td>Charles University, Czech Republic</td>
</tr>
<tr>
<td>C.M. Sperberg-McQueen</td>
<td>W3C, Black Mesa Technologies, USA</td>
</tr>
<tr>
<td>Theo Hoffenberg</td>
<td>Softissimo, France</td>
</tr>
<tr>
<td>Daniel Tapias</td>
<td>Sigma Technologies, Spain</td>
</tr>
<tr>
<td>Thomas Hofmann</td>
<td>Google, Switzerland</td>
</tr>
<tr>
<td>Alessandro Tescari</td>
<td>Pervoice, Italy</td>
</tr>
<tr>
<td>Keith Jeffery</td>
<td>European Research Consortium for Informatics and Mathematics, UK</td>
</tr>
<tr>
<td>Rudy Tirry</td>
<td>European Union of Associations of Translation Companies, Lionbridge, Belgium</td>
</tr>
<tr>
<td>Stefan Kreckwitz</td>
<td>Across, Germany</td>
</tr>
<tr>
<td>Hans Uszkoreit</td>
<td>DFKI, Germany</td>
</tr>
<tr>
<td>Claude de Loupy</td>
<td>Syllabs, France</td>
</tr>
<tr>
<td>Andrejs Vasiljevs</td>
<td>Tilde, Latvia</td>
</tr>
<tr>
<td>Elisabeth Maier</td>
<td>CLS Communication, Switzerland</td>
</tr>
<tr>
<td>Michel Vérel</td>
<td>Vecsys, France</td>
</tr>
<tr>
<td>Daniel Marcu</td>
<td>Language Weaver, USA, Romania</td>
</tr>
<tr>
<td>Alex Waibel</td>
<td>Carnegie Mellon University, University of Karlsruhe, USA, Germany</td>
</tr>
</tbody>
</table>
Three Vision Groups

- Vision Group **Translation and Localisation** (technical documentation, official bulletins, GUI localisation, games, services etc.)
  - Target stakeholders: large users of translation services, (machine) translation, software companies, game companies, localisation industry

- Vision Group **Media and Information Services** (audiovisual sector, news, digital libraries, portals, search engines etc.)
  - Target stakeholders: media industries, search engine providers, archives

- Vision Group **Interactive Systems** (mobile assistance, dialogue translation, call centres, etc.)
  - Target stakeholders: mobile software and service providers, telecom industry, call centres

http://www.meta-net.eu
Vision Group Meetings

- **Vision Group Translation and Localisation**
  - July 23, 2010, Berlin, Germany
  - September 28, 2010, Brussels, Belgium
  - April 7/8, 2011, Prague, Czech Republic

- **Vision Group Media and Information Services**
  - September 10, 2010, Paris, France
  - October 15, 2010, Barcelona, Spain
  - April 1, 2011, Vienna, Austria

- **Vision Group Interactive Systems**
  - September 10, 2010, Paris, France
  - October 5, 2010, Prague, Czech Republic
  - March 28, 2011, Rotterdam, The Netherlands
Mission of Technology Council

- About 25-35 experts of the European LT and linguistic landscape.
- Tasks of the Council:
  - Discussion of accumulated and condensed findings and proposals.
  - Definition of cross cutting strategies.
  - Prepare recommendations for joint action.
  - Editing and adoption of the Vision Paper.
  - Drafting of the Strategic Research Agenda.
  - Planning the organisational structure needed for keeping the momentum beyond the lifetime of the Network of Excellence.
  - Turning the planning and mobilisation process into a sustainable implementation programme.
- Next slide: members of the META Technology Council
Lines of Action

- SRA Communication Campaign
- Data Liberation Campaign
- Federated Treebanks Initiative
- European Language Technology Service Platform
LT is a Core Data Technology!

European Data Forum
Dublin, 10 April 2013

BY
PHILIPPE WACKER
Secretary General
“Data is the new oil”
Neelie Kroes, 26 March 2013

Do we have a drill?
The Data Challenge

- Europe is content-rich!
- Does Europe have a drill to mine its content?
  - Does the drill work only on English ground?
  - Is Europe a (single) market for content?
  - Where is Europe’s competitive advantage in content/data?
  - Will Europe manage to come out as a global leader of the data revolution?
- Mining data in separate language silos? Reinventing the wheel for each (language-based) data market?
Europe has two cards to play:

- **Multilingual**: the European Language Cloud
  - FREE data, i.e. make it language-transparent, language-independent
  - Digital Single Market for the 1b+ speakers of ALL European languages
  - Leverage multilingualism into global competitive advantage
- **Added-value**: focus on the real needs of economy & society
  - CONSOLIDATE in “verticals” e.g. retail, pharma, tourism, agro, transport, finance, media, etc.
  - … and, yes, by all means, drill the government big data oceans to deliver (open) public data SERVICES!
Language Technology is the Missing Link in the European Digital Agenda!
LT-Innovate is the Forum for Europe's Language Technology Industry. LTi gathers >150 LT suppliers in 24 countries. The Language Technology industry generated an aggregate turnover of €19.3bn in 2011, with a yearly growth rate in excess of 10%.

www.lt-innovate.eu

Twitter: @LTInnovate @LangTechNews

Blog: www.ltinnovate.blogspot.be

LangTechNews Service: www.lt-innovate.eu/ltnews

Discover the world of LT: http://bit.ly/13UAaST

Jochen Hummel
Chairman, CEO ESTeam
jh@lt-innovate.eu

Philippe Wacker
Secretary General
phw@lt-innovate.eu
+32 2 219 0305

THE FORUM FOR EUROPE’S LANGUAGE TECHNOLOGY INDUSTRY
Language Infrastructures

Building Bridges to Foster the Adoption of the Multilingual Web

Hans-Ulrich von Freyberg
Cocomore AG
Bridges between What?

Data providers:
- CMS
- DAM
- PIM
- DB
- CS

LT Applications:
- Summarize
- Extract
- Enrich
- Analyse
- Check
- Translate
- Retrieve
What Kind of Bridges?

Open standards

Standardized interfaces
MultilingualWeb: Starting to Build Bridges

MultilingualWeb has defined a localization metadata standard. It also provides reference implementations that show the standard in action.

MultilingualWeb brings key players together: content producers, localization industry, data world and language technology research & providers.

MultilingualWeb has reached out to the Web community at large, with W3C as the home of standardization efforts.

MultilingualWeb receives funding by the European Commission (project name MultilingualWeb-LT (LT-Web)) through the Seventh Framework Programme (FP7) in the area of Language Technologies. Grant Agreement No. 287815.
Economic Benefits of LT Infrastructures

- **Enabling business**
  - Retail
  - Export
  - Media

- **Driving productivity and demand**
  - Localization and LSPs
  - LT providers
  - Data technology providers