Vision Group
Media and Information Services
Preliminary Findings

Philippe Wacker
EMF – the Forum of e-Excellence, Belgium

META-FORUM 2010, Brussels
Philippe Wacker is Executive Director of EMF, the Forum of e-Excellence [www.emfs.eu](http://www.emfs.eu)

EMF is the European cross-stakeholders' network promoting excellence in the digital economy. Around the core group of winners of the European Seal of e-Excellence, EMF gathers a broad group of stakeholders in the digital economy: associations and clusters, large companies, research organisations, public entities, investors, individual advisors, etc.

EMF members are the drivers of the digital economy. Their success is based on a careful blending of innovation and market needs!
Fields: audiovisual sector, news services, digital libraries, portals and vortals, search engines, social networks etc.

Stakeholders: audiovisual and media industries, web and search engine providers, archives, etc.

Technologies: speech processing, subtitling, text simplification, media mining, multilingual data processing, topic identification, content classification and structuring, information extraction, summarisation, multidimensional analytics, authoring, LT-enabled digitisation, mono/multilingual/multimedia search, semantic search, unregulated language and discourse processing, etc.

Organizers
- Stelios Piperidis (ILSP, Greece)
- Margaretha Mazura (EMF, Belgium)

Meetings
1. Paris, 10 September 2010
2. Barcelona, 15 October 2010
Goals:

- Focus on LT enabled Media and Information Services
- Language & content processing technologies as communication & information facilitators
- Requirements that Media and Information related sectors have from LT and the role LT will play.

A clear view on LT in Media and Information Services will:

- Highlight their importance in a sector which is becoming more and more important from a financial point of view,
- Trigger financial and strategic support (by the EC and the EP, among others), to give Europe the chance of becoming the global player.
Elements of a Vision Statement
Premises

- Media and information services are highly language dependent in a multilingual Europe (63 languages!)
- Millions of hours of archives have been accumulated and are being conserved at high cost – these constitute a gigantic information and knowledge resource
- A multi-polar world implies that no single language will be dominant – more and more content is produced in different languages (e.g. Chinese, Korean, Russian)
- Most European media and information service providers remain monolingual or face substantial hurdles in developing truly multilingual strategies
- The pool of qualified (multi)linguists is shrinking while the e-skills gap is broadening
- Access to (relevant) information and knowledge remains (often) conditional on translation
- The potential of language technologies remains largely untapped
Vision statement

- A sustained investment in R+D+I in language technologies will leverage European media and information services and contribute to creating millions of new jobs.

- The systematic build-up and reinforcement of the European language technology industry by a well-targeted deployment of leading edge R&D capabilities throughout the various specific product, service and process value-chains will truly empower the emergence of a multilingual content and service economy, generate societal benefits, engender export opportunities, etc.

- The European institutions are uniquely placed to spearhead this opportunity as the Member States remain mostly concerned with the promotion of their respective national languages.

- Language technologies are an essential cornerstone of the knowledge economy.
Domain specific Visions
I need to know now...

- European citizens need to accurately know about X: <who, what, where, when, why, what others say about X, ...>

- Information overload handling and Social stream mining
  - An efficient way of exploiting millions of available knowledge bases, social & real time streams is required.

- Life logging
  - Gather information on a massive scale and exploit it when someone is looking for something
  - Technologies able to gather all concepts & associated content and/or knowledge related to every discipline need to be implemented
Improved information navigation and presentation

- Quality, coverage and robustness of text mining technologies, further exploited in search and in information navigation and presentation
  - Robust, wide coverage language analysis (parsing, etc) in all community languages should be pursued.

- Genre-aware LT applications
  - Able to react and behave in a way that is best suited to a given communication situation.
  - Research advances needed in formalizing and incorporating text type knowledge in LT applications.

- Semantic Annotation
  - Extend research from sentences to discourses, from documents to dialogues, from artificial to natural interaction.
  - Include multimodal, multifunctional, interpersonal communication, and cross-modal interaction to facilitate accessibility.
Intelligent media channels speaking our language

- Federated Audiovisual Search
  - Provide intelligent answers to everyday’s questions.
  - Innovative technologies in intelligent ways of recognizing and identifying objects, persons and actions are required.

- Multimedia multilingual subtitling
  - Foster the access to information for ethnic minorities and facilitate information exchange between EU countries.

- Deliver more efficient online advertising
Personal assistants

- Personalized task-centered interactive information assistants
  - Know or adapt to: what the user knows or has already asked for, the user’s language, education and level of expertise.
  - Richer understanding of information seeking should be a priority.

- Robust speech-enabled executive assistants
  - Automatic minutes production from meetings, automatic indexing for voice search, speech transcription and translation for videos.
  - Research advances in Speech Recognition, transcription and synthesis should be pursued.

- Voice Control instead of traditional GUIs
  - Simpler in use, provide trustful services.
  - Need to upgrade old-fashioned, complex-to-use GUIs with new functionalities
Domain specific Needs
Demand and Observations

- Capture of web content [a collection of everything, text, images, videos etc., like archive.org] at a European level
- Improvement of speech technologies
- Advances in Audiovisual Search
- Deeper understanding of text content
- Large-scale text modelling
- Genre and text-type based layers in language analysis and generation.
- Dialogue and Interaction modelling
In a Nutshell: Topics with Visionary Potential

- Domain specific
  - Information agents
    - Enabling citizens to know accurately and instantly
    - Improved information navigation and presentation
  - Intelligent language aware media channels and related services
  - Personal information assistants

- Domain independent (not discussed in this short presentation)
  - Eliminate language barriers for consumers and SMEs.
    - Promote the Language Resources ecosystem.
    - Adopt a common infrastructure to ensure interoperability.
    - Cost-effective porting of LT services and solutions across domains.
    - Enable rapid progress in basic technology for semantic annotation and search.
  - Develop synergies among industry and academia.