META Seal of Recognition and META Prize Award Ceremony – LT Fireworks

Nicoletta Calzolari (CNR) and Georg Rehm (DFKI) on behalf of the META Technology Council

META FORUM 2012  A Strategy for Multilingual Europe
Brussels, Belgium, June 20/21

Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME, CESAR, METANET4U, META-NORD (grant agreements no. 249119, 271022, 270893, 270899).
META Seal of Recognition
META Seal of Recognition

- Awarded to software products and services that actively contribute to the European Multilingual Information Society.
- Awarded by the META Technology Council.
- At META-FORUM 2012 we’re awarding the Seal of Recognition for the second time.
- 22 nominations suggested by META, META-NET Vision Group and META Technology Council members.
- 11 META Seals of Recognition (voting in early June).

http://www.meta-net.eu
META Technology Council

- Ca. 30 experts of the European LT landscape.
- Discusses and reviews META-NET findings & proposals.
- Defines cross cutting strategies.
- Prepares recommendations for joint action.
- Prepared the Strategic Research Agenda (SRA).

http://www.meta-net.eu
META Seal of Recognition

- Google Translate
- Personal Translator by LinguaTec
- Loquendo TTS
- Reverso by Softissimo
- Lucy LT by Lucy Software & Services GmbH
- SVOX TTS
- Media Speech by VecSys
- Systran7 by Systran
- TAUS Data Association
- Vocable Karaoke by VecSys

On behalf of the META Technology Council

Hans U. Schleer
Co-ordinator of META-NET
DFKI, Germany

Georg Rehm
Network Manager of META-NET
DFKI, Germany

www.meta-net.eu
Audioma-Box
by PerVoice SpA

THE META TECHNOLOGY COUNCIL PROUDLY AWARD THE META SEAL OF RECOGNITION TO

In recognition of software, products, and services which actively contribute to the European Multilingual Information Society.

Alessandro Tescari

www.meta-net.eu
META – FORUM 2012
META Seal of Recognition Event

Brussels, June 20th, 2012

Pervasive Voice Recognition
Pervoice: a new experience in ASR

- Pervoice brings to the market a new experience in Automatic Speech Recognition, based on the latest technological developments
  - Large Vocabularies (> 300K)
  - Speaker Independent
  - Spontaneous Speech Recognition
Technology

- Internationally recognized quality Speech Recognition engine from FBK
- Handles multiple languages for specific sectors
- Implemented with multichannel recognition, localization, personalization and automatic punctuation
• Remote Transcription: www.Flyscribe.it

• Transcription Workflow: PASS (PerVoice AudioText Synchro Suite)

• Transcription Systems:
  – Audioma-Box
  – Audioma-RT

• Subtitling Solutions: PSW (Pervoice Subtitling Workstation)
Future developments

- Wikivoice
  - A platform for self made language models
- Today’s Word
  - LM Automatic Updating System
Any sufficiently advanced technology is indistinguishable from magic

Arthur C. Clarke
The META Technology Council proudly award the META Seal of Recognition to
European Media Laboratory (EML)

In recognition of software, products, and services which actively contribute to the
European Multilingual Information Society.

Siegfried Kunzmann

Hans Uszkoreit
Co-ordinator of META-NET
DFKI, Germany

Georg Rehm
Network Manager of META-NET
DFKI, Germany

www.meta-net.eu
Automatic Transcription of Speech

From Applied Research to the Market

Dr. Siegfried (Jimmy) Kunzmann
Manager R&D
**Objective:** Automatic transcription of voicemails to SMS / e-mail / mobile device

- Register mobile number at “Sipgate One” → [http://sipgate.de/](http://sipgate.de/)
Summary

• Focus of the European Media Laboratory
  - Provide robust transcription technology, tools, platform
  - Partner to expand language portfolio
    - License tools, provide expertise, collaborate
  - Provide EML transcription technology to system integrators
    - Enable on-premise installation und customization

• Good quality server-based transcription requires
  - Access to lots of REAL application data
  - Rapid customization to application domains
  - Broader language coverage

Thank You 😊
Mobile Voice Messaging & Search

- **Objective:** Fast and convenient messaging & search on mobile devices
- **Languages:** German, Italian
- **Messaging:** Dictation, SMS
- **Search:** search, navigation, internet movie database, Wikipedia, …
Extract!

by Textkernel

The META Technology Council proudly award the META Seal of Recognition to

Extract!

by Textkernel

In recognition of software, products, and services which actively contribute to the European Multilingual Information Society.

Jakub Zavrel

www.meta-net.eu

On behalf of the META Technology Council

Hans Uszkoreit
Co-ordinator of META-NET
DFKI, Germany

Georg Rehm
Network Manager of META-NET
DFKI, Germany
Semantic Technology to power the multi-lingual job market

Extract! and other Textkernel HR products

Jakub Zavrel, CEO, Textkernel, <zavrel@textkernel.com>
Textkernel:

- Spinoff from R&D in machine learning and language technology
- Founded 2001, offices in Amsterdam (HQ), Frankfurt, 42 employees; strong R&D focus
- Core technology: understanding unstructured text data.

Market:

- Multi-lingual tools (15 languages) to extract CVs and jobs
- Search! + Match! to connect people and jobs
- Jobfeed: largest real time DB for job market analysis
- Customers: UWV, CBS, Adecco, Randstad, Monster, Stepstone, SAP, Unisys, Axa, Philips, etc. (250 direct, 2000+ indirect)
Barriers in Recruitment

Candidate: Why do I have to jump through so many hoops when I apply for a job?

Recruiter: Why is it so hard to search, attract and capture the right candidates?

Organisation: Am I using the full potential of my candidate database and other sources? How can I get more out of my marketing budget?
Textkernel reverses the traditional application process: Extract information from the CV, pre-populate the application form.

- Save up to 90% time in processing CV’s into recruitment systems.
- Easy aggregation of searchable information without the data entry.
Roadmap for next years

• Sell Extract, Jobfeed and Search & Match across Europe
• Continue 30% revenue growth

• Match people to jobs on a large scale based on a deep understanding of skills, experience, preferences, and culture
• Learn from feedback

• Make the European multi-lingual labor market transparent for organizations and job seekers
IPTranslator
by PLuTO

The META Technology Council proudly award the META Seal of Recognition to IPTranslator by PLuTO

In recognition of software, products, and services which actively contribute to the European Multilingual Information Society.

Heidi Depraetere
(on behalf of Páraic Sheridan)

www.meta-net.eu
PLuTO Project - IPTranslator.com

Heidi Depraetere

META-FORUM 2012, Brussels, June 20

Funded under the EU ICT Policy Support Programme
PLuTO Project

- **Objective:** develop an online solution for patent translation
- **Term:** 3 year funded project: April 2010 → April 2013
- **Languages:** EN-FR, EN-PT, EN-DE, EN-ES, EN-RU, EN-JA, EN-ZH
- **Partners:**
  - Dublin City University, Ireland
  - ES Team, Sweden
  - CrossLang, Belgium
  - Dutch Patent User Group, The Netherlands
**REVIENDICACIONES**
1. Utilización de uno o más compuestos seleccionados del ácido elágico, sus ésteres, los sales de ácido elágico y de sus ésteres, y sus mezclas, como agente anti-pellicular.

2. Utilización según la revendación 1 para tratar las patologías asociadas a la proliferación de levaduras del género Malassezia sobre el cuero cabelludo.

3. Utilización según la revendación 1 o 2, caracterizada en que los sales son

**Claims**
1. Use of one or more compounds selected from the group consisting of elagic acid, its esters, salts of elagic acid and its esters, and mixtures thereof, as anti-dandruff agent.

2. Use according to claim 1 for treating skin conditions associated with proliferation of yeasts of the genus Malassezia on the scalp.

3. Use according to claim 1 or 2, characterized in that the salt or salts are selected from salts of elagic acid or elagic acid ester mixtures.
Future...

Translation Focus ➔

Patent Workbench Focus ➔

IPTranslator

paronpro
Title:
PEDAL SIMULATOR

Abstract:
A pedal simulator includes a cylinder mechanism for absorbing a fluid pressure generated in a pressurization chamber of a master cylinder by operating a brake pedal. The cylinder mechanism includes a housing, a piston slidably held in the housing and defining a fluid chamber for introducing a working fluid thereinto, and a plurality of spring members for urging the piston...

Claims:
What is claimed is:
1. A cylinder mechanism comprising:
a housing;
a piston slidably held in said housing and defining a fluid chamber for introducing a working fluid thereinto;
a plurality of disc springs for urging said piston toward said fluid chamber, at least two of said disc springs forming a pair and the bottom ends facing each other; and

2. The cylinder mechanism of claim 1, wherein the fluid chamber is a cylinder disposed in said housing and having a bore for said piston slidably held in said housing.
IRSTLM Toolkit by FBK

The META Technology Council proudly award the META Seal of Recognition to

IRSTLM Toolkit by FBK

In recognition of software, products, and services which actively contribute to the European Multilingual Information Society.

Bernardo Magnini
(on behalf of Marcello Federico)

On behalf of the META Technology Council

Hans Uszkoreit
Co-ordinator of META-NET
DFKI, Germany

www.meta-net.eu

Georg Rehm
Network Manager of META-NET
DFKI, Germany
IRSTLM

A toolkit for the estimation, representation, and computation of statistical language models

Features

- advanced smoothing methods for large LMs
- approximate smoothing for gigantic LMs
- pruning and quantization of LMs
- efficiently storing LMs on disk and in memory
- several LM adaptation methods
- compatible with LMs created with other tools
- integrated inside MOSES and FBK’s ASR
- open source LGPL license:

  http://sourceforge.net/projects/irstlm
Evolution

1992
- Command and control of a robot (1K words)
1995
- Dictation of medical reports (10Kw)
1997
- Spoken data entry (10Kw, recursive LMs)
2000
- Broadcast news transcription (60Kw, 5M 3-grams)
2004
- News translation (2Mw, 200M 4-grams)
2006
- Translation of speeches (4.5Mw, 350M 5-grams)
2008
- NIST MT task (6G 6-grams, on-disk creation)
2011
- Translation of TED talks (multiple, hybrid LMs)

IRSTLM
• more than 10,000 downloads
• many from companies
  Apptek, Applied Language Solutions, Translated, Pervoice,
  Digital Silk Road, Symantec Ltd, Asia Online, …
• and public institutions
  JRC and DGT (EC), World Intellectual Property Organization (UN)
Credits

Marcello Federico
Nicola Bertoldi
Mauro Cettolo

Chris Dyer
Fabio Brugnara
Arianna Bisazza
Christian Hardmeier
Hieu Hoang
Language Weaver

In recognition of software, products, and services which actively contribute to the European Multilingual Information Society.

Radu Soricuț

On behalf of the META Technology Council

Hans Uszkoreit
Co-ordinator of META-NET
DFKI, Germany

Georg Rehm
Network Manager of META-NET
DFKI, Germany

www.meta-net.eu
Ecosystems of (MT) Value

Radu Soricut, Senior Research Scientist, SDL
SDL BeGlobal: Easy to Start, Easy to Learn, Easy to Grow

- **Access** over 80+ direct language pairs
- **Secure** your intellectual property within the corporate enterprise using SDL BeGlobal
- **Integrate** with your internal systems and department-level projects with our API
- **Manage and control** corporate terms and brandings
- Leverage **reporting** to understand unknown translation “hotspots”
- **Perform ad-hoc translation** through SDL BeGlobal’s Desktop and our NEW SDL BeGlobal Translator
- Leverage SDL BeGlobal’s expanding library of **pre-integrated offerings**
- **Expand** your use for chat, knowledge base, “ad-hoc” translations, email and anything else you can think of

Includes:
- Access to 80+ Direct Language Pairs
- SDL GlobalConnect
- SDL BeGlobal Online
- SDL BeGlobal Translator
- SDL BeGlobal Desktop Widget
- SDL Brand and Terminology Management
- SDL BeGlobal REST API
- TrustScore™
- Touchpoints
- Reporting
- Account Management
SDL BeGlobal: Continuous Improvements

13,747 MT engines built in 2011 by the SDL customer delivery team

Example: Use MT to support delivery of HT quality across 30 languages via Post-Editing

- Brazilian Portuguese: 81%
- Danish 94%
- Int. Spanish 98%
- Simplified Chinese 60%
- Spanish 80%
- Finish 62%
- Japanese: 59%
- Hungarian: 28%
- German 75%
- Dutch 78%
- Turkish: 34%
- Indonesian 67%
- Italian 76%
- Norwegian 98%
- Traditional Chinese 58%
- Arabic: 41%
- Korean: 65%
- Polish: 78%
- Russian: 78%
- Thai: 54%
- Hebrew: 54%
- Swedish 90%
- French 96%
- Greek 85%
- Romanian 71%
- Korean: 65%
- Czech: 56%
- Hebrew: 54%
It’s time to change the conversation

Instead of talking about technology components, we need to talk about ecosystems of value. That is what benefits customers and end users the most.

- Is the MT system usable as is? How easy is it to plug it into my enterprise ecosystem?
- Does the MT system deliver on my immediate business goals?
- How easy is to upgrade the MT system and have it take advantage of the full enterprise ecosystem in which it operates? Is the MT ecosystem enabled globally?
- How can we make MT relevant to the broadest group of end- and power-users in the enterprise?
PHONEXIA s.r.o.

In recognition of software, products, and services which actively contribute to the European Multilingual Information Society.

Radim Kudla

www.meta-net.eu
Goal:
Help clients to extract automatically maximum of valuable information from spoken speech.

Focus: speech technologies
• Speaker ID, Gender ID, Language ID...
• Speech transcription, Keyword spotting...
• Custom R&D

Since 2006, tight collaboration with Brno University of Technology (BUT Speech@FIT):
– 20 researches moves the technology forward
– 15 people for development, sales, marketing and technical support
Example: Multilingual speech transcription and keyword spotting system
1) How to index and search in large quantity of multimodal data
speech transcription, speaker id, gender id, language id, age ...

2) How to get the relevant information
There is more information spoken, but also more redundant and useless information.

3) How to integrate with approaches developed for text
THE META TECHNOLOGY COUNCIL PROUDLY AWARDS THE META SEAL OF RECOGNITION TO
Asia Online

In recognition of software, products, and services which actively contribute to the European Multilingual Information Society.

Kirti Vashee, Dion Wiggins
(video)

www.meta-net.eu
Voxsigma
Software Suite by
Vocapia Research

The META Technology Council proudly award the META Seal of Recognition to

Voxsigma Software Suite by Vocapia Research

In recognition of software, products, and services which actively contribute to the European Multilingual Information Society.

Joseph Mariani
(on behalf of Bernard Prouts)

www.meta-net.eu

On behalf of the META Technology Council

Hans Uszkoreit
Co-ordinator of META-NET
DFKI, Germany

Georg Rehm
Network Manager of META-NET
DFKI, Germany
Presentation of Vocapia Research & the VoxSigma® Software Suite

prouts@vocapia.com

June 20/21, 2012
Who Are We?

- Vocapia Research is a R&D company and a software publisher in the speech and language technology field
- Founded in 2000, licensed CNRS-LIMSI technology in 2002
- Privileged partnership with LIMSI, a French CNRS laboratory
- Participation in national and international research projects
- Located in the scientific pole of the Saclay Plateau, France
VoxSigma® Software Suite

- Large vocabulary speech-to-text capabilities in multiple languages (Arabic, Dutch, English, French, German, Greek, Italian, Mandarin, Polish, Portuguese, Russian and Spanish)
- Audio segmentation and partitioning, speaker identification, language recognition audio/text alignment
- Designed for professional users needing to transcribe large quantities of audio and video documents such as broadcast data or conversational telephone speech data
- Tailored solutions to match customer needs
- Transcribed with LIMSI TrecVID data (since 2002) and MediaEval (since 2010)
- Ranked first in STT evaluations: Ester (French), NBEST’08 (Dutch)
- Most popular applications: media monitoring, teleconference transcripts, speech analytics, security applications, automatic subtitling
Perspectives

- Retain close ties with research
- Continue state-of-the-art research and development
- Extend SaaS offer (support more languages, more tasks)
- Cover all European languages by end of 2013
- Increase our European partnership (such as via European research projects)
KantaMT
by Xcelerator.ie

The META Technology Council proudly award the META Seal of Recognition to KantaMT by Xcelerator.ie.

In recognition of software, products, and services which actively contribute to the European Multilingual Information Society.

Tony O’Dowd (video)

www.meta-net.eu

On behalf of the META Technology Council

Hans Uszkoreit
Co-ordinator of META-NET
DFKI, Germany

Georg Rehm
Network Manager of META-NET
DFKI, Germany
The META Technology Council proudly award the META Seal of Recognition to XFST by Xerox

In recognition of software, products, and services which actively contribute to the European Multilingual Information Society.

Lauri Karttunen (video)

On behalf of the META Technology Council

Hans Uszkoreit
Co-ordinator of META-NET
DFKI, Germany

On behalf of the META Technology Council

Georg Rehm
Network Manager of META-NET
DFKI, Germany

www.meta-net.eu
XFST
Xerox Finite-State Toolkit

Lauri Karttunen
with
Ronald M. Kaplan, Todd Yampol, Pasi Tapanainen,
Kenneth R. Beesley, Hervé Poirier, André Kempe,
Tamás Gaál, and many others
1993 – present and future
What is XFST?

• A development tool for compiling finite-state networks from
  - strings
  - text files
  - xfst regular expressions
  with calculus operations (union, intersection, composition, …)

• A utility for applying finite state networks to text for
  - rewriting
  - tokenization
  - morphological analysis and generation
  - pattern matching
What is XFST used for?

• Since its inception in 1993 XFST has been used to create lingware for dozens of languages
  English, Finnish, French, German, Italian, Spanish, …
  Arabic, Basque, Georgian, Irish, Turkish, …
  Mongolian, Rotokas, Sámi, Zulu, Xhosa, …

• Industrial users of XFST and its commercial version include
  Temis
  Basis Technology
  SAP
  Microsoft
Future of XFST

• In the near future Xerox will release the source code of XFST and its enhanced industrial version freely for non-commercial use under a click-through license.

• The current executable is available at

• Other resources:
  Finite State Morphology
  Kenneth R. Beesley and Lauri Karttunen
  CSLI Publications, 2003
  http://www.fsmbook.com/
META Prize

- Awarded to outstanding products or services supporting the European Multilingual Information Society.

- At META-FORUM 2012 we’re awarding the Prize for the second time.
META Prize: Winners 2011

- 1st META Prize
  - **Moses** – For outstanding and innovative contributions to statistical machine translation.

- 2nd META Prize:
  - **Jibbigo Mobile Translators, Jibbigo** – For outstanding mobile voice translators bringing speech translation to mobile devices.
  - **Voxalead, Exalead** – For outstanding audio visual search and transcription software and services.
  - **ELRA (European Language Resources Association)** – For outstanding long term commitment to the preparation and distribution of language resources and technologies.
  - **LDC (Linguistic Data Consortium)** – For outstanding long term commitment to the preparation and distribution of language resources and technologies.

http://www.meta-net.eu
The Technology Council settled upon a balance between:

- Machine Translation
- Interactive Systems
- Media and Information Services

- Research vs. industry
- Europe vs. global players
- Open source vs. commercial software
META Prize – 2012

- Awarded to outstanding products or services supporting the European Multilingual Information Society.
- 19 nominations.
- And one clear winner.
- One META Prize.

http://www.meta-net.eu
JRC Optima Activity
Language Technology

For outstanding use of Language Technology to fight information overflow and to overcome language barriers.

The META Technology Council recognises that this work is an important cornerstone on the way to enabling the European Multilingual Information Society.

On behalf of the META Technology Council

Hans Uszkorct
Co-ordinator of META-NET
DFKI, Germany

Georg Rehm
Network Manager of META-NET
DFKI, Germany

www.meta-net.eu
Joint Research Centre

The European Commission’s in-house science service

www.jrc.ec.europa.eu

Serving society
Stimulating innovation
Supporting legislation
Europe Media Monitor (EMM)

Erik van der Goot & Optima team
Some EMM statistics

- Gathers approximately **150,000 new news articles per day**
- In ~ **50 languages**
- **From ~ 3750 news portals** world-wide (roughly 10,000 pages/RSS feeds)

- Classifies all news according to **hundreds of subjects and countries**.
- (~1400 category definitions, ~35,000 key word patterns and pattern combinations)
- **24 / 7**: Runs 24 hours per day, 7 days a week.

- Started in 2002
- **Up to 1.5 Million hits per day** for public websites (EMM NewsBrief, MediSys, NewsExplorer)
- But more importantly, 25,000 visitors/day on the public sites

- Developed, built, maintained and run by the JRC.

19 June 2012
The importance of multilingual information gathering
Various **multilingual parallel corpora** in up to 22 languages (23 official EU languages minus Irish), e.g.:
- JRC-Acquis (total of over 1 billion words; available since May 2006);
- DGT-TM (total of 57 million translation units; developed by DGT; since 2007, extended in May 2012);

**Multilingual multi-label categorisation software** JRC EuroVoc indexer:
- Trained for 22 languages (released in May 2012)

**Multilingual named entity resource JRC-Names**:
- 27 different scripts and many more languages;
  Up to 413 spellings for a single name;
- Collected by analysing up to 100,000 news articles in up to 20 languages since 2004; partially augmented through Wikipedia.
Muammar Gaddafi

19 June 2012
Congratulations to all winners!