Mining and Understanding of multilingual content for Intelligent Sentiment Enriched context and Social Oriented interpretation

Stefanos Vrochidis
Deputy Project Coordinator
Scientific Manager
Information Technologies Institute Centre for Research and Technology Hellas
MULTISENSOR in fp7

• ICT Call 10 (FP7-ICT-2013-10)
  • Challenge 4: Technologies for Digital Content and Languages
    • Objective ICT-2013.4.1(a) Content analytics and language technologies

• MULTISENSOR
  • STREP
  • 9 partners from Europe (Greece, Spain, Germany, Belgium, Bulgaria)
  • Coordinated by Information Technologies Institute, Centre for Research and Technology Hellas (CERTH)
  • Budget ~ 4,2Mil Euros and EC funding ~ 2,95Mil Euros
  • Expected starting date: November 2013
  • Duration: 3 Years
MULTISENSOR for Society

- Current Situation
  - Great increase in multilingual and multimedia content
  - Consumption of large amounts of unreliable information
  - Linguistically and geographically encapsulated areas
- Financial Crisis is a characteristic example
  - Contradictory and unreliable information causes:
    - Nervousness of politics
    - Insecurity in the population
    - Unstable ground for international investments

- There is a need for:
  - Unified access to multilingual and multicultural economic and news story material across borders
  - Sentiment, context-aware, spatiotemporal and semantic interpretation
  - Semantic integration of content and extraction of hidden meanings
  - User tailored summarisation

Wroclaw, October 1, 2013
MULTISENSOR concept

- MULTISENSOR envisages
  - Multidimensional Semantic Content Integration
  - To support
    - International Media monitoring
    - SME Internationalisation
MULTISENSOR Scientific Objectives

SO1. Mining and content distillation of heterogeneous multimedia and multilingual data
  • Speech recognition, machine translation, name entity extraction
  • Text analysis, concept extraction and linking
  • Multimedia concept and event detection

SO2. User- and context-centric analysis of heterogeneous content
  • Context analysis and representation
  • Sentiment extraction
  • Social media analysis, community detection, information propagation

SO3. Multidimensional content integration and retrieval
  • Topic-based classification and topic/event detection
  • Semantic content integration
  • Multimodal Indexing and Retrieval

SO4. Semantic reasoning and intelligent decision support services
  • Semantic representation, ontology framework
  • Hybrid reasoning (multi-thread reasoning, parallel inferencing, spatiotemporal reasoning.
  • Reasoning-based decision support

SO5. Multilingual summarisation and presentation of the information to the user
  • Concept-driven deep summarisation
  • Multilingual surface-oriented summarisation
MULTISENSOR Use Cases

Pilot Use Case 1: International media monitoring

- Journalists need to master continuous heterogeneous data-streams (news agencies, web portals, blogs, social networks)
- Professional clients of media monitoring portals require direct access to comprehensive, targeted, business and consumer information
  - Support journalists and media monitoring companies in retrieving, integrating and summarising heterogeneous information

Pilot Use Case 2: SME Internationalisation

- SMEs consider internationalisation as the main way out of this crisis
- SME internationalisation requires to retrieve spending habits of consumers, economic fundamentals of the countries, consumer behaviour, etc.
  - Support SMEs to internationalise
MULTISENSOR Architecture

Semantics Reasoning (WP5)

- Summarization and Content Delivery (WP6)
- Intelligent Decision Support (WP5)
- Multimodal Retrieval (WP4)

Knowledge Repositories and Databases (WP5, WP7)

Content Alignment and Integration (WP4)

User-centric Content Analysis (WP3)

Data Mining and Content Extraction (WP2)

Ontologies, Vocabularies

Middle-tiers

Front-ends

Delivery Interfaces

Business Logic

Content Layer

Data Layer

Wroclaw, October 1, 2013
MULTISENSOR Consortium

Centre for Research and Technology Hellas (CERTH)
- Project Coordinator (Ioannis Kompatsiaris)
- Scientific Manager (Stefanos Vrochidis)
- Multimedia analysis and retrieval, Semantic content integration

Universitat Pompeu Fabra (UPF)
- Summarisation, text generation, concept extraction from text

Fundacio Barcelona Media – Yahoo! Research (BM-Y!)
- Social media analysis, context and sentiment extraction

LINGUATEC (LT)
- Speech recognition, machine translation, name entity extraction
MULTISENSOR Consortium

**Everis**
- Technical Manager (Enric Staromiejski Torregrosa)
- System development and integration

**Pressrelations (PR)**
- Media monitoring company, content provider
- Use case creation, evaluation, system development, exploitation

**Ontotext**
- Semantic representation, reasoning, decision support (WP5)

**Deutsche Welle (DW)**
- Content provider, user case creation, evaluation, dissemination

**PIMEC**
- Organisation of SMEs
- Use case creation, evaluation, dissemination, exploitation

Wroclaw, October 1, 2013
MULTISENSOR Impact

- Impact on Enterprises and especially on SMEs
  - High SME participation including PIMEC
  - Technology take up by SMEs and larger enterprises
  - Services to be used by SMEs (internationalisation support)

- Improvements on Language Technologies
  - Advancements in speech recognition, concept extraction and linking
  - Indirect advancements in MT by providing deep semantic speech and text analysis
  - Advancements in text generation and summarisation techniques
  - Application of Language Technologies for content aggregation, sentiment and context extraction

- Impact on Language Technologies Market
  - Bring closer LT and multimedia analysis domains
  - Open source tools
  - Novel techniques published
  - Freeware tools
  - Commercial Tools
  - New and improved services using LT (SME internationalization tools, media monitoring, journalism)
MULTISENSOR & LT community

• MULTISENSOR partners are active members of the LT community (Linguatec, UPF (TALN group))

• MULTISENSOR is interested in:
  • Following the advancements in LT domain
  • Sharing developed techniques and resources during and after the project
  • Closely cooperating with projects in the area that have similar objectives (exchange information, organize and participate in common events)
  • Reusing and advancing the outcome of LT projects that fit in the LT tasks of the workplan
  • Participating in LT events to disseminate results and establish cooperations
Thank you!

Stefanos Vrochidis
MULTISENSOR
Deputy Project Coordinator
Scientific Manager
stefanos@iti.gr