MARCH 6, 2014

Future@SystemX

Multimedia and Multilingual Analysis for Open Sources

Demonstration of MediaCentric Solution

Olivier Mesnard and Nabil Bouzerna

Project Objectives

- **To develop an experimental platform for the analysis of unstructured content (text, audio / video)**
  - Integration of the software components provided by project partners for the main sectors of language automatic processing industry (transcription, translation, information extraction, search...)
  - Development of innovative applications prototypes for information retrieval and open sources monitoring based on those components

- **To take into account users needs uncovered by existing applications**
  - Extension to social networks data
  - Extension to multilingual data
  - Reduction of costs and delays to adapt processings to a new domain or a new language

Work to be done

- **At software component level**
  - Improve the robustness of processing on noisy data (amateur video, blogs, NFWC…)
  - Integrate new languages
    - Define a process for the limited cost (time and resources) development of linguistic resources, based on learning from targeted corpus.
- **At application level**
  - Interconnect components, share a metadata repository or indexes
- **Deployment and infrastructure**
  - Anticipate scaling (from a few hundreds of thousands of documents to hundreds of millions of documents) and establish an architecture and deployment strategy that facilitates the scaling.
- **Immediacy**
  - Process data in constrained time (indexes refreshing strategy)
- **User interface and interaction**
  - Innovation in visualization and interactivity (also to address the need for scalability)

Demonstration of MediaCentric Solution