

LIRICS

Linguistic Infrastructure for Interoperable Resources and Systems

- ▶ **Kick off meeting presentation**
- ▶ **Proposal N° 22.236**
- ▶ **Presented by Laurent Romary
(INRIA, France, chair of ISO-TC37/SC4)**

Expected results of the project

- Europe being a mosaic of languages, the processing of multilingual linguistic data concerns a lot of people in Europe
- And these data need to be improved, extended, processed, merged, used and re-used
- **But today there exists no established standard to enable interoperability and re-use of multilingual data**
- For this reason, these tasks form a timely and costly part of daily work of Europe's industry

Expected results of the project

To lower this cost, LIRICS will:

- Provide Europe with a set of industry validated standards for language resource management ratified within the project lifetime
- Facilitate the acceptance of these standards by providing an open-source reference implementation platform, related web services and test suites
- Gain full industry support and input to the standards development via the Industry Advisory group and demonstration workshops
- Provide a pay-per-use business model for use by industry validated during the project

Linguistic Infrastructure created

- Standards following the ISO-TC37/SC4 milestones
The deliverables will be direct inputs to the ISO ballots
 - a Lexical Markup Framework (ISO 24613) committee draft will be submitted to ISO ballot at M18. And DIS ballot at M27.
 - improvement and testing of Morpho-syntactic Markup Framework WD (ISO 24611)
 - NWIP for a syntactic annotation model
- ISO Data Category Registry (ISO 12620) strengthened with descriptions in morpho-syntax and semantics to be endorsed by ISO Thematic Domain Committee
- LIRICS web site and mailing list

Users & Benefits for users

- **Market:**
 - translation
 - localised tools
 - linguistic data based solutions
- **Users:** NLP solution providers, lexicon publishers and multilingual end users
- **Benefits for users:** better interoperability, lower cost of operation, more languages available, better approach to asian & semitic languages

Business Case

- Reference implementation platform
 - all LIRICS defined ISO standards will be defined on the basis of web services in order to support distributed NLP resources
 - support « try before you buy » paradigm which enables NLP companies to give temporary access and charge on per-usage basis
 - provide open-source reference implementation of wrappers for lexicons, morphological analysers, syntactic parsers and semantic annotators

Participants & roles: consortium

The LIRICS consortium bring together leading experts in the field of Natural Language Processing via participation in ISO committees

- INRIA (F) specialist in standardisation
- DFKI (D) sp. in morpho-syntax & syntax processing
- USFD (UK) provider of the GATE open source platform
- CNR-ILC (I) sp. in language resources
- UW (A) sp. in terminology management & language codes
- Util (NL) sp. in computational semantics
- MPI (D) sp. in meta-data
- Unis (UK) sp. in language resources
- IULA-UPF (E) sp. in lexicons & grammars

Participants & roles: Ind. Ad. Gr.

- For the standards to have impacts, LIRICS will ensure their usability by consulting with a group of industrial users
- The Industry advisory group will be consulted to identify priorities and requirements
- At present, 21 members:
 - ▶ NLP solution providers like Systran, Sinequa, Temis or Morphologic
 - ▶ Lexicon publishers like Longman-Pearson
 - ▶ End users like EADS-CCR, British Telecom, Telefonica Invest-Des. or HP
- Membership will be expanded