IULA Web services for text data

Núria Bel
IULA - TRL
Universitat Pompeu Fabra

INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA - Barcelona

Institut Universitari de Lingüística Aplicada
Technologies of Language Resources Group - TRL
Universitat Pompeu Fabra (Barcelona)

- IULA, created in 1994, is a research and research-training centre with researchers from Departments of Translation and Language Sciences, Information Technologies and Communication.
- TRL Group, started in 2006, is working on technologies for the production of Language Resources.
- Web services deployed for **enabling** (internal and external) **researchers** (academia & industry) **to exploit text data**

Currently available:
40 IULA Web Services for text data

Types of services:

1. Text cleaning and normalization:
   - Prepare data for exploitation (conversion, cleaning, anonymization)
2. Quantitative information and statistical analysis:
   - Analytic figures on words in texts, and about their relevance
3. Annotation / Enrichment services:
   - Adding semantics: Named Entity Recognition, Semantic classification, Relations (who does what).

IULA web services include own and 3rd party tools freely available (FreeLing, MALT, ..) but are offered as:

Our users are not interested in the backstage: researchers in Humanities and Social Sciences, Language Technologies, etc.

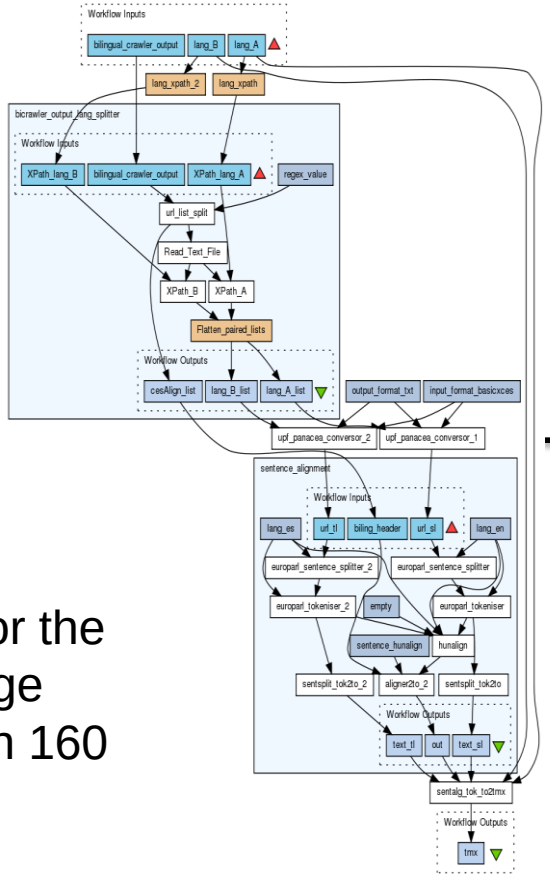
- Single interaction point for manipulating text data
- Simplified & based on “one operation only” criterion
- No installation, no maintenance, no machine dependencies for users
- WS can be combined in workflows: we can provide quick solutions.
- WS can be used as trustable experiment independent tools and to guarantee experiment replication

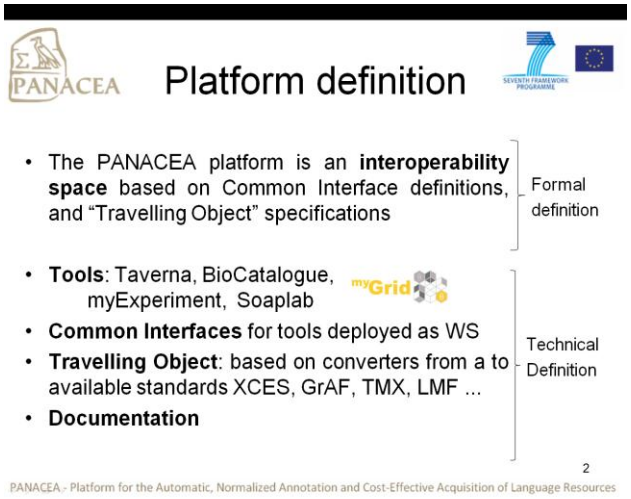
Web Services are results of participation in different projects:

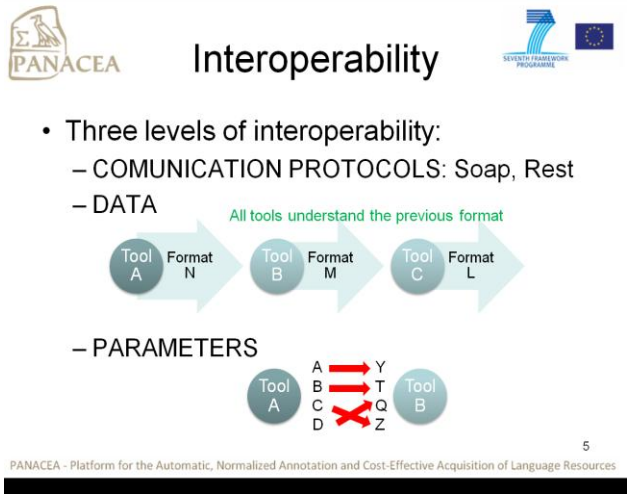
CLARIN
Common Language Resources and Technology Infrastructure

METANET4U

PANACEA





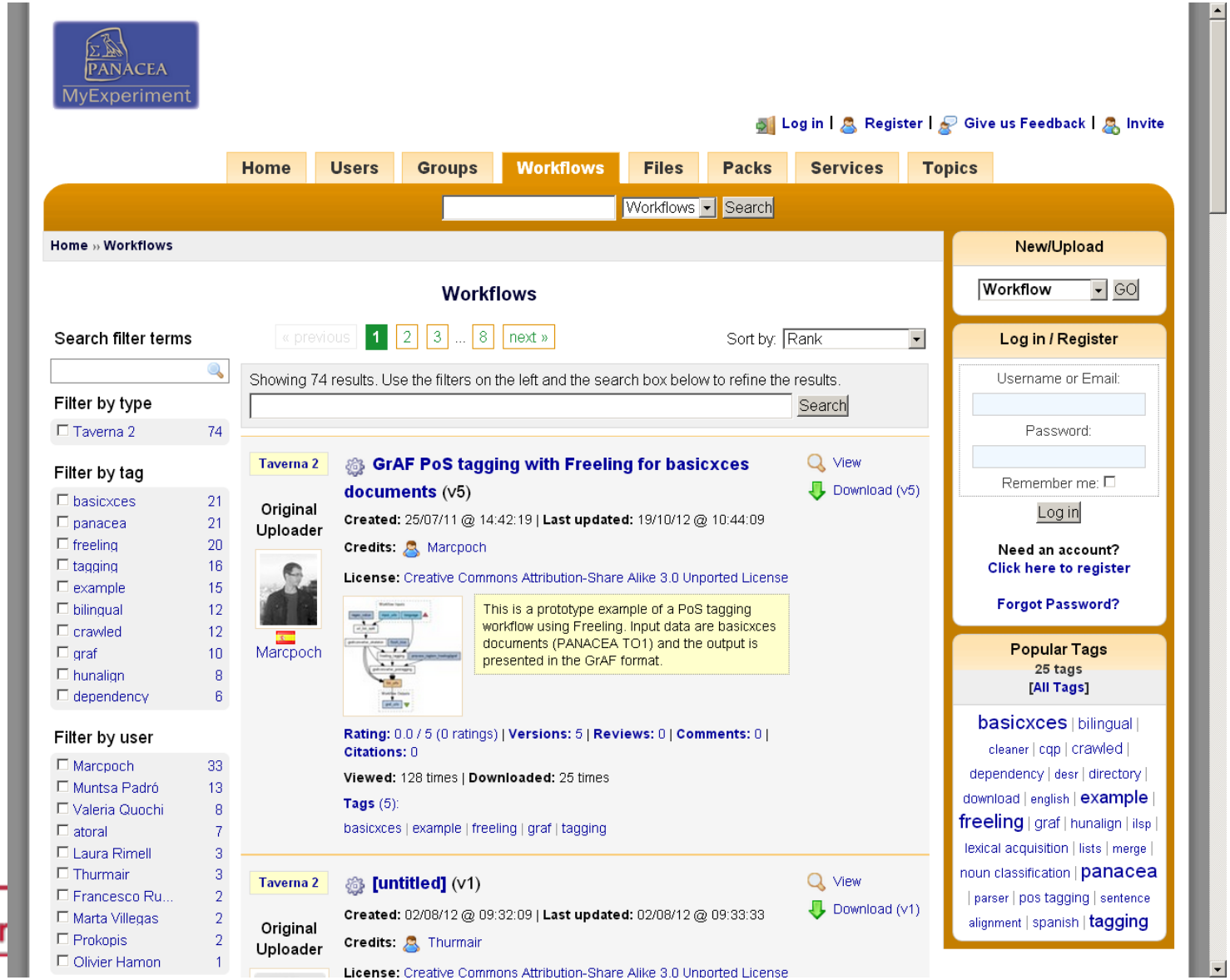


Distributed platform for the production of Language Resources (more than 160 WS from all partners)

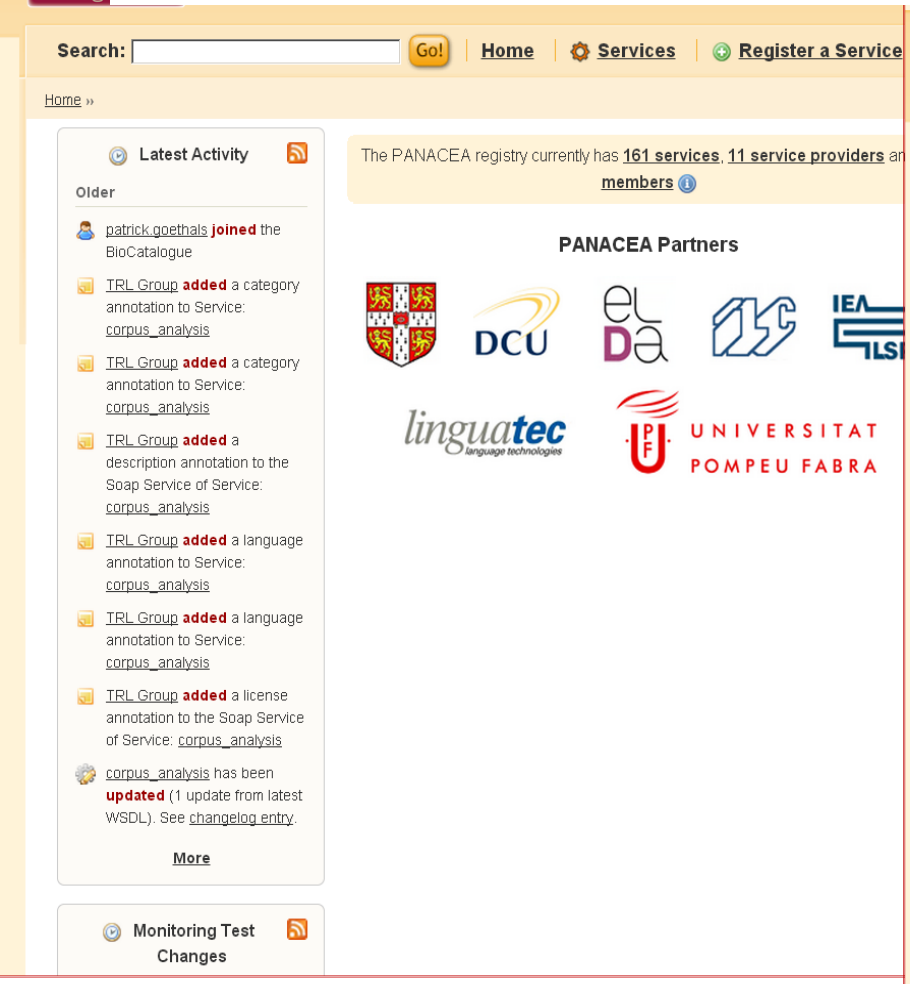
upf. iula gruptrl

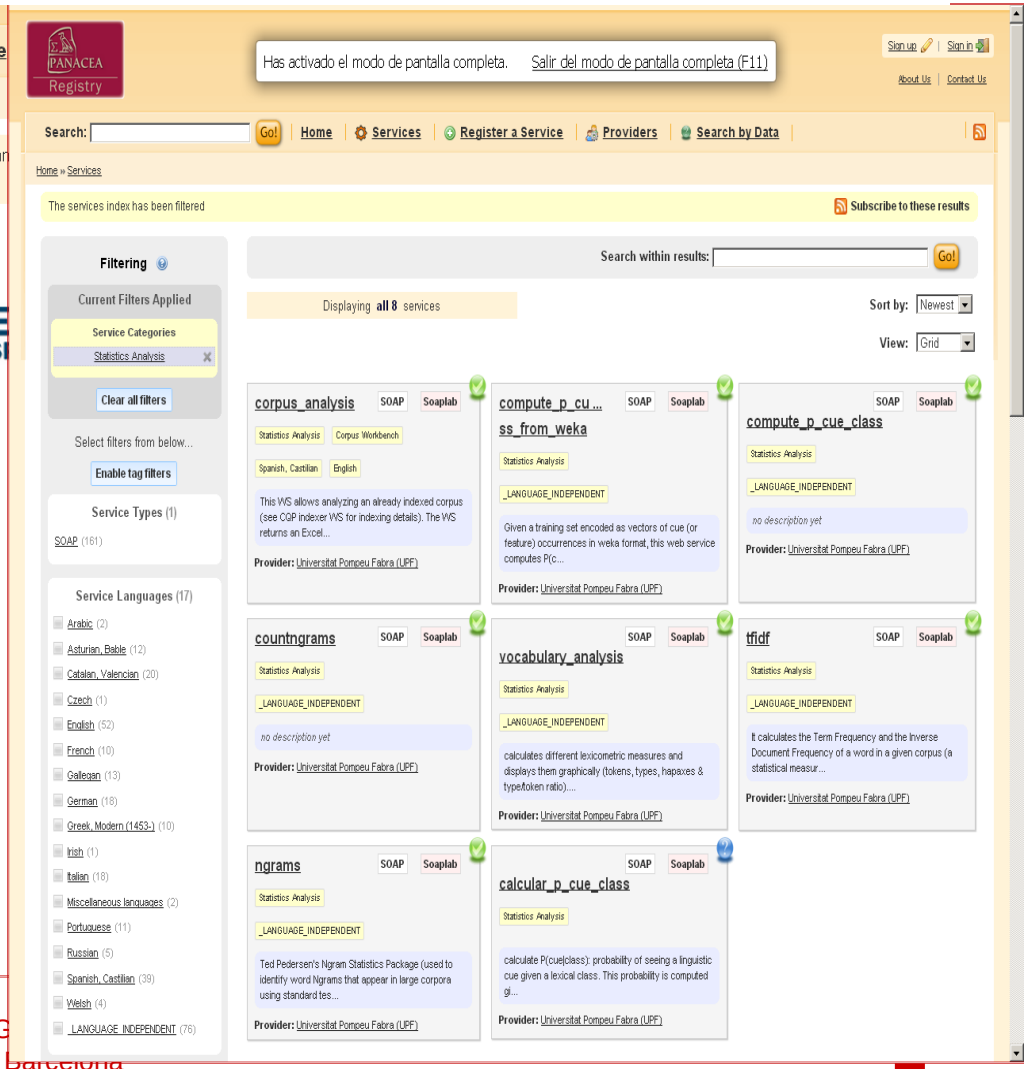
INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA - Barcelona

About 70 workflows in myexperiment.elda.org

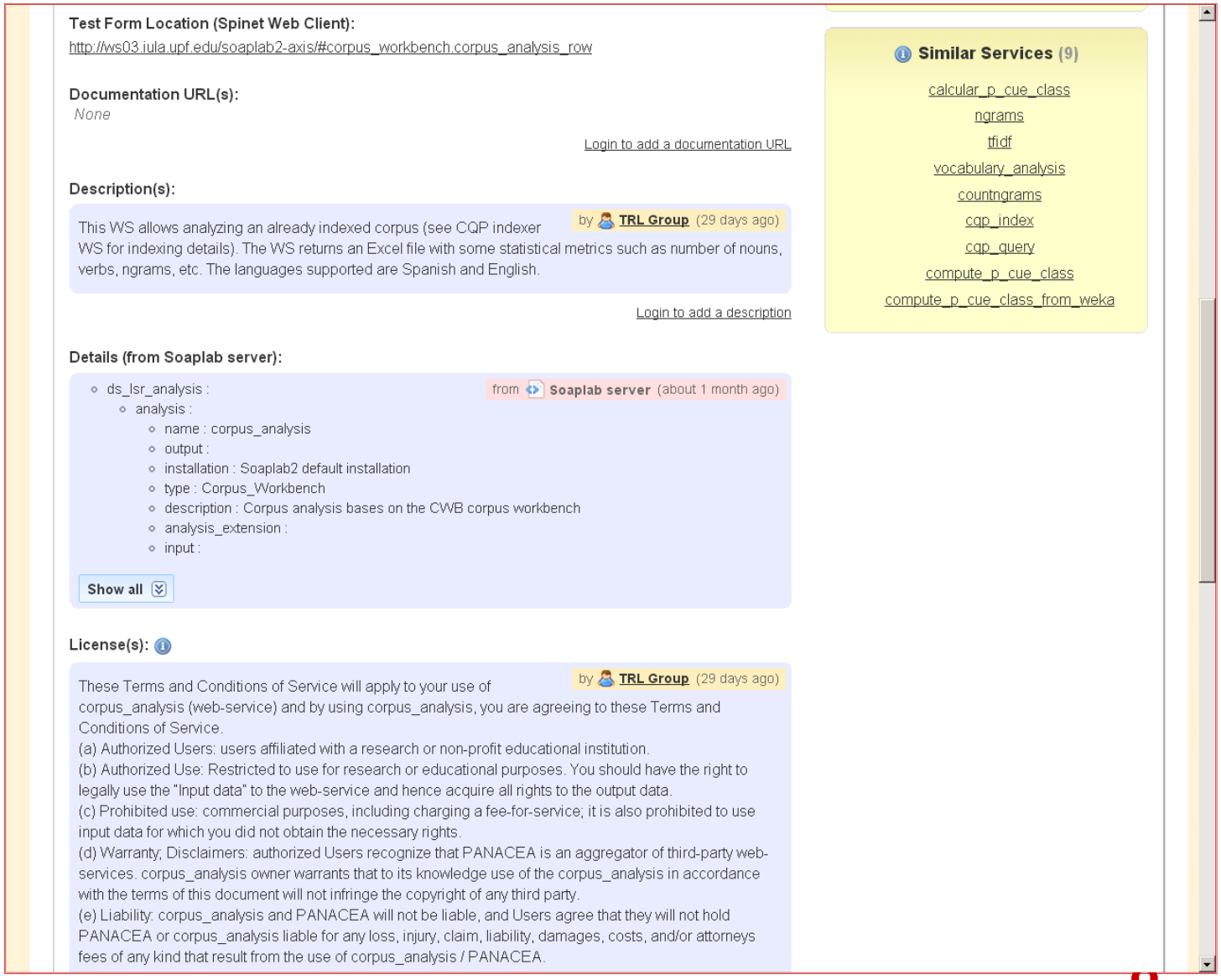


PANACEA REGISTRY (registry.elda.org), instance of BIOCATALOGUE





Details of Web Service terms of use. Workflows are treated as resources, hence Creative Commons.



Test Form Location (Spinet Web Client):
http://ws03.iula.upf.edu/soaplab2-axis/#corpus_workbench.corpus_analysis_row

Documentation URL(s):
None

Login to add a documentation URL

Description(s):

by **TRL Group** (29 days ago)

This WS allows analyzing an already indexed corpus (see CQP indexer WS for indexing details). The WS returns an Excel file with some statistical metrics such as number of nouns, verbs, ngrams, etc. The languages supported are Spanish and English.

Login to add a description

Details (from Soaplab server):

from **Soaplab server** (about 1 month ago)

- ds_lsr_analysis :
  - analysis :
    - name : corpus_analysis
    - output :
    - installation : Soaplab2 default installation
    - type : Corpus_Workbench
    - description : Corpus analysis bases on the CWB corpus workbench
    - analysis_extension :
    - input :

Show all

License(s):

by **TRL Group** (29 days ago)

These Terms and Conditions of Service will apply to your use of corpus_analysis (web-service) and by using corpus_analysis, you are agreeing to these Terms and Conditions of Service.
(a) Authorized Users: users affiliated with a research or non-profit educational institution.
(b) Authorized Use: Restricted to use for research or educational purposes. You should have the right to legally use the "Input data" to the web-service and hence acquire all rights to the output data.
(c) Prohibited use: commercial purposes, including charging a fee-for-service; it is also prohibited to use input data for which you did not obtain the necessary rights.
(d) Warranty; Disclaimers: authorized Users recognize that PANACEA is an aggregator of third-party web-services. corpus_analysis owner warrants that to its knowledge use of the corpus_analysis in accordance with the terms of this document will not infringe the copyright of any third party.
(e) Liability: corpus_analysis and PANACEA will not be liable, and Users agree that they will not hold PANACEA or corpus_analysis liable for any loss, injury, claim, liability, damages, costs, and/or attorneys fees of any kind that result from the use of corpus_analysis / PANACEA.

Similar Services (9)

calcular_p_cue_class
ngrams
tfidf
vocabulary_analysis
countngrams
cqp_index
cqp_query
compute_p_cue_class
compute_p_cue_class_from_weka

upf. iula gruptrl

Dissemination of available IULA web services and language resources

INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
UNIVERSITAT POMPEU FABRA - Barcelona

FUTURE

- Supported until end 2014. After, just passive support if no ...
- Innovation projects with industry: Text Analytics in different scenarios: Sentiment Analysis, Consumer behaviour, etc.
- Approaching data sources for collaboration: visualize together text and exploitation services
- Research Infrastructures for Humanities and Social Sciences:

Unió Europea
Fons europeu
de desenvolupament regional
Una manera de fer Europa

DASISH

Generalitat de Catalunya

CLARIN ERIC
Common Language Resources and Technology Infrastructure

More information

- www.panacea-lr.eu
- Registry.elda.org & myexperiment.elda.org
- www.iula.upf.edu
- nuria.bel@upf.edu

T H A N K S !