

IULA Web services for text data

Núria Bel

IULA - TRL

Universitat Pompeu Fabra

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:

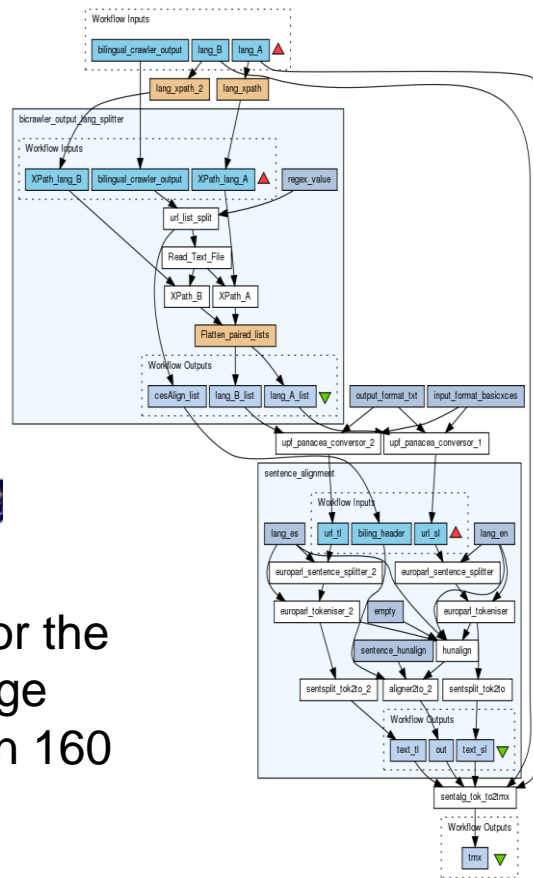
1. Text cleaning and normalization:
 - Prepare data for exploitation (conversion, cleaning, anonymization)
2. Quantitative information and statistical analysis:
 - Analytic figures on words in texts, and about their relevance
3. Annotation / Enrichment services:
 - Adding semantics: Named Entity Recognition,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



e-IRG - Dublin, Núria P. ...
 Distributed platform for the production of Language Resources (more than 160 WS from all partners)



INSTITUT UNIVERSITARI DE LINGÜÍSTICA APLICADA
 UNIVERSITAT POMPEU FABRA - Barcelona



- The PANACEA platform is an **interoperability space** based on Common Interface definitions, and "Travelling Object" specifications Formal definition
- **Tools:** Taverna, BioCatalogue, myExperiment, Soaplab myGrid
- **Common Interfaces** for tools deployed as WS Technical Definition
- **Travelling Object:** based on converters from a to available standards XCES, GrAF, TMX, LMF ...
- **Documentation**

PANACEA - Platform for the Automatic, Normalized Annotation and Cost-Effective Acquisition of Language Resources



- Three levels of interoperability:
 - COMMUNICATION PROTOCOLS: Soap, Rest
 - DATA
 - All tools understand the previous format
 - Tool A Format N → Tool B Format M → Tool C Format L
 - PARAMETERS

Tool A	A	→	Y	Tool B
	B	→	T	
	C	→	Q	
	D	→	Z	

PANACEA - Platform for the Automatic, Normalized Annotation and Cost-Effective Acquisition of Language Resources

About 70 workflows in myexperiment.elda.org

PANACEA MyExperiment

Log in | Register | Give us Feedback | Invite

Home Users Groups **Workflows** Files Packs Services Topics

Workflow Search

Home » Workflows

Workflows

Search filter terms: Sort by: Rank

Showing 74 results. Use the filters on the left and the search box below to refine the results.

Search

Taverna 2 **GrAF PoS tagging with Freeling for basicxces documents (v5)** View Download (v5)

Original Uploader

Created: 25/07/11 @ 14:42:19 | **Last updated:** 19/10/12 @ 10:44:09

Credits: Marcpoch

License: Creative Commons Attribution-Share Alike 3.0 Unported License

Marcpoch

This is a prototype example of a PoS tagging workflow using Freeling. Input data are basicxces documents (PANACEA TO1) and the output is presented in the GrAF format.

Rating: 0.0 / 5 (0 ratings) | **Versions:** 5 | **Reviews:** 0 | **Comments:** 0 | **Citations:** 0

Viewed: 128 times | **Downloaded:** 25 times

Tags (5): basicxces | example | freeling | graf | tagging

Taverna 2 **[untitled] (v1)** View Download (v1)

Original Uploader

Created: 02/08/12 @ 09:32:09 | **Last updated:** 02/08/12 @ 09:33:33

Credits: Thurmair

License: Creative Commons Attribution-Share Alike 3.0 Unported License

Filter by type

- Taverna 2 74

Filter by tag

- basicxces 21
- panacea 21
- freeling 20
- tagging 16
- example 15
- bilingual 12
- crawled 12
- graf 10
- hunalign 8
- dependency 6

Filter by user

- Marcpoch 33
- Muntsa Padró 13
- Valeria Quochi 8
- atoral 7
- Laura Rimell 3
- Thurmair 3
- Francesco Ru... 2
- Marta Villegas 2
- Prokopis 2
- Olivier Hamon 1

New/Upload

Workflow GO

Log in / Register

Username or Email:

Password:

Remember me:

Need an account?
[Click here to register](#)

[Forgot Password?](#)

Popular Tags
25 tags
[\[All Tags\]](#)

[basicxces](#) | [bilingual](#) | [cleaner](#) | [cqp](#) | [crawled](#) | [dependency](#) | [desr](#) | [directory](#) | [download](#) | [english](#) | [example](#) | [freeling](#) | [graf](#) | [hunalign](#) | [itlp](#) | [lexical acquisition](#) | [lists](#) | [merge](#) | [noun classification](#) | [panacea](#) | [parser](#) | [pos tagging](#) | [sentence alignment](#) | [spanish](#) | [tagging](#)

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



Search: [Home](#) [Services](#) [Register a Service](#)

[Home](#) »

Latest Activity

Older

- patrick.goethals **joined** the BioCatalogue
- TRL_Group **added** a category annotation to Service: [corpus_analysis](#)
- TRL_Group **added** a category annotation to Service: [corpus_analysis](#)
- TRL_Group **added** a description annotation to the Soap Service of Service: [corpus_analysis](#)
- TRL_Group **added** a language annotation to Service: [corpus_analysis](#)
- TRL_Group **added** a language annotation to Service: [corpus_analysis](#)
- TRL_Group **added** a license annotation to the Soap Service of Service: [corpus_analysis](#)
- [corpus_analysis](#) has been **updated** (1 update from latest WSDL). See [changelog entry](#).

[More](#)

[Monitoring Test Changes](#)

The PANACEA registry currently has **161 services**, **11 service providers** and **1 members**

PANACEA Partners



Has activado el modo de pantalla completa. [Salir del modo de pantalla completa \(F11\)](#)

[Sign Up](#) | [Sign In](#)

[About Us](#) | [Contact Us](#)

Search: [Home](#) [Services](#) [Register a Service](#) [Providers](#) [Search by Data](#)

[Home](#) » [Services](#)

The services index has been filtered

[Subscribe to these results](#)

Filtering

Current Filters Applied

Service Categories

[Statistics Analysis](#) X

[Clear all filters](#)

Select filters from below...

[Enable tag filters](#)

Service Types (1)

SOAP (161)

Service Languages (17)

- Arabic (2)
- Asturian, Bable (12)
- Catalan, Valencian (20)
- Czech (1)
- English (52)
- French (10)
- Gallean (13)
- German (18)
- Greek, Modern (1453) (10)
- Irish (1)
- Italian (18)
- Miscellaneous languages (2)
- Portuguese (11)
- Russian (5)
- Spanish, Castilian (39)
- Welsh (4)
- [LANGUAGE_INDEPENDENT](#) (78)

Search within results:

Displaying **all 8** services

Sort by: [Newest](#)

View: [Grid](#)

corpus_analysis

Statistics Analysis [Corpus Workbench](#)

[Spanish, Castilian](#) [English](#)

Provider: [Universitat Pompeu Fabra \(UPF\)](#)

compute_p_cu...

Statistics Analysis

[LANGUAGE_INDEPENDENT](#)

Provider: [Universitat Pompeu Fabra \(UPF\)](#)

compute_p_cue_class

Statistics Analysis

[LANGUAGE_INDEPENDENT](#)

Provider: [Universitat Pompeu Fabra \(UPF\)](#)

countngrams

Statistics Analysis

[LANGUAGE_INDEPENDENT](#)

Provider: [Universitat Pompeu Fabra \(UPF\)](#)

vocabulary_analysis

Statistics Analysis

[LANGUAGE_INDEPENDENT](#)

Provider: [Universitat Pompeu Fabra \(UPF\)](#)

tfidf

Statistics Analysis

[LANGUAGE_INDEPENDENT](#)

Provider: [Universitat Pompeu Fabra \(UPF\)](#)

ngrams

Statistics Analysis

[LANGUAGE_INDEPENDENT](#)

Provider: [Universitat Pompeu Fabra \(UPF\)](#)

calcular_p_cue_class

Statistics Analysis

Provider: [Universitat Pompeu Fabra \(UPF\)](#)



INSTITUT UNIVERSITARI DE LINGÜÍSTICA I ULTRA LINGÜÍSTICA
UNIVERSITAT POMPEU FABRA - Barcelona

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):

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.


by [TRL Group](#) (29 days ago)

[Login to add a description](#)

Details (from Soaplab server):

- ds_isr_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 :

from [Soaplab server](#) (about 1 month ago)

[Show all](#) 

License(s):

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.

by [TRL Group](#) (29 days ago)

- (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](#)

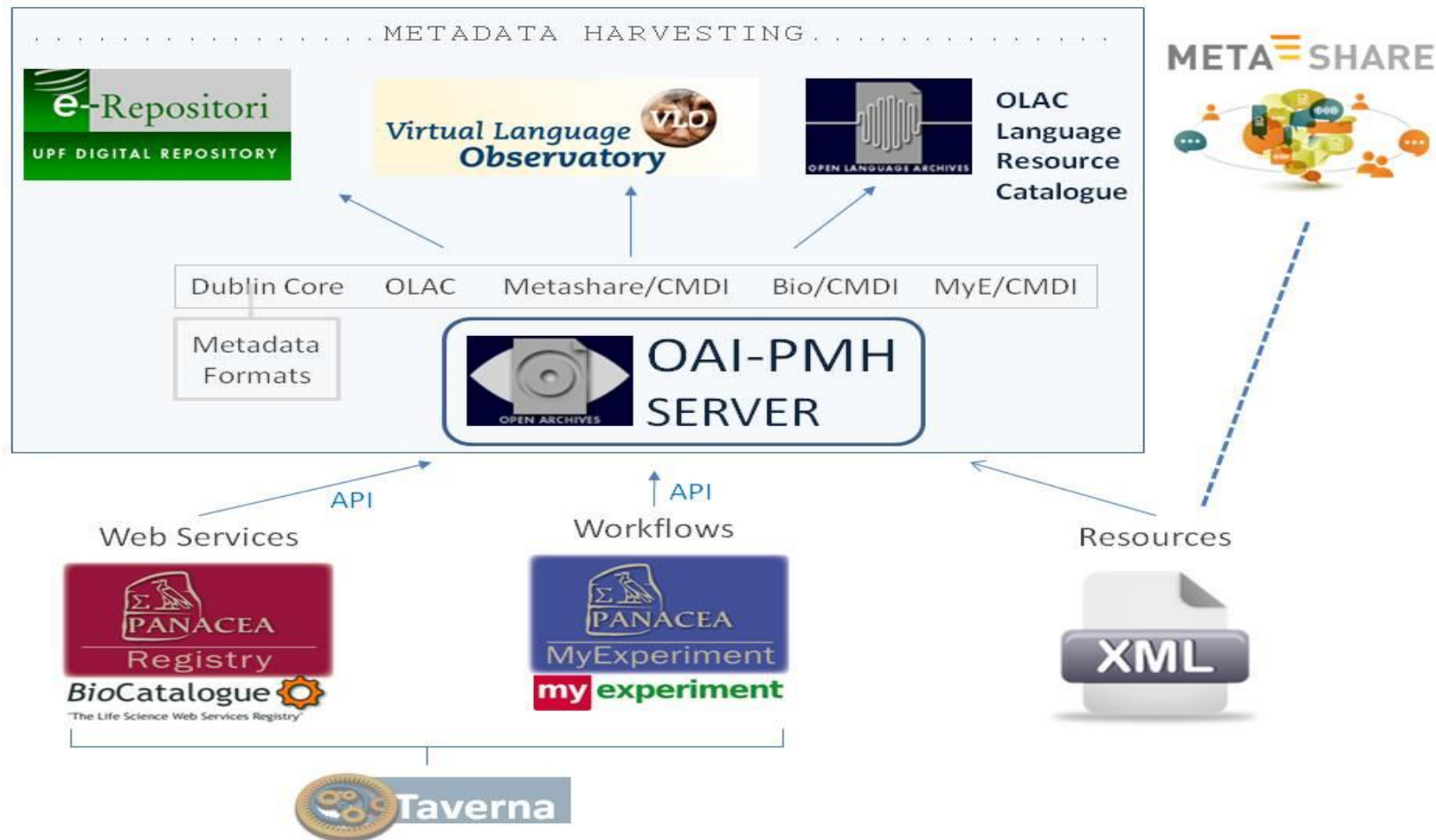
[cqp_index](#)

[cqp_query](#)

[compute_p_cue_class](#)

[compute_p_cue_class_from_weka](#)

Dissemination of available IULA web services and language resources



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



Generalitat de Catalunya

CLARIN ERIC

Common Language Resources and Technology Infrastructure



INSTITUT UNIVERSITARI DE LINGÜÍSTIC,
UNIVERSITAT POMPEU FABRA - Barcelon...

More information

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

THANKS!