

FLaReNet

Fostering Language Resources Network

EU –ECP-2007-LANG-617001



e Content plus

Núria Bel – IULA -- Universitat Pompeu Fabra – Barcelona, SPAIN





WHAT?



- FLaReNet is a EU-eContentPlus thematic network whose objective is the definition of a common European strategy that supports the creation of and the access to Language Resources and Technologies for all the European languages (at least).
- To foster a common vision in the decision making levels about the importance of these resources



What are LR's



- Particular language data in one or another form that inform research, technologies, etc.
- Dictionaries, texts, aligned texts, bilingual glossaries, annotated texts, etc.





FLaReNet's Mission



- FLaReNet promotes consensus among stakeholders about priorities for the domain as well as for identifying strategic goals in the short and long terms.
- Main result is to supply the Commission, national organizations and the industry with an agreed program of actions that helps to guarantee the development and access to the language resources required by the multilingual Europe.



Working Groups on




- **WP2:** The **Chart for the area of LRs and LT** in its different dimensions
- **WP3:** Methods and models for **LR building, reuse**, interlinking, maintenance, sharing, distribution...
- **WP4:** **Harmonisation** of formats and standards
- **WP5:** Definition of **evaluation and validation protocols** and procedures
- **WP6:** Methods for the **automatic construction and processing** of LRs



Members



- Sept-2009 network was made of **81 Institutional Members from 31 countries**, and **251 Individual Subscribers**
- Coordinated by
 - Istituto di Linguistica Computazionale “Antonio Zampolli”, CNR,
 - And a Common Steering Committee with representatives of
 - Evaluation and Language Resources Distribution Agency, ELDA;
 - Institute for Language and Speech Processing, Athena RC;
 - Institute for Multilingual and Multimedia Information, CNRS;
 - Universitat Pompeu Fabra;
 - Wien University and
 - Utrecht University.
- Also with an International Cooperation group



Collaborative Web Page

Submission #358 | FLaReNet - Mozilla Firefox

Fitxer Edita Visualitza Història Adreces d'interès Enllaços Ajuda

http://www.flarenet.eu/?q=node/267/submission/358

Més visites Primers passos Darreres notícies

Main menu

- Welcome to FLaReNet
- Network Organisation
- Working Groups
 - Grant Agreement & Annexes
 - Deliverables
 - Join FLaReNet
- Events
- News
- Contact Us

nuriabel

- Contact
- My account
- Search
- Site Map
- Working Groups
- Create content
- Log out

Recent comments

- Reply Test
5 weeks 10 hours ago
- Creation of a think-tank on
Language Resources Acquisition

Form: WG6 - Questionnaire - Strategies and Recommendations

Submitted by nuriabel

Thursday, 30 July, 2009 - 07:10

83.60.28.250

First Name: *

Nuria

Family Name: *

Bel

Questions: *

- ☒ 1. Which are the most promising techniques under study?
- ☐ 2. Why are the current automatic production techniques promising?
- ☐ 3. How they behave with respect to coverage, robustness and accuracy?
- ☐ 4. Can current techniques for automatic production be tuned for offering less coverage but more precision?
- ☒ 5. To what extent are automatic production techniques dependent on annotated data?
- ☐ 6. What is more cost-effective: proliferation of domain tuned resources or construction of unique general large language resource per language and application?
- ☐ 7. What is the cost of the supply of language resources for deploying language independent, domain independent applications?
- ☒ 8. What is the cost of the shortage of language resources for deploying language independent, domain independent applications?
- ☐ 9. Can the web 2.0 and the collective intelligence be used in the production of LRs better than pure automatic methods?
- ☐ 10. Are there methods to evaluate the quality of the resulting LRs?

Motivations (optional):

The problem created by shortage must be known to evaluate to what extent the investment in new methods is worth.



Workshops



- 2009 Wien

***The European Language Resources and
Technologies Forum:
Shaping the Future of the Multilingual Digital
Europe***

- 2010 Barcelona (February, 11 and 12)

***The future of Language Resources,
Language Resources of the Future***



And other activities ...



- **Relations with parallel International Initiatives**
 - SILT – NSF US
 - Cyberling – US
 - LDC – US
 - Language Grid – Japan NICT & Thailand NACTEC
 - AFNLP – Asia
 - ISO – International
 - ...
- **Relations with Projects**
 - Relations with CLARIN, EuroMatrix, KYOTO, ...
 - Synergy with some new projects: T4ME, Panacea, Multilingual Web, ...
 - FlareNet output used in the conceptual definition of T4ME
 - Involved in defining the ORI
- **Actions for the ORI in the new NoE T4ME**



More information at


www.flarenet.eu

Get involved!



BUT

- Survey of Language Resources ..
- Who knows what is out there?
- Is it a reasonable question?



Quoting our D6.1a



The compilation of information for this first survey was harder than expected because of the **lack of documentation** for most of the resources surveyed. Besides, **the availability of the resource itself is problematic**: Sometimes a resource found in one of the catalogues/repositories **is no longer available** or simply **impossible to be found**; sometimes it is **only possible to find a paper reporting on some aspects of it**; and, finally, sometimes **the information is distributed among different websites, documents or papers at conferences**. This made it really difficult to carry out an efficient and consistent study, as the information found is not always coherent (e.g. not every corpus specifies the number of words it has) and sometimes it even differs from the one found in different catalogues/repositories.



A Proposal



- In fact ...NOBODY knows ..
- and we need to know ...
- A LR is a valuable resource, but not if nobody knows about it,
- Danger: there is no offer, so ... is there no demand?
- This is the risk when proposing new strategic policies



TheHarvestingDay.eu



WHAT? The harvesting day will be a day in which a robot will collect Basic Metadata Description (BAMDES) describing resources and tools, as published at their web sites.

WHY? To allow LR developers to enhance and ensure the visibility of their language resources and tools

WHEN? The first harvesting day will take place on XX XXXX, and will be then repeated periodically in an automatic manner.

WHERE? You need to install Zebra in your server (self executable package available) and make your resources visible.

WHO? Every resource and/or tool developer/provider is invited to participate. Harvesting results will be provided to the main resource and tools catalogues and observatories (ELDA/ELRA, CLARIN, T4ME...).

Mozilla Firefox

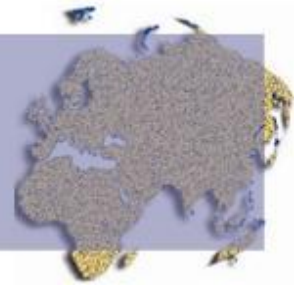
Fitxer Edita Visualitza Historial Adreces d'interès Eines Ajuda

file:///C:/Users/nuria/AppData/Local/Temp/UPF-IULA-resources-1.xml

interprete XSLT

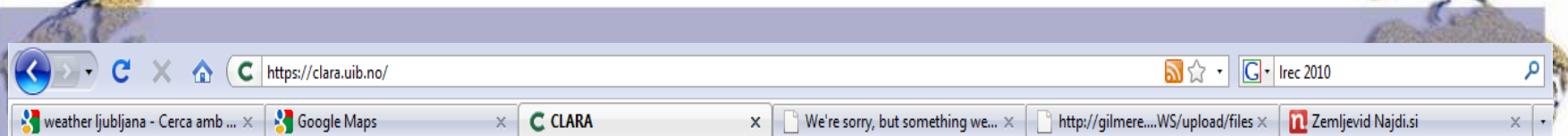
Pàgina d'inici del Mozilla Firefox x Error de càrrega de la pàgina x Gmail: correo electrónico de Google x file:///C:/Users/...-resources-1.xml x

```
<oai:metadataNamespace>http://www.iula.upf.edu</oai:metadataNamespace>
</oai:metadataFormat>
</ListMetadataFormats>
- <ListRecords metadataPrefix="clarin">
- <oai:record>
- <oai:header>
  <oai:identifier>oai:UIB.es:TDX-0713104-100204</oai:identifier>
  <oai:datestamp>2004-07-13</oai:datestamp>
</oai:header>
- <oai:metadata>
- <LexicalResource xsi:schemaLocation="http://www.iula.upf.edu ClarinSchema3.xsd">
  <ResourceName>Banco de neologismos 2004-2007</ResourceName>
  <ResourceType>LexicalResource</ResourceType>
  <URL>http://cvc.cervantes.es/obref/banco_neologismos/</URL>
- <OwnerInstitution>
  Institut Universitari de Lingüística Aplicada / Instituto Cervantes
</OwnerInstitution>
  <ContactEmail>iulasecretaria@upf.edu</ContactEmail>
  <Language languageID="spa">es</Language>
  <LexiconType>monolingual</LexiconType>
  <Size sizeUnit="entries">24227</Size>
  <CoverageType>general</CoverageType>
  <EncodingFormat>HTML</EncodingFormat>
  <InformationContained>neologisms</InformationContained>
  <InformationContained>morphology</InformationContained>
</LexicalResource>
```



Prepare for... the harvesting day!!!

Enhance, guarantee the visibility of your resources and tools!



CLARA

Common Language Resources and their Applications — a Marie Curie ITN

[NEWS](#) [ABOUT](#) [VACANCIES](#) [TRAINING PROGRAM](#) [FELLOWS](#) [CALENDAR](#)

MPI announces two ESR vacancies

20 January, 2010, 15:29

The Max Planck Institute for Psycholinguistics (MPI) has two vacancies for two ESR (early stage researcher) positions in language technology, in particular *linguistic infrastructures, interoperability, standards and semantic web*. See the announcement under Vacancies.

[Comment](#)

Two ESR positions announced at UCPH

20 January, 2010, 15:08

The University of Copenhagen is announcing two ESR vacancies, one in Multimodal Communication Modeling and one in Semantic Annotation of Corpora. See the announcement under Vacancies as long as these vacancies are open.

[Comment](#)

Terminology course in Bergen postponed



CLARA is an Initial Training Network in the Marie Curie Actions financed by the EC under FP7.

META

[Register](#)

[Log in](#)

[Entries RSS](#)

[Comments RSS](#)

[WordPress.org](#)



Ljubljana, February 3-4, 2010

FLaReNet goes to 1st International
Sketch Grammar Workshop

