

Open Grey for Language Technology: a ride on the network

*Gabriella Pardelli, Manuela Sassi, Sara Goggi
Istituto di Linguistica Computazionale "Antonio Zampolli", ILC
Consiglio Nazionale delle Ricerche CNR- Pisa, Italy*

***GL14 Fourteenth International Conference on Grey Literature
National Research Council, Rome, Italy 29-30 November 2012***

Starting Questions

- ❑ Where to search for Language Technology (LT) documents?
- ❑ How to retrieve LT documents?
- ❑ Is the extensive experience in LT important for retrieving information?

The first decade of the new millenium has suddenly witnessed first the growth and then the rapid increase of the so-called “social networks” which totally transformed the way information is transmitted: nowadays the World Wide Web looks like an enourmous collection of documents inter-connected and linked to the various search engines by sharing the same paradigm (Web 2.0).

The selection, conservation and storage of digital content apparently makes the users’ fruition easier : but is this assumption really true?

To formulate appropriate and effective queries for a search is a difficult task for users and requires a careful terminological selection for obtaining the most from an Information Retrieval system



Information Retrieval is the academic discipline that studies the methodologies, tools, techniques, and languages for searching and retrieving relevant data for an information need.

Where to search for Open Access Language Technology documents?

When a user tries to retrieve information on a given topic from online repositories, there are several possibilities to formulate a query; for instance, given the query “Language Technology”, the web replies with about 878.000.000 results in 0,26 seconds (September 25, 2012 at 4.30 p.m.):

Academic articles for **language technology**

...of the state of the art in human **language technology** - *Cole* – Cited by 577

...:Stirring up trouble about **language, technology** and ...-*Postman*–Cited by 158

...Information extraction as a core **language technology** - *Wilks* – Cited by 69





1. [Language technology - Wikipedia, the free encyclopedia](http://en.wikipedia.org/wiki/Language_technology) [en.wikipedia.org/wiki/Language_technology](#) - [Traduci questa pagina](#)
Language technology is often called human *language technology* (HLT) or natural language processing (NLP) and consists of computational linguistics (or CL) ...
2. [Welcome to Language Technology World — LT World](http://www.lt-world.org/) www.lt-world.org/ - [Traduci questa pagina](#) 4 Feb 2012 – A portal on the range of *technologies* that deal with human *language*. News, conferences, projects, organisations, systems, and resources.
3. [CELI - Language and Information Technology](http://www.celi.it/) www.celi.it/ ... l'analisi del linguaggio è diventata un fattore di successo dei nostri Clienti. I *Language Specialists* madrelingua di CELI lavorano in più di 30 lingue diverse.
4. [Language Learning & Technology - Home](http://lt.msu.edu/) lt.msu.edu/ - [Traduci questa pagina](#) Online journal devoted to *technology* and *language* education research for foreign and second *language* educators. Full text of articles available.
5. [UNITN | Human Language Technology and Interfaces](http://www.unitn.it/ateneo/.../human-language-technology-and-interfaces) www.unitn.it/ateneo/.../human-language-technology-and-interfaces Le Tecnologie del Linguaggio (Human *Language Technologies*, HLT) ci permettono oggi di interagire a voce con vari servizi automatici, ad esempio per ...
6. [PDF] [Language Technology A First Overview - DFKI](http://www.dfki.de/~hansu/LT.pdf) www.dfki.de/~hansu/LT.pdf - [Traduci questa pagina](#) Formato file: PDF/Adobe Acrobat - [Visualizzazione rapida](#) di H Uszkoreit - [Citato da 9](#) - [Articoli correlati](#) *Language technologies* are information technologies that are specialized for dealing ... are also often subsumed under the term Human *Language Technology*.
7. [DFKI Language Technology lab](http://www.dfki.de/lt/) www.dfki.de/lt/ - [Traduci questa pagina](#) Die Deutsche Forschungszentrum für Künstliche Intelligenz GmbH mit Sitz in Kaiserslautern und Saarbrücken ist auf dem Gebiet innovativer ...
8. [CMU - Language Technologies Institute](http://www.lti.cs.cmu.edu/) www.lti.cs.cmu.edu/ - [Traduci questa pagina](#) CMU/LTI offers MS and PhD programs in *Language* and *Information Technologies*.
9. [Language Technology](http://www.lang-tech.org/) www.lang-tech.org/ - [Traduci questa pagina](#) LangTech is the european forum dedicated to communities and organisations involved in the development, deployment and exploitation of *Language* and ...
10. [Immagini relative a language technology](#) - [Segnala immagini non appropriate](#)
11. [FBK | Human Language Technology](http://hlt.fbk.eu/) hlt.fbk.eu/ - [Traduci questa pagina](#) FBK - Fondazione Bruno Kessler. Human *Language Technology*. FBK > IT. News. 10 Sep 2012. Demo Paper accepted at ISWC 2012. 03 Sep 2012 ...

My findings...

From this generic query it is possible to retrieve only:

- ✓ 9 portals
- ✓ 2 open access publications (Cole and Wilks)
- ✓ 1 review (Portman)

But only 2 documents from this set satisfy the user.....

Lets' try with queries formulated by an expert user:

1 Query: ACL Anthology

• <http://aclweb.org/anthology-new/C/C65/>



2

Query: Language Technology

The web answers... ..

<http://tenenga.com/download/TECA02-02.pdf>

Google custom search



Circa 11.900 risultati (0,16 secondi)

[<bold>Language, Technology, and Society Richard Sproat</bold>...](#)

Formato file: PDF/Adobe Acrobat

Language, Technology, and Society. Richard Sproat. (Oregon Health & Science University). Oxford: Oxford University Press, 2010, xiii+286 pp; hardbound, ...
aclweb.org/anthology-new/J/J11/J11-1011.pdf

1. [Draft for a road map on human language technology](#)
Formato file: PDF/Adobe Acrobat
motto "How will language and speech technology be used in the information ... Advances in human **language technology** will offer nearly universal access to on- ...
www.aclweb.org/anthology-new/W/W02/W02-1302.pdf
2. [Australasian Language Technology Association \(ALTA\)](#)
The Australasian **Language Technology** Association (ALTA) was founded at the 5th Australasian Natural Language Processing Workshop, in Canberra, ...
aclweb.org/anthology-new/docs/alta.html
3. [Evangelising Language Technology: A Practically-Focussed ...](#)
Formato file: PDF/Adobe Acrobat
Evangelising **Language Technology**: A Practically-Focussed Undergraduate Program. Robert Dale, Diego Mollá Aliod and Rolf Schwitter. Centre for Language ...
www.aclweb.org/anthology-new/W/W02/W02-0104.pdf
4. [LTP: A Chinese Language Technology Platform](#)
Formato file: PDF/Adobe Acrobat
LTP (**Language Technology** Platform) is an integrated Chinese processing platform which includes a suite of high perfor- mance natural language processing ...
aclweb.org/anthology-new/C/C10/C10-3004.pdf
5. [Spoken Language Technology: Where Do We Go From Here?](#)
Formato file: PDF/Adobe Acrobat
Spoken **Language Technology**: Where Do We Go From Here? Roger K. Moore. 20 20 Speech Ltd. Malvern, UK. Recent years have seen dramatic ...
6. [aclweb.org/anthology-new/P/P00/P00-1003.pdf](#)



- 7: [*Does Language Technology Offer Anything to Small Languages?](#)
 - Formato file:PDF/Adobe Acrobat
 - Does Language Technology Offer Anything to Small. Languages? Nick Thieberger. PARADISEC, University of Melbourne/. University of Hawai'i at Manoa...
 - aclweb.org/anthology/new/U/U07/U07*1002.pdf
- 8: [*Lexicons for Human Language Technology](#)
 - Formato file:PDF/Adobe Acrobat
 - Lexicons for Human Language Technology. Mark Liberman. Linguistic Data Consortium. University of Pennsylvania. Philadelphia, PA 19104-6305 myl@unagi...
 - aclweb.org/anthology/new/H/H94/H94*1004.pdf
- 9: [*Letter to the Editor: Language Technology for Beginners](#)
 - Formato file:PDF/Adobe Acrobat
 - Language Technology for Beginners. Ronald A. Cole 1. (University of Colorado). I am writing in response to Varol Akman's review (Computational Linguistics, ...)
 - www.aclweb.org/anthology/new/J/J99/J99*4012.pdf
- 10. [*Overview of the ARPA Human Language Technology Workshop](#)
 - Formato file:PDF/Adobe Acrobat
 - ARPA Human Language Technology Workshop. Madeleine Bates, Chair, Editor. BBN Systems & Technologies. 70 Fawcett Street. Cambridge, MA 02138. 1.
 - aclweb.org/anthology/new/H/H93/H93*1001.pdf

•10 open access Grey Literature documents found!

How to retrieve LT documents

Given the fact that the enormous amount of data available on the web is difficult to query from a semantic point of view, the human interpretation is always needed - - but which are the assumptions/conditions for making an effective query?

- being updated on the state-of-the-art;
- being skilled at navigating on the web portals;
- being able to understand the data.

“Noise” or “Silence”...

Performance of an information retrieval system can be measured by the following coefficients:

- ✓ *Precision*: proportion of relevant data retrieved from the total data retrieved
- ✓ *Recall*: extent of relevant data retrieved from the total data relevant in the database.

These coefficients measure two different factors:

- Noise* = non-relevant data retrieved
- Silence* = relevant data that have not been retrieved from the data base.

Retrieval models compute the degree to which certain elements answer to a query: a good model should be able to maximize recall and precision and minimize, respectively, “silence” and “noise”.

Concluding Remarks

The web is both a knowledge repository and a knowledge dispenser



- Create innovative paradigms for information retrieval
- Establish features for semantic search on web portals
- Achieve precision & recall

Nowadays knowledge extraction is only possible if:

1. the know-how of the state-of-the-art is updated
2. appropriate queries are formulated by selecting effective terms from the dedicated portals.