naturalbrowsing

exploring complexity

Natural Browsing is a research project funded by POR PUGLIA 2007/2013 Asse I Linea 1.1 Azione 1.1.2 Bando "Aiuti agli Investimenti in Ricerca per le PMI"

In every organization the biggest part of its knowledge is held by persons and e-documents (web pages, spreadsheets, word documents, database, emails, etc.). Such heterogeneous data are hardly usable by the organization members.

The Natural Browsing project comes from the idea of extracting and formalizing all this knowledge, making it accessible to all the organization members, so that many advantages can be taken from the users-system interaction in order to refine and increase the knowledge base.

Scenario

Browsing and searching are made through search engines which returns links that are strictly connected to the word without emphasizing the user participation.

The most part of these information has an unstructured form so, if there is a common substratum for any data sources, it will be represented by human language.
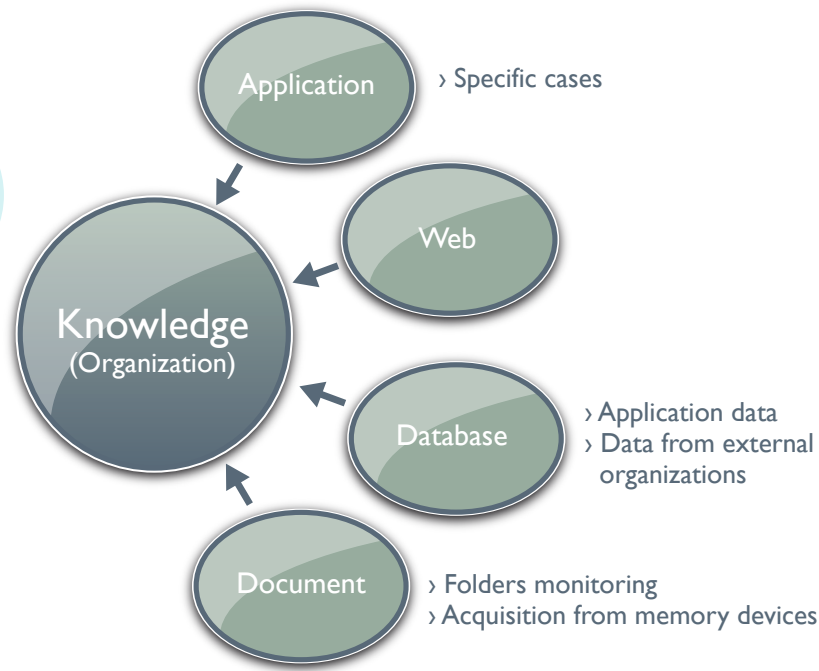
Nowadays there are different technologies for formalizing information extracted from textual documents in a data structure capable of describing semantics of every single word. So that, the only possible common element is a formalization of semantics through ontologies.

REGIONE PUGLIA

This project has also the fundamental academic contribution of:

› *University of Bari, Informatics Department;*

› *University of Bari, "Michelangelo Merlin" Physics Department;*

› *Polytechnic of Bari - Mechanical and Management engineering Department*

Application › Specific cases

Web

Knowledge
(Organization)

Database › Application data
› Data from external
organizations

Document › Folders monitoring
› Acquisition from memory devices

Making comprehensible and easy a consultation of knowledge in any organization.

Targets

The planning proposal consists in searching and structuring innovative processes of data extraction from heterogeneous sources, semantic enrichment and browsing. In the near future these actions will allow users to use internet by means of natural language.
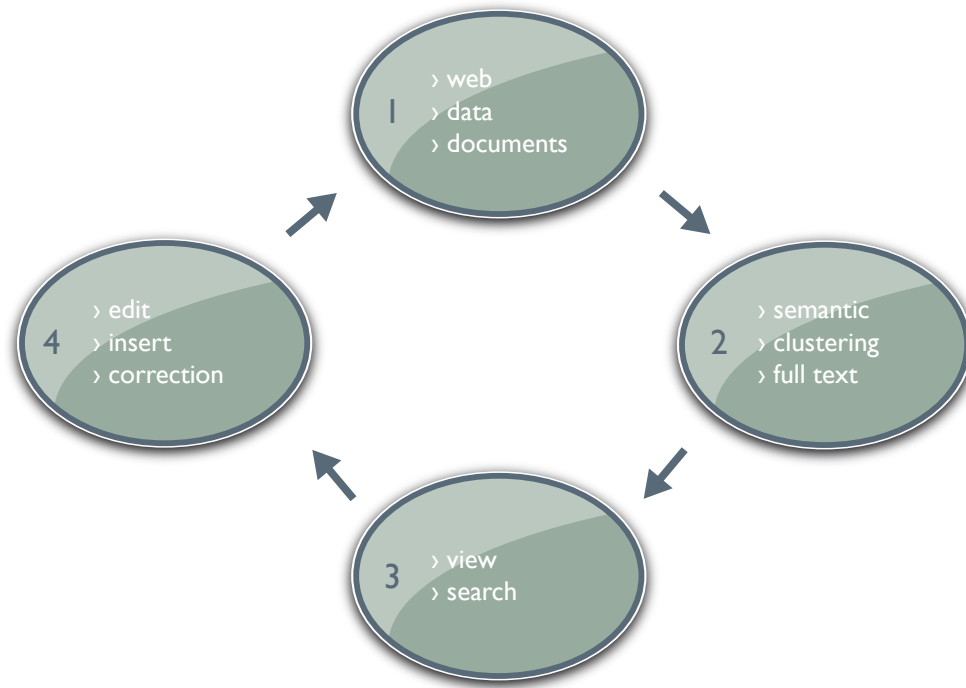
Natural Browsing aims are:

› state of art analysis;

› studying of a gathering process for data, information and contents coming from heterogeneous sources (documental repository, forum, web pages, blog, etc.) following a reference standard model (normalization in a common semantic format);

› realization of a consolidation system for semantic information in a repository;

› creation of a "semantic dictionary", namely a semantic enrichment process, which acts like a navigation infrastructure based on semantics;

› convoy in a single repository;

› individuation of a typical navigation model on semantic net;

› selection of processes for collecting new contributions from users;

› realization of demonstrative prototypes.

Data are usually heterogeneous:

› *Documents*
  Monitoring folders and memory devices

› *Web*
  Blog, Forum, Intranet applications

› *Database*
  Application data, data from external organizations

› *Applications*
  Specific cases

natural**browsing**
exploring complexity

Circle diagram nodes:

1 › web / › data / › documents

2 › semantic / › clustering / › full text

3 › view / › search

4 › edit / › insert / › correction

It will be possible to have another interaction level between the portal and its users and this will add other elements to the information found, for example locations, events and their georeferentiation.

Working procedures

The Natural Browsing projects supposes the definition of four processes:

1. finding information from different informative sources, then normalization in RDF language of these data following semantics rules, and, last, consolidation in a repository;

2. semantic enrichment on the informative unit relations and creation of the semantic net;

3. users navigation in the extracted data and further informative enrichment;

4. cyclic repetition of the second process, in order to allow semantic relation and semantic net to grow, considering also the users enriched contents (process number 3).

The first, second and fourth process are totally clear for users because the knowledge base that can be consulted keeps enriching with new and better structured information thanks to system processes that have no interaction with the human component. These processes are services prototypes, instead of web services or agents ones.

The user navigation (process number 3) is improved by a split in data elaboration. Each user profile has a precise access to an application web prototype for navigating in data, preferring a classic method, for example consulting categories, instead of being guided by the ad hoc semantic net built since the initial request in natural language.

natural**browsing**
exploring complexity

The Region of Puglia offers an excellent field of study for Natural Browsing future developments.

Application fields

From the beginning the planning proposal has been directed towards an applicability in different contexts and for supporting different situations in which it's important to emphasize data sources and the different contributions coming from navigation and consultation processes:

> managing and highlighting knowledge in an organization;

> bettering decision procedure in Public Administrations which want to test different ways of partecipatory democracy;

> support for all the organization members in acquiring knowledge through a single access point.

Developments

News and improvements are available on web site:

www.naturalbrowsing.it