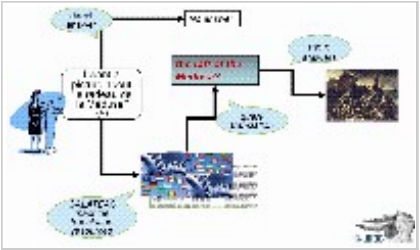
GALATEAS: Generalized Analysis of Logs for Automatic Translation and Episodic Analysis of Searches

GALATEAS offers content providers an innovative approach to understanding users' behaviour by analysing search engine transaction logs and facilitates the development of multilingual content access.



With the growth of digital libraries and digital library federation (as well as partially unstructured collections of documents such as web sites), a large set of vendors is offering engines for retrieving contents and metadata via search requests by the end user (queries). In most cases these queries are just unstructured fragments of text in a specific language. The first service offered by GALATEAS (LangLog) is focussed on getting meaning out of these lists of queries and it is addressed to library/federation/site managers. Contrary to mainstream service in this field, GALATEAS services will not considered standard structured information of web logs (e.g. click rate, visited pages, user's paths inside the document tree) but the information contained in queries from the point of view of language interpretation. By subscribing LangLog federations administrator and managers will be able to answer questions such as: as "Which are the topics which are most commonly searched in my collection, according to a certain language?"; "how do these topics relate with my catalogue?"; "Which named entities (people, places) are more popular among my users?". The second problem addressed by GALATEAS is the one of Cross Language Information Retrieval (CLIR) i.e. the capability of typing a query in one specific language and retrieving documents which are available in different languages. The CACAO consortium is already successfully providing services for indexing and searching over digital libraries and metadata repositories. During commercial exploration for marketing CACAO it emerged that certain institutions prefer to keep indexing and searching at their premises (using their own favourite search engine) and would be perfectly satisfied with a service of plain query translation.The second service offered by GALATEAS (QueryTrans) has the ambitious and innovative goal of providing the first web translation service specially tailored on query translation.

Project ID card

- Funded under: The Information and Communication Technologies Policy Support Programme
- Area: CIP-ICT-PSP.2009.5.1 - Multilingual Web: Machine translation for the multilingual web
- Total cost: €3.70m
- EU contribution: €1.85m
- Project reference: 250430
- Execution: From 01/04/2010 to 31/03/2013
- Project status: Closed
- Contract type: The Information and Communication Technologies Policy Support Programme PB Pilot Type B

Links

- GALATEAS project

Contact information

LAGOS Nikolaos
Chemin de Maupertuis 6
38240 Meylan
FR
Tel. +33 (0) 4 76 61 51 92
Fax. +33(4) 76 61 51 99
E-mail: **nikolaos.lagos@xrce.xerox.com**

Participants

Coordinator: XEROX SAS, FRANCE

Participating partners:

Organisation	Country
XEROX SAS	FRANCE
OBJET DIRECT SAS	FRANCE
HUMBOLDT-UNIVERSITAT ZU BERLIN	GERMANY
UNIVERSITA DEGLI STUDI DI TRENTO	ITALY
GONETWORK SRL	ITALY

CELI SRL	ITALY
UNIVERSITEIT VAN AMSTERDAM	THE NETHERLANDS
THE BRIDGEMAN ART LIBRARY LIMITED	UNITED KINGDOM

Last update: 15/04/2013

See this factsheet online:
http://ec.europa.eu/information_society/apps/projects/factsheet/index.cfm?project_ref=250430