Xerox Research Centre Europe coordinates EU CACAO project to provide cross-language access to online catalogues and libraries

Uxbridge, UK, May 19, 2008 – The Xerox Research Centre Europe is coordinating a project, funded by the European Union, that will develop an innovative approach to accessing, understanding and navigating multilingual textual content in digital libraries and Online Public Access Catalogues (OPACs).

CACAO (Cross language Access to Catalogues And Online libraries) is a consortium which comprises European academic and industrial research institutions as well as libraries.

The new and unique capabilities to be developed in CACAO lie at the heart of the Lisbon Strategy (European Council, Lisbon, March 2000) which defines how the European Union will strive to become the most competitive and dynamic knowledge-based economy in the world. CACAO is part of the Research and Technological Development which is the European Union’s main instrument for funding research in Europe to achieve Lisbon Strategy goals.

Specifically, CACAO is part of the eContentplus program supporting EU-wide co-ordination of library, museum and archive collections and the preservation of digital collections to ensure the availability of cultural, scholarly and scientific assets for future use.

Part of this program is the development of a European digital library, built on initiatives such as the European Digital Library (EDL) and The European Library (TEL), which addresses the multilingual and multicultural aspect of making digital content more accessible, usable and exploitable.

“It has been one of the major missions of the Xerox Research Centre Europe, since its creation in the 90s, to develop Smarter Document Management SM technologies that are able to augment documents with the ‘smarts’ necessary to perform context-sensitive analyses and transformations of documents in any language and remove the multi-lingual and multi-cultural barriers to promote better knowledge exchange.”, says Monica Beltrametti, Vice President and Director of XRCE. “We are glad that CACAO gives us the opportunity to apply some of our Smarter Document Management SM technologies to support Europe’s knowledge-based economy.”

The CACAO project aims to provide users in the EU with a powerful multilingual digital library and library catalogue search tool to enable them to understand, navigate, and access the full range of content available across Europe, regardless of language. CACAO goes beyond the limited single-language keyword searches to analyse the semantic content of a query. According to Frédérique Segond, Principal Scientist and Manager, XRCE Parsing & Semantics research group, “It will dramatically simplify and speed up time-consuming search for books and other documents in online library catalogues.” CACAO combines natural language processing techniques with existing information retrieval systems to enable end users simply to type queries in their own language to retrieve documents and objects in any available language. CACAO will also provide end users with a search tool that is multilingual, selective, smart and easy to use.
It is estimated that knowledge workers spend up to 30 percent of their time searching for business-related information in electronic documents. And, to an increasing extent, people today are using a multitude of languages other than English to communicate over the Internet. This move away from all-English on the Web will also be true for documents stored in European libraries. Likewise, users’ interest in documents published in languages other than their native language or English is also expected to rise.

XRCE’s long-standing Open Innovation policy has enabled its researchers and engineers to work with a wide variety of public and private research institutions in Europe, often under European Union and government contracts. XRCE has often coordinated such projects.

Xerox established its European research centre in Grenoble, France in the early 90s to create innovative document technology and drive the Xerox corporate transition to becoming a services-led technology business. The Centre is at the heart of many of the components of Xerox’s Smarter Document Management technology suite, such as hybrid text and image categorisation, XML document conversion, and linguistic analysis tools already in use by Xerox for a variety of mission critical services for its customers.

CACAO Project partners:
Celi S.R.L, Italy
Libera Università di Bolzano, Italy
Polska Akademia Nauk Biblioteka Kornicka, Poland
Cité des Sciences et de l’Industrie, France
Gonetwork S.R.L., Italy
Magyar Tudomanyos Akademia Nyelvtudomanyi Intezet, Hungary
Georg-August-Universitaet Goettingen Stiftung Öffentlichen Rechts, Germany
National Széchenyi Library, Hungary
Xerox Research Centre Europe, France

For more information on the Xerox Research Centre Europe, visit: www.xrce.xerox.com
For more information on CACAO, check out the project’s web site: http://www.cacaoproject.eu/home

-XXX-

About Xerox Europe
Xerox Europe, the European operations of Xerox Corporation, markets a comprehensive range of Xerox products, solutions and services, as well as associated supplies and software. Its offerings are focused on three main areas: offices from small to large, production print and graphic arts environments, and services that include consulting, systems design and management, and document outsourcing.

Xerox Europe also has manufacturing and logistics operations in Ireland, the UK and the Netherlands, and a research and development facility (Xerox Research Centre Europe) in Grenoble, France. For more information visit, www.xerox.com.

Xerox®, the Xerox wordmark and the spherical connection symbol are trademarks of Xerox Corporation in the United States and/or other countries. Prices, features, specifications, capabilities, appearance and availability of Xerox products and services are subject to change without notice.

Media Contacts
Kevin Perlmutter, Xerox Europe, +44 (0) 1895 845 217, kevin.perlmutter@xerox.com