Grid Search and Categorisation Engine

GRACE

Scope

GRACE is the first practical GRID application, designed as a comprehensive retrieval system, tailored specifically to the needs of researchers in any field. It is a unique solution which introduces the concept of knowledge domains, federated search & distributed processing to information retrieval. It collects categories & presents information to the researcher through a simple and user friendly interface. The knowledge domains are defined within loadable ontologies in the form of domain specific Thesauri. High Energy Physics ontology has been used for evaluating the application.

GRACE is based on the ability of distributed systems such as the Grid to offer controlled and authorised sharing of resources – computational, storage, human. This is precisely what the concept of the knowledge discovery makes possible in the domain of the information retrieval (IR). It offers appealing and relevant content (documentation, knowledge basis, etc.), but also allows dynamic document publishing and storage.

Innovation and Functionality

GRACE uses an innovative approach in the integration of multiple content sources. The GRACE technology systematically harvests relevant information from those sources, applies strong natural-language processing to that information, and uses indexing to create a Knowledge Domain. Each Knowledge Domain is a complete virtualisation of relevant content sources, and incorporates the underlying semantics encapsulated in relevant ontologies.

These ontologies are used to query the content sources and index them using association with key terminology. This then allows the extraction of only the required information from extensive content sources, using the GRACE technology. The same ontologies can be used as a normal index, for querying, browsing and the presentation of search results.

The utilisation of strong natural-language processing methods allows GRACE to offer an unprecedented level of functionality to knowledge workers. Functionalities include: Semantic Indexing, Named Entity Extraction, Automatic Extraction of Relevant Lexical Patterns, Inference Searching, and Automatic Creation of Semantic Maps. All of this can be achieved by maximising the use of the existing resources, rather than requiring additional investments into computing infrastructure. The first actors in the GRID enabled information retrieval tools will be those who already possess huge data/document repositories, whatever the format.

continued overleaf V

Contract number
IST-2001-38100

Type of project
Cost-shared Research and Technological Development Action

Project coordinator
Telecom Italia

Contact person
Maurizio Cecchi
Telecom Italia
Via G. Reiss Romoli 274
10148 Torino
Italy

maurizio.cecchi@telecomitalia.it

Project website
http://www.grace-ist.org/

Maximum Community contribution to project
EUR 1 890 000

Project start date
1 September 2002

Duration
30 months

continued overleaf V

A partial example of some of the potential users involved could include the following:

1) Universities
2) Public and Private Research Institutes and Organisations
3) Central and Local Public Administration, such as national and regional health systems or registry offices for personal information records

Positioning

The GRACE project offers an innovative information retrieval system that uses Grid technologies for just-in-time, flexible allocation of data and computational resources. The GRACE search engine is built upon Grid Middleware and infrastructure developed by projects such as the LHC Computing Grid (LCG), the European DataGrid, and now, the EGEE project.

GRACE needs significant computational resources, and therefore Grid technologies, as a result of the computationally intensive tasks for the application of very strong natural-language processing methods for indexing purposes.

Target Users and User Benefits

The GRACE project is based on the principle that a content management system must not replace or unnecessarily enlarge existing resources, but rather, to allow a user to maximise their use of those resources.

Many applications can make use of the technology developed by GRACE. Examples of these applications include information update alarms, graphic visualisation of semantic maps, and multilingual applications. The variety of available functionalities allows users to design strong and reliable content management systems that are most suitable for their information needs.

Maturity and Availability of Tools

The GRACE project will be completed in March 2005. Instructions on how to install and configure both the GRACE application and the Grid infrastructure are accessible from the project website. The GRACE application already comes with a thesauri and classification scheme for High Nuclear Physics.

Compliance with Standards

The GRACE project has collaborated with the Grid Information Retrieval (GIR) Working Group of GGF, whose goal is to establish a specific set of requirements, an architecture, and detailed specifications for a particular Information Retrieval (IR) system within the Open Grid Services Architecture. A specific workshop dedicated to GRACE have been organised in GGF and the main finding has been included in GIR outcomes. Full exploitation of GRACE capabilities will be enabled after WSRF completion.

Interoperability

The GRACE project has adapted the Grid Middleware that was developed by the LCG project, which in turn uses the Middleware that was released by the European DataGrid project. The current release of the GRACE technologies is 100% compliant with the LCG-2 release.

The GRACE testbed benefits from the general services provided by GILDA (the EGEE dissemination Grid infrastructure), including the GILDA Virtual Organization, which gathers together all those wanting to experience Grid computing onto the GILDA testbed.

Project Partners

Organisation name and country

EUROPEAN ORGANISATION FOR NUCLEAR RESEARCH (CERN)	CH
GL2006EUROPE	UK
SHEFFIELD HALLAM UNIVERSITY	UK
STOCKHOLMS UNIVERSITET	SE
TELECOM ITALIA	IT
UNIVERSITAET STUTTGART	DE