Towards a Grid Enabled Knowledge Management Services

Jawed Siddiqi , Babak Akhgar and Mehrdad Naderi
School of Computing and Management Sciences
Sheffield Hallam University, UK
Email: {b.akhgar, m.a.naderi, j.i.siddiqi}@shu.ac.uk

ABSTRACT

Our research aims to further our understanding of grid enabled knowledge management services. We do so by developing the novel notion of knowledge management services that are derived from a synthesis of current taxonomies in knowledge management and grounded by categories obtained through an elicitation exercise involving domain experts. The latter activity is part of GRACE - Grid enabled Search & Categorisation Engine - an EU IST Framework V funded project. The paper is an invitation for partnership collaboration to realise Grid enabled Knowledge Management Services.

Keywords: Grid, Knowledge Management, KM Services

1. INTRODUCTION

Computing has witnessed an increasing pressure to evolve new abstractions and concepts that enable new modes of interaction that allow applications to access and share resources and services across distributed heterogeneous wide area networks. Work under two umbrella phrases have contributed significantly towards these required concepts and abstractions, namely, Semantic Web and Grid Computing. Following Foster et al. [2, 3] we characterise the aim of such work is to provide desired qualities of service, however measured, on resources assembled dynamically from enterprise systems, service provider systems and customer systems involving diverse, de-centralised and distributed hardware, software and human resources.

The computer began as stand-alone device for computation it has evolved significantly through various stages of interactive distributed processing that include: interactive processing, cluster computing and meta-computing. The term Grid computing, should therefore, be seen as part of an evolution of that work. For our purposes we characterise this computing paradigm in terms of a globally distributed heterogeneous network that is scalable and operates in a seamless manner to dynamically enable co-ordinated resource sharing and problem solving. From a virtual organisation viewpoint it is useful to categorise Grid computing into two different perspectives computational grid and access grid. Much of the past work and the current emphasis is on the computational grid (i.e. in which a group of scientists are able to access and share "unlimited" processing capability). Our focus is on the access grid comprising of a set of extensible services that can be dynamically aggregated in a variety of ways to meet the needs of virtual organisations, which themselves can be defined by the service they operate and share and possibly configured as a service.

The rapid growth of the Internet and the impressive success in the number of people using the World Wide Web has transformed the computer from a computational device to an entry point to a world-wide network of information exchange and business transactions. More recently web services is an important development that address the needs of heterogeneous distributed computing by enabling application/services that are self-contained and self-describing to be published, located and invoked across the web. They are part of an evolution of the vision created by Tim Berners-Lee of a Semantic Web that will enable automated information access and use based on semantics of data that can be machine processed. The semantic web is, therefore, an extension of the current web, in which information is given well defined meaning better enabling computers and people to work collaboratively. This realisation will enable a quantitatively different level of service. The key underpinning for semantic web is ontologies because they provide machine processable semantics for the data that is communicated between different actors (people or software agents). Many definitions exist for ontology for our purposes we take it as an explicitly defined characterisation of a shared conceptualisation. Some characterisations are formally defined and take us into axiomatic domain theories whilst others are simply a thesaurus of natural language terms. Though both are at quite different levels of formalisation, they do share the common goal of providing the required vocabulary of terms and relations to enable the communication of common conceptualisations.

Initially the Internet was first used for communication primarily amongst academics and Grid technology was first developed to enable scientists world wide to collaboratively share resources. We envisage that by combining the work in Grid Computing and Semantic Web internet use will become more important in commercial setting that include enterprise application integration and business to business partner collaboration. As Foster et al [1]. point out " Just As the WWW began as technology for scientific collaboration and was adopted for e-business, we expect a similar trajectory for Grid technologies". However, the work in both areas appears largely to be separate and bringing them together offers considerable potential that would enable Virtual Environments for Distributed Collaboration(VEDC) and Enterprise Application Integration (EAI). One clear motivation for both these provisions is to offer organisations real or virtual the ability to manage knowledge.

Some have coined the umbrella phrase Semantic Grid for the convergence of work in these two areas. Indeed as DeRoure et al. [4] point out that the visions offered by Grid Computing and Semantic Web have much in common, they can however be distinguished by the emphasis they respectively place on computation and semantics. Building on their perspective on the evolution of these two key areas we have articulated a two dimensional progression as follows. Grid computing focus has been the emphasis on increasing the computational power of distributed heterogeneous processing through a logical progression along a meta-computation dimension that involves; distributed computing, cluster computing, peer to peer computing, and Grid. Whilst the emphasis in Semantic Web has led to an increase in the richness and capability of the semantics through a logical progression along a semantics dimension that includes metadata, ontology, formal semantics and inference.

Our current research aims to describe the key requirements for provision of knowledge management (KM) services within a Grid environment. In this paper we present a generic conceptual model for a canonical set of requirements for the development of KM services within a Grid environment. This model is influenced by our research in an EU funded project GRACE (www.grace-ist.org) under "Information Society Technology Programme FP5. The GRACE project is an European Union (EU) RTD Framework V funded project. Partners in the project are Telecom Italia LABS, European Organisation for Nuclear Research (CERN), Sheffield Hallam University, Virtual Self, Stockholm University Library and Stuttgart University Library. It aims to deliver GRid enabled seArch and Categorisation Engine by making terabytes of information that already exists and is distributed on vast amounts of geographically distant locations highly accessible. Our focus and contribution in GRACE toolkit development is the identification, elaboration, validation and evaluation of the necessary grid enabled application technology to enable the next generation KM services to build upon core Grid services and focus on the design of KM technologies for knowledge worker, communities and organisations alike. The project aims to deliver the following results:

1. **Development of a distributed search engine**. An existing open source 3rd party solution will be enhanced and adapted to both Grid technology and the categorisation engine. The search engine will be enhanced by the categorisation output to allow for more accurate indexing and relevancy ranking.
2. **Adaptation of the categorisation engine** to the Grid infrastructure. The required development activities will make the categorisation fully distributed in order to minimise network latency and communication bottlenecks.
3. **Extensive testing and evaluation** of the distributed search and categorisation engine through utilising its experimental implementation. This will be performed on several beta-sites utilising “real world” documents from both organisational resources and scientific networks, especially through the 3 Scientific Libraries in Geneva, Stuttgart and Stockholm. Specific DATAGRID project Test Beds will be used.

2. Towards Grid enabled Knowledge Services

A first step toward characterising our understanding of knowledge management (KM) services and their potential application within VO as applied in organisational context. It is necessary for us to first look at some of core theoretical vocabularies of KM, namely data, information and knowledge. Siddiqi and Akhgar

[5] provide a simple summary of the main components of knowledge management as: **Data**: set of facts or observations that can be computationally processed. **Information**: human interpretation of the data, which will vary depending on viewpoint and manipulation of the data. **Knowledge**: an abstraction of a learning process which can be viewed as value added information".

The distinctiveness in the above perspective is the focus on the acquisition and assimilation processes, these involve in operational terms: search, categorise disseminate, share, harvest etc. This perspective is derived from our contextual understanding of organisations and communities who now require their information recourses to everyone everywhere, but they can also better harvest improvements and acquiring knowledge from everyone everywhere. Hence for our purposes we adopt Siddiqi and Akhgar's [6] definition of knowledge management as "a process of creating value added information (i.e. knowledge) so that the information available to all its users to help them do their work more effectively".

Based on this definition we characterise KM services in their most abstract computational form as the necessary component based applications required to make knowledge available to knowledge workers that require it, where they require it, when they require it in a virtually instantaneous manner and in the form in which they require it in order to increase performance and competitive advantage. An immediate advantage that accrues from this definition is that it enables a novel logical alignment for underpinning integration between KM services and existing services within a Grid environment in the context of VO realised through the paradigm of Service Based of Architecture (SBA).

2.1 Taxonomical Classification of KM Services

In order to produce this characterisation of these KM services we triangulated three literature sources to produce a taxonomic classification. The three sources were: Akhgar and Siddiqi's [6] taxonomy, Gartner Advisory groups [7] and the taxonomy produced by the European Knowledge Management Forum (EKFM) [8]. These three taxonomies were synthesised to produce a closely aligned taxonomical classification for our KM services. The classification is as follows: Gather, Contribute, Organise, Distribute, Collaborate, and Refine. These are elaborated upon in the proceeding section of this paper.

The GRACE requirements were elicited through a collaborative requirements engineering process involving an intensive distributed elicitation through a requirements inquiry cycle with domain specialists involving: three internal stakeholder groups and a focus group of twenty external organisations. For a more detailed explanation of the formal framework that was employed to ensure consistency in elicitation, classification, validation, conflict resolution, prioritisation, and definition and specification of the requirements. For the full set of requirements see GRACE project Web site. The resultant set of requirements was classified into categories elicited from the domain experts, these categories were mapped into the triangulated taxonomy. A partial mapping of categories to elements of the taxonomy to provide an illustration is as follows:

Gather:
Information resources harvesting such as information choice, access protocols (e.g. OAI and HTTP) and logical grouping of information.
Search activities such multi formatting of document (e.g. PDF and RDF), multiple information sources, uniform query syntax, quick search functionality, meta-data search capability and automated summary generation

Contribute:
User profiling includes following: user profile for each register user, persistence of pervious search result for each register user and sharing of user profile based on authorisation restriction.
Personalisation includes: creation of "my collection", manual editing of my collection and automated update of my collection.

Organise:
Classification includes: user specified classification schemes, dynamic classification and ontology-based merger of search results.
Categorisation includes following : automated inference of key phrases in the documents, dynamic categorisation of information resources, automated labelling of document clusters based on the key phrases, and clustering of documents based on existing taxonomies and ontologies.
Meta-data includes following: filtering, ordering and grouping of the original meta-data, version control and recognition of open standard meta-data format.

Distribute:
Client Access include: multiple concurrent users, complying with W3C standard, and access through a defined API.
User interface include: multiple user interface, screen help, progress indicator and automated generation of inferred key phrases.

Multilinguality include: multiple display (e.g. in English and German), cross lingual queries support, and support for translation of meta-data and documents.
Security include: anonymous and register user types, utilisation of Grid security policies and authentication for subscribe information resources.
Collaborate: To provide the necessary collaborative environment such user communities include community of practices, expert domains, user groups, corresponding documents to roles, search result based profiling, sharing and viewing of data, and configuration of group data access.

Refine: To provide a taxonomy for the organisation of information, more significantly allow knowledge workers to refine or analyse the contents of the knowledge base in different ways specifically provide lift and normalisation, semantic bridging (based on target ontology) and execution (through evaluation of semantic bridges).

Contextualisation of this taxonomical structure in the context of VO and specifically for our research implies three required conditions to progress towards a model for gird enabled knowledge services, they are:

First, these KM services are not stand alone, but rather they should be viewed and built on each other in compliance with: globally distributed resources; grid core services and grid base-line services. Moreover, they deliver value in proportion to their level of mutual integration and interaction based on aggregation of their components which governs the ontological drill down of taxonomies (e.g. naming convention), delegated authentication credentials, and time limitation.

Second; all the facets of these KM services from service based architecture perspective are enabled and their Quality of Services are heavily dependent on at least two other services. They are: data centric systems that store and/or manipulate enormous volumes of data and information with their interface definition and networking technologies and protocols that enable not only the transmission of data and information (e.g. HTTP and XML) but also multiple bindings for single interface through transport protocol and data encoding format (i.e. WSDL) and dynamic service creation.

Third, in the context of our characterisation of VO the principal characteristic of KM services is that any given scenario should act in accordance with the notion of heterogeneity, distributed transparency, replication transparency, authorisation and access transparency and adaptability characteristics view of Grid discussed earlier realised through core and base-line grid services available to the application layer.

3. CONCLUSION

The paper's primary contribution is to present the notion of KM services. These are formed from the synthesised taxonomy and grounded via requirements elicited, that have been classified in terms of KM categories, elicited from domain experts. The status of the research on the GRACE project is to propose a service based model which provides a framework to guide the toolkit that is currently under development.

The motivation to provide an early report of the results obtained so far is to provide an opportunity for KM and Grid communities to act as potential clients for the GRACE project. The knowledge management community could participate by providing scenarios that could be elaborated in terms of the KM model and subsequently executed on the Grace toolkit. The grid community can assess the feasibility of launching KM services on Grid platforms. Successful execution of the toolkit and its deployment on a grid platform would be a positive and significant step towards a grid enabled KM services. We invite interested reader to visit the website and contact us to collaborate in our mission to realise a grid enabled KM as a first step towards a semantic grid.

REFERENCES

[1] Foster, I. and Kesselman, C. (eds.) (1999), “The Grid: Blueprint for a New Computing Infrastructure”, Morgan Kaufmann.

[2] Foster, I. Kesselman, C., Tuecke, S. (2001), “The Anatomy of the Grid: Enabling Scalable Virtual Organizations”, Intl J. Supercomputer Applications.

[3] Foster, I. Kesselman, C. Nick, J.M. Tuecke, S. “The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration”,

[4] De Roure, D. Baker, M.A. Jennings, N.R. Shadbolt, N.R. (2003), “The Evolution of the Grid”,

[5] Siddiqi and Akhgar (2002), Source Code as Strategic Information based Resource, The British Computer Society Quality Special Interest Group's 10th Annual International Conference 'SQM

[6] Akhgar and Siddiqi (2001); A framework for the delivery of web-centric knowledge management applications, Internet Computing IC'2001. Vol 1 CSREA Press.

[7] Gartner Advisory groups (2001). Gartner Group Report on KM.

[8] European Knowledge Management Forum (EKFM 2001). IST Program