China's E-Science Knowledge Grid Environment

Hai Zhuge, *Chinese Academy of Sciences*

The Internet and World Wide Web are milestones in the history of information sharing. Scientists are increasingly relying on them to support their research.[1] They can communicate with each other through email and Net meetings and post their experiment data and research results on their personal Web pages or organizations' Web sites. They can also retrieve technical reports and papers from familiar online digital libraries or new ones they discover through search engines such as Google. But Web pages' exponential growth and intrinsic characteristics prevent people from effectively and efficiently sharing information. Much effort to solve this issue has achieved limited success.[2,3] And, because Web pages can't reflect machine-understandable semantics, the Web has difficulty supporting intelligent services.[1,4]

Knowledge is the basis of realizing intelligent services. As Vannevar Bush pointed out, sharing and inheriting knowledge can be challenging.[5] The Knowledge Grid, according to Fran Berman, will be a mechanism that can synthesize knowledge from data through mining and reference methods and enable search engines to make references, answer questions, and draw conclusions from masses of data.[6] The Knowledge Grid will exploit research toward the next-generation Web, using it to build a more efficient and effective intelligent application platform.

The next-generation Web

To overcome the Web's shortcomings, scientists are working toward the next-generation Web, with most research focusing on two approaches. The first is to improve the existing Web, which includes the Semantic Web, Web Services, and Web intelligence.

The Semantic Web's goal is to establish machine understandable Web resources. Researchers in this area plan to accomplish this by creating ontology and logic mechanisms and replacing HTML with markup languages such as XML, RDF, OIL, and DAML.[7]

Web Services aim to provide an open platform for developing, deploying, managing, and interacting among globally distributed e-services based on Web standards such as UDDI (universal description, discovery, and integration) and WSDL (Web Services Description Language, www.w3.org/TR/wsdl). Web Services also let developers integrate services that reside and run in different places. The intelligent agent technique is an important means of implementing active Web Services.[7]

Web intelligence uses AI and information-processing technologies, including text mining and information extraction, to enable the Web to support some intelligent services.[8]

The second approach aims to establish a new application platform independent of the Web: the Grid. The global Grid (www.gridforum.org) promotes sharing, managing, coordinating, and controlling distributed computing resources, such as machines, networks, data, and any devices. The Grid's goal is to enable compatible devices to be plugged in anywhere on the Grid and be guaranteed the required services, just as the power grid does with electricity.[9] In fact, researchers borrowed the term "grid" from the power grid to refer to computing-power clustering when the concept appeared in 1995. Grid computing almost excludes Web technologies.

Taking the Grid a step further, the *Semantic Grid* attempts to incorporate the advantages of the Grid, the Semantic Web, and Web Services to extend the Grid in semantics and enhance the Semantic Web's computing power (www.semanticgrid.org). The Grid architecture already has shifted to the service-oriented Open Grid Services Architecture, where we can see some Web Services features.[10] Computer scientists are also exploring ideal computing models and resource organization models.

What is the Knowledge Grid?

Traditional natural language processing, speech and handwriting recognition, scientific computing, formal semantics, and security—these aren't major concerns of the Knowledge Grid. The Knowledge Grid will go beyond traditional and improved information retrieval, filtering, and mining and question-answering techniques.

The Knowledge Grid is an intelligent, sustainable Internet application environment that enables people or virtual roles

1094-7167/04/$20.00 © 2004 IEEE
Published by the IEEE Computer Society

Relevant Research and Plans in China

Established in July 2001, the China Knowledge Grid Research Group (http://kg.ict.ac.cn) has been pursuing the best solutions for three fundamental issues of the future interconnected environment—normally reorganizing, semantically interconnecting, and dynamically clustering and fusing globally distributed resources.[1–3] The group's significant research progress includes:

- Implementing an experimental Knowledge Grid system, which includes a multidimensional knowledge space model and a knowledge operation language and its implementation mechanism. The mechanism features a 3D semantic-space browser and a knowledge flow model for realizing effective knowledge sharing in virtual organizations.[4]
- Establishing a theory for normally organizing semantic resources. The theory integrates the multidimensional resource space model and the semantic-link network model. Key to the theory are the semantic normal forms based on orthogonal classification semantics and semantic links. The theory supports a single semantic point of access, semantic clustering, and complete knowledge services.
- Developing algorithms and software tools for building semantic-link networks, ranking resources in the networks, matching between semantic views, and refining knowledge. They also developed algorithms and tools for analogical reasoning and computing models on semantic-link networks, semantic profiling, and defining complex service requirements by using component-based technique.
- Creating a single semantic image technique based on normalization theories and using it to realize Web-based image clustering and service integration. They also proposed semantic-based approaches for integrating knowledge flows and data flows, and proposed the Active Document Framework to enable active information services.[5]
- Proposing a unified resource model and future interconnected-environment model with social and ecological characteristics.[6,7]
- Making progress on Knowledge Grid applications in cooperative research and education.

The China Knowledge Grid Research Group leads a cooperative team working for the China Semantic Grid Research Plan—a five-year research plan supported by the Ministry of Science and Technology of China.[8] The National Science Foundation of China has also launched a five-year e-science research plan.

References

1. H. Zhuge, "Semantics, Resources and Grid," *Future Generation Computer Systems*, vol. 20, no. 1, 2004, pp. 1–5.
2. H. Zhuge, *Knowledge Grid*, to be published, World Scientific, 2004.
3. H. Zhuge, "A Knowledge Grid Model and Platform for Global Knowledge Sharing," *Expert Systems with Applications*, vol. 22, no. 4, 2002, pp. 313–320.
4. H. Zhuge, "A Knowledge Flow Model for Peer-to-Peer Team Knowledge Sharing and Management," *Expert Systems with Applications*, vol. 23, no. 1, 2002, pp. 23–30.
5. H. Zhuge, "Active E-Document Framework ADF: Model and Tool," *Information and Management*, vol. 41, 2003, pp. 87–97.
6. H. Zhuge, "Clustering Soft-Devices in the Semantic Grid," *Computing in Science and Eng.,* vol. 4, no. 6, Nov./Dec. 2002, pp. 60–62.
7. H. Zhuge, "Eco-Grid: A Harmonious Evolved Interconnection Environment," to be published in *Comm. ACM*.
8. H. Zhuge, "Semantic Grid Research in China: Scientific Issues, Infrastructure and Methodology," to be published in *Comm. ACM*.

(mechanisms that facilitate interoperation among users, applications, and resources) to effectively capture, publish, share, and manage explicit knowledge resources. It also provides on-demand services to support innovation, cooperative teamwork, problem solving, and decision making. It incorporates epistemology and ontology to reflect human cognition characteristics; exploits social, ecological, and economic principles; and adopts the techniques and standards developed during work toward the next-generation Web.

Efforts toward the next-generation Web provide the Knowledge Grid with many possible techniques and implementation platforms. The Grid isn't the only platform for realizing the Knowledge Grid. In fact, the Knowledge Grid should absorb the Grid's ideals (the meaning of "grid" in the Knowledge Grid is broader than in Grid computing).

The Knowledge Grid's distinguishing characteristics

Five characteristics set the Knowledge Grid apart from other technologies.

First, people can access and manage knowledge distributed around the world through a single semantic entry point without knowing the required knowledge's locations.

Second, relevant knowledge distributed around the world can be intelligently clustered and fused to provide on-demand knowledge services with underlying reasoning and explanations. To help achieve this, knowledge providers should provide metaknowledge (knowledge of using knowledge). A type of uniform resource model will help encapsulate knowledge and metaknowledge to realize active clustered and fused knowledge services.

Third, people or virtual roles can share knowledge and enjoy reasoning services in a single semantic space (on the basis of mapping, reorganizing, and abstracting between semantic spaces) where no barriers to mutual understanding exist. The Knowledge Grid also enables pervasive knowledge sharing.

Fourth, the Knowledge Grid can gather knowledge worldwide and guarantee proper knowledge closure (a minimum complete knowledge set) for solving problems. To achieve this goal, we need to create new knowledge organization models.

Finally, in the Knowledge Grid environment, knowledge isn't statically stored; it can dynamically evolve to keep up-to-date. This means the Knowledge Grid's knowledge services can continuously improve during use.

Issues in Knowledge Grid research

Research in this field concerns five issues: The first is theories, models, methods, and mechanisms for capturing and representing knowledge. Capturing knowledge on the Knowledge Grid occurs in two ways. In one method, people obtain knowledge from each other directly or from knowledge resources others publish, and they publish their new knowledge on the Knowledge Grid. In the other method, the Knowledge Grid obtains knowledge from resources such as data, text, and image through extracting, mining, inducing, deducing, synthesizing, and so on. The Knowledge Grid should be able to assist people or virtual roles to effectively capture and publish knowledge in a machine- and human-understandable form (or one that's easily translatable to a human-understandable form). So, we should build an open primitive set to realize knowledge representation. These primitives should be able to represent multigranularity knowledge and derive new knowledge by operations on the primitives.

The second issue is knowledge visualization and innovation. This mainly concerns an intelligent user interface (for example, a semantic browser or a knowledge browser) that lets people share knowledge with each other visually. Two ways to bridge the gap between knowledge representation and visualization are cognitive maps and the semantic link network (www2003.org/cdrom/papers/poster/p148/P148-Zhuge/P148-Zhuge.htm). The interface should implement the Knowledge Grid's distinguishing characteristics and be able to inspire knowledge innovation through analogy and induction mechanisms and organizational rules.

The third issue is effective propagation and management of knowledge in dynamic virtual organizations. This goal requires us to eliminate redundant communication between team members. The China Knowledge Grid Research Group proposed a knowledge flow network to achieve effective knowledge sharing in dynamic virtual teams. (See the sidebar "Related Research and Plans in China" for additional work in this field.)

The fourth issue is effective knowledge organization, evaluation, refinement, and derivation. Knowledge should be organized in semantic normal forms to guarantee efficient retrieval and correct operations. The Knowledge Grid should be able to eliminate redundant knowledge and refine the contained knowledge to maintain a reasonable expansion of useful knowledge. It should also be able to derive new knowledge from existing well-represented knowledge, use cases, and raw knowledge material such as text.

The final issue is knowledge association and integration. The Knowledge Grid should be able to associate and integrate knowledge resources of different levels (for example, concepts, axioms, rules, and methods) and different domains to support cross-domain analogy, problem solving, and scientific discovery.

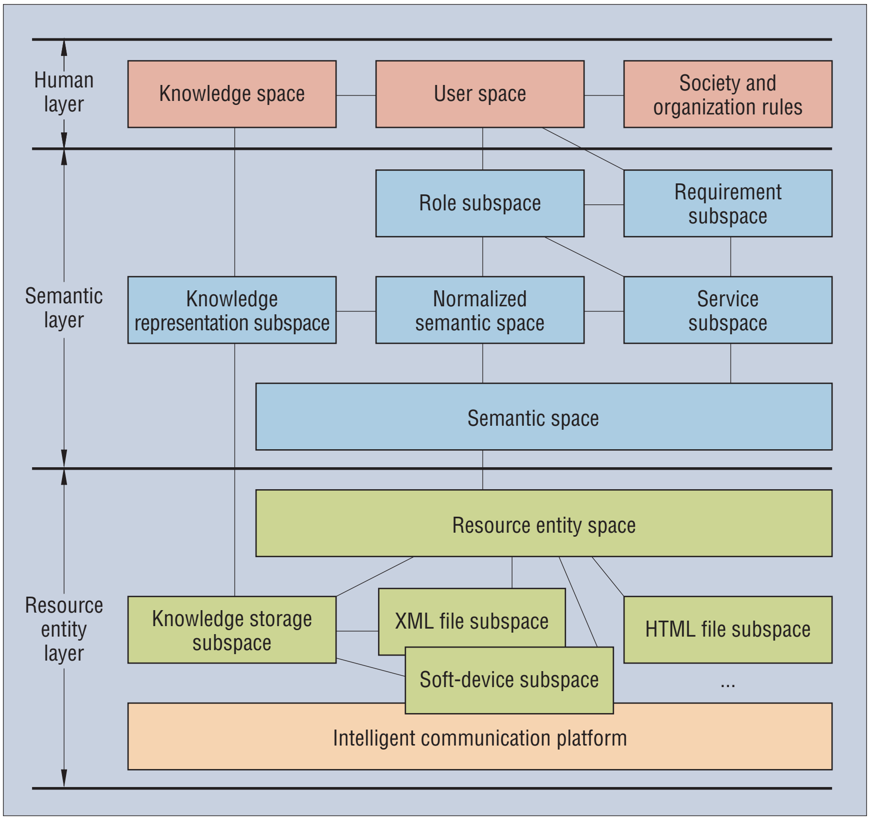

Figure 1. A multispace architecture model of the Knowledge Grid.

Knowledge Grid architecture: A multispace model

The Knowledge Grid architecture comprises three layers (see Figure 1).

The *human layer* reflects the Knowledge Grid's social and human behavior characteristics. It includes a *knowledge space* containing explicit knowledge from all participating human users, a *user space* containing user information, the *society and organization rules*, and a viewpoint of social value for evaluation. Humans use natural language to convey explicit knowledge in certain media, such as text files. The Knowledge Grid might require users to provide an attachment of metaknowledge, background knowledge, and common sense that current media omits.

The *semantic layer* includes the *knowledge representation subspace*, which expresses users' knowledge in a machine-understandable form, and the *role subspace*, where users play versatile roles according to their intentions. This layer also contains the *requirement subspace*, where users and services express requirements through roles, and the *service subspace*, where all services are autonomous and self-represented. The system evaluates services according to their usefulness under a viewpoint of social value.

The knowledge representation subspace is organized in terms of the *semantic space* and the *normalized semantic space*. The semantic space can be normalized to achieve such qualities as completeness, integrity, effectiveness, and correctness. The semantic link, ontology, and name space belong to the semantic space.

The semantic layer keeps human users from directly interacting with the *resource entity layer*, so they don't need to be concerned with the resources' form and location. It can also isolate any changes

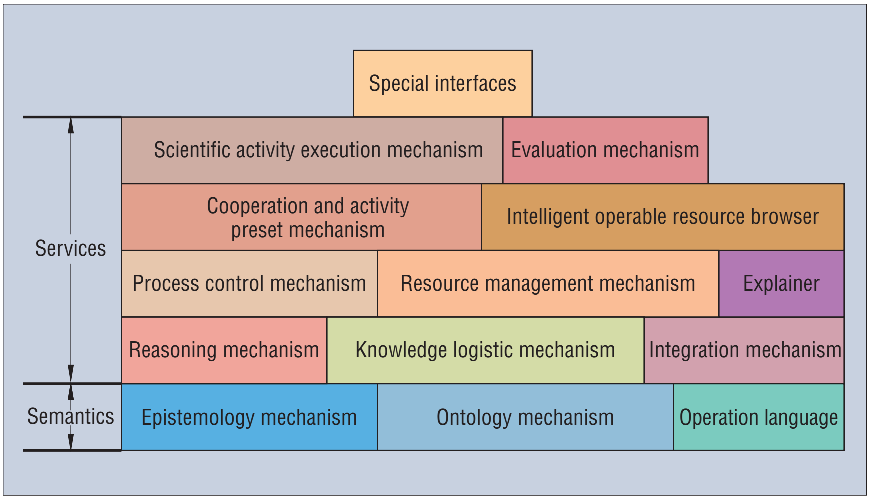

Figure 2. IMAGINE's main services and their relationships.

in the resource entity layer. So, the semantic layer can remain stable as the communication platform and representation basis evolve, such as through updates to markup languages.

The resource entity layer contains the *intelligent communication platform* and the *resource entity space*, which includes the *knowledge storage subspace*, the *XML file subspace*, the *HTML file subspace*, the *soft-device subspace*, and so on. (The Knowledge Grid should take advantage of and be compatible with the Web.) The knowledge storage subspace is achieved through the primitives defined in the semantic space. The China Knowledge Grid Research Group has proposed a soft-device model for generalizing and encapsulating various types of resources, including a reasoning mechanism, knowledge resources, and data resources. The intelligent communication platform incorporates the advantages of client-server, Grid, and peer-to-peer computing, and supports mobility and correctness to implement the Knowledge Grid's distinguishing characteristics.

In the Knowledge Grid architecture, any user or service can select a role with respect to the application field and enter a requirement in the requirement space. The services in the service space will actively search the requirement space to find the suitable requirement and then provide the best service for the user (or the virtual role) requiring the service. Services acting as brokers select the best service or compose relevant services to provide a uniform service. Service integration involves *data flow integration* and *knowledge flow integration* to achieve a single semantic image in interactions among services. This is especially true for those intelligent services that carry out reasoning according to the knowledge in the knowledge space.

Comparing the Web and the Knowledge Grid

Both the Web and the Knowledge Grid are useful ways to access information. But, as the following example shows, the Knowledge Grid will enable better information services.

Users seeking medical information on the Web can use search engines to locate suitable Web sites or rely on their memory of URLs for hospital or health sites. The Web contains more than 82 million health sites, and time-consuming, site-by-site browsing and search results with large amounts of useless information often annoy users. Furthermore, users might be confused by the various doctors' opinions and concerned that their consultations are based on dated knowledge. Those researching related symptoms usually can obtain consultations on separate symptoms. Moreover, the whole search and consultation process might miss some experts, especially those who specialize in uncommon diseases.

The Knowledge Grid will dramatically improve this situation. It will accurately and completely locate all relevant knowledge, cluster and synthesize the consultation results, and actively push the results to users according to their semantic profiles describing their illnesses.

Users will obtain an explanation of the results with underlying reasoning based on the clustered knowledge. During reasoning, the system will also consider relationships between symptoms. The consultation results will also adapt to profile changes. Knowledge provided by different doctors worldwide will be refined, checked, and evaluated on usefulness, consistency, and up-to-date information.

The Knowledge Grid will derive new knowledge from existing knowledge, patient feedback, and medical textbooks, papers, and other related information. Patients will also be able to enter their symptoms at a single semantic point and receive an instant consultation. The results will include several treatment options, taking into consideration such factors as cost, staff skill level, waiting time, and road traffic.

China's e-science Knowledge Grid environment

The Knowledge Grid also has advantages in scientific research. China's e-science Knowledge Grid environment IMAGINE (Integrated Multidisciplinary Autonomous Global Innovation Networking Environment) is putting Knowledge Grid ideals into practice.

Basic requirements

Scientific research combines research resources and research activities. Research resources include scientific documents, such as technical reports and papers, experiment data, and instruments. Research activities include conducting surveys and experiments; collecting references; and analyzing, discussing, evaluating, and submitting results. Cooperation between scientists intentionally or unintentionally happens through resource sharing. Research activities can be loosely or tightly coupled and can belong to one person or a research group. Also, the personal or team activity flow, data flow, and knowledge flow connect relevant research activities.

High-level scientific research activities involve knowledge, methodology, cognition, and creative thinking. IMAGINE aims to speed up the process and increase knowledge inspiration, generation, innovation, propagation, fusion, and management in cooperative research.

Services

IMAGINE provides two main service categories, each aiming to increase efficiency and effectiveness. Basic services help scientists perform research activities. Advanced services manage activity flow, data flow, and knowledge flow and aid scientific discovery.

The following additional services support the two categories by performing specific functions (see Figure 2):

- *Special interfaces* support domain-specific human-computer interaction.
- A *scientific activity execution mechanism* executes and monitors activities submitted by the interfaces.
- An *evaluation mechanism* evaluates scientific research results according to a set of criteria, such as citation rates, comments, reviews, invited talks, and awards.
- A *cooperation and activity preset mechanism* accepts the definition of the team, including setting up research groups and resource-sharing activities.
- An *intelligent operable resource browser* supports common scientific activities such as document retrieval.
- A *process control mechanism* monitors and controls the predefined cooperation processes.
- A *resource management mechanism* uniformly manages research resources.
- An *explainer* explains the semantics of operations submitted by the interfaces and browsers according to the operation language's semantics.
- A *reasoning mechanism* provides reasoning services according to requirements, activity, and relevant knowledge.
- A *knowledge logistic mechanism* actively discovers knowledge requirements and knowledge sources, planning knowledge provision chains and carrying out knowledge provision.
- An *integration mechanism* provides semantic-based resource integration services including knowledge integration, data integration, and file integration. It can gather relevant resources to form complete content that matches the requirements of multiple views.

The epistemology and ontology mechanisms and operation language belong to the semantic space.

Incorporating epistemology and ontology

Many scientists are investigating and building ontology mechanisms for objectively explaining objects' semantics and their ties. An ontology reflects people's consensus on semantics. In fact, epistemology and ontology are two inseparable profiles of the unified human cognition process. The epistemology mechanism is a semantic description tool that reflects human subjective cognition. Different people can have different epistemology of the same object or event. The mechanism helps humans and agents understand, generate, and describe new knowledge as they're sharing resources.

The China Knowledge Grid Research Group uses a kind of *dynamic hierarchical semantic-link network* to reflect a resource provider's epistemology. The network uses a set of semantic-link primitives to connect nodes (resources, concepts, and networks) under time constraints. It can carry out four basic types of reasoning:

- Dynamic-link reasoning based on the transitivity of semantic links
- View-matching reasoning based on the semantic-inclusion relationship between network views
- Analogical reasoning based on structure mapping between network views
- Abstraction of a given semantic-link network

We've developed a semantic-link-making tool for users to conveniently describe their understandings of provided resources and background knowledge. The epistemology will accompany the resources, record different understandings, and continue to evolve as people's understandings deepen.

The Knowledge Grid is the next-generation Web's high-level infrastructure. It provides on-demand knowledge services through incorporating human behavior and cognition, various knowledge resources, machine-understandable semantics, and entity resources.

The Knowledge Grid infrastructure supports e-science through a set of relevant application services and semantic resources. It not only makes accurately sharing research resources and knowledge more convenient, but it also supports cooperative research and scientific discovery. The cooperation between intelligent application services and the Knowledge Grid forms an Intelligent Grid Environment that in turn provides intelligent and active services.

Acknowledgments

China's National Grand Fundamental Research 973 Program and National Science Foundation supported research for this article.

References

1. J. Hendler, "Science and the Semantic Web," *Science*, vol. 299, no. 5606, Jan. 2003, pp. 520–521.
2. L.A. Adamic and B.A. Huberman, "Power-Law Distribution of the World Wide Web," *Science*, vol. 287, no. 5461, Mar. 2000, p. 2115.
3. J. Kleinberg and S. Lawrence, "The Structure of the Web," *Science*, vol. 294, no. 5548, Nov. 2001, pp. 1849–1850.
4. T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web," *Scientific Am.*, vol. 284, no. 5, May 2001, pp. 34–43.
5. V. Bush, "As We May Think," *The Atlantic Monthly*, vol. 176, no. 1, 1945, pp.101–108.
6. F. Berman, "From TeraGrid to Knowledge Grid," *Comm. ACM*, vol. 44, no. 11, Nov. 2001, pp. 27–28.
7. J. Hendler, "Agents and the Semantic Web," *IEEE Intelligent Systems*, vol. 16, no. 2, Mar./Apr. 2001, pp. 30–37.
8. N. Zhong, J. Liu, and Y. Yao, "In Search of the Wisdom Web," *Computer*, vol. 35, no. 11, Nov. 2002, pp. 27–31.
9. I. Foster, "Internet Computing and the Emerging Grid," *Nature*, 7 Dec. 2000, www.nature.com/nature/webmatters/grid/grid.html.
10. I. Foster et al., "Grid Services for Distributed System Integration," *Computer*, vol. 35, no. 6, June 2002, pp. 37–46.

Hai Zhuge is a professor at and the executive director of the Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences. Contact him at zhuge@ict.ac.cn.