 Hai Zhuge

Semantic Grid: Scientific Issues, Infrastructure, and Methodology

Creating a new interconnection environment incorporating the Internet, sensor networks, mobile devices, and the interconnection semantics.

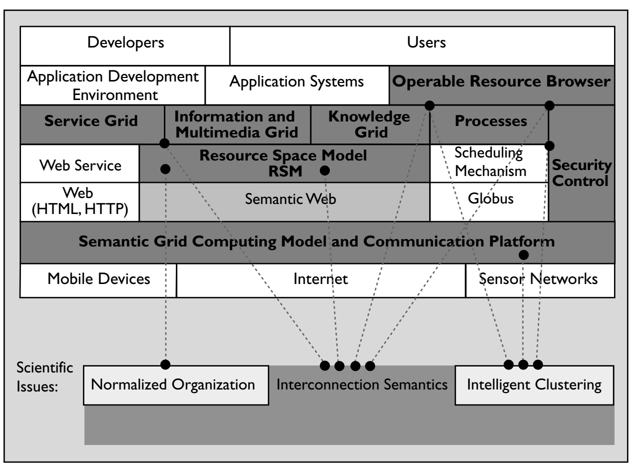
Database theory and systems, a seminal development of information management, has profoundly affected computer applications for nearly 40 years. The Internet is a brilliant innovation, with unprecedented social and economic influence, but the lack of a theoretical foundation has limited its full potential.

The Internet and the Web share four characteristics: rapid expansion of versatile resources and users, microstructured and macro-scale-free resource organization, inequality of information reputation (distribution) [1, 4], and human-understandable, not machine-understandable, semantics. These characteristics cause the difficulty in accurate, effective, efficient, and safe use of globally distributed resources.

IT professionals are attempting to create a new interconnection environment by adding machine-understandable semantics [2] for more efficiently sharing, managing, coordinating, scheduling, and controlling distributed computing resources [3].

The Semantic Grid

The Semantic Grid is an Internet-centered interconnection environment that can effectively organize, share, cluster, fuse, and manage globally distributed versatile resources based on the interconnection semantics.



Infrastructure of the Semantic Grid and its scientific issues.

Scientific Issues. *Interconnection semantics* is the study of the semantics in the interconnection environment for supporting intelligent applications by meaningfully interconnecting resources in semantic spaces where machines and humans can understand each other. Different from the semantics of natural and programming languages, interconnection semantics concerns:

- Single Semantic Image. Mapping versatile resources into a single common semantic space to enable resource utilization to be independent from their type and location.
- Transformation and consistency between semantic spaces: classification semantics, layout semantics, logical semantics, process semantics, and concurrent semantics.
- Realize semantic-based storage and retrieval in scalable large-scale network environments.
- Semantically rich interface between humans and resources and between resources.
- Evolution of semantics and the impact and rule.

Normalized resource organization. The study of organizing resources in semantic normal forms to eliminate redundant, disorder, and useless resources to ensure the correctness and accuracy of resource operation, and to realize complete and effective resource sharing. This is the issue of abstracting a disorder system and then appropriately reconstructing it in a normalized space. The reconstruction should avoid information loss and guarantee the correctness of operations. The simplification of complex systems is a fundamental task of science.

Intelligent clustering and fusing pertains to the study of self-organization and complex system optimization, specifically concerning:

- Coupling: to couple versatile resources and "push" them to those needs according to the interconnection semantics.
- Clustering and fusing: relevant resources dynamically get together and fuse to provide appropriate on-demand services based on constraints and multilevel semantic understanding.
- Autonomous and collaboration: to support autonomous and collaboration activities, to reflect evolution information, and to evolve according to changing situations.
- Efficiency, trust, and security: to organize resources for balancing between the efficiency, trust, security, and quality of services.
- Ideal computing model: the Grid focuses on scheduling and controlling resources in a stable middle-scale network. Peer-to-peer is able to effectively share files in a dynamic and large-scale network. Due to the lack of semantics, it is difficult for them to support high-level intelligent applications. The ideal model should take advantage of the Grid and peer-to-peer to support intelligent applications based on the interconnection semantics.

Infrastructure. The Semantic Grid should include the eight components labeled in boldface type in the figure here, where the Semantic Resource Space Model semantically and normally organizes versatile resources. The design needs the theory (normal form theory and integrity constraint theory), design criteria, method, and development tools. A resource operation language is needed to correctly operate resources. The communication basis of the Semantic Grid is the incorporation of the Internet, sensor networks, and mobile devices.

The Service Grid, Knowledge Grid [5], Multimedia Information Grid, and the cooperative processes are in the high level of the Semantic Grid. The Semantic Resource Space Model is responsible for managing all these resources. The Service Grid is the logistical and intelligent pushing mechanism based on the organization and integration of distributed services. The Multimedia Information Grid is the organization and the provider of multimedia information. The cooperative processes support the definition, management, and integration of business processes. The process definition and verification concerns behavior semantics, workflow, and knowledge flow [6]. Security control guarantees the safety of the Semantic Grid operation. The developers work with the application development environment, and the end users work with the application systems and the operable resource browser, which can obtain the complete support from the Semantic Grid during use.

Methodology. *Cross the boundaries of disciplines*. The establishment of the Semantic Grid requires research across epistemology, linguistics, culture, art, cognitive science, and system methodology.

Abstraction and specialization. Abstraction investigates common characteristics and rules of versatile resources by generalization, and then proposes the uniform model and method. Specialization investigates the special rules of resources to properly integrate and couple resources.

Balance the interests of different aspects. The Semantic Grid should take advantage of both centralization and decentralization, and balance autonomous, self-organization and normal organization. The mobility, completeness and correctness of services should also be balanced.

Inherit and innovation. The Semantic Grid should inherit current technologies. Any Internet and Web applications should be able to seamlessly migrate onto the Semantic Grid environment. It should absorb the advantages of the Grid, Semantic Web, and Web services, and go beyond their scopes by adopting a new computing model, communication platform, and normal resource organization model.

Application in China: Dunhuang Culture Grid. Along with the basic research (see kg.ict.ac.cn/publications), the China Knowledge Grid Research Group is conducting applications in China's e-science and e-culture. The group is developing the Dunhuang Culture Grid for exploring, exhibiting, and protecting ancient Dunhuang cave culture, which represents the world's culture heritage in the desert of west China. In the near future, people around the world will experience the glorious ancient culture in the form of color statues and wall paintings in over 900 caves as well as the precious calligraphies with rich content services via the Internet. **C**

References

1. Adamic, L.A. and Huberman, B.A. Power-Law distribution of the World Wide Web. *Science*, *287*, 24 (2000), 2115.
2. Berners-Lee, T., Hendler, J., and Lassila, O. Semantic Web. *Scientific American 284, 5* (2001) 34–43.
3. Foster, I. Internet computing and the emerging grid. *Nature 408*, 6815 (2000).
4. Kleinberg, J. and Lawrence, S. The structure of the Web. *Science 294*, 30 (2001) 1849–1850.
5. Zhuge, H. China's e-science knowledge grid environment, *IEEE Intelligent Systems 19*, 1 (2004) 13–17.
6. Zhuge, H. *The Knowledge Grid*. World Scientific Publishing Co., 2004.

Hai Zhuge (zhuge@ict.ac.cn) is the chief scientist of the China Semantic Grid project (No. 2003CB317001), a five-year project of the China National Basic Research and Development Plan Program.
