Bull. Nov. Comp. Center, Comp. Science, 22 (2005), 13–22
© 2005 NCC Publisher

An intellectual network system for information support of innovative activity (innovative portal)*

A.S. Alekseev, M.B. Ostapkevich, S.V. Piskunov

Abstract. The architecture of an innovative portal based on a fine-grain modularity and an assembly technology is presented. The portal resources and services for the interaction between participants of the innovative activity within a common information space are realized. The investigation of the system and the creation of its prototype have shown that the informational, computational, and communication capabilities of the network infrastructure should be increased. The scalability of the portal calls for its installation in the network of Supercomputer Centers, RAS is based on the GRID technology. The implementation of such a scheme would make it possible to carry out complicated innovative projects by using the entire set of supercomputers.

1. Introduction

The network information technologies provide acquisition, transmission, processing, and integration of knowledge, whose function is similar to that of the blood vessels and the nervous system in the rapidly developing world of innovative economy.

The construction concepts of a network system for the support of innovative activity as well as peculiarities of its architecture are discussed in [1–3] and on the Web site http://sinin.nsc.ru. Let us recall the basic statements of the project. The reason for the development is an ever increasing understanding of the fact that "it is the ideas, but not the resources that are of value" [4]. Currently, the need for the construction of the innovative structure at the national level is perceived (see publications at http://www.mag.innov.ru). The network system ought to provide the widest and "transparent" information link between scientists, engineers, and investors. On the one hand, the network information system (hereinafter referred to as innovative portal) includes tools for the construction and integration of various information resources (data and knowledge bases) related to innovations. On the other hand, the innovative portal has flexible tools for personal participation with different roles in the innovative activity. All the tools proposed provide the organization of an active information environment that decreases the information gap between science and industry. These tools provide a conventional support for innovative activity in the form

*Supported by the Russian Foundation for Basic Research under Grants 03-07-90302, 05-07-98003, IP 111 SB RAS, IP 21.2 RAS.

of hi-tech developments for potential investors or providing help in finding an investor for scientific institutions: they create an active environment for solving complicated research R&D tasks by virtual (network) teams.

The innovative portal combines an information portal and a corporate one and a constructor of portals as follows. The system, being an information portal, can generate a corporate portal by submitting a request with a list of specifications. The corporate portal generated provides the team members with services and applications depending on their roles (in the project) and settings according to specifications in the request. Many teams can simultaneously be formed. Thus, numerous specific corporate portals can co-exist. These interact by means of a dedicated protocol. The system has a distributed architecture and is located on a set of servers interconnected by a high-speed network.

Many of the information and the corporate portals have been developed and are under construction in Russia and worldwide. They integrate information resources and unify the access to them in various fields of science, culture, healthcare, industry, trade, politics, and education. They are based on data and knowledge management systems. The portal construction technology is well developed. Nevertheless, no systems were found (either in Internet, or in publications) that integrate heterogeneous information resources and, at the same time, help to organize innovative projects and realize them by Internet-based teams. It seems very important to unite the computational and the informational resources in the SB RAS and organizations of the Novosibirsk region within the system being constructed. This would allow one to make use of the innovative potential of the SB RAS as well as on the regional scale in order to pass from projects based on natural resources into the hi-tech ones.

The system is under development: its architecture is being refined, new subsystems and modules are being constructed to bring new services to the users. A description of the system architecture is presented in this paper. The changes made in the system after previous publications are discussed. It is especially important, that the scope for the further development of the system as distributed system of automated simulation for innovative projects is outlined. The distributed system must become a part of the Federal Network of Supercomputer Centers for the informational and the computational support of innovative activity.

2. Architecture of the innovative portal

2.1. Requirements for the portal functionality. The destination of the portal outlined in the introduction determines the requirements discussed in detail in [1–3]. We will emphasize only those that have the greatest influence on the portal architecture. The portal must provide:

1. Online access to great amounts of informational resources (papers, books, patents, thesis, preprints, legal acts etc.) with the help of various classifications (adjustable by the user to a specific application);
2. Acquisition of information from heterogeneous distributed sources of knowledge by a user-defined set of criteria;
3. Uniting the innovative activity participants into virtual teams, providing them with a comfortable (as if they are at the same physical location) and adjustable (according to their role) environment for communications, and with access to documents and results of simulation, etc.;
4. Network team members with a tool (a portal constructor) to organize their own corporate portal (with tuning by the prototype) within the innovative portal and with tools (editors) to construct personal workplaces and create information collections (depending on the role in the innovative process);
5. Support of "simulative" innovative design, which implies a wide and ever increasing use of mathematical simulation results.

Let us clarify the latter item. The idea of "simulative" design is to substitute the construction of a physical prototype of an object (automobile, airplane, nuclear bomb, informational/computational network, exploitation of oil and gas provinces, dams, industrial enterprises, medicaments, etc.) for a object simulation software model. The simulation based on an object model is performed for studying the behavior of real objects. Such a kind of design is the only choice if physical tests are impossible for technical, ecological, legal, or other reasons. Even if physical tests are feasible for some objects, a shorter design time (manifold) and the quality perfection (because many situations can be simulated, but only a limited number of physical tests at the last phase of the design process can be carried out) dictate the use of the simulative design.

2.2. A description of the portal architecture. The description is presented in [2, 3]. The development of the architecture was carried out to meet these requirements. The portal is constructed with the help of a library for parallel and distributed applications with an open architecture. The library was developed at the ICM&MG SB RAS, and is called the Dynamically Configurable Modular System (the DCMS) [5, 6]. The portal is divided into the following three levels: the DCMS, the portal kernel, and the portal kernel extensions. When the library was created, the technology of module construction was refined, and now it can be used not only in the DCMS, but also in the entire portal project. The technology is based on a fine-grain modularity. Usually, a module is divided into a few files with

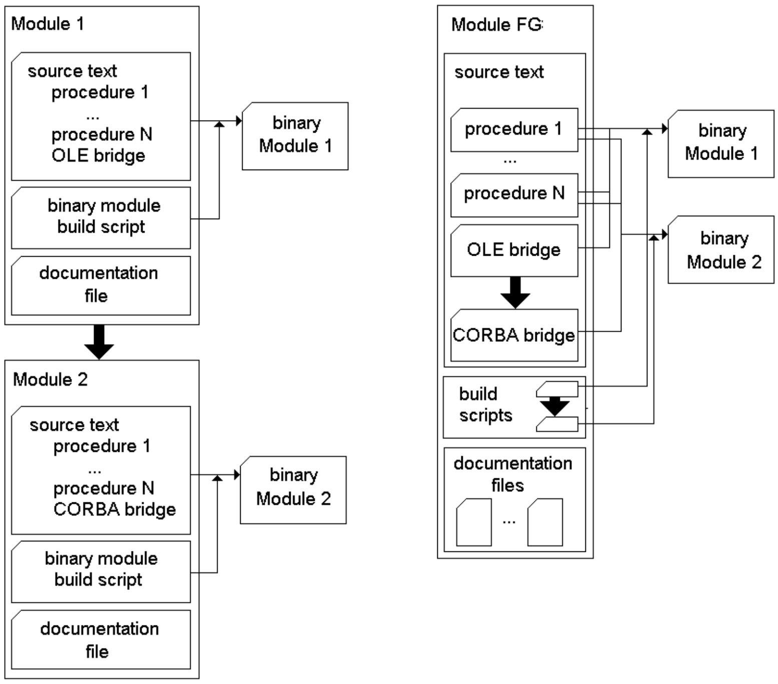

Figure 1. Conventional and fine-grain structures of the module

implementation and description. In the fine-grain modularity, the division is done into pieces with interfaces, descriptions of variables, single procedures, documents, and scripts for the construction of binary modules.

In addition, the unified naming for procedures, variables, and modules is used. Each identifier contains: a) the full module name and its version, b) the object type, c) the meaningful name. Such an approach makes it possible to assemble modules from the source text fragments according to an assembly programming paradigm [7, 8]. The visible drawback of this approach, which is the big number of source files, is fully compensated by its advantages: 1) the usage of small fragments facilitates, the formation of new modules, and the reconstruction of existing ones, 2) many routine operations with modules in the process of their design (from the binary module compilation to generation of different kinds of documents) are readily automated. Practical implementation of the the DCMS has demonstrated these above advantages. A sample of the new module construction is shown in Figure 1 (the left is a classic module, the right is a module with fine-grain assembly). In the case of the classic module, in order to implement the CORBA support, a meaningful modification must be applied to the source text. In the assembly case, a separate file with the CORBA support is added, while fragments with the meaningful code remain intact.

The key DCMS features are [6]: a) event-driven interface support, b) usage of a single data format for data in the virtual memory, c) unified access to data in the external memory and in the network, d) fine-grain modularity. The first one gives the user a local module design technology. The second one relieves the user from the memory resource management, while the third one helps to avoid the protocol and the data format support when designing an application. The latter permits one to ignore the module semantics when modifying a modular structure. Altogether they provide an open architecture, the ease of maintenance for applications constructed, a high degree of independence of the APIs standaeds, network protocols and data formats.

The dependence of the DCMS on the OS used is minimized by separation of platform-dependent functions into a dedicated group, which is reimplemented for each new platform. The experience has shown that the size of this function group is tiny in comparison with the size of the rest of the DCMS. Currently, list of supported platforms is Win32 and UNIX. The library is portable to any 32/64 bit platform under OS that has dynamically linked the library support and the TCP/IP protocol. The DCMS is fully implemented, although its further modification is possible. The implementation is primarily based on the ANSI C programming language. The source text size exceeds two MB. The number of files exceeds 3000. The DCMS provides functions for data management in the virtual memory, and for the management of local or remote objects. Samples of objects are the news, papers, pages, projects, etc. The library provides a means for using external modules with the help of the event-driven interface.

The basic set of portal modules (the portal kernel) is created with the help of the technology proposed. The modules form several groups. The modules for object manager extensions implement the data format support, protocols, and interface for data transfer and access. The modules for the text and binary representation of hierarchical data, as well as the modules for data access within a local file system and the Oracle DBMS, are implemented. The modules for processing the user request are created for the following operations: object creation/deletion, folder creation/deletion, object search by a string key, and object edition. A series of modules for generation and modification of an abstract document is available. An abstract document is a data object to represent an information to the user. The abstract document also contains the visualization parameters. The representation of data in the form of an abstract document eliminates the dependence of the portal on the Web interface and the markup (HTML/XML) languages. The abstract document is kept in the internal format and contains atomic elements such as a text, references, pictures, and composite elements, e.g. a table, a list. The abstract document is translated into the HTML content (the HTML pages sent to a browser) implemented in the HTML generator module, which can be used with any browser that supports the HTML 4.01. The portal kernel contains modules for the user authorization and data security. Their main functions are the verification of the user login, the creation of an active session for the user after successful login, the destruction of the closed session, the verification of the session validity for any operation requested by the user. The described modules provide the users with the functionality of the information portal. The modules are implemented in the ANSI C. The size of the source text is about 1MB.

The kernel modules: 1) provide the construction of a flexible user interface that is visually adaptable by the user to his applications without need for programming the new modules and modifying the existing ones; 2) give a means to describe interfaces with the user application software (e.g. mathematical and simulation systems); 3) can interoperate with other information and corporate resources because of the ability to add a support for new protocols, interfaces, and formats (on the basis of the fine-grain modularity and an open architecture of the object manager).

In addition to the above modules and those in [2,3], some kernel extension modules are created. They provide the following services: a news line, a real time communication forum, gathering and visualization of the portal use statistics, and the user activity log.

A subsystem to construct the thematical portal service is currently being implemented and tested (http://portal.sscc.ru). It is based on the DCMS and contains components for: 1) the object loading/saving (the news article, pages, projects, etc.) and its modification; 2) construction of a tree and list views of the object contents and directory structure; 3) visualization of titles, values, operation panels for the items in trees and lists; and 4) sorting out of tree and list view elements. Applications are constructed on the basis of the following components: 1) the directory navigator (helps to move within classifications and their subfolders, create new classifications and edit their structure); 2) the system editor (edits objects of any type at a low level up to atomic values); 3) the application editor (creates and visually edits new objects on the basis of typical templates, such as a news article, a project, etc. The interface for the following services is constructed with the help of the following applications: 1) the associative object search; this service can be used in different problem areas since a set of search attributes is extensible and can be set immediately before the search in the explicit form or be kept in the user preferences as implicit setting; 2) the journal for the status information and system events; 3) the construction of the news lines and insertion of articles.

A sample of the application editor construction on the basis of the assembly technology [7, 8] (however, in contrast to the fine-grain modularity, fragments are coarse-grain components rather than fine-grain source fragments) is shown in Figure 2. The assembly program sets the order of exe-

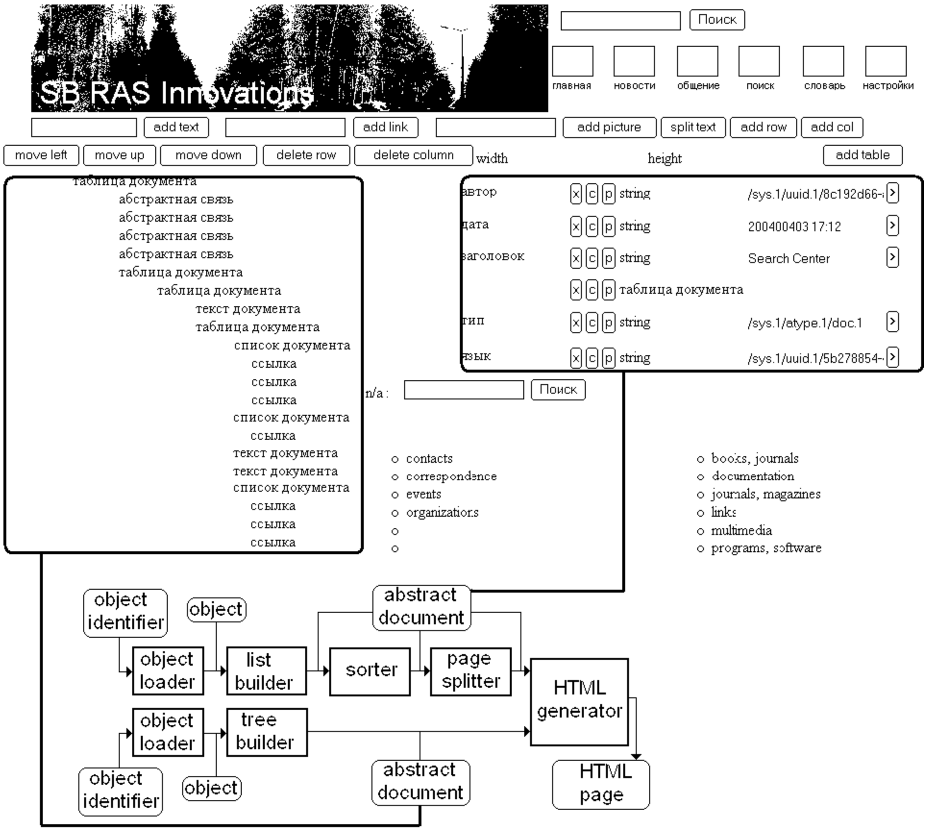

Figure 2. An assembly of portal application

cution of components, initializes the input parameter (the object name) for the object loading component, and connects the output parameters of the preceding components to the input parameters of the following components. For example, a detailed list is formed by the following composition of components: 1) the loading object, 2) the construction of a detailed list view (in the form of an abstract document), 3) the list sorting, 4) the list cutting into pages, 5) the list insertion into a composite abstract document together with the other parts: the logotype, the toolbar, and the tree view.

The innovative activity ontology was developed at the Institute of Economics, SB RAS as part of the joint project (see publications at http://sinin.nsc.ru). "The online workplace of innovative activity participant" (http://swb.sinin.nsc.ru/) has been implemented and is being tested in this institute, but so far it has not been integrated yet into the innovative portal.

An expert system for the interactive support of a technological innovative chain for a new kind of production was developed but not yet integrated [9, 10]. In order to construct chains, the problem of access to data and knowledge bases was solved: a model, the architecture and the main modules of the integration system of distributed heterogeneous data sources were developed. This integration system provides sematic search within heterogeneous information resources associated with innovations. The architecture and functioning of the system for integration of distributed hetegeneous structured data sources are based on the multiagent approach, which permits one to dynamically include new data sources and simultaneously process many requests to different sources. The structured sources are considered to be information sources. Different relations, object data bases are examples of such sources. Another example is weakly structured resources in XML, RDF, OWL, or DAML+OIL formats. The system components are designed: 1) for supporting the models of problem areas and data sources; 2) for structures and forms of the behavior of various agents that participate in the search for information within the area associated with a certain problem data from many data sources.

All subsystems will be integrated into the innovative portal on the basis of an event-driven interface, a unified user's data representation in the form of an abstract document and the use of network protocols (primarily, the HTTP one) at the next stage. All user materials will be integrated into a single information space SININ to create stable information flows through the network innovative infrastructure to be used in everyday innovative activity of the SB RAS organizations and other participants of the innovative process.

3. Portal development

The development of the system structure and the creation of its services has shown that the system has a distributed architecture and is tolerant towards the hardware and the software platforms, interoperable, and can easily be adapted to the user needs. In general, this means that the system is scalable. The system scalability permits us to outline the following tendency of its development. In order to fully implement the potential of the "simulative" design, one should place the system in the network of RAS supercomputers interconnected by the GRID. This realization of the system would make it possible to support a complicated innovative project cooperatively on all supercomputers. This would provide such a project with huge informational and computational resources that it needs. This need is a result of the feature of innovative activity that requires a maximum decrease of the design time: from formulating a task to designing a product. A product in a wide sense can be anything from a physical product to a technology, a model of a natural phenomenon, etc. This design time can take months, even days, but, certainly, not years or decades. According to the foreign experts's estimates, such a kind of project design that needs simulations would require computations with hundreds and even thousands TFlops. Subtasks that jointly form a project can be executed in supercenters that have relevant resources, applications, and professionals for this subtask. For example, the researches from the ITAM SB RAS could simulate gas thermo-dynamic processes for the project of space vehicle construction at the Siberian Supercomputer Center of the ICM&MG SB RAS. Simulation for oil and gas industry can be done for a project of exploiting the East Siberia natural resources. In addition, one can hardly create a new project without full text access in the online mode to the information (articles, reports, books, patents, theses, preprints, etc.). This information can be geographically distributed among many sources. It would be convenient to concentrate it in supercomputer centers on the regional basis. The system of storage, search, and processing of information at the Siberian Supercomputer Center can be an example. It is a mirror of the electronic library at SPSTL SB RAS and provides access to the full text electronic information from approximately 6000 foreign and Russian journals. The mirror size is approximately 1 Tb. Full copies can be obtained at workplaces (personal workstations or corporate networks), figuratively speaking, in the real time mode. Such systems can be constructed in those supercomputer centers that for now do not have them, as connected into a single storage system, and integrated with the innovative portal.

Evidently, an adequate growth of the information and hardware structure at the Siberian Supercomputer Center (and, possibly, at other centers) should accompany such a portal development. The size of electronic storage should be increased up to 10 Tb. It should contain information from regional sources. The computational capabilities have also to be increased. The fast network links for the connection with SB RAS institutions should be established for computations of big-size problems in the supercomputer environment.

4. Conclusion

The innovative portal proposed can be an important step in its transition into a distributed system for computer simulation of innovative projects within the federal innovative infrastructure based on a supercenters network. The implementation of such a system can assist in rising the hi-tech industry in our country. This would create competitive technologies worldwide for the search for natural resources, environmental protection, application software design, designing the new equipment on the basis of fundamental scientific results in computational mathematics, physics, and technology.

References

[1] Alekseev A.S., Parinov S.I., Piskunov S.V. Intellectualized network system of information support for innovative infrastructure at SB RAS // ECO.— Novosibirsk, 2004.— Vol. 5, No. 359,— P. 76–85.

[2] Alekseev A.S., Parinov S.I., Piskunov S.V. About project “Intellectualized network system of information support for innovative infrastructure at SB RAS” // Infosphere. — Novosibirsk, 2003. — Vol. 3, No. 18. — P. 7–10.

[3] Alekseev A.S., Ostapkevich M.B., Piskunov S.V. On the structure and functions of intellectualized network system for innovative activity support // Int. Conf. “Communications-2004”. Proc. 8th Int. Conf. “The problems of information network functioning”. — Novosibirsk, 2004. — P. 32–40.

[4] Makarov V.L. The economics of knowledge: lessons for Russia // Messenger of RAS. — 2003. — Vol. 73, No. 5.

[5] Ostapkevich M.B. The DCMS library for open architecture applications // NCC Bulletin. Series Comp. Sci. — Novosibirsk: NCC Publisher, 2004. — Iss. 21. — P. 99–111.

[6] Ostapkevich M.B., Piskunov S.V. The design of software tools for parallel and distributed application construction with open architecture // Int. Conf. “Communications-2004”. Proc. Int. Seminar “Computational methods and optimization tasks solutions”. — Novosibirsk, 2004. — P. 130–137.

[7] Valkovsky V.A., Malyshkin V.E. The Synthesis of Parallel Programs and Systems on Computational Models. — Novosibirsk: Nauka, 1988.

[8] Kraeva M.A., Malyshkin V.E. Assembly technology for parallel realization of numerical models on MIMD-multicomputers // Int. J. Future Generation Computer Systems. — 2001. — Vol. 17. — P. 755–765.

[9] Zagorulko Yu.A., Bulgakov S.V. The usage of anthologies for innovative chain construction in the system of innovative activity support in the region // Proc. 6th Int. Conf. “The problems of control and simulation in complex systems”. — Samara: Samara SC RAS, 2004. — P. 328–333.

[10] Bulgakov S.V. An approach to multiagent system construction for meaningful search of from heterogenous structured data sources // Proc. 9th National AI Conf. with International Participants — KAI’2004. — Moscow: Fismatlit, 2004. — Vol. 2. — P. 706–714.