The HPC Virtual Consultant: Improving Access to Online Knowledge

Sally D. Haerer
Northwest Alliance for Computational Science & Engineering
Oregon State University

Abstract

Funded by a grant through the Programming Environment & Training (PET) program at the Naval Oceanographic Office Major Shared Resource Center (MSRC) at Stennis Space Center, under contract with Logicon Inc., the Northwest Alliance for Computational Science & Engineering (NACSE) at Oregon State University has entered into a collaboration to create a web-based, intelligent, search-and-query interface that will provide a portal to a variety of excellent online documentation sites around the country. This will allow programmers and technical consultants alike to learn just one common access point, one navigational hierarchy and information strategy, and one search mechanism, yet have access to a vast amount of information and expertise from distributed, collaborating HPC sites. Although predominantly targeted to assist DoD computational scientists utilizing MSRC facilities, this portal will help service HPC users and consultants everywhere.

General Overview

HPC programmers encounter critical problems in developing and optimizing their codes, and these problems must be addressed in a timely manner. In many instances, the user requires specific advice or assistance on architectures, algorithms, standards, optimization techniques, or other coding practices. This can be sought in several ways, such as contacting a technical consultant, seeking advice from a colleague, or tapping various online documentation resources. Although useful documentation is available at a large number of Web sites (local sites, MSRC sites, vendor-supported sites, other HPC centers), most users are familiar with only a limited number of such resources.

The main value of online documentation is this: it is there when the user needs it without the schedule restrictions and availability of other human resources. Even when one has access to experienced technical consultants, it is sometimes the case that the consultants themselves need to spend significant time searching through documentation and examples in order to find the right strategy to solve a complex user problem. That search might be local or remote, depending upon the quality of documentation, the familiarity of the consultant with various documentation systems, and his/her searching strategies.

As an example, suppose a user is struggling with asynchronous I/O in a model where performance improvement is crucial. Depending on the hour of the day, technical assistance may or may not be available through a consultant. If that consultant does not have first-hand expertise with asynchronous I/O performance, or if the researcher is working off-hours, some type of documentation search will be required. The process of searching documentation can be tedious and difficult. An HPC programmer may spend many hours in a fruitless search for the appropriate source of information. An HPC consultant may also spend significant amounts of time locating the resources needed to help solve a user's problem. This is not just an issue of human costs; during the search process, a possibly critical problem remains unsolved.

There are many HPC sites in the United States that already provide online help and references into online vendor documentation. Most of these sites also have the advantage of FAQs (Frequently Asked Questions), hints on common programming problems, and feature articles on particularly difficult user coding tasks. The problem is that each site has a different Web organization, documentation hierarchy, searching mechanism, and user interface design. To use a site effectively, the user must learn site-specific strategies for navigating successfully through the mass of other information leading to the location where an answer may be found. Each user, whether computational scientist or technical consultant, must acquire that expertise on his/her own.

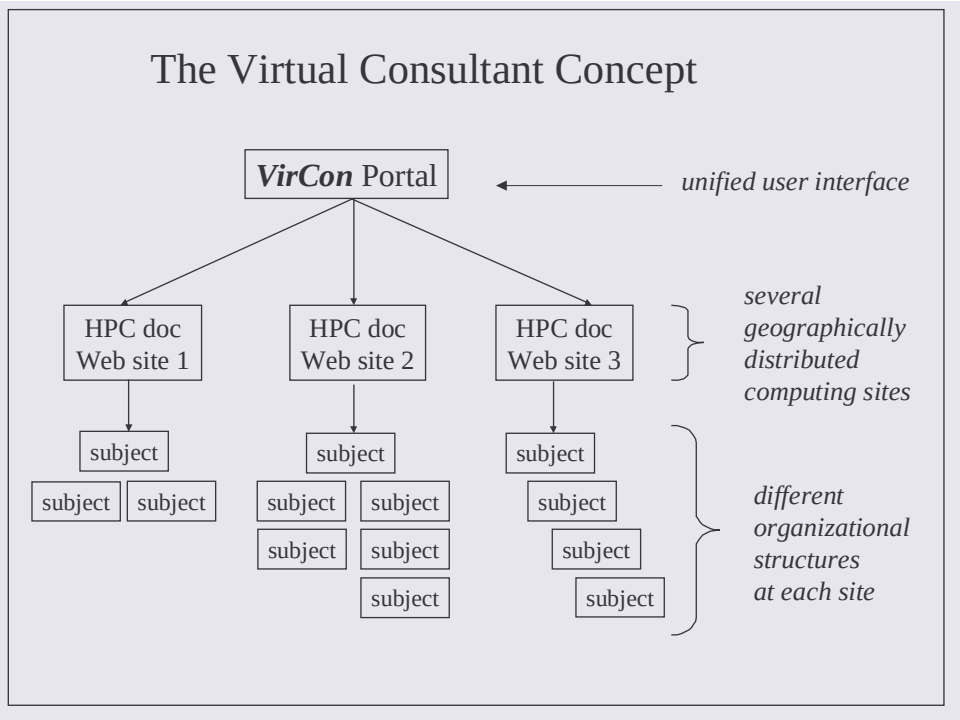
The Northwest Alliance for Computational Science & Engineering (NACSE) at Oregon State University is designing and developing an HPC "Virtual Consultant"; that is, an intelligent search-and-query interface that will provide a portal to a variety of online documentation sites. This will allow programmers and technical consultants alike to learn just one common access point, one navigational hierarchy and information strategy, and one search mechanism, yet have access to a vast amount of information and expertise from distributed sites.

By providing consolidated access to a wide variety of quality Web resources, the Virtual Consultant will allow users, as well as consultants, to benefit from materials they would not be able to easily find on their own. By removing the need to learn and traverse multiple Web sites, more time can be spent directly on problem-solving. Furthermore, the development of a model for a common, distributed web-based documentation strategy means that other targeted search-and-query interfaces could be engineered in the future with much less lead-time and resources than this venture.

Returning to our example, the HPC programmer (or technical consultant) who is struggling with the issue of asynchronous I/O techniques and performance would be able to effectively utilize several good sources distributed throughout the United States to discover options, techniques, tools, utilities, suggested strategies, and/or command syntax for the problem at hand, on the particular target architecture of interest -- all using a single point of access and a simple user interface.

Components of the Virtual Consultant

Our project, tagged as ***VirCon*** by our development team, attempts to offer an alternative to the HPC community by interfacing with a variety of quality documentation sites within our industry. Here is a simplified picture of *VirCon:*



The basic components of the HPC Virtual Consultant include:

- Selected, distributed, high-quality HPC documentation sites located around the country.
- Each of these HPC sites have an information structure that is unique and serves the individual purposes of that site regarding its mission and its local needs, requirements, and specifics. It is intended that each site's unique information structure remain intact as we build a common interface to access the data held therein.
- A common user interface that ties together and utilizes this vast quantity of information.

This user interface is the component that brings the project together in a fashion that is usable to our target audience. This interface is being designed, developed, and implemented by NACSE staff. The form that it takes is being created using a proven User-Centered Design strategy involving identification of the user audience and their needs, and iteratively testing and improving a prototype to hone the interface to match user specifications. The production interface will evolve from this strategy.

VirCon contains a *search-and-query interface* that accesses, selects, and retrieves appropriate information based on user input. It then formulates all query results and displays them as Web-links directing users to the pertinent document locations at these distributed sites. This display still maintains the common look-and-feel of the *VirCon* portal, while providing credit to the sites which contributed relevant information.

This project is being developed with a restricted number of architecture platforms and subject areas, but once the methodology has been established for this limited source, it can be expanded to encompass other areas with significantly less effort. Both the metadata and the search-and-query system will be designed, implemented, and documented specifically to facilitate such extensions. The search-and-query interface is browser-based and is able to accommodate users on UNIX, Macintosh, Windows-98, and Windows-NT platforms.

Collaboration Sites

NACSE has identified several HPC sites known for the quality of their online information, and which users have already found to be helpful for problem-solving. We have enlisting the cooperation of key HPC professionals from these sites to participate in the planning and implementation of this project. Meetings and workshops are designed to include technical consultants and online information specialists from each site in order to orchestrate the requirements of target users and the overall data organizational structure. Through this interaction, NACSE developers can better understanding and formulating metadata needed to facilitate information maintenance and retrieval activities. The collaborating representatives will also assist with the *VirCon* interface usability and target audience identification.

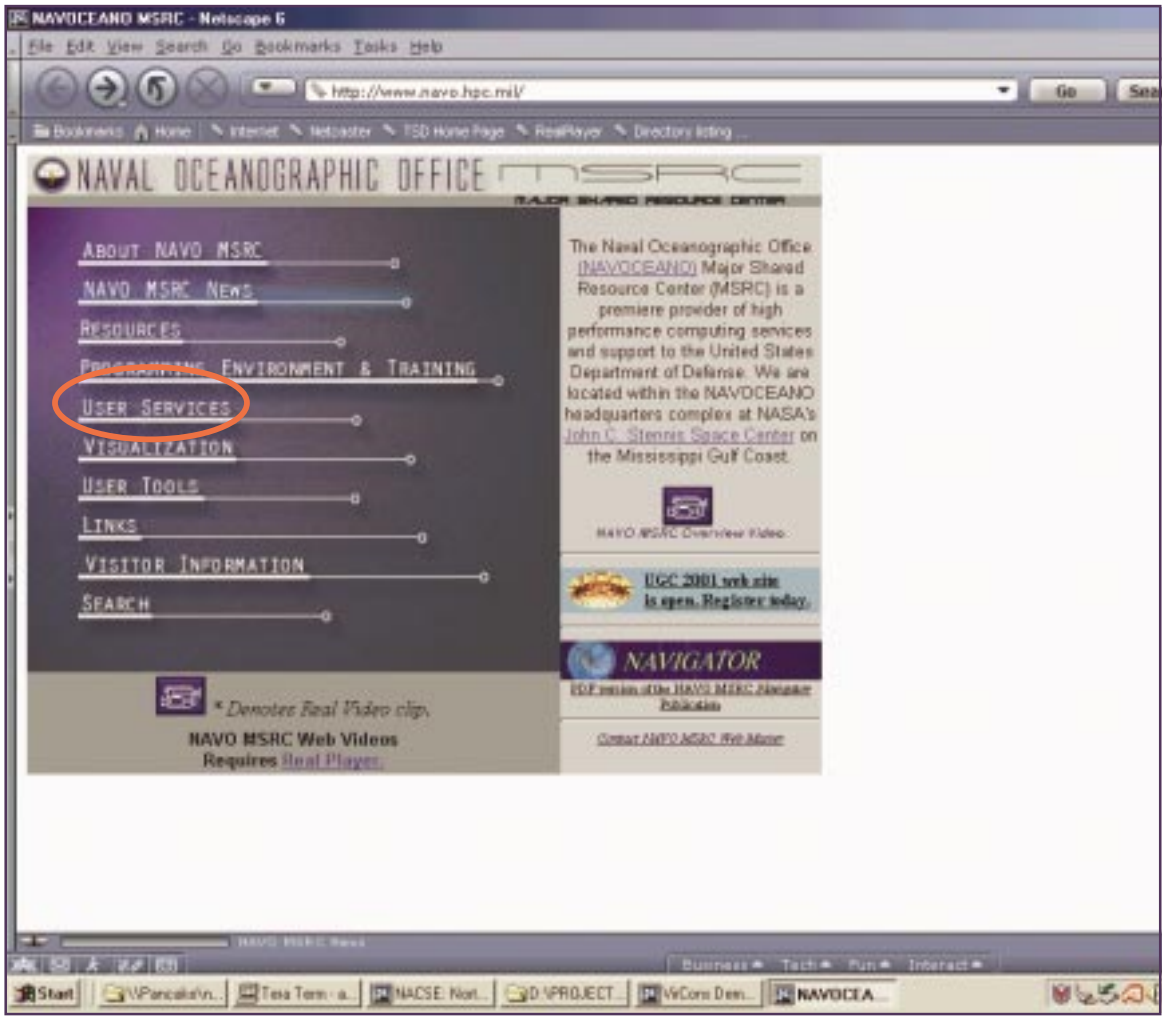
During the entire project, feedback and interaction will be critical components to ensure that the *VirCon*'s organization and functionality meet the needs of the HPC community in an effective manner.

The participating sites for this project are as follows:

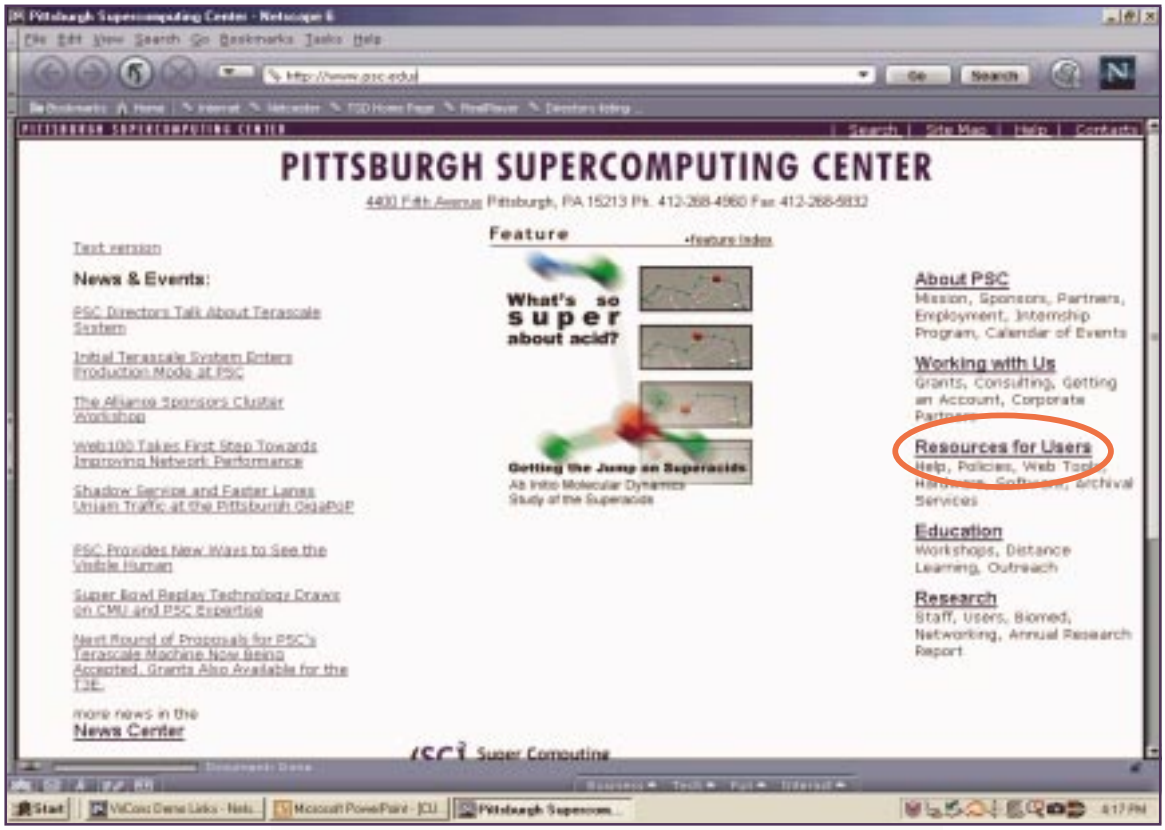
- ***Naval Oceanographic Office MSRC***
- ***Arctic Region Supercomputing Center***
- ***Ohio Supercomputing Center***
- ***Pittsburgh Supercomputing Center***
- ***NCAR's Scientific Computing Division***
- ***San Diego Supercomputer Center***

Sample Web-sites

In order to further demonstrate the complexities that currently must be overcome if a user wishes to take advantage of the vast information at leading Web sites, we would like to offer just a few simple examples. These are screen-dumps of the main Web pages to three selected partner HPC sites; each is attractive, but has a very different look-and-feel, information hierarchy, and purpose. For the reader's convenience, we have placed a red circle as the hint of where a user might want to look next for needed information:



Example 1: Naval Oceanographic Office



Example 2: Pittsburgh Supercomputing Center

Example 3: San Diego Supercomputer Center

While each of these example sites is excellent in its own right, a user with the single goal of finding specific information to help answer a programming problem will have to learn where to go to find any particular topic. As demonstrated in these three examples, the next link might be called "User Services", or it might be called "Resources for Users", or it might be under a broader category like "Computing". In other words, the path to the desired answer will vary for each site.

Once the general documentation link is located, the user must navigate through the abundance of information to the specific topic of the problem being sought. This is clearly a tedious process given that the search is entirely different for each HPC site. In contrast, the *VirCon* project is developing a strategy by which a user may state a specific problem area only once, and then all relevant links are presented in a common format from all participating sites.

Intelligent Search and Retrieval Techniques

A variety of search and retrieval techniques are being investigated to improve the search results for our target users. For this project, we are attempting information retrieval over a distributed information system. Rather than a simple information retrieval approach, we are incorporating a user-active, interaction with the documentation resources. It is not being designed as just a content-based or keyword approach, but rather one utilizing collaborative interaction in order to refine the retrieval. Human, as well as automatic, categorization of the combined content of the various HPC Web-sites will enable a portal for convenient user searching across this collection of distributed sites for answers to HPC program-related questions. The goal of the project is to go beyond the basic keyword/phrase style of search matching; we are using more advanced Computer Science methods of information retrieval. We feel that a combination of several proven techniques is the answer to providing reliable, targeted Web links in response to various user queries.

Automatic collaborative filtering systems have been under development for some time and offer the opportunity to connect a user's interest with certain documents through a human rating scheme. The more people use the system and provide document ratings, the more useful the system's document recommendations can become.

Term frequency analysis is a technique that measures how often all terms (words) in a document occur and compares term frequencies between documents to determine their estimated similarity. For instance, if two documents both use certain words more often than others, then we may deduce that these two documents bear some similarity. These measurements of similarity can be used by *VirCon* in two ways:

- To find similar documents – if a user has indicated that a certain document is useful, *VirCon* might be able to then suggest other documents that are similar.

- To match free-form user queries with documents - if we allowed a user to enter a free-form query (a few sentences describing a problem), we may be able to measure the similarity between their text and other documents. This can lead to a good list of possible "answer" matches.

Relevance feedback is another method used for searching. With this approach, we allow the user to make a query and then select a few documents as being close or useful to finding an answer to the query. Then, *VirCon* may be able to find more relevant search results based on this combination of the original search query plus the selected documents that the user identified as useful.

The guidance and direction of Dr. Jon Herlocker (Oregon State University Department of Computer Science) has been vital for this project. His primary research interest is in the field of collaborative filtering.

Summary

The efficiency and effectiveness of DoD programmers and technical consultants are the focus of this project. This project is leveraging the large amounts of online HPC documentation already available at multiple sources within DoD and throughout the broad HPC community, offering a centralized view of these distributed sites. The resultant information portal will inevitably service the entire HPC community of computational scientists and technical consultants alike.

By providing consolidated access to a wide variety of Web resources in a usable, common interface with an effective navigation strategy and an intelligent search-and retrieval mechanism, the *Virtual Consultant* is being designed to allow users to benefit from materials they would not be able to find on their own. By removing the need to learn and traverse multiple Web sites, more time can be spent directly on problem-solving. The goal is simply a reduction in the time-to-solution.

Furthermore, the development of a model for a common, intelligent, distributed web-based documentation strategy means that other targeted search-and-query interfaces could later be engineered with much less lead-time and resources than this venture.

Funding Source

Operated by Logicon Inc., a Northrop Grumman company, the Naval Oceanographic Office Major Shared Resource Center (MSRC) is a premiere provider of high performance computing services and support to the United States Department of Defense. The Programming Environments and Training (PET) program has been established as a component of this MSRC to enhance the productivity of DoD HPC users as research scientists and engineers. Under their funding, the Northwest Alliance for Computational

Science & Engineering (NACSE) at Oregon State University has entered into this collaboration with other sites to create the *HPC Virtual Consultant* project.

Acknowledgements

NACSE would like to acknowledge the following people and institutions for their critical contributions to this project:

From Oregon State University

- Cherri Pancake, Professor and Intel Fellow of Computer Science and Director of NACSE – Principle Investigator
- Tim Holt, NACSE Database Development Engineer – Lead Software Developer
- Jon Herlocker, Assistant Professor of Computer Science – Collaborator and Researcher for searching, matching, and retrieval techniques
- Sally Haerer, Associate Director of NACSE – Project Coordinator

From Participating Partner Sites:

- Arctic Region Supercomputing Center
- Naval Oceanographics Office MSRC
- Ohio Supercomputer Center
- Pittsburgh Supercomputing Center
- San Diego Supercomputing Center
- Scientific Computing Division at NCAR

Research References

Herlocker, J., Konstan, J., and Riedl, J., Explaining Collaborative Filtering Recommendations. *Proceedings of the ACM 2000 Conference on Computer Supported Cooperative Work*, December 2-6, 2000.

Pancake, C., Newsome, M., and Hanus, F. J., 'Split Personalities' for Scientific Databases: Targeting Database Middleware and Interfaces to Specific Audiences. *Future Generation Computer Systems*, #16, 1999. pp. 135-152.

Rocha, L., and Bollen, J., Biologically Motivated Distributed Designs for Adaptive Knowledge Management. *Santa Fe Institute Series in the Sciences of Complexity*. Oxford University Press, 2000.

Pancake, C., Improving the Usability of Numerical Software through User-Centered Design. In *The Quality of Numerical Software: Assessment and Enhancement*, ed. B. Ford and J. Rice. Chapman & Hall, London, pp. 44-60.

Mandel, T., *The Elements of User Interface Design*. John Wiley & Sons, Inc., 1997.