A Framework for Web-based Research Support Systems

Y.Y. Yao

Department of Computer Science, University of Regina

Regina, Saskatchewan, Canada S4S 0A2

E-mail: yyao@cs.uregina.ca

Abstract

The objective of research support systems (RSS) is to support and improve research, which may be viewed as a counterpart of decision support systems (DSS) for scientists. Web-based RSS (WRSS) assist scientists in the research process on the Web platform. WRSS are based on the assembling, integration, and adaption of existing computer technology and information systems for the purpose of research support. A framework of WRSS is presented by focusing on research activities and phases, as well as the technology support needed. The emphasis is on the conceptual formulation of WRSS. Different systems are linked to various research activities, and a pool of support sub-systems is established. As an illustrative example, Web-based information retrieval support systems (WIRSS) are discussed. The results of WRSS may lead to new and viable research tools.

1   Introduction

The impact of computer technology on research can be felt by every scientist. Computer software and information systems exist that support scientists in many activities, such as communication, literature search, data analysis and manuscript preparation. As new technologies evolve and existing technologies expand, a scientist needs to adjust accordingly and make full use of them when carrying out research.

With the fast growth of the Web and easy availability of information on the Web, we have arrived at a new information age. The Web provides a new medium for gathering, storing, processing, presenting, sharing, and using information. There is a tremendous amount of online materials, such as articles, journals, newspapers, databases, digital libraries, and so on. There are also many products, tools and services on the Web, such as news groups, downloadable software, document delivery systems, and so on. Scientists face many challenges in using Web-based information resources, such as information overload, misinformation, fees, poorly designed navigation, retrieval, and browsing tools [6]. The huge quantity of available software systems, tools and techniques, such as those used for data analysis, simulation, graphics, and document preparation, makes it more difficult for scientists to select the right tools and techniques for solving a particular problem. How to support scientists to meet such challenges is an important issue.

In the management context, decision support systems (DSS) have been studied extensively to support and improve decision making for managers [12]. To a large extent, DSS apply existing computer technologies to build systems that support management decision making in an organization. The development of computer science affects the design and implementation of DSS, as new computer technologies have been constantly added to DSS. For example, the intelligent agents have been used in DSS [12]. The development of the Web leads to Web-based decision support systems [11].

Following DSS, we consider the concept of research support systems (RSS) for scientists [8, 15]. Such systems will support scientists in finding the relevant information, choosing the right tools and producing the effective presentation of research results. The development of RSS on the Web results in Web-based RSS (WRSS). The main objective of this paper is to present a framework of WRSS by focusing on research activities and phases, as well as the technology support needed. Different information systems are linked to various research activities, and a pool of support sub-systems is established.

The study of WRSS aims at a less explored domain of application, rather than new technologies and methodologies. Many computer systems have been implemented to support various research activities. For example, retrieval systems and search engines support literature search, statistical packages support data analysis, and word processors and text editors support manuscript preparation. However, there is a lack of study of such systems in a common framework. The introduction of RSS thus provides such a framework in which various systems are assembled and integrated to serve the same goal of research support. DSS can be

viewed as a product from the combination of computer science and management science. Likewise, RSS results from combining computer science and research methodologies.

The phrase "research support systems" has been used by many authors and appeared in many places [8]. By querying the search engine Google (http://www.google.com/) with the exact phrase "research support systems" on July 25, 2003, we obtained 327 hits. The search of the exact phrase "research support system" produced 558 hits. Although there is no consensus on an exact definition of RSS, the support functionality such systems facilitate is clearly identified. In this paper, we restrict our discussion of RSS to a narrow context concerning only the research activities.

The discussion of the paper briefly summarizes the initial and scattered ideas on RSS. The framework for RSS is examined by focusing on characteristics and functionalities of RSS, as well as existing computer technologies related to these functionalities.

2   Research Process and Methods

Research is a highly complex and subtle human activity, which may be difficult to formulate formally. On the one hand, it seems impossible to give any formal instruction in how to do research. On the other hand, some lessons and general principles can be learned from the experience of scientists. There are some basic principles and techniques that are commonly used in most types of scientific investigations [2]. The study of research methods adopts the view that scientists follow a fairly systematic process in scientific investigation [1, 2, 5, 9, 10].

Scientific research is effected by the perception of science and the purpose of science. There are many definitions of science.

> "Generally, science is the search for understanding of the world around us. Science is the attempt to find order and lawful relations in the world. It is a method of viewing the world." [10]

> "Science is a process of formulating specific questions and then finding answers in order to gain a better understanding of nature." [5]

> "Science is an attempt to understand the universe and humanity's relationship to nature." [3]

Martella *et al.* [10] summarized several purposes of science, i.e., to describe and predict, to improve or manipulate the world around us, and to explain our world. The results of the scientific process may provide a description of an event or phenomenon. The knowledge obtained from research may help us to make predictions about what will happen in the future. Research findings may be useful for us to make an improvement in the subject matter. More specifically,

research findings are used to determine the best or most effective interventions to bring about desirable changes. Finally, scientists develop models and theories to explain why a phenomenon occurs.

Geisler [4] pointed out that science encompasses the entire spectrum of research universe, including exploratory, basic, applied, and directed basic research. Research leads to technology development and useful innovation. A classical linear model of research activity can be schematically described as [4]:

Generation of ideas $\longrightarrow$
  Exploratory (Basic) research $\longrightarrow$
    Applied research $\longrightarrow$
      Development $\longrightarrow$
        Design testing and prototyping $\longrightarrow$
          Engineering.
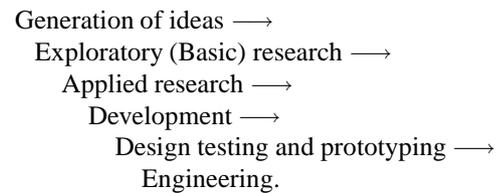
While early stages in the linear model focus on science, the later stages focus on technology.

Scientific research is commonly divided into steps or phases, such as exploration, problem definition, hypothesis formulation, experimentation, hypothesis testing, verification and validation, and application [1, 2, 10]. We adopt the model of the research process from Garziano and Raulin [5], and combine it with other models [7, 10]:

- **Idea-generating phase**. The objective is to identify a topic of interest to study. It may also be referred to as the preparation [2] or the exploration phase. Initial ideas can emerge from vague thoughts and in very non-scientific ways [5]. Curiosity, interest, enthusiasm, intuition, imagination, tolerance of uncertainty, diversity, and communication with colleagues are some of the critical ingredients in idea generation [2, 9]. Literature search and reading also play an important role in this phase [2, 9].

- **Problem-definition phase**. The objective is to precisely and clearly define and formulate vague and general ideas generated in the previous phase. Problem definition involves careful conceptualization and abstraction. The success in problem definition increases the probability of a successful research project. With respect to a precisely defined problem, it is relatively easy to find related and solved problems, as well as potential solutions.

- **Procedure-design/planning phase**. The objective is to make a workable research plan by considering all issues involved, such as expected findings and results, available tools and methodologies, experiments, system implementation, time and resource constraints, and so on. This phase deals with planning and organizing research at strategic level [2].

- **Observation/experimentation phase**. The objective is to observe real world phenomenon, collect data, and carry out experiments. Depending on the nature of the research disciplines, various tools and equipment, as well as different methods, can be used.

- **Data-analysis phase**. The objective is to make sense out of the data collected. One extracts potentially useful information, abstraction, findings, and knowledge from data. Statistical software packages can be used.

- **Results-interpretation phase**. The objective is to build rational models and theories that explain the results from the data-analysis phase. It is necessary to investigate how the results help answer the research question, and how this answer contributes to the knowledge of the field. The connections to other concepts and existing studies may also be established.

- **Communication phase**. The objective is to present the research results to the research community. Communication can be done in either a formal or an informal manner. Books and scientific journals are the traditional communication media. Web publication is a new means of communication. Oral presentation at a conference, or discussion with colleagues, is an interactive means of communication.

It is possible to combine several phases into one, or to divide one phase into more detailed steps. The division between phases is not a clear cut. Moreover, the research process does not follow a rigid sequencing of the phases. Iteration of different phrases may be necessary [5]. Although the described research process is adopted from the ones used for social sciences, the basic principles are applicable to all sciences.

3 Characteristics of Web-based Research Support Systems

The design and implementation of viable research support systems depends on a clear understanding of research processes and methods. The study of RSS is of interdisciplinary nature, combining research methods and computer technology. Some main characteristics of RSS are outlined in this section.

RSS as a new application of computer technology to research process. The impact of computer technology on research in almost every discipline of science is evident. Scientists use many software packages and tools in research. However, most of the packages and tools is designed for general purpose, rather than just for supporting research. It is therefore useful to re-examine them under the new light of research support. By focusing on a particular context, it is possible to obtain more insights and make modification to existing systems. The study of RSS is the application of computer technology in a new domain.

Active role of RSS. Many textbooks on research methods cover how to use computers and software in research [5, 7]. In this case, scientists play an active role. In contrast, RSS will play a more active role to assist scientists.

Granziano and Raulin [5] made a clear separation of research process and content:

> "The particular observations made vary from one discipline to another because each discipline is interested in observing and understanding different phenomena. But the basic processes and the systematic way of studying problems are common elements of science, regardless of each discipline's particular subject matter. It is the process and not the content that distinguishes science from other ways of knowing, and it is the content – the particular phenomena and fact of interest – that distinguishes one scientific discipline from another." [5]

From the quotation, we can identify two basic characteristics of RSS.

Rationality and feasibility. Research process is relatively discipline independent. One can therefore make the underlying assumption that the research process is common to all sciences, which establishes the rationality and feasibility of RSS.

Separation of general research process and discipline specific knowledge. Research content is discipline specific. RSS must have at least two related and separate components. A discipline independent component that supports the general research process, and a discipline (domain) specific component that supports research in a particular discipline. The separation of general methods and discipline specific activities and knowledge is similar to the separation of the inference engine and domain specific knowledge base in expert systems [12].

Additional characteristics of RSS are given below.

Conceptual organization of RSS. Several arrangements of RSS facilities can be considered. One can order the facilities of RSS according to the sequence of research phases. Alternatively, one can arrange the facilities of RSS in a hierarchical manner based on the complexity of research activities, from simple and well structured to complicated and ill-structured.

Personalized support. Scientists commonly work on related but different problems. Even for the same problem, different scientists work on different views or aspects. Personalization is a necessary feature of RSS. In addition to general support, RSS should be able to provide support on a personal level. In fact, personalized support may be of more value to a scientist.

Adaptation over time. RSS must be adaptive over time to incorporate new tools and use new technologies. The systems should have a learning ability to improve their performance based on interaction with scientists.

Web platform. The Web serves as an effective and convenient platform for the implementation of RSS. The Web effectively links many different types of entities in a simple way. Information, software packages, and tools on the Web can be used to support research. The Web provides an easy way to store, share, search and publish research results. The Web provides a good user interface to many systems, and facilitates communication. The flexibility, distributivity, acceptability, availability, and simplicity make it suitable for RSS.

4 Functionalities of RSS and Related Computer Technologies

To support a large spectrum of research activities, RSS must be flexible and have many functionalities. This section summarizes these functionalities and the required computer technologies.

Profile management. The profile management may be broadly divided into research project profile management, and scientist profile management at both group level and individual level. A main component of profile management is the knowledge base, which serves as the basis of RSS. Profile management module collects, organizes, and stores all relevant information about a research project and the scientists involved. The profiles can be stored in various formats, such as databases, text documents, or XML documents. The profiles can be used to achieve personalized and project specific support.

Resource management. Many types of resources exist for supporting research. Examples are human resources, tool resources, and information/knowledge resources. A reservoir of experts can be constructed based on, for example, scientist profiles. Whenever there is a problem during research, it is possible to identify a group of experts whom to be consulted. The main functions of human resource management are the maintenance of expert reservoir, and the matching and retrieval of a useful group of experts. The reservoir of experts may be a virtual one, which consists of links to other systems, databases, or scientists' home pages. The tool resource and information/knowledge resource management modules maintain different types of objects, but have similar functions. The information resources are combined from many sources, such as libraries, digital libraries, and the entire Web. Database, knowledge base, information retrieval, and agent technologies can be used. Web search engines can be used for retrieval.

Data/knowledge management. Typically, research involves the collection and processing of a large amount of data. RSS must have a module to record the useful data, information and knowledge during the entire research process. The module must contain some data/knowledge operation and retrieval facilities. Database and information retrieval systems can be used.

The profile, resource, and data managements form a solid basis of RSS. Consider now the following specific supporting functionalities:

- **Exploring support**. In the early stage of research, a scientist may have a vague idea and may not be aware of the works of fellow researchers. Exploration thus plays an important role. There are many means of exploration, such as browsing databases, libraries, and the Web. A scientist's profile may be useful in focusing the exploration areas. If the Web is used for browsing, the historical data can be tracked. The collected data can be analyzed using machine learning and data mining tools to provide a scientist useful information and hints. The profile can also be updated. Currently, Web browsers are a useful exploration tool. Their functions need to be expanded for providing support to research.

- **Retrieval support**. Once a scientist forms relatively solid ideas, it is necessary to search literature to find relevant information. Retrieval support assists retrieval related activities, such as browsing, searching, organization, and utilization of information [14, 15, 17]. A more detailed discussion on retrieval support will be given in the next section.

- **Reading support**. Reading critically and extensively is important, especially in the preparation stage [2, 9]. The advances in digital libraries and electronic publications make the reading support a necessity. Software packages exist so that a reader can add book marks, make notes, link different parts of an article, and make logical connections of different articles. A reading support system needs to assist a reader in actively finding relevant materials, as well as constructing cognitive maps of the materials read. Reading support systems can be combined with exploring and retrieval support systems. Machine learning and text mining methods can be used to assist a reader by learning from the reading history. Agent technology can be used to actively look for useful information and periodically inform scientists with new information. On-line dictionaries may also be useful in reading support.

- **Analyzing support**. Successful analyzing support depends on tool management. It is necessary to help a scientist find the right tool for a particular problem in analyzing data. In addition, the system should also assist a scientist in using a tool. An explanation feature may be needed, which answers the question why

a particular tool is used. If the functions of tools are described as plain text, information retrieval systems can be used to find the right tool. Computer graphics and visualization may be useful in analyzing support.

- **Writing support**. There are many writing support software tools, such as word-processor and typesetting software. Many packages come with additional functions, such as spelling-checking, grammar-checking, and various other agents. A writing support system should also contain some functions mentioned in the retrieval support systems. For example, a writing support system can find relevant articles based on the text written by a scientist and suggest possible references.

A research support system consists of many sub-systems to support different activities. The sub-systems share common data and knowledge bases. As one can not have a clear classification of research activities, it is difficult to have a clear classification of different types of support sub-systems.

5 Web-based Information Retrieval Support Systems

Each support sub-system described has its special features. As an illustration, we examine the retrieval support sub-systems.

On the Web platform, Web-based information retrieval support systems (WIRSS) provide models, languages, utilities, and tools to assist a scientist (user) in exploring, searching, analyzing, understanding, and organizing a document collection and search results [14, 15, 17, 18]. These tools allow the user to explore both semantic and structural information of each individual document, as well as the entire collection.

Three related types of models need to be considered in IRSS. Documents in a document collection serve as the raw data of IRSS. The document models deal with representations and interpretations of documents and the document collection. The retrieval models deal with the search. The presentation models deal with the representation and interpretations of results from the search. A single document model, a retrieval model, or presentation model may not be suitable for different types of users. Therefore, IRSS must support multi-model, and provide tools for users to manage various models.

The *document models* allow a multi-representation of documents. Granular computing (GrC), a methodology for computing at different levels of abstractions, will play an important role in the construction of document models [13, 16, 19]. One can use a hierarchical granulation of document collection, namely, a layered and multi-resolution

representation of documents. The same document is represented in less detail at a higher level than at a lower level. With the multi-model capability, a user can create different logical views of a document collection and logical links between documents, as well as compare and investigate various views. The resulting in-depth knowledge can help the user to locate and infer useful information. The recent development of XML enables us to describe both the structural and semantic information of a document. Such information makes the construction of multi-document models an achievable goal.

The *retrieval models* provide languages and tools to assist a user to perform tasks such as searching and browsing. IRSS should provide multi-strategy retrieval. A user can choose different retrieval models with respect to different document models.

The *presentation models* allow a user to view and arrange search results, as well as various document models. The same results can be viewed in different ways by using distinct presentation models. Moreover, a user can analyze and compare results from different retrieval models.

The usefulness or the relevance of each information item (i.e., a group of documents, a document, and parts of a document) can only be determined by the user. Without the involvement of a user in the analysis and organization of a document collection, one can not expect the user to provide a meaningful query. The user involvement is particularly important in the exploratory type of searching and browsing. An IRSS is highly interactive so that a user can make decisions at various stages. A user plays a more active role in the process of finding useful information.

IRSS are supported by existing computer technologies:

- **Expert systems (ES)**. A well established practice in expert systems is the separation of knowledge and inference engine. While the inference engine is logic based and problem independent, the knowledge base is domain specific. One can build knowledge-based IRSS by separating the knowledge base and management sub-system. Another feature of expert systems is the explanation facility. An expert system not only provides a solution, but can also explain why and how the solution is arrived. It is reasonable to insist on the explanation functionality of an IRSS.

- **Machine learning, data mining and text mining**. By applying algorithms of machine learning, data mining, and text mining to documents stored in an IRSS, one may discover patterns and extract knowledge useful to a user. Such functionalities are particularly useful to users who are interested in exploratory searching and browsing. A user can track trends in a particular area or discover emerging topics from the constantly changing document collection. A user may also discover

links between different documents or research areas.

- **Computer graphics and data visualization**. In many cases, a user may not want details about particular documents, but a general feeling, before going to a more in-depth analysis. With the granulation of document collection, it is possible to provide a user with granulated view, in which details are omitted. Most current information retrieval systems (IRS) present search results in the form of a ranked list of individual documents. In an IRSS, a user should be able to use graphics and visualization tools to view a particular document model. Visualization enables a user to perform high level inference and analysis.

- **Intelligent information agents**. Agents have been used by many IRS to collect information and interact with users. The potentials of agents need to be further explored for IRSS. A user should be allowed to construct a personalized agent to interact with an IRSS. The autonomy and learning capabilities of agents make them attractive to both IRSS and users.

6 Conclusion

The combination of computer technology with other disciplines of sciences normally produces useful products. Decision support systems are based on the combination of management science and computer science. By combining computer technology and research methods, we examine the notion of research support systems. The focus is on the assembling and integration of existing computer systems for the support of scientific research. Some preliminary and scattered ideas on the topic are discussed. The initial exploration suggests that WIRSS may play a major role in future study on research methods.

References

[1] Adams, G.R. and Schvaneveldt, J.D. *Understanding Research Methods*, Longman, New York, 1985.

[2] Beveridge, W.I.B. *The Art of Scientific Investigation*, Vintage Books, New York, 1957.

[3] Bohm, D. and Peat, F.D. *Science, Order, and Creativity*, 2nd edition, Routledge, London, 2000.

[4] Geisler, E. *Creating Value with Science and Technology*, Quorum Books, Westport, Connecticut, 2001.

[5] Graziano, A.M and Raulin, M.L. *Research Methods: A Process of Inquiry*, 4th edition, Allyn and Bacon, Boston, 2000.

[6] Hoggan, D.B. Challenges, strategies, and tools for research scientists: using Web-based information resources, *Electronic Journal of Academic and Special Librarianship*, **3**, 2002.

[7] Hult, C.A. *Researching and Writing Across the Curriculum*, Wadsworth Publishing Company, Belmont, California, 1990.

[8] Tang, H. Web-based research support systems, Manuscript, 2003.

[9] Ladd, G.W. *Imagination in Research: An Economist's View*, Iowa State University Press, Ames, Iowa, 1987.

[10] Martella, R.C., Nelson, R. and Marchard-Martella, N.E. *Research Methods: Learning to Become a Critical Research Consumer*, Allyn and Bacon, Boston, 1999.

[11] Power, D.J. and Kaparthi, S. Building Web-based decision support systems, *Studies in Informatics and Control*, **11**, 291-302, 2002.

[12] Turban, E. and Aronson, J.E. *Decision Support Systems and Intelligent System*, Prentice Hall, New Jersey, 2001.

[13] Wong, S.K.M., Yao, Y.Y., and Butz, C.J. Granular information retrieval, in: *Soft Computing in Information Retrieval: Techniques and Applications*, Crestani, F. and Pasi, G. (Eds.), Springer, Berlin, 317-331, 2000.

[14] Yao, Y.Y. Information retrieval support systems, *Proceedings of FUZZ-IEEE'02*, 773-778, 2002.

[15] Yao, Y.Y. Granular computing for the design of information retrieval support systems, in: *Information Retrieval and Clustering*, Wu, W., Xiong, H. and Shekhar, S. (Eds.), Kluwer Academic Publishers, 2003.

[16] Yao, Y.Y., Song, K. and Saxton, L.V. Granular computing for the organization and retrieval of scientific XML documents, *Proceedings of the Sixth International Conference on Computer Science and Informatics*, 377-381, 2002.

[17] Yao, J.T. and Yao, Y.Y. Web-based information retrieval support systems: building research tools for scientists in the new information age, *Proceedings of the IEEE/WIC International Conference on Web Intelligence*, 2003.

[18] Yao, J.T. and Yao, Y.Y. Information granulation for Web based information retrieval support systems, *Proceedings of SPIE*, Vol. 5098, 138-146, 2003.

[19] Zadeh, L.A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, *Fuzzy Sets and Systems*, **19**, 111-127, 1997.