



The University of Adelaide
Information Technology Division

GPO Box 498
Adelaide SA 5001
Australia

F:COM/3

8 April 1991

Australian Vice-Chancellors Committee
Special Projects Sub-committee of
the AARNet Advisory Board
GPO Box 1142
CANBERRA ACT 2601

*AARNet PROJECT
SUBMISSION*
12

Dear Sirs,

Application for AARNet Funding of
Guru - A Distributed Text Database

Enclosed is a response to the AV-CC memorandum regarding AARNet Project funding dated 19 March 1991.

This project has the support of the South Australian AARNet Advisory Board and is sponsored by The University of Adelaide.

We would recommend that funding be allocated to the whole project but that the project be reviewed after the successful completion of phase 1. We are unsure what additional load a fully functional system will place on AARNet and so we will attempt to perform some traffic analysis during phase 1 to ensure that the system will not detrimentally affect AARNet if phase 2 is deployed.

Yours sincerely,

Peter L Nissen
Director
Information Technology Division

12 APR 91: John - I have a file for AARNet Special Projects. do you want project funding to go on this?

Application for AARNet Funding of

Guru
A Distributed Text Database

Ross N. Williams
Renaissance Software

30 March 1991

Executive Summary

Thanks to the introduction of fast, reliable, networks such as AARNet, the factor now limiting the utility of the world internet is the inability of users to locate network resources just seconds away. While the network allows users to send mail, copy files, and read news, there is no automated facility for tracking down the resources that one might wish to use. Without such a facility, most of the network is functionally invisible.

This is a proposal for a project called *Guru* to create a keyword-indexed, distributed text database running on AARNet. The database will allow a list of keywords to be associated with any document on the network. Users will then be able to locate such indexed documents using keywords. Experience with an earlier program called *Guru Classic* has demonstrated the effectiveness of the simple keyword mechanism.

The goals of the project are:

- To rapidly implement a simple *Guru* system using *news* as a transport.
- To perform research into large distributed text databases.
- To produce a detailed design document for a worldwide database.

The first of these goals is certain to be achieved (it is technically simple) and will quickly provide an extraordinarily powerful network search facility. The remaining work will lay the foundation for an efficient worldwide database.

The project will be managed by the University of Adelaide.

Funding of \$35 000 in 1991 is sought from AARNet to enable this project to proceed.

Examples of Services Provided

Keyword searches are extremely powerful. In a working Guru system, a user wishing to search the internet would type in a list of keywords and the Guru system would respond by sending all registered documents (or their descriptions) whose keyword lists contain *all* of the specified keywords. Thus, in just three words the request

network australian database

narrows the search down to only those messages concerned with Australian network databases. The highly selective power of keywords allows sensible searches to be performed of extremely large unorganized databases. The flexibility of keyword searching allows the keyword mechanism to be used to search for a diverse range of services. The examples below indicate the range of services that are possible.

Locate people: If everyone on the network placed a message into the Guru system about themselves, it would be very easy to locate people by their name, location, or interests. For example, the query "person compression" would give a list of all the people on the network who were interested in compression. The query "john hanggliding" might locate someone called John you vaguely remember and who used to go hang gliding.

Locate Files: Guru would assist in the location of files on a diversity of topics.

National Help System: Guru would provide a nationally integrated help system. Users would enter messages describing the problems they were having and the solutions that they found to those problems. For example, the query "sun virtual memory problem" might locate all the other people on the network with virtual memory problems on their sun computer.

National Distributed Archive: A common theme in computer systems is that information with no home tends to get lost or irretrievably archived. Every day throughout the network, thousands of users destroy enormous amounts of useful information simply because there is nowhere coherent to put it. Guru will solve this problem by providing an unorganized, public place to locate such information, and would eventually become a national distributed archive. All this is possible because there is no centralized database: each user stores on his own disk the files he wishes to make public.

Project Evaluation Guidelines

This section addresses the guidelines specified as being used to evaluate proposals.

(a) The project should have national benefit and if possible involve more than one organization.

This project will have a fast, direct, and sustained national and international benefit. The effect of even the simple implementation proposed will be throw the light switch on the network and make everything visible.

Only one organization (the University of Adelaide) will be involved in the development of the database. However, once running, most internet users in Australia (and later the world) could become involved with presenting their documents to the database to be indexed.

(b) equipment is not funded unless it is to be purchased and owned by AARNet.

This project does not require the purchase of any equipment.

The word "database" conjures up images of hundreds of disk drives whirring away in a big room somewhere. Guru is the exact opposite. Instead of storing all the files in one place, Guru leaves the files in it's user's directories and merely carries indexing information. The result is that no central Guru equipment is required.

(c) the project may be developmental or feasibility related in nature. One of the aims of this program is to foster a body of knowledge in advanced network technology.

This project provides the best of both worlds. The first phase is to develop a simple working system to provide an immediate benefit. This working system will then provide a concrete foundation for the research and development of a more efficient worldwide database. This powerful combination of theory and practice is bound to lead to an increased understanding of advanced networks; and to top it all off, for the first time, users will be able to *find* this information!

(d) project funding is based solely on the merit of each project. There is no geographic component in the allocation of funds.

This proposal stands on its merit.

Organizational Details

Project name: The AARNet Guru Project (*Guru* for short).

Organization: The University of Adelaide.

Activity: The project will consist of two phases. Phase one is the implementation of a simple Guru system using *news* as a transport. Phase two is research into the problems involved with implementing a much larger and more sophisticated text database.

Management: The project would be managed by Mark Prior (Computing Officer) of Information Technology, the University of Adelaide. Mark would provide an administrative base for Ross Williams who would be employed under contract to execute the project.

System status: The Guru system resulting from this work would be placed in the public domain.

Projected Budget: A salary of \$45,000 was chosen as approximate for a short-term contract of the level of creative design, implementation, and research involved in this project.

Phase One:	3 months salary	\$ 11,250
	Project Management and use of resources	\$ 1,000
Phase Two:	6 months salary	\$ 22,500
	Project Management and use of resources	\$ 2,000
Contingencies:		\$ 1,250
Total Requested		\$38,000

Technical Details of a News-Based Guru (Phase I)

Architectural overview: The Guru database will be implemented by a set of participating Guru nodes. Each node will contain a daemon that processes requests carried by a special guru newsgroup `aus.guru` in the *news* system. To issue a request, users invoke their local Guru program and type in a request. The local Guru program then posts the request to `aus.guru`. As the request filters through the network, the Guru daemons of participating nodes search their local databases for matches. Matching messages (or a description of them) are mailed to the user who originally placed the request.

Node architecture: On each participating node, Guru is implemented by a Guru daemon process and by one or more index files that act as a registry for documents held on the machine. Because most users on the network can both read news and send mail, Guru will be able to be installed by any enthusiastic user without the intervention of a system programmer with root privileges. This flexibility will assist the spread of Guru through the network. The Guru daemon itself will consist of a non-terminating process that accepts requests to enter specified files into the local database, and which processes requests in `aus.guru`, responding to requests for which it finds a match.

Message architecture: A message is a text file containing special fields at the top of the file. In addition to the date and a one line summary, the special fields will also contain the list of keywords associated with the message. To enter a message into the Guru, a user would simply give the command

```
guru add filename
```

The Guru program would then ensure that the file is readable by everyone on the machine and would then send a message to the local Guru daemon requesting that the file be registered.

Appendix: Guru Classic

This proposal has arisen from earlier work by Ross Williams in 1988 in which he implemented a single-node text database now referred to as Guru Classic. Guru Classic allowed text messages to be indexed using the keyword mechanism described earlier. It was originally designed to assist in undergraduate teaching and was highly successful in that role. Since then, it has been used in a variety of other contexts.

Guru Classic has been installed at the following sites:

- University of Adelaide Computer Science Department.
- University of Adelaide Mathematics Departments.
- Flinders University Computer Science Department.
- Australian National University at Mount Stromlo Observatory.
- Australian National (Railways).

Guru Classic took just seventy hours to design, implement, and install.