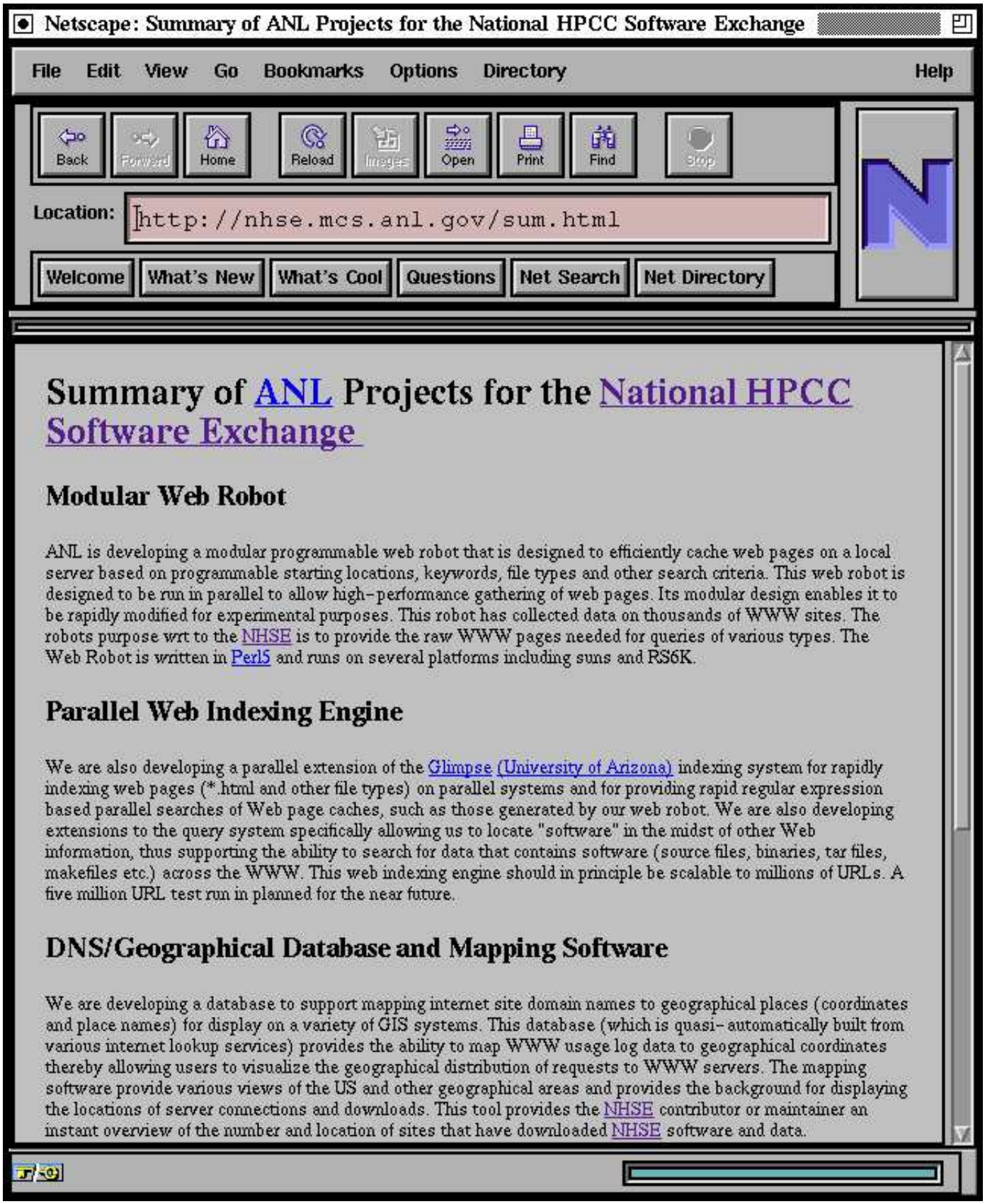
Netscape: Summary of ANL Projects for the National HPCC Software Exchange
File Edit View Go Bookmarks Options Directory Help
Back Forward Home Reload Images Open Print Find Stop
Location: http://nhse.mcs.anl.gov/sum.html
Welcome What's New What's Cool Questions Net Search Net Directory
Summary of ANL Projects for the National HPCC Software Exchange
Modular Web Robot
ANL is developing a modular programmable web robot that is designed to efficiently cache web pages on a local server based on programmable starting locations, keywords, file types and other search criteria. This web robot is designed to be run in parallel to allow high-performance gathering of web pages. Its modular design enables it to be rapidly modified for experimental purposes. This robot has collected data on thousands of WWW sites. The robots purpose wrt to the NHSE is to provide the raw WWW pages needed for queries of various types. The Web Robot is written in Perl5 and runs on several platforms including suns and RS6K.
Parallel Web Indexing Engine
We are also developing a parallel extension of the Glimpse (University of Arizona) indexing system for rapidly indexing web pages (*.html and other file types) on parallel systems and for providing rapid regular expression based parallel searches of Web page caches, such as those generated by our web robot. We are also developing extensions to the query system specifically allowing us to locate "software" in the midst of other Web information, thus supporting the ability to search for data that contains software (source files, binaries, tar files, makefiles etc.) across the WWW. This web indexing engine should in principle be scalable to millions of URLs. A five million URL test run in planned for the near future.
DNS/Geographical Database and Mapping Software
We are developing a database to support mapping internet site domain names to geographical places (coordinates and place names) for display on a variety of GIS systems. This database (which is quasi-automatically built from various internet lookup services) provides the ability to map WWW usage log data to geographical coordinates thereby allowing users to visualize the geographical distribution of requests to WWW servers. The mapping software provide various views of the US and other geographical areas and provides the background for displaying the locations of server connections and downloads. This tool provides the NHSE contributor or maintainer an instant overview of the number and location of sites that have downloaded NHSE software and data.