VIAS

VIAS – A Pathway to Knowledge.

The National Center for Supercomputing Applications presents VIAS, a domain specific information retrieval, archival, and processing system. Specifically designed to provide continually updated information on highly dynamic topics, VIAS is ideally suited for use with volatile information sources like the World Wide Web. It continually crawls the Web, monitors mailing lists and USENET newsgroups, and archives all of the information it finds on its topics of interest. However, *VIAS goes beyond simple archival of the knowledge it finds, to the autonomous generation of new knowledge and automatic categorization of existing knowledge*. A library of state-of-the-art proprietary metadata algorithms is continuously executed against the archives, providing a rich source of new knowledge to be queried or viewed in realtime.

WHY IS IT NEEDED?

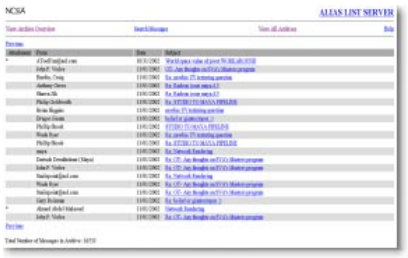
As the popularity of the World Wide Web explodes, there is increasingly an overabundance of information available. Information, however, does not provide a direct path to discovery, it is merely the *medium* from which desired knowledge may be extracted. Software tools must be used to selectively aggregate this information and intelligently process it to unlock its secrets.

Maintaining such collections over time is complicated by the volatility of the World Wide Web. Sites are posted, only to be changed or removed at any time, and the constant stream of electronic mail messages and USENET newsgroup postings are lost forever the instant they are sent.

VIAS was designed to address the unique needs of knowledge-based research on highly dynamic topics over time. By uniting mailing list and newsgroup archivers with web crawlers, VIAS is able to maintain ceaselessly updated archives of topically oriented information. A library of processing tools are brought to bear upon these archives, shedding new light upon them and unleashing their hidden potential.

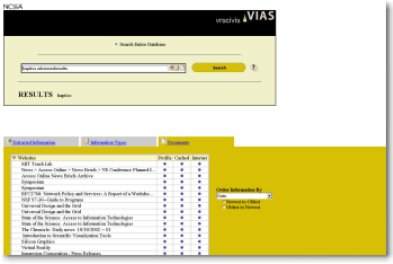
HOW DOES IT WORK?

The VIAS system operates in three distinct phases. In the first phase, a user sets up a VIAS archive to gather information in his or her area of interest. This is done by defining the scope of interest and seeding the system with potential sources of information. VIAS then initiates a search of these information sources, including the World Wide Web, electronic mailing lists, and USENET newsgroups. Each potential document is stripped of navigation bars, advertisements, and other extraneous text, scanned for relevancy, and archived if appropriate.



The second phase consists of subjecting the archives to a series of algorithms in which company names, acronyms, and other metadata are discovered and placed in the archives. Both phases are done outside of realtime to allow thorough searching and rigorous metadata extraction.

Finally, the archives are queried via any web browser. Search terms are entered, desired metadata is indicated, and the results are delivered in realtime to the browser.



HOW MIGHT IT BE USED?

As a scientist, one might be interested in finding the names of other researchers who are working on problems of a specific type, the names of companies that offer particular products, or a list of bibliographic references related to a new area of interest. The World Wide Web, electronic mailing lists, and USENET newsgroups all provide related information, but it quickly becomes a scavenger hunt to try to find the desired names or references. With VIAS, the researcher merely needs to type the query into a web browser and select the types of information to be returned. VIAS responds with the requested metainformation and the URLs or archived messages from which it originated.

WHERE CAN I LEARN MORE?

Check out the VIAS website at:

http://vias.ncsa.uiuc.edu

NCSA