

## The Soft Pages Project

1 V-6

Th. Johannsen \* G. Mansfield † K. Takahashi ‡ S. Noguchi §

## 1 Introduction

With the growing of Internet and its community, electronic publishing has become a frequently used way of announcing research results, standards, and the like. Most of these documents are kept in files on various file servers, mainly available by anonymous ftp or FTAM. In fact, anonymous ftp is one of the most popular applications of the Internet.

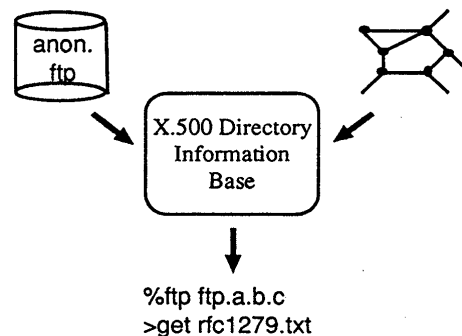
Existing file information tools (like WAIS, WWW, Archie) have concentrated on finding the documents/files for given search criteria. By the means of these tools users are enabled to find the piece of information they have been looking for and a way to retrieve these documents.

However, the problem of finding and retrieving a document is solved only from the users point of view. For network managers it is relevant where users get their files from. FTP causes 30 to 50 % of backbone traffic. This trunk route traffic (using rare bandwidth in most cases) can be reduced without limiting the service for users. The idea is to make more use of copies of documents held at file servers near the user and thus shifting ftp traffic to local and less-crowded national networks. Usually, a network user will not know all alternative sites for file transfer and go straight to a well known server to get the files he wants. To change this behaviour the user must be provided with additional information which is best done by giving him a powerful new file access tool.

## 2 File Server Evaluation

The Soft Pages Project gives a suggestion how to retrieve a document in a network traffic optimized manner. The strategy is to find out how the user's site is connected to various ftp servers in terms of speed / bandwidth, charges, network load on that connection etc. Therefore, for each Internet connection (line) a cost index is calculated from these parameters which carries the information whether it is "cheap" or "expensive" to use this particular line. Then several alternatives (i.e. connections to several ftp servers) are compared by their cost index.

By this way we can evaluate file servers with respect to the properties of connections from the users site to these file server sites. The network configuration information necessary for cost calculation and line evaluation is stored with the X.500 Directory and therefore globally available.



## 3 Information about Files

Information about documents / files is kept by Soft Pages in the Directory, as well. For each file server, names and attributes of files are stored and updated periodically. This provides global access to Archie-like information for all registered file

---

\*Thomas Johannsen (Tohoku University, Sendai)

†Glenn Mansfield (AIC Systems Laboratories, Sendai)

‡Kaoru Takahashi (Tohoku University, Sendai)

§Shoichi Noguchi (Tohoku University, Sendai)

servers and, furthermore, opens the way to store document description together with the file name. Users will search the Directory for the name or other attributes of files they want to retrieve. The answer they get from the Directory will contain the address of the ftp server holding such a file.

If a document is stored at two or more sites, the site with the lowest cost index (which naturally will be the "nearest" in network terms) is chosen for retrieval. This selection is done by the Soft Pages User Client and might be hidden from users.

Using the Directory, document search can be much more powerful than Archie because it is not restricted to file name matches but might be run for keywords as well. As soon as worldwide unique document identifiers are available, they will be held as attributes to document (file) entries in the Directory and increase the value of the database.

## 4 X.500 Extensions

In order to set the databases up, new X.500 objects had to be defined.

There are objects for

**network configuration :**

- network
- line
- node
- port

**IP address look-up :**

- IP group (network address)
- IP address (host address)

**file name look-up :**

- file server
- file

## 5 Status of the Pilot Project

Tohoku University, AIC Systems Lab. and WIDE are working on the Soft Pages Project. A pilot consisting of two Directory System Agents (DSAs)

is running at AIC Systems Lab. and can be accessed from other participants. Work has been done to provide an X-Windows based Soft Pages User Agent.

The Directory Information Base of this pilot holds entries for nearly all (connected) IP networks in Japan and provides file information for roughly 25 registered anonymous ftp sites from Japan.

## 6 Future Plans

There are several extensions planned in order to transfer Soft Pages from the pilot stage to a stable and broadly available service. More networks will get involved by setting up DSAs of their own. The most important part will be to put more and more detailed information about network configuration into the systems knowledge. Evaluation of network connections is useful only when based on accurate configuration data.

Furthermore, network configuration information collected for Soft Pages and put into the Directory can be used for network management (especially configuration management) and for network information services as provided by Network Information Centers (NICs). These applications are subjects of study.

## 7 References

CCITT:

The Directory - Overview of Concepts,  
Models and Services  
Recommendations X.500 - X.521

Johannsen, Th., Mansfield, G., Noguchi, S.:

The Soft Pages Project -  
Optimized Document Retrieval  
Paper for Network Services Conference '92,  
Pisa, Italy, November 1992

G. Mansfield, Th. Johannsen:

The Soft Pages Project -  
Activity Report 1991 / 1992 for WIDE  
Technical paper, July 1992  
ftp.tohoku.ac.jp:pub/spp/wide.ps