Integrating Internet Resource Discovery Services with Open Hypermedia Systems

Rupert Hollom
Wendy Hall

CSTR 93-18

Department of Electronics and Computer Science
University of Southampton
Southampton SO9 5NH

Abstract

Over the past few years the number of people accessing the Internet and the quantity and variety of resources available through this medium has increased dramatically. To enable easier access to these information stores systems have been developed to partially automate the location and retrieval of any required part of this data reserve. These utilities can, at present, be used in conjunction with existing hypermedia systems as peripheral parts rather than as an integrated item. This paper will discuss these systems and investigate methods by which they can be used, and how they may increase the effectiveness of hypermedia systems, such as Microcosm, if they can be made an integral part of such software environments.

ontents

1. Introduction.

Although the Internet has been under construction for just over twenty years until recently the main areas of activity have been those of physical connectivity, i.e. spreading the territory of the Internet, data transmission integrity, speed and capacity. The result of these efforts has been a marked increase in the areas using the Internet. The number of users connecting to the Internet has also risen dramatically as an indirect result of the increased availability and reliability of the service. It has been estimated that one million machines are connected interactively and a further several hundred thousand are periodically connected (electronic mail and network news) daily to the Internet [Schwartz, M.F., Emtage, A., Kahle, B., Neuman, B.C. (1992)].

As a result of this increased utilisation the volume of information available through this world-wide network has now reached hundreds of terabytes, the American Library of Congress holds approximately twenty-five terabytes in its archives alone [Stein, R.M. (1991)].

The time taken for a user to browse such vast tracts of data would be unacceptable to all but the most fool-hardy, so to try to enable Internet users to find the resource that they require services are being developed that give them a simple interface to locate and retrieve resources.

Whilst the Internet has been burgeoning there has been extensive interest in the fields of hypertext and hypermedia, and although it is not a recent idea, the area where these two developing technologies meet is certainly an exciting one. An early exponent of the use of hypertext together with data storage and retrieval techniques was Nelson with project Xanadu [Nelson, T.H. (1988)], other attempts include KMS, based on the ZOG system developed at Carnegie-Mellon [Akscyn, R.M., McCracken, D.L., Yoder, E.A. (1988)] and Intermedia developed at Brown University's Institute for Research in Information and Scholarship (IRIS) [Yankelovich, N., Haan, B.J., Meyrowitz, N.K., Drucker, S.M. (1988)]. The main difficulties with these systems is that they used a certain amount of 'mark-up' within the documents, therefore the original integrity of the document was lost. Although Intermedia held the links separately, the documents were still marked to indicate link positioning. Systems such as Microcosm hold the links separately so there is no alteration to the document, which allows links to be placed in documents where the author has read-only access.

At present Microcosm can be used with existing Internet Resource Discovery Systems by adding them as viewers to the system [Hill, G., Wilkins, R., Hall, W. (1992)]. This scenario makes it difficult for the user to link the information being accessed in the hypermedia environment with the Internet resource bases being queried. For example a user could not select a piece of text within the hypermedia environment and automatically query an Internet resource. There is a need therefore that such resource discovery systems become an inherent part of any hypertext system that is going to be of more than cursory usefulness within the wider context of the Internet.

2. Overview of Internet Resource Discovery Services.

There is an ever increasing number of Internet Resource Discovery Services (IRDS's), this report will not attempt to cover every one, but rather outline a number of the better known ones. These will include Alex, Archie, Gopher, Indie, Prospero, Unique Resource Locators, Wide Area Information Servers (WAIS), World Wide Web (WWW). Not all of these systems are true discovery tools, some are methods of imposing a user's view onto the structure of the Internet. All these systems share the common goal of simplifying the task of accessing information. These systems not only access resources containing material relevant to the fields that enabled their existence but to virtually every other discipline imaginable. They also enable users to impose their own personal view on this global data store. The order in which these services are discussed in this paper is purely arbitrary and is not meant as a classification in any way.

2.1 Alex

Alex [Cate, V. (1992)] is a file system, developed at the School of Computer Science at Carnegie-Mellon University. It provides users with transparent read access to files on anonymous Internet FTP sites. With this method applications can also access the files on any of the anonymous FTP sites without having to log on to the site. The Alex file system uses a cache to try to get reasonable response times, within the cache details such as machine names, directory information and the contents of remote files are stored. Alex uses a soft consistency mechanism to guarantee that only updates that have occurred within 5% of the reported file's age on the FTP site might not be reflected locally. For example if a file had been resident on an FTP site for 20 days only changes that had occured within the last day would not be seen on the Alex server. At present Alex is implemented as an NFS server as shown in figure 1, this means that the Alex server can be mounted as a logical drive by any machine that uses NFS.

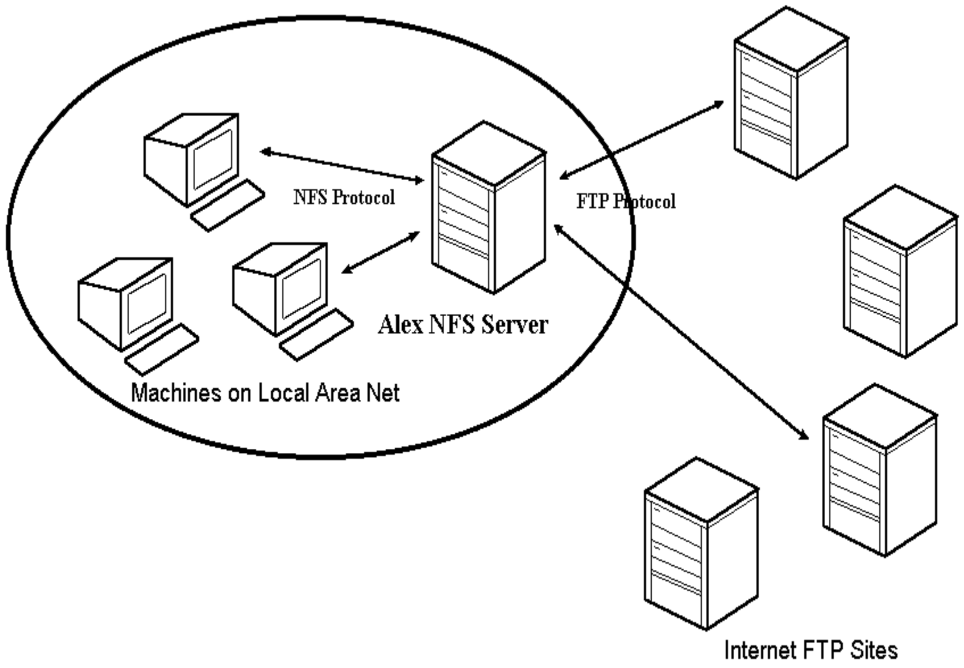

Figure 1 : The logical structure of Alex.

2.2 Archie

Archie [Emtage, A., Deutsch, P. (1992)] maintains a database of files that are retrievable by anonymous FTP from sites that are scattered all over the world. A user can query the Archie database in two ways, either by searching for a particular filename within the database or by searching the database with reference to programs that perform a particular required function. The Archie database is updated monthly by performing a recursive directory listing of each of the archive sites that are registered with the system. The Archie service, like Gopher (which is described in section 2.3), is divided into client and server portions. The server application is divided into three main constituent parts, the data gathering component (DGC) and data maintenance component (DMC) maintain the database with respect to the filenames of the FTP sites. The third part of the server is the user access component (UAC) which is where the clients access the database. The major difference between the Gopher service and the Archie service is that the former finds the related data files for the user and, if required, retrieves them from the remote store, whereas the later can only notify the user as to the location of the files; it is then the responsibility of the user to procure them.

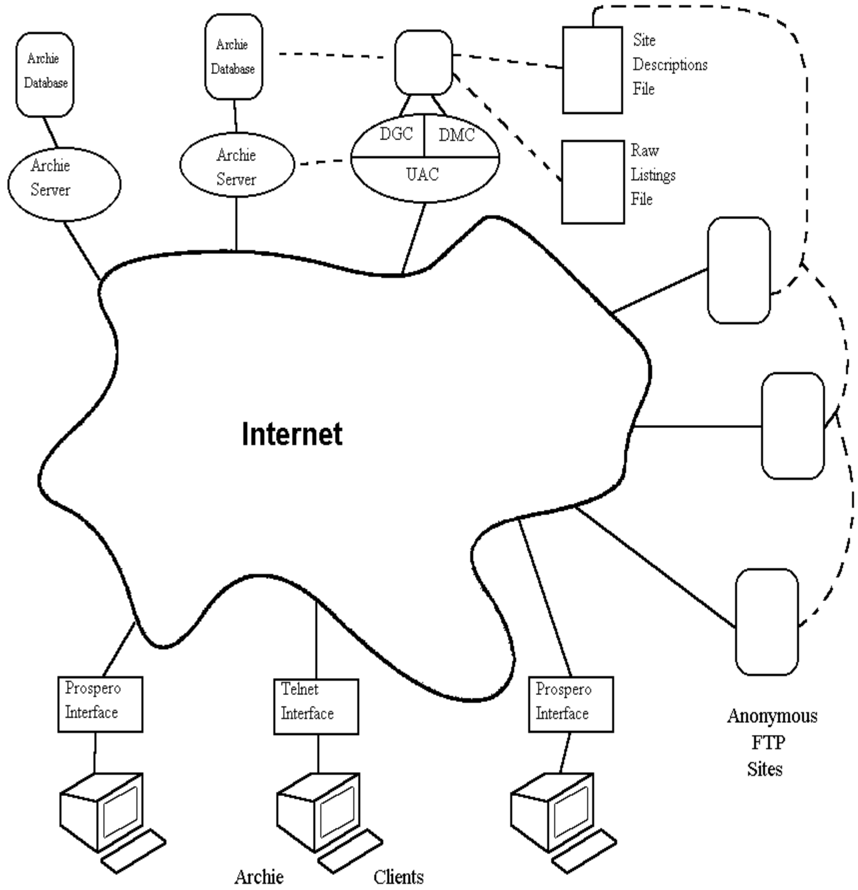

Figure 2 : The Basic Archie architecture

The basic Archie architecture is shown in figure 2, as can be seen from the diagram there is more than one Archie server, currently there are thirteen replicated servers around the world and the user can choose the site that is geographically closest to them. There are three possible ways by which Archie database can be accessed; these are telnet, e-mail and more recently the Prospero interface (Prospero is covered in more depth in section 2.5). The telnet interface has been found to be rather intensive on the servers resources so the other two methods are preferable from the point of view of the site at which the server is located. To try to maintain consistency between the database strewn over the world there is a central database in Montreal, Canada which regularly checks the FTP sites. The other sites update their databases from this master. It has been estimated that fifty percent of all Internet traffic to and from Montreal is directly related to the Archie update mechanism.

2.3 Gopher

The Internet Gopher service is implemented as a group of autonomous clients and servers operating within the Gopher information space. This space can be thought of as a generalised directed graph or hierarchy of information. Leaf nodes within this hierarchy are documents and the intermediate nodes are directories or indices. The Gopher service utilises its own protocol [Alberti, R., Anklesaria, F., Lindner, P., McCahill, M., Torrey, D. (1992)] which is implemented on top of TCP-IP (Transmission Control Protocol/Internet Protocol), the basic architecture of Gopher is shown in figure 3.

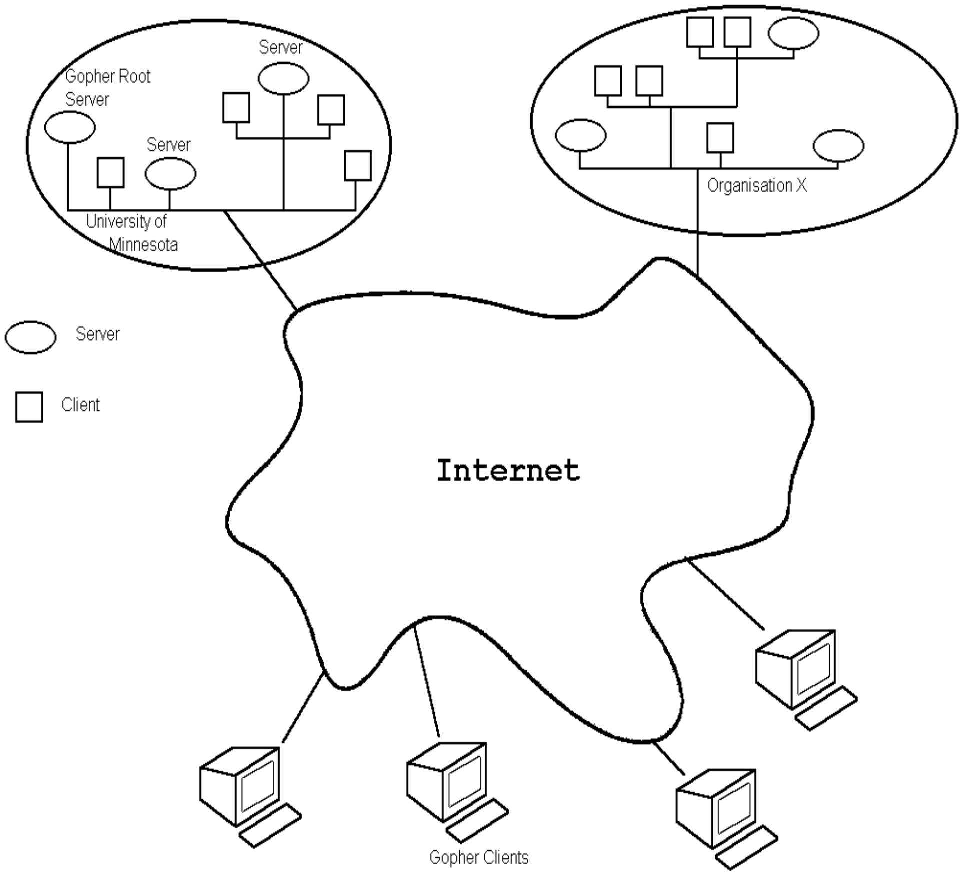

Figure 3 : Gopher.

As mentioned previously the Gopher service is based upon a hierarchy of information and the root of this tree is stored at the University of Minnesota on the host rawBits.micro.umn.edu. This is the default directory that is retrieved by a Gopher client when first invoked. It is possible, however, to alter this default directory to one that is more applicable to the user's requirements. For example it would not be sensible to set the default directory to one that is stored on a machine in New Zealand if the user were located in Southampton; instead it would be much more sensible to use the directory that is stored on host gopher.ed.ac.uk .

The Gopher architecture allows for a hierarchy of servers so that there could be a top-level server for an organisation and then various lower-level servers for the departments within the organisation. This allows the user to gradually hone the search until the required resource is located. The service is available in two forms, the first is a series of menus through which the user navigates picking entries of interest so that the Gopher client can retrieve the next level of the menu structure until eventually the information is found. The second method is a full-text search implemented by utilising special Gopher search servers which hold full-text inverted indices of subsets of the documents stored in the a Gopher server. A Gopher search server can be set up to index more than one normal Gopher server so that any particular logical area, i.e. field of interest, can be covered by one search server even though the documents may not necessarily be located on any one server. Recent Gopher clients also allow access to information stored on WAIS, Archie and FTP servers as well as the Gopher servers.

2.4 Indie

Indie, or to give it its proper title Distributed Indexing [Danzig, P.B., Ahn, J., Noll, J., Obraczka (1991)], [Danzig, P.B. Li, S.-L., Obraczka, K, (1992)] is a resource discovery tool that draws together the Internet's resource discovery structure. The basic structure of Indie is similar to WAIS, although the terminology used is different. There is one directory of services (this is actually replicated a number of times) and any number of broker databases. These brokers index data from various sources, which include their own database, data stored in other brokers, and data available from other sources such as Archie. The various copies of the directory of services are all equal; there is no master copy from which all the others get updated, instead when a new client or broker registers, it can do so with any of the directory of services and the algorithm for Indie's update and recovery guarantees that all the replicas of the directory will eventually learn of the change.

2.5 Prospero

The Prospero file system allows users to create customised views of a global file system [Neuman, B.C. (1992)]. Prospero is not actually a system by which users can search the Internet for that data they require but rather a method by which they can select a view of the Internet that they find most useful. In this respect it is similar to the Andrews File System [Howard, J., Kazar, M., Menees, S., Nichols, D., Satyanarayanan, M., Sidebotham, R., West, M., (1987)] and the Alex file system [Cate, V., (1992)]. The user can build for themselves a virtual file system in the typical hierarchical structure with the files and directories of most interest to them near the root, so that they have shorter path names, and those of decreasing interest getting further out into the 'branches' and so having longer filenames.

2.6 Unique Resource Locators

Unique Resource Locators [Berners-Lee, T.J., (1993)] are not a method of searching for Internet resources or indexing them, but are a technique for describing the location of a particular file. This system has been devised to enable the various different formats of Internet Resource Storage to be uniquely distinguished. It also needs to be extensible for any future services. The system has, at present, been mainly used in World Wide Web and has a full Backus-Naur Format description.

2.7 Wide Area Information Server (WAIS)

The WAIS project was a joint venture between four companies, Thinking Machines Corporation, KPMG Peat Marwick, Apple Computer, Inc., and Dow Jones & Company. Since its inception a new company, WAIS Inc., has been formed to market the WAIS products for different computer platforms that have been produced from this collaboration. Two of the major goals of the original project were :

- Provide users with a uniform, easy-to-use, location transparent mechanism to access information.
- Allow a user at a workstation to catalogue and view information from a large number of sources. [Kahle, B. (1989)]

The WAIS model is based on the typical client-server design and is shown diagrammatically in figure 4.

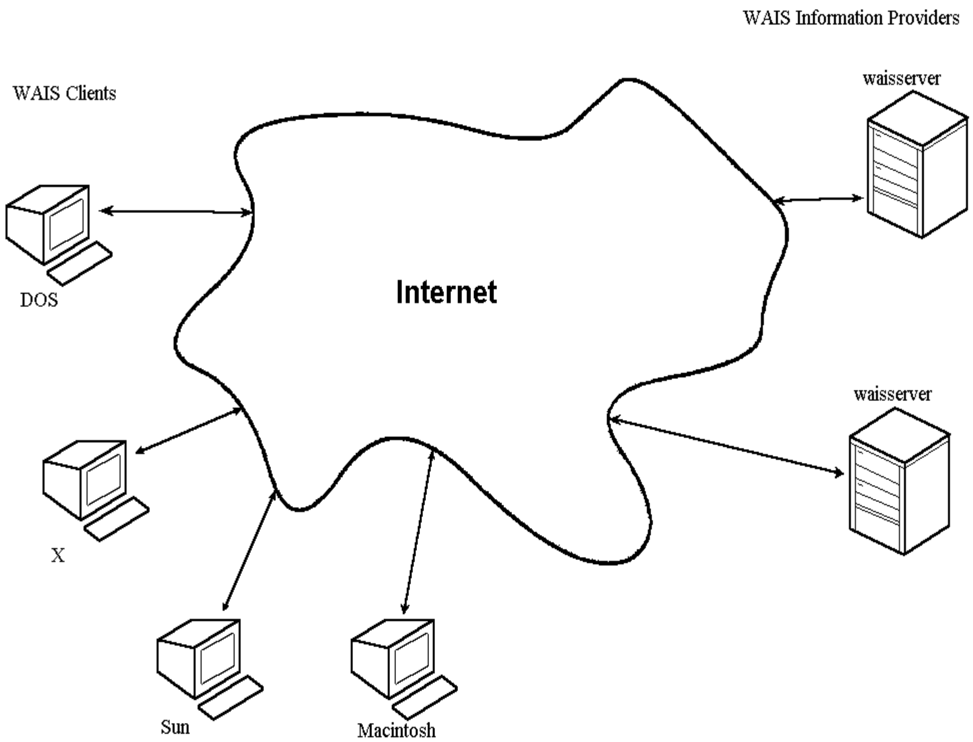

Figure 4 : WAIS Client-Server Design.

Each server keeps a complete inverted index of all the documents within its database and hence can use a full text retrieval system when a query is lodged with the server. The server can then respond with a set of all the relevant documents which are selected from the database using a word weighting algorithm to find the best matches. The set can also contain the names of other servers that have registered with the server being queried. This is unlikely unless the query was directed at the directory-of-servers server with which all WAIS servers must be registered for them to be publicly accessible. WAIS can therefore be seen as a set of decentralised indices all being accessed transparently to the user.

The client application displays the set of matched documents, which may be any format, (e.g. postscript, text, graphics, animations, etc.), The user selects the required document(s) to be retrieved from the database for display. If a particular document proves to be especially interesting the user can utilise a feature called relevance-feedback, this enables the user to select a document, or section of a document, and to re-run the query so that other documents similar to the one selected are also returned in the set. This selection process ranks the documents in terms of the number of words in common.

The protocol used for communication between the client and server(s) is an extension of the NISO Z39.50 protocol [Lynch, C. (1991)] so that other services that also wish to communicate with the WAIS servers have a standard with which they can conform to ensure compatibility.

In the brief time that the WAIS project has been running it has already proved to be quite a success. There were over 225 publicly registered databases as of June 1992 and over 6000 hosts with an estimated 10,000 users accessing those servers, each of which have a "specialised subject".

2.8 World Wide Web (WWW or W3).

The original concept of the World Wide Web was developed at CERN as an aid to the high energy physics community.

"The World Wide Web initiative encourages physicists to share information using wide-area networks." [Berners-Lee, T.J., Cailliau, R., Groff, J.-F., Pollermann, B., (1992a)]

The Web allows 'pages' of information to be displayed and within these pages there are hypertext links to other pages within the system. The documents at these end points need not necessarily be on the same server as the document from which the link originated but this is all transparent to the user. New pages can be added to the system and then links made from existing pages that are relevant to the new addition. Links can also be made from the new page to existing documents. This means that the user can browse through this environment following any links that they find interesting and possibly finding that new links have been added since their last visit to a particular document.

Thus this model merges the techniques of information discovery on the Internet and hypertext. The user has no need, and in most case no wish, to know the underlying mechanics when a link is followed or where the information is coming from. Instead they are interested in the content of the information. It can therefore be said that the World Wide Web organises the information available via the Internet into a distributed hypertext model with a client application running on the users/browsers machine and various servers around the globe providing the information required.

When the original idea of the World Wide Web was being considered it was decided that having a purely hypertext based system would not be flexible enough for all tasks that would be undertaken, since in quite a few instances it would not be obvious which of the hypertext links to follow to find the particular information. To this end the system was designed and built with two separate discovery models available :

- one based on the hypertext paradigm of following links from highlighted sections of text.
- the other based upon the flat search paradigm for accessing indices in the information space.

The benefits of adopting both these approaches is that it gives the World Wide Web user access to other Internet resources that cannot be easily formatted into hypertext form, such as Gopher servers, WAIS databases, Network News groups, and anonymous FTP sites as well as the World Wide Web servers. This, together with the architecture of the World Wide Web is shown graphically in figure 5.

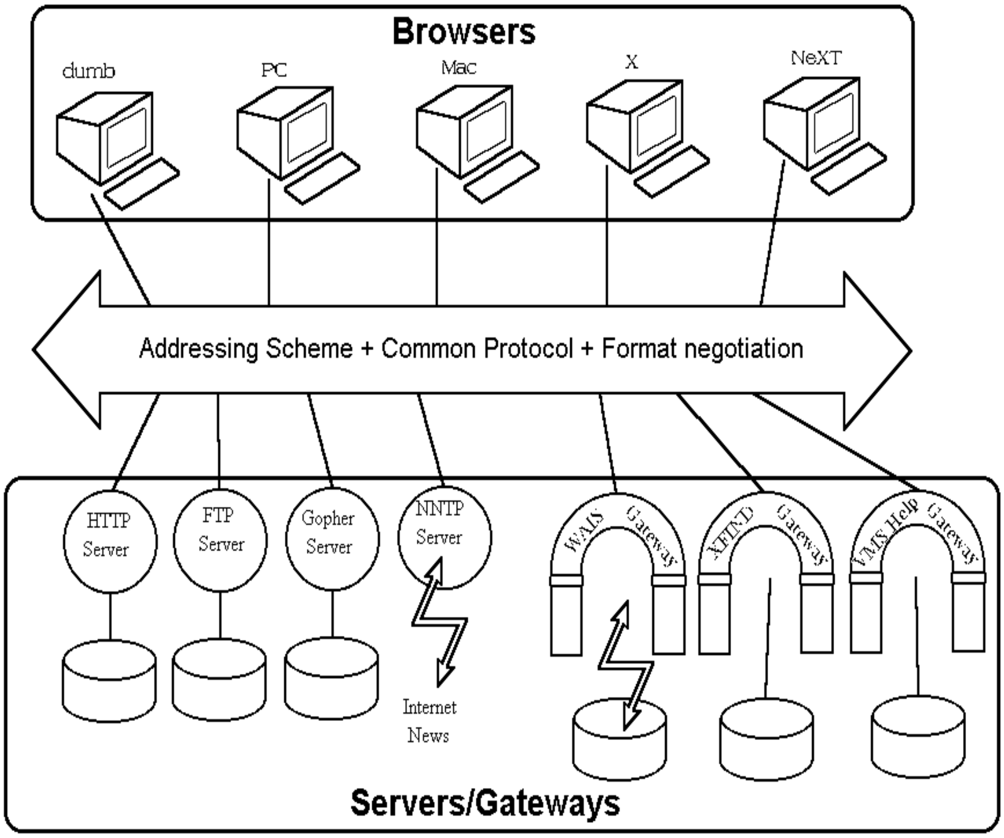

Figure 5 : World Wide Web Architecture.

When a client application is first being installed a default cover page can be specified which will be retrieved and displayed whenever the application is started. There is a standard front page available on the CERN server and this gives access to the three discovery trees currently supported by the World Wide Web, the three trees are :

- Classification by subject/server type

- High-energy physics (as this was the field that the World Wide Web was originally set-up to support. It features prominently in the information stored on the system, especially the CERN server)

- Classification by organisation

To allow links to be embedded within the documents accessible by the World Wide Web a form of SGML (ISO 8879:1986) is used, called Hypertext markup Language, or HTML. Markup is used to indicate the position of a link in the document and also the page to which it is linked. The description of the end point of the link is specified using a Unique Resource Locator (URL). These are discussed later in this paper. If the user wishes to follow a link they simply click with a mouse button on the area of highlighted text. The document at the end of the link is then retrieved, using a Hypertext Transfer Protocol (HTTP). A new protocol was used to give World Wide Web servers features that were not available via existing protocols with adequate performance for following hypertext links.

On the whole the idea that "one view encompasses all systems" [Berners-Lee, T.J., Cailliau, R., Groff, J.-F., (1992b)] seems to have been reasonably successful.

3 Using currently available Internet Resource Discovery Systems with Open Hypermedia Systems.

One of the major problems of authoring a hypermedia application is the task of collecting the resources from which the application is constructed. It would therefore seem sensible to use Internet resource discovery systems in conjunction with hypermedia systems, thus alleviating to some extent the task of finding suitable documents.

The Internet resource discovery systems that the rest of this paper will concentrate on are the WWW and WAIS. However the techniques discussed here are applicable to all of the discovery systems covered previously, but not to systems, such as Prospero, that organise the user's view of the Internet. Systems such as Prospero would have to be implemented at an operating system level. The hypermedia system would therefore automatically use such systems directly.

At the University of Southampton in the Image and Media Lab. an open hypermedia system, called Microcosm [Fountain, A., Hall, W., Heath, I., Davies, H. (1990)], [Davies, H., Hall, W., Heath, I., Hill, G., Wilkins, R. (1992)] has been developed and it is this system that will be considered in the remainder of this paper.

3.1 Using existing Internet Resource Discovery Systems with Microcosm.

Microcosm has been implemented as a set of autonomous interacting processes running under Microsoft Windows 3.1. The core of the system is implemented using a filter based model [Hill et. al. (1992)]. Each process in the filter performs a specific task. Also included in the system are a number of viewers. It is with these that the browser/author interacts with the system. There are different viewers for the different file types, e.g. a text viewer, a bitmap viewer, etc.

If one of the previously mentioned discovery systems were to be added to Microcosm in their "raw" state then it would have to be as a viewer, because a filter must be able to accept Microcosm messages, act upon them if need be, and then pass the messages on to the next filter in the chain. None of the Internet systems has any degree of tailorability so it would not be possible for them to accept an incoming message nor to send the message to the next filter.

The different Microcosm viewers fall into one of three categories of Microcosm "awareness". Specially written viewers such as the text viewer are fully aware so they can interact with the rest of the Microcosm system on all levels. The next tier down are the partially aware viewers. These are usually mainstream applications that have some degree of programmability included so they can be altered to understand some of the Microcosm messages and interact with the system to some extent. The lowest level is that of unaware viewers. For example Windows notepad cannot be altered at all to use the standard Microcosm messages but it can be started by Microcosm with a specific document. To pass information out to the hypermedia system notepad must rely upon Microcosm monitoring the clipboard for any changes and then the appropriate action can be taken. It is into this last group that all the existing resource discovery systems are cast, because Microcosm does support external applications it makes it a reasonably simple task to use programs that are unaware. This means that the author could make a link from a piece of text, or an area of a bitmap, to the discovery system so that when the hypermedia browser follows the link the discovery system would be started.

As mentioned earlier there is no possibility of two way communication between Microcosm and the Internet resource discovery system so the resources thus discovered would not be directly available to the hypermedia application. They would have to be saved using the discovery system and then imported into Microcosm which makes the whole operation rather circuitous. Another problem with this approach is the lack of a common interface between the hypermedia system and the discovery system so that it would be all too easy for the browser or author to become very confused between the two systems. It would be much better if the two systems were properly integrated.

3.2 Integrating Internet Resource Discovery Systems into Microcosm.

The filter based approach of Microcosm allows for new link creation methods to be added with a minimum of difficulty. At present Microcosm has a "Compute Link" filter [Li, Z., Hall, W., Davies, H., (1992)] and it is envisaged that the Internet resource discovery systems would be implemented in a similar manner to this. The user would select a block of text and then select "Discover Links". A message would be built by the text viewer and passed along the filter chain until a filter capable of acting upon the message received it. This would be the discovery filter and depending upon the method implemented it would contact the appropriate servers in an attempt to find relevant documents. If any suitable documents were found then the author/browser would be able to select the required one which would be retrieved and displayed.

This raises some interesting problems :

- How to locate the resource ?

- How to retrieve the documents ?
- How to display the documents ?

The technical method of solving each of the above problems is covered by their various protocols. The main question is the point at which the different operations should be implemented within Microcosm. As intimated in the previous paragraphs a new filter would have to be written to locate suitable resources that might hold relevant documents. A first attempt at such a filter is currently in progress and is based upon the WAIS discovery methodology. Once documents have been found they then need to be retrieved from the remote server for display on the local machine. The logical place for the document retrieval functionality would be in the Document Management System (DMS) portion of Microcosm, which could perform the necessary transportation tasks to make a copy of the document on the local machine. Once this had been completed a message would be dispatched to the appropriate viewer indicating the new document to be displayed. In most cases the documents are purely textual so the standard Microcosm text viewer could be used but if the system were to be widened to allow the system to access other systems such as World Wide Web and Gopher new viewers would need to be written to cope with the specialised document structures and layout.

If the World Wide Web system were to be fully integrated into Microcosm then not only would a new viewer have to be written to cope with the HTML format that World Wide Web documents use but it would also have to be able to extract the linking information contained within these documents. This would have to then pass the information on to the DMS which would retrieve the document specified as the end point of the link.

4. Why Internet Resource Discovery Services should become an integral part of Microcosm.

Microcosm is a resource based hypermedia system. In its present form the resource needs to be on the local machine, or a least upon logical drives mounted upon the local machine, although a distributed version is currently being written. The next logical step is to widen the scope of the resource base and if Internet Resource Discovery systems are fully integrated into Microcosm then the available resource base effectively becomes the entire Internet with all its wealth of documents and information.

The extensible nature of Microcosm allows new ideas such as these to be seamlessly integrated into the system so that the author/browser can interact with the system in the usual manner with no knowledge that the document may be coming from a distant server or from the local hard disc. Existing systems require a plethora of applications to discover new resources, link them into the hypermedia application and browse them. Integrating resource discovery into Microcosm gives the user a consistent interface with which to work, hence lowering the cognitive overheads imposed by many different applications. The user can devote more intellectual effort to the content of the hypermedia application so increasing productivity and the applicability of Microcosm to all fields.

Another benefit of such a strategy is that it would allow the browser more flexibility in exploring the subject area of the hypermedia application. If a new aspect of the subject occurred to the user as they were browsing/exploring the system then related documents could be located and built into the users personal view of the system for future reference even if the original author had not thought to explore that particular avenue.

The SERC funded SuperJANET project promises to have a pervasive network between institutions that can deliver data at speeds in the range 10Mbs to 155Mbs. When the network is in place the long vaunted promise of digital video and sound deliverable over networks will be truly possible. The scope of SuperJANET will not be as far reaching as the Internet but will still allow UK institutions to interchange and access remote hypermedia applications in reasonable time frames. The quantity and diversity of resources available to the author/user will blossom so making the task of locating the required documents even more troublesome than at present.

It has been shown by projects such as WWW and WAIS that discovery systems are a valuable addition to the tools available to the user. The number of people choosing to use them indicates this. If a unified system could be produced under which many of the different methods could seamlessly operate the popularity of them would dramatically increase.

5. Future Work

As mentioned previously in this paper a WAIS filter for Microcosm is presently being written. Hopefully other protocols will follow allowing Microcosm to use additional on-line discovery systems as well. Further into the future is the possibility of semi-intelligent 'agents' linked to Microcosm that will search the databases for documents that are relevant to the text selected, taking into account the meaning and context of the words rather than simply performing a word count comparison.

Also connected to the ideas outlined in the paper is the possibility to control a Microcosm session remotely. This would be a particularly useful aid for tutors, enabling them to demonstrate particular aspects of an application that they feel are important to the students. The first version of this will be written to work over a local area network but eventually it should be possible to alter the software so that the remote machines can be located anywhere that there is a network connection.

6. Conclusion.

As the amount of information available through the Internet has grown and become more accessible to users with little or no knowledge of networks, systems such as those discussed within section 2 of this paper have become necessary and hence have slowly come into being.

This paper has presented ideas for the integration of these services with open hypermedia systems, such as Microcosm, so that industrial strength hypermedia systems can be created and utilise the entire gamut of resources available via the world's networks. This will enable a richer environment to be constructed in which to build hypermedia applications and also enhance hypermedia's applicability to more areas of knowledge.

Also with the speed of data transmission over the global network ever increasing it will soon be possible to have a central store of digital video, sound, etc. and then deliver it on request in real-time over the network, although the band width required would be rather high. Collaboration on a massive scale will become a possibility with such networks, allowing for a much broader base of available applications. It is imperative, therefore, that discovery tools such as those mentioned in the body of this paper be incorporated into hypermedia systems as soon as possible. Thus allowing users to concentrate on the more important, and interesting, task of creating the application as opposed to finding the material with which to construct it.

It would, however, be wrong to suggest that Internet Resource Discovery systems are a panacea to the difficulties of locating resources. It is still extremely difficult to locate suitable diagrams for a particular topic because there is no universally accepted classification system for pictures. Research is continuing in these areas so in the not too distant future automatic location of pictures and digital video should also be a possibility, so allowing truly global hypermedia applications to be produced.

7. References.

Akscyn, R.M. McCracken, D.L., Yoder, E.A., (1988), "KMS : A Distributed Hypermedia System for Managing Knowledge in Organisations", *Communications of the ACM, Vol. 31, No. 7, July, pp. 820-835*

Alberti, R., Anklesaria, F., Lindner, P., McCahill, M., Torrey, D., (1992), "The Internet Gopher protocol : a distributed document search and retrieval protocol", *On-line documentation, Spring*

Berners-Lee, T.J., Cailliau, R., Groff, J.-F., Pollermann, B., (1992a), "World Wide Web : An Information Infrastructure for High-Energy Physics", *Proceedings International Workshop on Software Engineering and Artificial Intelligence for High Energy Physics, La Londe, France.*

Berners-Lee, T.J., Cailliau, R., Groff, J.-F., (1992b), "The World Wide Web", *Computer Networks and ISDN Systems, Vol. 24, No. 4-5, pp. 454-459.*

Berners-Lee, T.J., (1993), "Unique Resource Locators", *Internet Draft, IETF URL Working Group, Expires September 30, 1993.*

Cate, V., (1992), "Alex - A Global Filesystem", *Proceedings of the Usenix File Systems Workshop, pp 1-11.*

Danzig, P.B., Ahn, J., Noll, J., Obraczka, K., (1991), "Distributed Indexing : A Scalable mechanism for Distributed Information Retrieval", *Proceedings of the 14th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, October, pp. 220-229.*

Danzig, P.B., Li, S.-H., Obraczka, (1992), "Distributed Indexing of Autonomous Internet Services", *Journal of Computer Systems, Vol. 5, No. 4.*

Davies, H., Hall, W., Heath, I., Hill, G., Wilkins, R., (1992), "Microcosm : An Open Hypermedia Environment for Information Integration", *ECHT '92, Milan, December, pp. 181-190.*

Emtage, A., Deutsch, P., (1992), "Archie - An electronic Directory Service for the Internet", *Proceedings USENIX Winter Conference, January, pp. 93-110.*

Fountain, A., Hall, W., Heath, I., Davies, H., (1990), "MICROCOSM : An Open Model for Hypermedia with Dynamic Linking", *Hypertext : Concepts, Systems and Applications. The Proceedings of The European Conference on Hypertext, INRIA, France, November.*

Hill, G., Wilkins, R., Hall, W., (1992), "Open and Re configurable Hypermedia Systems : A Filter-Based model", *Computer Science technical Report, University of Southampton, UK, CSTR 92-12.*

Howard, J., Kazar, M., Menees, S., Nichols, D., Satyanarayanan, M., Sidebotham, R., West, M., (1987), "Scale and Performance in a Distributed File System", *ACM Transactions on Computer Systems, Vol. 6, No. 1, Jan., pp 51-81.*

Kahle, B., (1989), "Wide Area Information Server Concepts", *Thinking Machines Technical Memo DR89-1, Cambridge, MA : Thinking Machines Corp.*

Li, Z., Hall, W., Davies, H., (1992), "Hypermedia links and information retrieval", *Proceedings of the 14th British Computer Society Research.*

Lynch, C., (1991), "The Z39.50 Information Retrieval Protocol : An Overview and Status Report", *Computer Communication Review, ACM SIGCOMM, Vol. 21, No. 1, pp. 58-70*

Nelson, T.H., (1988), "Managing Immense Storage", *Byte, Vol. 13, No. 1, pp. 225-238*

Neuman, B.C., (1992), "Prospero : A Tool for Organising Internet Resources", *Electronic Networking : Research, Applications, and policy, Vol. 2 No. 1, pp. 30-37.*

Schwartz, M.F., Emtage, A., Kahle, B., Neuman, B.C., (1992), "A Comparison of Internet Resource Discovery Approaches", *Computing Systems, Vol. 5, No. 4*

Stein, R.M., (1991), "Browsing through Terabytes", *Byte, Vol. 16, No. 5, May, pp 157-164*

Yankelovich, N., Haan, B.J., Meyrowitz, N.K., Drucker, S.M., (1988), "Intermedia ; The Concept and Construction of a Seamless Information Environment", *Computer, Vol. 21, No. 1, Jan., pp. 81-96*