To appear in Computer Networks and ISDN Systems, Vol. 27, No. 6, April 1995, Proc. Third International World-Wide Web Conference, WWW'95, Darmstadt, Germany, April 1995.

Serving Information to the Web with Hyper-G

Keith Andrews, Frank Kappe, and Hermann Maurer

Institute for Information Processing and Computer Supported New Media (IICM)
Graz University of Technology
A-8010 Graz, Austria.
{kandrews,fkappe,hmaurer}@iicm.tu-graz.ac.at

Abstract

The provision and maintenance of truly large-scale information resources on the World-Wide Web necessitates server architectures offering substantially more functionality than simply serving HTML files from the local file system and processing CGI requests.

This paper describes Hyper-G, a large-scale, *multi-protocol*, distributed, hypermedia information system which uses an object-oriented database layer to provide information structuring and link maintenance facilities in addition to fully integrated attribute and content search, a hierarchical access control scheme, support for multiple languages, interactive link editing, and point-and-click document insertion.

Keywords: Hyper-G, hypermedia, information system, multi-protocol, automatic link maintenance, server architecture.

1 Introduction

The tremendous popularity of the World-Wide Web (W3) is encouraging ever more individuals and institutions to set up servers of their own. However, the initial euphoria of seeing ones own home page complete with inline mug-shot is often followed by the sobering experience of manually maintaining hyperlinks, massaging directory trees of HTML, and continually reindexing files. Once datasets exceed a few hundred documents, more advanced server architectures are called for which offer substantially more support for the structuring and automatic maintenance of information than simply serving HTML files from the local file system.

This paper describes Hyper-G [AK94, FM94], a large-scale, distributed, hypermedia information system which provides such structuring and maintenance facilities, in addition to fully integrated attribute and content search, a hierarchical access control scheme, support for multiple languages, interactive link editing, and point-and-click document insertion. Hyper-G is multi-protocol enabled – it can serve information to W3, Gopher, and native Hyper-G clients.

2 Hyper-G

Hyper-G combines the intuitiveness of top-down hierarchical navigation with the immediacy of associative hyperlinks and the power of focussed attribute and content searches. The basis for these three tightly-coupled navigational facilities is the Hyper-G data model shown in Figure 1.

Documents may be grouped into aggregate *collections*, which may themselves belong to other collections and which may span multiple Hyper-G servers, providing a unified view of distributed resources. A special kind of collection called a *cluster* is used to form multimedia and/or multilingual aggregates (the appropriate language version of a document to be displayed is selected according to the user's language

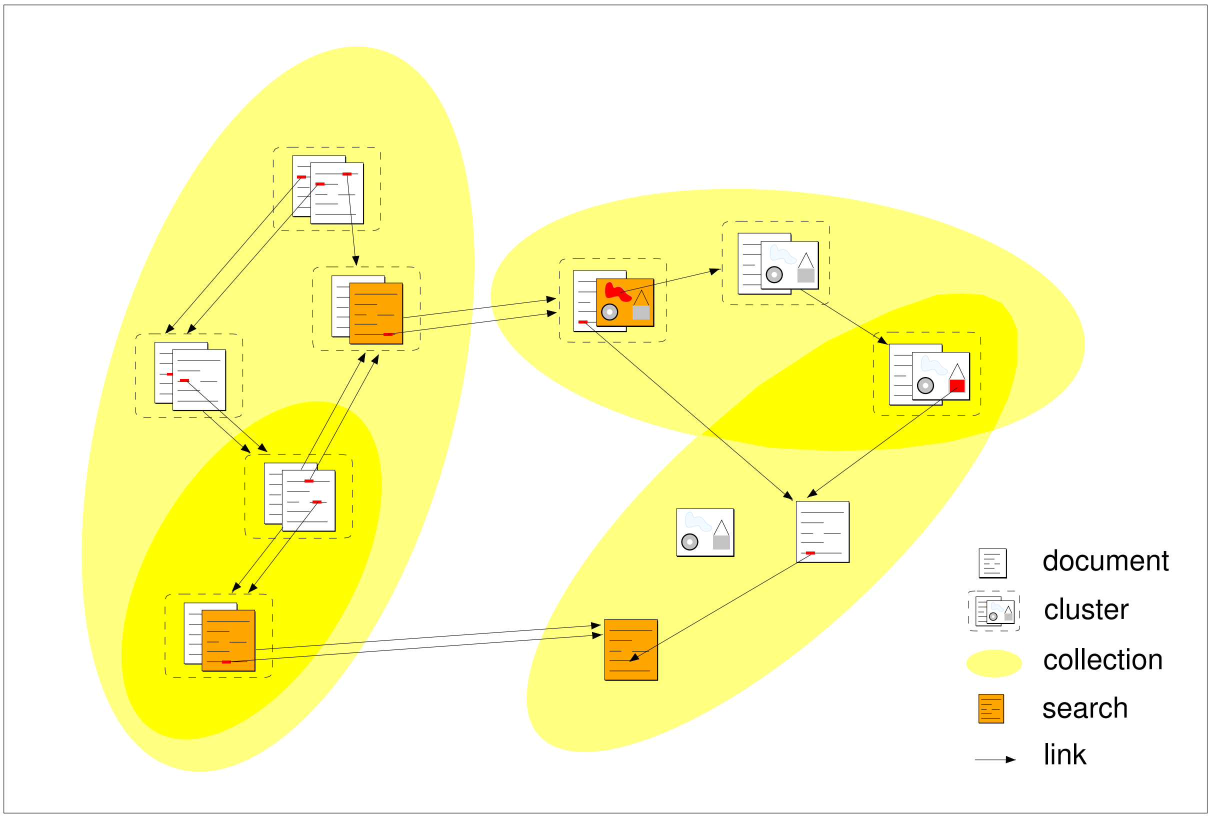

Figure 1: Hyper-G Data Model

preference setting). Documents and collections may belong to multiple parent collections, opening up the possibility of providing multiple views of the same information (for example, incoming mail sorted by date, author, and subject). Collections are typically provided with an introductory text, the *collection head*, which is displayed automatically when a collection is accessed and often contains links to specific parts of the collection.

Hyperlinks in Hyper-G connect a *source anchor* within one document to either a *destination anchor* within another document, an entire document, or a collection. Links are not stored within documents but in a separate link database: hence links are not restricted to text documents, they can be followed backwards, they are updated and deleted automatically when their destination moves or is deleted (no "dangling links"), and they are easy to visualise graphically.

Scalable performance is achieved using a document naming scheme that allows replication and caching with weak consistency. Integrity of the database across server boundaries is maintained using a scalable flood algorithm [Kap95].

Hyper-G has fully integrated search facilities: every document and collection is automatically indexed upon insertion into the database – no extra indexing steps are required. Both attribute (author, title, keywords, etc.) and full text (content) searches are supported, including boolean combinations and term truncation. Searches may be *focussed*, restricted in scope to particular sets of collections which may span multiple servers.

Both anonymous and identified users are supported, with access rights assignable on a per document or per collection basis to user groups or individual users. Identified users have "home collections" within which to organise personal documents and keep pointers to resources.

As can be seen in Figure 2, Hyper-G is a multi-protocol system. When accessed by a Gopher client, the Hyper-G server maps the collection hierarchy into a Gopher menu tree (hyperlinks cannot be represented in Gopher). A synthetic search item is generated at the foot of each Gopher menu to allow searching the corresponding collection. When accessed by a W3 client, each level of the collection hierarchy is converted to an HTML document containing a menu of links to its members. Hyper-G text documents

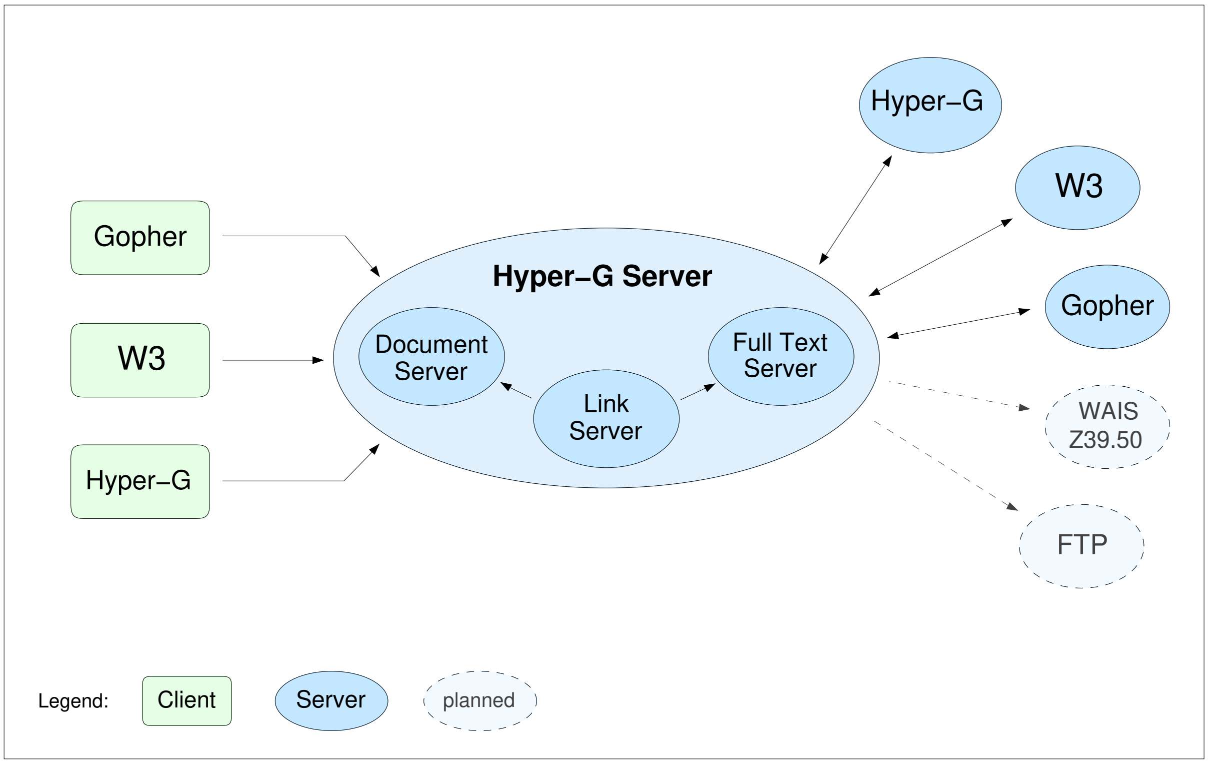

Figure 2: The Architecture of Hyper-G

are transformed on-the-fly into HTML documents, any hyperlinks they might have are merged in at the appropriate places. Additional Hyper-G functionality such as user identification, language preference selection, and searching is implemented via HTML forms which are accessible at any time.

The Hyper-G server is able to store pointers to remote objects on Gopher and W3 servers. This allows the incorporation of information on remote non-Hyper-G servers (almost) seamlessly. Interoperability with WAIS (Z39.50) and FTP servers is planned. The Hyper-G server is currently available for most common Unix platforms.

3 Information Provision with Hyper-G

Server administrators have a wide range of tools available to insert, maintain, and manipulate the information content of a Hyper-G server. The `hgadmin` administration program allows the interactive creation of user accounts and groups. Utility programs, converters, and perl scripts are available to insert pre-existing data in a variety of formats. The `hifexport` and `hifimport` (Hyper-G Interchange Format [IIC]) utilities allow entire collections to be extracted from and exchanged between Hyper-G servers and other applications.

Hyper-G has its own SGML text format, Hyper-G Text Format (HTF) [Kap93], which is very similar to HTML – so much so, in fact, that we are considering adopting HTML 3.0 once it is standardised. External and inline images are usually JPEG, GIF, or TIFF although any common format may be used (the server does not care about the physical format of a document, clients simply have to be able to understand that format).

New information is typically prepared interactively using the editing capabilities of one of the native Hyper-G clients. Both the Harmony client for X Windows and the Amadeus client for MS-Windows support point-and-click insertion of documents and interactive link editing. Figure 3 shows a new document being inserted and Figure 4 a link being created with Harmony.

Information is usually structured first into (overlapping) collections and subcollections; both collections and documents may be *re-used* by virtue of their membership of multiple collections. Hyperlinks

Figure 3: Document Insertion

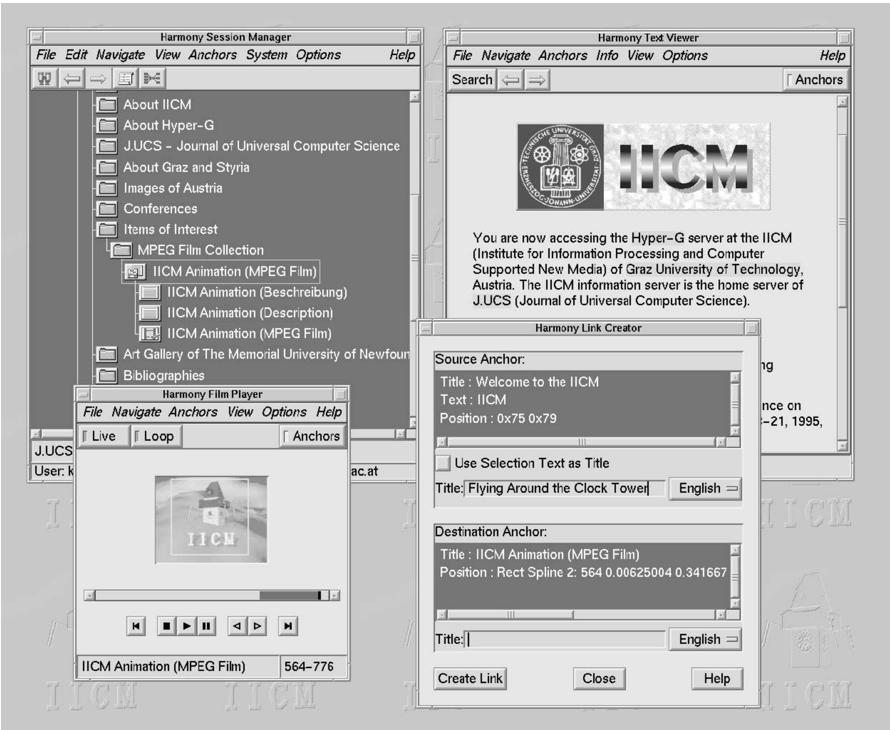

Figure 4: Interactive Link Creation

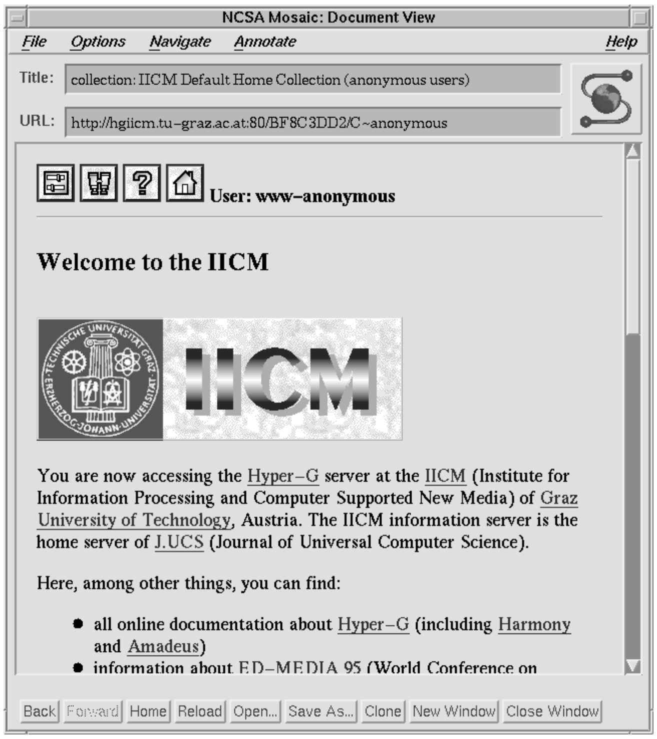

Figure 5: The IICM Welcome Page

are used for cross-references, orthogonal to the collection structure.

4 Serving to the Web with Hyper-G

The Hyper-G server responds to HTTP requests with a HTML representation of the requested information, preserving as much as possible of the rich functionality of Hyper-G in the process. Figure 5 shows the welcome page of the IICM Information Server through the eyes of Mosaic.

Pages are divided into four logical components separated by horizontal rules:

- The *Menu Bar* at the top of the page has four buttons, Options, Search, Help, and Home, respectively.
- The *Text Area* displays a text document (and also the head of a collection, if present).
- The *Collection Area* is present only for collections and displays links to the members (documents, clusters, and subcollections) of the collection.
- The *Attributes Area* at the bottom of the page provides meta-information about the document or collection, such as author, creation time, access rights, etc. (which attributes are displayed is configurable).

Figure 6 shows a collection about the Graz School of Music And Drama containing four subcollections and two clusters belonging to user "mmis" (the collection has no collection head, hence the text area is absent).

Clicking on the Options button brings up the Hyper-G Options panel (a HTML form) shown in Figure 7. This is where users can identify themselves to the system (to gain better access rights), see who else is online, change their language preference, etc.

The Hyper-G Search panel shown in Figure 8 is the interface to Hyper-G's powerful attribute and content search facilities. Search may be restricted in scope to the current collection (and its subcollections), or may span the whole local server.

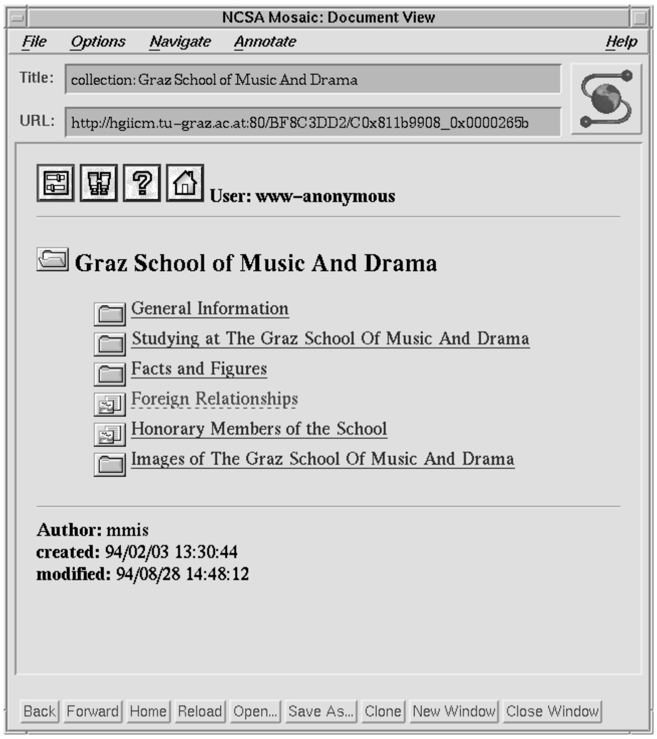

Figure 6: A Simple Collection

Figure 7: The Hyper-G Options Panel

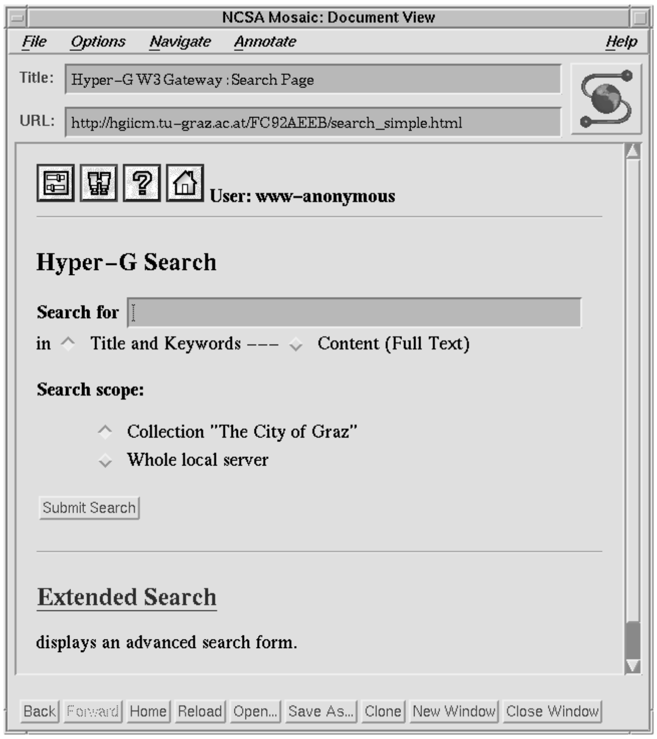

Figure 8: The Hyper-G Search Panel

5 Concluding Remarks

The structuring and ease of maintenance of *large* information bases is one of the major strengths of Hyper-G. The IICM Information Server currently holds about 75,000 documents, including online documentation about Hyper-G, information about Graz and Austria, and up-to-date information on the ED-MEDIA 95 conference in Graz in June 1995. Graz University of Technology's Hyper-G Server contains personnel and course information as well as the complete ACM SIGGRAPH and HCI bibliographies (fully searchable). The European Space Agency (ESA) uses Hyper-G to organise and publish its earth observation data. Figure 9 shows the welcome page of the ESA Earth Observation Guide and Directory Service (GDS), which currently comprises some 45,000 documents. The individual icons used by Hyper-G can be customised to suit a particular application – note the replacement of the standard Hyper-G icons by customised ESA icons. Other users of Hyper-G include numerous universities and companies, the Museum of New Zealand, and the German Mathematics Association, to name just a few.

Hyper-G also is being used as the basis for a major new electronic publishing venture. The Journal of Universal Computer Science (J.UCS), supported by Springer Verlag, is the first high-quality, fully-refereed, fully-citable scientific journal to depend primarily on Internet distribution [MS94]. The pilot issue is already available at several sites, the first regular issue will be available world-wide at the end of January 1995.

Further information about Hyper-G and Harmony and installation details may be retrieved by anonymous ftp from `ftp.iicm.tu-graz.ac.at` in directory `/pub/Hyper-G` or from the IICM Information Server under `http://info.iicm.tu-graz.ac.at/`.

Acknowledgements

Financial support of Hyper-G by the Austrian Ministry of Science, JOANNEUM RESEARCH, and the European Space Agency is gratefully acknowledged.

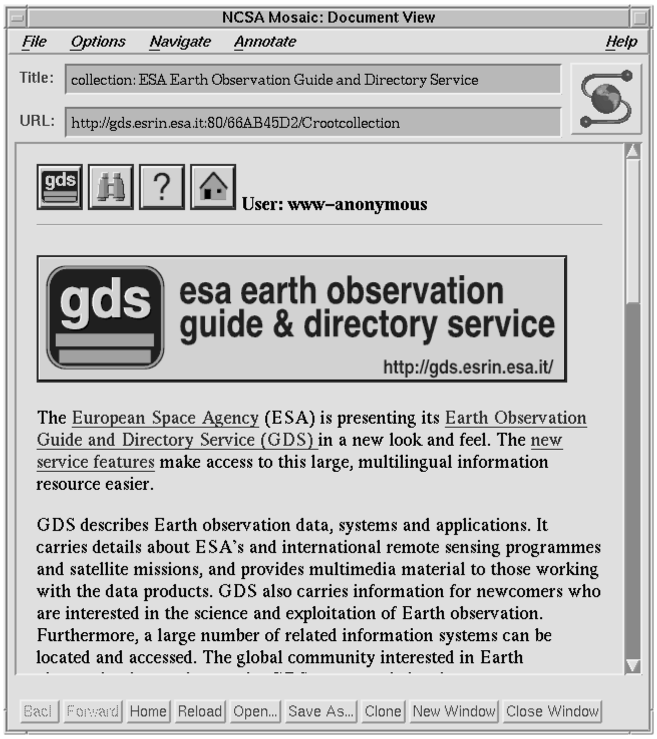

Figure 9: The ESA GDS Welcome Page

References

[AK94] Keith Andrews and Frank Kappe. Soaring through hyperspace: A snapshot of Hyper-G and its Harmony client. In Wolfgang Herzner and Frank Kappe, editors, *Proc. of Eurographics Symposium on Multimedia/Hypermedia in Open Distributed Environments*, pages 181–191, Graz, Austria, June 1994. Springer.

[FM94] Barry Fenn and Hermann Maurer. Harmony on an expanding net. *Interactions*, 1(4):26–38, October 1994.

[IIC] Hyper-G technical documentation. Available at URL `http://info.iicm.tu-graz.ac.at/Ctechnical`, Graz University of Technology, Austria.

[Kap93] Frank Kappe. Hyper-G text format (HTF). Technical report, December 1993. Available by anonymous ftp from `ftp.iicm.tu-graz.ac.at` in file `/pub/Hyper-G/papers/HTF.ps`.

[Kap95] Frank Kappe. A scalable architecture for maintaining referential integrity in distributed information systems. *J. UCS*, 1(2), February 1995. Available at URL `http://info.iicm.tu-graz.ac.at /Ca_scalable_architecture_for_maintaining`.

[MS94] Hermann Maurer and Klaus Schmaranz. J.UCS – the next generation in electronic journal publishing. *Computer Networks for Research in Europe*, 26:S63–S69, 1994. Supplement to Vol. 26 of Computer Networks and ISDN Systems.