Building application dependent hypertexts[1]

Rainer Hammwöhner, Marc Rittberger

Universität Konstanz

Informationswissenschaft

Postfach 5560

D-78434 Konstanz

Abstract

The Konstanz Hypertext System offers a domain-specific developmental environment for the construction of large hypertexts. Through its flexibility, the structuring means employed in the Konstanz Hypertext System offers an instrument which permits one to respond directly to the demands relevant to specific applications in the construction of hypertexts. Especially the integration of information obtained from external resources is emphasized. After a discussion of the information sources which can be connected to the KHS a short introduction to the hypertext model of the KHS is provided. The role of structuring means in the integration of external information is pointed out. The scope of possible applications and the flexibility of the system are demonstrated by the following three comprehensive examples: resource discovery of online databases, management of electronic mail and the compilation of an issue of an electronic journal.

1 Introduction

The time seems to have come for hypertext systems to change from small systems limited to special fields of application to distributed systems containing extensive stocks of information, capable of being used by many different users, and relevant for various different applications and purposes. The interactive and organizational forms of information collection made possible by hypertext extend by far the forms usual in classical information systems. Because of the complexity of the information spaces and the extent of the information collections to be dealt with, the automatic construction of hypertexts is becoming increasingly important. The generation of links and typing, cf. Allan 1995, Nanard & Nanard 1995 and, well, searching in hypertexts, cf. Croft & Turtle 1993, Agosti et al. 1995, to identify starting points play an important role in the construction of hypertexts.

In the following article we will describe the capabilities of the Konstanz Hypertext System (KHS) in connection with specialized methods which support users in the construction of hypertexts adequate for the goals of various applications. A section about external information sources, which can be employed for the automatic construction of large hypertexts in the KHS, addresses the principal aspects and problems of the integration of external information in open hypertext systems. We will give a short outline of the model of

---

[1] This paper has been published in Information Processing & Management. Vol. 33, No. 2, 1997, pp. 243-254.

the KHS. Concluding, three selected examples are presented which show from various viewpoints the construction of hypertext structures using the KHS.

2 Integration of external Information sources in hypermedia

The integration of external information sources in hypermedia systems can be examined from two viewpoints. On the one hand, local information products can be enhanced and backed up through enrichment with external information. On the other hand, the user is offered the possibility of simplified and unified access to external information services. A connection to the discussion of gateways for online databases becomes thereby apparent, cf. Kuhlen 1995, pp. 298-299 and 319-330. The central aspects of such an integration can be summarized as follows:

- The construction of new hypertexts is facilitated if the integration of data from existing files, hypertexts or databases is supported by the hypertext system.

- Through open hypertext systems the integrated employment of application-specific, often disparately structured, databases is made possible. This is the focus of research within the ESPRIT project HIFI, cf. Cavallaro 1995.

- The above-mentioned hypertext systems of the HIFI project have as their object the integration of organization-internal databases. A further logical development of this concept also draws in organization-external databases described in Aßfalg et al. 1993, so to open up access to the commercial information market.

- The rapid growth of value-added services on the Internet has led to the development of several worldwide distributed information services structured on the lines of hypertext principles, first Gopher, then the World Wide Web and finally Hyper-G, whereby the latter has, despite a high level of development, not become as widespread the two first-named systems. As a consequence of this tendency to the globalisation of the information market and the integration of information service offerings, presence in this market and interconnectability with the respective competing systems through gateways in the meantime has already become a prerequisite for open hypertext systems like the KHS.

The above arguments underline the need for accessibility to external information by users of the KHS or other open hypertext systems. It is to be emphasized, however, that through its structuring and navigational possibilities, a system such the KHS enables all users to define an integrated view, one adapted to their personal information needs, of external information sources and thereby to acquire easier access to modern information service offerings.

2.1 Features of information resources

The information sources in question can now be distinguished on the basis of several criteria which are of immediate significance for the type of data integration:

- The frequency of change determines how often data stocks are brought up to date. The higher the updating rate, the more one aims at providing the latest data with each access which results in the hypertext containing reference information only.

- Many data suppliers charge fees for access. The higher these fees are, the more likely it is that one will have to accept that data will be integrated into the hypertext in the sense of a downloading and not be continually brought up to date.

- The exchange of data between hypertexts or application programs will require an adaptation of the data formats. However, more serious is that often the logical structure of a document must also be modified. This is especially apparent if data are

2

to be transferred between two hypertext systems with respectively different models, cf. Leggett & Killough 1991.

- Communication between a hypertext and a data source must draw on a normed procedure which is defined by a communication protocol. The access procedures which are permitted by this protocol (context-free search operations or context-sensitive navigational operations, etc.), are especially important for the design of the integration process.

- Usually data exchange processes are synchronous with user interaction, i.e., external data sources provide information only upon request. However, there are also data sources with asynchronous processes. The arrival of e-mail is, e.g., completely independent of any sort of user action.

- Not all data sources are accessible at all times. This may be due to maintenance periods, etc., but often, however, it depends on the utilization of network capacity.

Further qualitative attributes of data sources, , e.g., scope and coverage of the subject area, comprehensiveness, currency and timeliness, accuracy and consistency of available data, are relevant for the respective use, but without significance for the choice of an integration procedure.

2.2 Aspects of the integration process

Within the integration process several steps can be discerned each asking for its own methodology:

- The first step is choosing an appropriate data resource. This requires knowledge about the contents and quality of every data resource being in consideration. In the following we will give an example on choosing relevant online databases.

- After an information resource has been selected, a data connection must be established confirming to a specific communication protocol.

- If not all elements of the information resource shall be used, a selection criterion must be formulated and transferred.

- As a result, a set of relevant data items will be received which must be converted into units of the local hypertext.

- The newly created units must be connected between themselves and to the preexisting units of the hypertext.

- Of course, the construction of a hypertext from external resources cannot be thought of as a singular event happening at the beginning of a hypertext's lifetime and leaving the hypertext unchanged further on. Therefore update strategies must be provided.

Depending on the features of a data resource the implementation of each of these steps — with the exception of the first one — can differ substantially. A further source of complexity arises from the integration of the data originating from different information sources and the use of this information to support varying applications.

The next section will introduce the hypertext model of the KHS and show how this model helps to deal with the questions mentioned above.

3 The Konstanz Hypertext System

The KHS is an open hypertext-system designed to allow the integration of various application domains, the use of multiple information resources and parallel use by an - in principle -

arbitrary number of users. The unifying framework is supplied by a generic, application independent hypertext-model comprising a structure model which describes the structure of well-formed hypertexts with an interaction model which defines generic interaction styles. Both the structure model and the interaction model can be refined to suit the needs of special applications or individual users. Because the interaction model is of little consequence for the construction of hypertexts[2], only the structure model will be described in some detail.

3.1 Structure elements of the KHS

The simple unit-link structure of early hypertexts proved unable to provide sufficient orientational clues in large and complex hypertexts, cf. Halasz 1988. The KHS therefore employs additional structuring mechanisms that heavily base on the typing of hypertext objects. Typing of hypertext objects, cf. Hammwöhner & Kuhlen 1994, as employed by the KHS allows the stepwise refinement of structure and behaviour of hypertext objects. Media units are devoted to the representation of the information content of a hypertext. Typing of units means imposing restrictions on their internal structure, presentation and interactive behaviour. Forms as special semi-structured type of media units have some internal structure allowing access to named content fields. Any hypertext unit can be described by index terms such that index based search facilities can be provided.

Composite units are a special unit type containing no media information but units which may be of composite type themselves. The KHS hypertext model regards composite units as the backbone of the hypertext structure and as a means for structured navigation. Any unit within a hypertext - with the exception of a single, so-called "top level unit" - must be part of at least one composite unit. In practice it will be part of several units and thus poly-hierarchically arranged. Typed composite units make use of constraints defining restrictions on type and content of units which can be inserted. Thus, a structure built by composite units serves three purposes:

- Hierarchies — in contrast to webs — can be traversed in an unambiguous manner and thus provide a ordering principle for hypertext units.

- Features of units can be inherited within the hierarchy. Inheritance of access permissions provides an elaborated access control system necessary for applications in the field of cooperative work.

- As composite nodes define restrictions on the type of their subunits which themselves provide type specific functions, a decomposition of the hypertext into functional coherent parts belonging to distinct applications is achieved.

Links represent the relations between hypertext objects. The large number of possible connections which exist between two (or more) objects (hypertext units and links) necessitates an explicit differentiation of link types based on their function in the discourse. Depending on their type links may connect whole units, pieces of text (hot-words) within units, or sections within images (hot-area). The declarative semantics of a link-type is given by constraints on the type and content of the connected units and anchors. The procedural semantics describing the behavior of a link object during navigation is defined by type specific methods.

3.2 Basic principles of hypertext construction

Essentially hypertext construction is guided by structure itself. Any newly created hypertext by the KHS has some initial structure which is derived from type specific templates which

---

[2] A description of the interaction model is provided in Rittberger et al. 1994.

can be defined using a formal specification language. Expressions of this language can be assigned to composite unit types as well as to define an initial internal structure for any instance of this type. Assigning templates to hypertexts in the whole and composite units in detail can be used to achieve the pre-structuring of hypermedia at an appropriate scale. A hypertext devoted to cooperative work for instance will have a macrostructure providing composite units which serve as archives for mail or appointments and deadlines. Further areas of the hypertext are dedicated to the shared access to documents or pieces of software and bibliographic references.

The internal structure of a composite unit dedicated to an application field is controlled by the following means:

- Type specific constraints define the type of subunits which may be inserted. The subunit types themselves define further type constraints such that the whole structure is controlled.

- Another kind of constraints controls type and anchor content of application-specific links. Both kinds of constraints in combination can be thought of as a hypertext grammar.

- Where constraints impose restrictions on the internal structure of composite units, templates and type specific construction methods define explicit application-specific hypertext structures to be created at the time of unit instantiation.

These structuring facilities are helpful for the user wanting to create a hypertext intellectually. Template structures will provide outlines for the most common document types. The system will suggest destinations for possible linkages originating from the unit currently being in work. Even if sufficient conditions for the existence of links of a specific type cannot be provided, necessary conditions basing on type and content constraints can be useful for reducing the number of candidates as possible link destination. This helpful feature becomes essential in the case of automatic hypertext construction using data from external resources.

KHS represents external information resources as units of a specific type. Procedures for establishing a data connection, for data request and transfer are implemented as type specific methods. Additionally a parser appropriate for the processing of the arriving data stream is provided. The following process of data integration can be tailored to the requirements of application and data resource. Data resources emitting huge bulks of information which have to be split into several units are represented as composite units which are fit to integrate the received data as subunits. Data resources emitting single data elements at a time can be represented as media units containing the data element retrieved last. These differing approaches shall be explained by two sketchy examples.

The World Wide Web is an information system which gives access to one data element at a time. Search within the web usually is achieved by navigation. The standard representation of the web as an information source within the KHS is a non-composite media unit type. Instances of web units show the contents of one web page as addressed by a resource locator. The HTML-code of the web content is being parsed and converted to the KHS text format. Hot-words can be used for web navigation. In this case the content of the gateway-unit will be exchanged. This type of data integration obviously is appropriate for data accessible by navigation if no further processing of these data is necessary. So at any time the most up to date information is at hand and no local storage space is needed. This gateway-approach of course employs no means of automatic hypertext construction and therefore will not be described in more detail within this paper.

Any program can be thought of as a hypertext consisting of units containing source code and links defined by routine call references. Object oriented programs with their class inheritance

structure are even more hypertext-like. A KHS application dedicated to software-engineering is capable of considering Smalltalk-Images as data resources. The entry point of this application is a composite unit establishing the connection to the image. On request a preliminarily specified set of classes with their definitions and methods is loaded and converted into a hypertext-structure which is inserted into the application's root unit. Individual unit types are assigned to class and method definitions. Inheritance relations and call chains are represented as links. A hierarchy of composite units resembling the class and method categories of Smalltalk provides orientation for the programmer. If this conversion process takes place for the first time, all organizational structures have to be built from the scratch. Later on additional classes or methods can be loaded and inserted into the existing structure which can then be used as a software development-environment.

The example of software hypertexts is based on a information resource containing information which already is heavily structured. This is not the case e.g. with online-databases containing large spaces of unconnected information. Integration of information from these resources requires not only the interpretation of structure but also the heuristic definition of structure. Most natural to the user are access structures based on formal features of the information items just like author names of bibliographic references or arrival dates of email. To allow an efficient access to these data they are arranged into a balanced tree. There is a tradeoff between the depth of the tree influencing the average number of required navigation steps and the breadth of the tree influencing the number of items being displayed at a time. To maintain the balanced structure of a tree during the process of data integration is very costly. Thus a reorganization leading to a balanced structure is performed only if the deformation of the tree exceeds a certain threshold. An alternative to arranging the data into a balanced search structure is the insertion of data into a user-defined (poly-)hierarchical classification structure - similar to the folder structure of mail classification systems. This structure is not optimised with respect to efficient access but groups information objects which are thematically coherent or are used in the same application-context. Special aspects of theses approaches will be elaborated based on three comprehensive examples.

4 Applications in the KHS

With the aids discussed in the above sections a range of applications have been created, a selection of which we would like to present below. We will thereby present the organization of online database descriptions, work with e-mail and the construction of electronic journals using the KHS. The applications have their main focus in archiving, but also aim at the selection of online databases, reception and sending of e-mail, as well as electronic publishing. We will not thereby deal in depth with the topic of interaction with the KHS, but will instead describe the structure and organization of KHS hypertexts, which are supposed to support the aims. More details concerning the aspects of the applications not described here are given in Aßfalg et al. 1993, Rittberger 1994 and Rittberger et al. 1994. The three examples also describe alternative structuring procedures which are possible in the KHS in relation to different applications. In the choice of online databases the structure is created using content-based clusters which rest on statistical procedures. E-mail management is performed with the help of the knowledge base, and in the construction of an electronic journal mixed procedures are employed. Besides these substantively-based procedures, in all three cases formal criteria are additionally employed in order to build the hypertext structure, consisting of a poly-hierarchy and links.

4.1 Resource discovery

Besides the above-mentioned access to online databases, the KHS can also support the user in resource discovery with the selection of the online databases which may be relevant for a

search. The KHS supports users in selecting an online database with a hypertext which contains database descriptions of online databases of several international hosts. Users can start the selection process during the preparation of an online search. The selection of online databases depends on many relevant factors such as contents, sources, geographic and time coverage, update frequency and costs. All of these factors are accessible to users for each online database in the KHS, but we have placed the main emphasis on content-based support of database selection. To obtain the most recent database content, we used the database descriptions of three hosts, which are available in machine-readable form. KHS performs automatic searches in the meta-databases of the hosts DATASTAR, DIALOG and STN International and requests all available database descriptions of the three hosts. Every database description is converted into a so-called 'database description unit' containing a form with the available information in separate fields. A hypertext with all the database description units is built.

In integrating the data of the database descriptions from the three hosts, the units which arise from the individual database descriptions are bound into the structural hierarchy of the hypertext and are linked together. To arrange the new units which are created without information about a concrete context of the user or an external process (e.g., an online search), the KHS uses rule-based processes which are based on the described ordering principles. The online database descriptions are therefore inserted in the poly-hierarchy of the hypertext on the basis of ordering- and function-specific criteria, but not on the basis of authorization criteria:

Ordering: The database description units are filed in a hierarchical structure in a hypertext in alphabetical order according to their names. The aim of the alphabetical order is to construct a tree structure which contains all online database descriptions in a maximally compact and ordered form. A balanced tree is constructed in which only the leaves of the lowest hierarchical level contain the textual information and the inner units serve the construction of the structure. The KHS calculates the depth of the tree from the number of database description units which are represented in the lowest hierarchical level and the maximum number of units which should be included in one composite unit. The KHS constructs the hierarchical structure for the database description units in a tree with the optimal distribution so that the degree and height of the tree is as small as possible for the pregiven number of units.

Functional: Topically related database description units are inserted in semantically defined clusters which permit target-oriented access to the hypertext units. Based on the vector space model using inverse document frequency and the cosine measure, cf. Salton & McGill 1983, the respective substantively similar online database descriptions are interconnected with weighted links. To limit the number of links and to get a fully interconnected web of database description units a threshold for the link weight and a minimum number of at least ten links starting from every unit are introduced. Links exceeding the threshold are built as bi-directionally links whereas the other links are built as mono-directional links. Therefore we get a interconnected hierarchy of 755 online database description units with about 11 links per unit.

The described similarity specifications can be used to construct clusters through various processes familiar in information retrieval, cf. Rasmussen 1992. In hypertexts such cluster processes excellently serve either to locate further relevant units for a known unit or to identify a hypertext hierarchy relevant to a search. In order to do justice to the requirements for the selection of databases, none of the usual clustering processes were employed, since the sets should not be disjoint and should have a comprehensible size and should, for reasons of efficiency, be constructed based on the existing links.
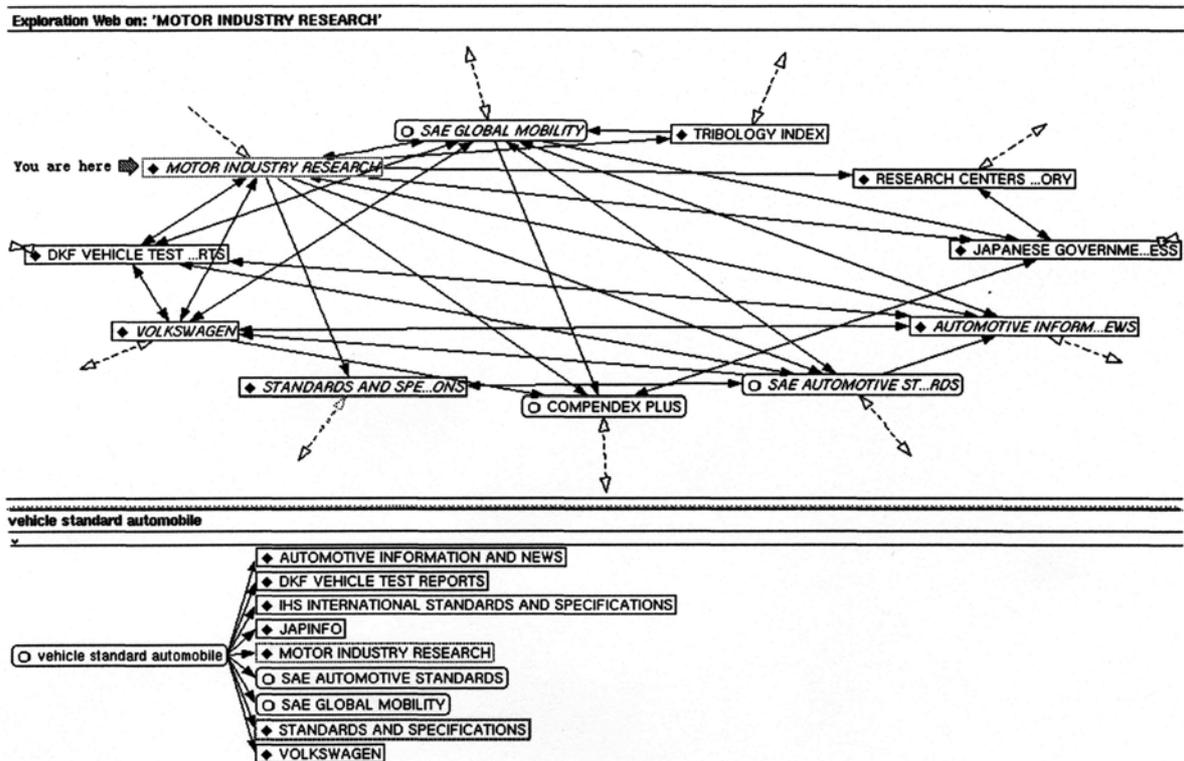
Figure 1: Cluster on database description units on information about vehicles, automobiles and standards and links starting at one of the units of the cluster.

The links of each individual database description unit are compared with the links of the database description units similar to the given unit. Those units are removed from these sets which do not appear cyclically in viewing the links between the units. According to the taxonomy of Allan 1995 the 'tangent links' are thereby removed. The clusters which are created are assigned as names the most important keywords from the indexing of the database description units.

Figure l shows the links and a part of the hierarchy of a specific database description unit viewed by KHS-Windows. The media units in the composite unit 'vehicle standard automobile' (lower part) describe online databases treating automobiles or standards and were grouped together as a cluster. Besides the hierarchy shown in figure l, the units also can be found in a hierarchy which is alphabetically organized by name of the units (e.g. 'Motor Industry Research' will be found under 'MO'). The upper part of the window shows the links starting from the unit 'Motor Industry Research' and the interlinking of the target units. As expected, most of the units which are linked to 'Motor Industry Research' can be found in the cluster viewed in the upper part of figure 1.

Poly-hierarchical structures are created in a hypertext which permit access to relevant database description units not only according to alphabetical, but also according to content criteria. Further, with the links, besides the poly-hierarchical structure, a content-based network-like structure is available for navigation.

4.2 Management of e-mail

KHS runs a process which looks for incoming mail using SMTP. Every mail is converted into a hypertext unit containing the structured information of a mail in a form with several

fields like sender, receiver, subject, etc.. Every incoming node has to be integrated into a poly-hierarchy of folders created by an author, e.g. the user of the KHS mail application. Management of arriving e-mail through classifying based of similarity of descriptor sets, cf. Hammwöhner & Kuhlen 1994, which was first used in the KHS, has been supplemented by a rule-based process. Rule-based constructions of archive structures were developed in connection with mail systems - the best known example may be the 'Information Lens', cf. Malone et al. 1987, however, also the 'MAFIA' system, cf. Lutz et al. 1990 or ISCREEN, cf. Pollock 1988. Classification processes are supplied by so-called agents, which are specific for the type of information (mail, schedule, etc.). A set of rules is assigned to each agent, which respectively consist in their conditional part of an example template, which is compared with the object to be classified, and an action, which is carried out in case of agreement. The action usually results in placing the data element in a specific folder. Thus when new mail arrives, the agent is activated to classify the mail, compares the newly arrived object with all the example forms of its rules and carries out the action parts of the applicable rules.
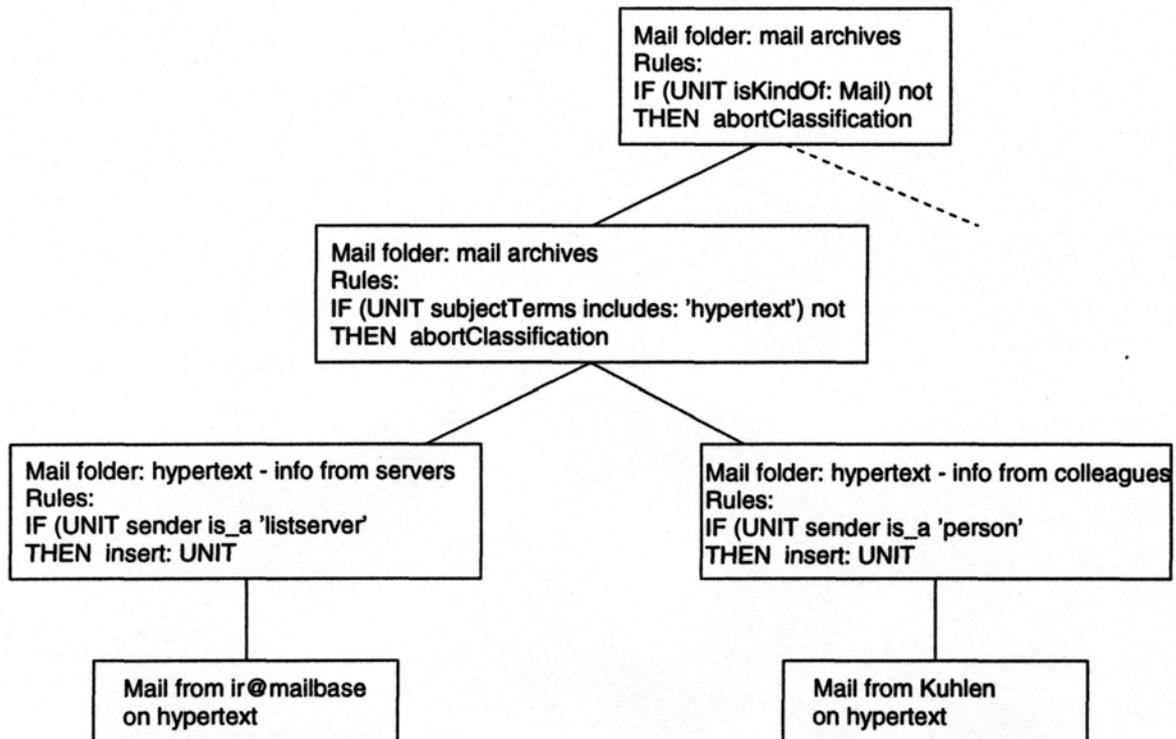
The KHS uses a more complicated approach, which rewards increased calculation effort with an improved structuring possibility. No rule set is assigned to an agent here, but only a specific type of link which leads to a set of agent specific rules and a start unit, which as a composite unit forms the root of the hierarchical structure to be viewed.

For this approach a knowledge base is used which is bound into the KHS through the use of typed links and units, so that all formally represented knowledge objects are present in the KHS as hypertext objects. This knowledge is represented using frames, rules, a limited type of constraint, tasks and access paths. Besides static knowledge, which is represented above all with frames, tasks which are defined by rules can be employed for modelling in the KHS, cf. Zink 1995.

Every mail unit has to be classified by running through the following steps:

- Starting from a composite unit, with the first run-through, this is a pre-given start unit, all rule objects which are reachable through the pre-given link type are collected into a rule set.

- Those rules are filtered out whose conditional part is fulfilled when the respective mail unit to be classified is applied as a parameter. The conditions can thereby relate to the attributes of the unit involved, but can also relate to the terminological or organization-specific knowledge, which is represented in the hypertext through frames, or the user profile of the user, who carries out the classification.

- The discovered rules are put in order using preference relationships which are represented by links connecting the rules. The action parts of the rules are carried out in the thereby given sequence. The following actions are currently possible:

  o Insert the unit in the relevant composite units.

  o Remove the unit from the composite unit. This rule is relevant for reclassification of existing data stacks. In this case a mail unit, which is subject to reclassification and is already part of some folders, can be inserted to additional folders by insertion rules and removed from others by deletion rules. Furthermore deletion rules can be used to remove obsolete data.

  o Execute the so-described classification processes with all subunits of the composite unit being currently in consideration using the rule sets assigned to these units.

- o Terminate the classification for this sub-hierarchy. An interruption of the classification processes is also given, if no conditional part of any rule can be fulfilled.
- o Starting from the current composite unit use another method of classification. This type of action permits to employ specialized classification processes like term or frame matching for sub-branches of the hypertext, without having to leave the rule-based overall frame.

```
┌──────────────────────────────────┐
│ Mail folder: mail archives       │
│ Rules:                           │
│ IF (UNIT isKindOf: Mail) not     │
│ THEN  abortClassification        │
└──────────────────────────────────┘

┌──────────────────────────────────────────┐
│ Mail folder: mail archives               │
│ Rules:                                   │
│ IF (UNIT subjectTerms includes: 'hypertext') not │
│ THEN  abortClassification                │
└──────────────────────────────────────────┘

┌──────────────────────────────────────────┐   ┌──────────────────────────────────────────┐
│ Mail folder: hypertext - info from servers│   │ Mail folder: hypertext - info from colleagues│
│ Rules:                                   │   │ Rules:                                   │
│ IF (UNIT sender is_a 'listserver'        │   │ IF (UNIT sender is_a 'person'            │
│ THEN  insert: UNIT                       │   │ THEN  insert: UNIT                       │
└──────────────────────────────────────────┘   └──────────────────────────────────────────┘

┌──────────────────────────┐                    ┌──────────────────────────┐
│ Mail from ir@mailbase    │                    │ Mail from Kuhlen          │
│ on hypertext             │                    │ on hypertext              │
└──────────────────────────┘                    └──────────────────────────┘
```

Figure 2: Rule-based classification - the aim is to distinguish mail on the topic of hypertext in terms of whether it stems from a list server or a person.

Figure 2 shows a small example of how an incoming mail unit is used in the rule-based classification processes. The starting composite unit is the top mail folder where an incoming unit is tested whether it is of type mail unit. The next folder in the hierarchy looks whether the subject field of the mail unit includes the term 'hypertext' and if it appears the mail unit's sender will be checked whether it is from a specific list-server or person or from a group of persons (e.g. colleagues). If a condition of a lowest-level mail folder is fulfilled, the mail unit can be inserted. In our case the mail unit is inserted in one of the two lowest-level mail folders in figure 2 or in both or, if none of the conditions serves the purpose, the KHS puts the mail unit in a mail folder for unclassified mail.

The described process causes the user of the KHS mail application to build a poly-hierarchy of composite units enhanced with condition rules due to his personal mails. The KHS itself classifies every mail unit and puts it into the relevant mail folder, therefore filling the structural, user-built framework with mail units which are a special type of media units.

The construction of the poly-hierarchy of the received and sent mail is accompanied by the automatic linking of mail units. Thus the actions familiar in the handling of e-mail tools, such as reply, include, etc., are realized in the KHS through appropriate links. Further, each mail is connected with the address unit corresponding to the sender or receiver, so that in this manner the correspondence with individual persons or groups can be determined.

For the next example, besides the KHS mail-tool, KHS schedule management and management of literature citations with the KHS, which are not discussed here but employ the same structuring and classification mechanisms as the two described examples, are drawn upon.

4.3 Management of an electronic journal

The dynamic development of new electronic communication media has created a new publication product, the electronic journal. The process of creating an electronic journal (in a slightly idealized form) seems highly appropriate to illustrate the possibilities of hypertext, since the planning and publication instruments there are identical. The refereeing process claims a high share of the administrative expenditure of a journal. An efficient support of precisely this process is essential, because there is considerable potential to increase currency and topicality. Since well-known authors must still be convinced of the standards of the new medium, however, the impression must not arise that the thoroughness of the evaluation, and thereby the level of the journal, would suffer through this increase in efficiency.

In the following section the process is outlined through which an electronic journal is created, and how it is supported by the KHS:

- The process begins with the receipt of an article by the journal editorial staff sent via e-mail to KHS in a pre-given format (e.g. HTML). The article, added as an attachment to the electronic mail, is automatically placed in an input file, and the author receives a confirmation of its delivery from KHS.

- The document is automatically converted into a hypertext-conforming format and segmented into hypertext units, which represent the different parts of the documents, e.g. authors' address, introduction, sections, conclusion, figures, tables, etc., by a special type of template. The result is checked and supplemented by an editor. The document is inserted in its own hypertext folder, which runs through all further processing steps with the document. Here all versions of the document, all comments and annotations, etc. are assembled. Literature citations contained in the document are compared with the central bibliography assigned to the journal, and literature citations are represented by links (see figure 3 step: conversion and segmentation).

- At the start of the evaluation phase the document is placed in a hypertext area which can be viewed by all editors and referees. Each member of the committee can comment on all submitted articles which contributes to objectify the evaluation (see figure 3 step: open referees).

- The KHS assigns the document to a number of referees on the basis of topical features and possible further criteria, which are handled by a rule-based process similar to the one mentioned for the classification of electronic mail. Each of the referees, who cooperate on an ongoing basis with the journal, has constant access to his own working area in the hypertext, but also external referees are temporarily assigned a work area. In these areas the article is classified and a schedule date is assigned to the document (see figure 3 step: referee process with named referees).

- The referees can view the document with KHS tools or also through a WWW browser (e.g., Netscape). They attach their comments to the document as annotations or debate via electronic mail on the subject of the documents. These annotations and mails are linked to the document by the KHS and are also accessible to the other referees and can be annotated themselves.

- After the completion of the refereeing process the document is again transferred to a work area which the editor is responsible for. He briefly checks what comments

should be made accessible to the author and sets up appropriate access rights so that the author can view the document and comments (preferably via WWW).
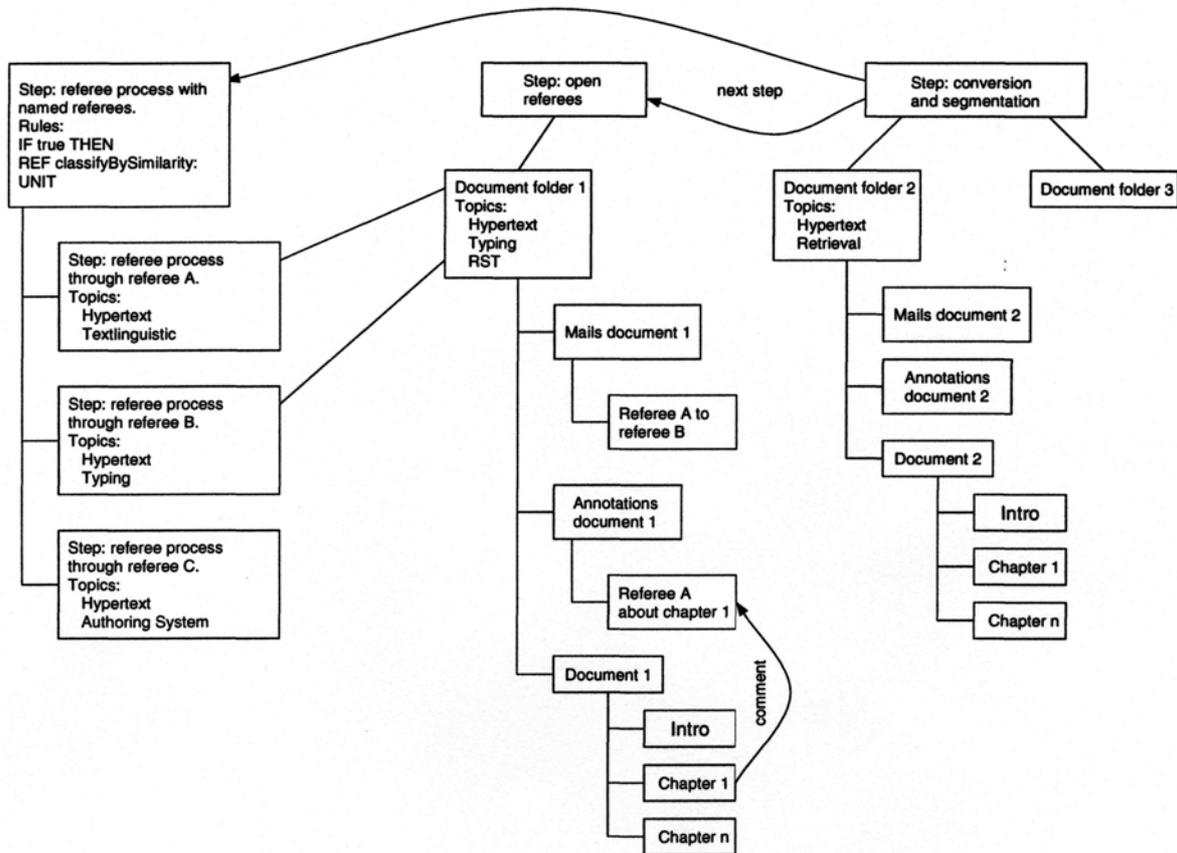
Figure 3: Steps of a referee process for an electronic journal.

- If the article is finally rejected, the document is inserted in an internal archive, otherwise the author now has an opportunity to revise the document and submit an improved version. This version may possibly go through all the above-described steps again handled by the KHS.

- If the article is accepted for publication, the final processing takes place:

    o The textual format is finally adapted to the format of the journal.

    o The KHS automatically creates links to other articles and authors in the journal on the basis of statistical or formal similarities or co-references, etc.

- The document is transferred for publication to an area of the hypertext which is accessible to the public or subscribers.

By publication in the WWW readers have access to all the navigational and search operations of the KHS, index- and keyword searches, navigation in documents, browsing on the basis of document similarities or through use of references and co-references, etc. Above all the electronic journal, in contrast to its printed predecessors, is not a medium which condemns its readers to purely passive response. It is rather to be assumed that readers will also be able to comment on and annotate the published articles. Such commentaries can thereby take on the extent of co-reports. In order to stimulate discussion, in some cases it can also be reasonable to make the referees' annotations accessible to the public, in part as the crystallization point

for further discussion. An electronic journal thus develops into an intellectual discussion forum.

In time, then, a large journal arises whose storage requires complex structures in order to adequately save the individual contributions and the hypertext units assigned to them. An extensive poly-hierarchical structure is created which is supported both automatically and also manually and which can be navigationally accessed through a large number of additional links. The 'Rundbrief Information Retrieval' of the GI can be regarded as a first attempt towards the construction and storage of an electronic journal with the KHS.

5 Conclusion

The automatic construction of hypertexts will achieve additional importance over the course of the next few years through the steadily growing flood of publications, particularly along the WWW. It is thereby important that hypertext is seen not only as an interface for traditional database applications, but rather the modelling- and interaction opportunities of hypertext are being taken advantage of and used in the production of hypertexts. Through its highly flexible architecture, the KHS offers a suitable application field in order to meet various different problems with an adequate, thus aim-oriented, problem-solving approach. The examples presented here are a first attempt at the automatic production and construction of large, multi-disciplinary and multi-functional hypertexts like those which will be created in the future. For selection of online databases a hypertext containing database description units is built automatically and the units are linked together with the help of statistical processes based on the vector space model. Electronic mail uses a hypertext where the poly-hierarchy of the composite units and the assigned classification rules are built by the individual user and the KHS fills the structure with media units and automatically links the units in the hypertext. Finally we reported on running an electronic journal, where different aspects of the application were supported by KHS using the named utilities to build a poly-hierarchical hypertext and handle the communication between editors, referees, authors and readers.

References

M. Agosti, M. Melucci and F. Crestani, 1995. Automatic authoring and construction of hypermedia for information retrieval. *Multimedia Systems,* Vol. 3, No. 1 , pp. 15-24.

J. Allan, 1995. *Automatic hypertext construction.* PhD thesis, Cornell University.

R. Aßfalg, R. Hammwöhner and M. Rittberger, 1993. The hypertext internet connection: E-mail, online search, gopher. In D.I. Raitt and B. Jeapes (eds.), *Online Information 93. 17th International Online Information Meeting, London,* pp. 453-464. Learned Information: London.

U. Cavallaro, 1995. HIFI - Hypertext interface to external databases. In W. Schuler, J. Hannemann and N. Streitz (eds.), *Designing user interfaces for hypermedia,* pp. 219-224. Springer: Berlin.

W.B. Croft and H.R. Turtle, 1993. Retrieval strategies for hypertext. *Information Processing & Management,* Vol. 29, No. 3 , pp. 313-324.

F.G. Halasz, 1988. Reflections on NoteCards: Seven issues for the next generation of hypermedia systems. *Communications of the ACM,* Vol. 31, No. 7 , pp. 836-852.

R. Hammwöhner and R. Kuhlen, 1994. Semantic control of open hypertext systems by typed objects. *Journal of Information Science,* Vol. 20, No. 3 , pp. 175-184.

R. Kuhlen, 1995. *Informationsmarkt. Chancen und Risiken der Kommerzialisierung von Wissen.* No. 15 in Schriften zur Informationswissenschaft. Universitätsverlag Konstanz: Konstanz.

J.J. Leggett and R. Killough, 1991. Issues in hypertext interchange. *Hypermedia,* Vol. 3, No. 3 , pp. 159-188.

E. Lutz, H. von Kleist-Retzow and K. Hoernig, 1990. MAFIA - an active mail-filter-agent for an intelligent document processing support. *ACM SIGOIS Bulletin,* Vol. 11, No. 4 , pp. 16-32.

T. W. Malone, J. Yates and R. I. Benjamin, 1987. Electronic markets and electronic hierarchies. *Communications of the ACM,* Vol. 30, No. 6 , pp. 484-497.

J. Nanard and M. Nanard, 1995. Hypertext design environments and the hypertext design process. *Communications of the ACM,* Vol. 38, No. 8 , pp. 49-56.

S. Pollock, 1988. A rule-based message filtering system. *ACM Transaction on Office Information Systems,* Vol. 6, No. 3 , pp. 232-254.

E. Rasmussen, 1992. Clustering algorithms. In W.B. Frakes and R. Baeza-Yates (eds.), *Data structures & algorithms,* Chapter 16, pp. 419-442. Prentice Hall.

M. Rittberger, R. Hammwöhner, R. Aßfalg and R. Kuhlen, 1994. A homogenous interaction platform for navigation and search in and from open hypertext systems. In *RIAO 94 Conference Proceedings. Intelligent multimedia information retrieval systems and management,* pp. 649-663, New York. Rockefeller University.

M. Rittberger, 1994. Support of online database selection in KHS. In M.E. Williams (ed.), *National Online Meeting'94, New York 10 -12 May,* pp. 379-387.

G. Salton and M.J. McGill, 1983. *Introduction to modern information retrieval.* McGraw-Hill: New York.

V. Zink, 1995. Die Wissensbasis im Konstanzer Hypertextsystem (KHS). Report 77-95, Informationswissenschaft, Universität Konstanz.