
# RIESCE: AN HYPERTEXTUAL TOOL FOR BROWSING INFORMATION PRODUCED IN THE BUILDING SECTOR BY PFED

*Nicolo Asaro*, Rossella Corrao*, Daniela Faconti*, Ornella Fiandaca*, Patrizia Grifoni** and Alberto Silvani***

*Department of Progetto e Costruzione Edilizia, University of Palermo - Palermo, Italy
**Istituto di Studi sulla Ricerca e sulla Documentazione Scientifica, Consiglio Nazionale delle Ricerche - Rome, Italy

***Abstract***

*The structure of a set of documents allows for navigation inside single texts. Yet it is often also an obstacle to comparability between different parts and concepts. Even document formatting is only a partial answer to the problem, since it fails to develop the matter of relationships between enunciation and results. It was need to retrieve information about the scientific findings made by Research Units (Operating Units) during the first three year period of the "Progetto Finalizzato Edilizia" (PFEd) for assessment and transfer purposes which led us to design and develop a system to facilitate the retrieval of the relevant information. We chose the building sector for this application because of its relative lack of previous experience of this type, the variety and complexity of documentation available and, last but not least, the general underestimation of research topics and results vis-a-vis the development of the sector. By making the suggestion and information inherent in its findings more available, in terms of method and ambit, as well as more explicit, the research has already achieved a significant result. In view of the prototype character of the experiment, the information will probably be adapted to produce an hypertext on the final results.*

## 1. Introduction

The socio-economic and cultural transformation of which we are at once spectators and actors is partly reflected in the renewed importance which is being attached to information in relation to economic, political and cultural life, and also in the way in which this is evolving and developing. In this context, data processing in the science most likely to foster the exchange and dissemination information among actors, be they social, cultural, economic and/or political structures of different types and importance or individuals.
The increasingly pressing need to exchange and transfer scientific and research findings has multiplied the use of information systems organising the data and results obtained in this way, facilitating navigation, access and information retrieval through user-friendly interfaces and multimedia system. All this has acquired new importance in view of the development of geographical networks. In fact electronic and hypermedia documents are "de-located" : they be consulted at a distance, transferred or "multilocated", "assembled" by activating connection between documents of different origins and locations.
Hence the widely felt belief that is necessary to build a system to support the consultation of information on research and its results in the context of the CNR "Progetto Finalizzato Edilizia" .
The system has been developed to further disseminate results in the research context, in a more strictly professional context, and finally, for teaching purposes among university students. All of which clearly points to a certain heterogeneusness among the users to which the system is addressed. This persuaded us to develop a PC-based system as opposed to an on-line system first. In fact, the development of networks is something of a recent phenomenon, and not all user groups can, or have a suitable level of information to access them. Problems concerning the length of documents introduced into the hypertext and questions of copyright also swayed us in favour a PC-based system. On top of that, since the building sector is a notably mobile one with a

high degree of geographical dispersion, we decided to grant a great deal of autonomy of action to single station. It was necessary to analyse the types of documents produced by Operating Units (OU) and the type of information involved in each document. This analysis allowed us to define the architecture of the information[1), and hence the keys to a documents library, a single document, a part of the document or to information about the OUs which took part in the research. It was thus possible to identify the functions of the system and to develop a prototype.

**2. Why a hypermedia system**

In view of the need to develop a tool to help disseminate knowledge and the results of research in the building sector in professional and academic ambits, it was necessary to establish choice criteria for the relevant set of documents. Due to the nature of the subject addressed, the documents differ by typology (texts, Images, tables), size and media used (paper, discs etc). Hypermedia seem to reflect the heterogeneous characteristics of the documents chosen, and also respond to the need, mentioned above, to facilitate the information retrieval and dissemination of in the sector.
Hypermedia systems are based on the notion of the electronic document [4]. In an electronic document, information is coded according to the typical (numeric) formats used by computers. The document thus becomes a coherent and finite set of structurable information, accessible to anyone sufficiently entitled.
The electronic document may be the product of the transfer in a numeric format of documents of another type, or it may be created directly as such.
Depending on the type of information they contain, it is possible to distinguish between documents with a text/character format and documents with an image format.
Note that the electronic medium allows us to emphasise some types of document use as opposed to others, whenever the onus is placed on information retrieval functions according to logical-associative and conceptual processes. An electronic document may be a synthesis created from more than one source; a catalogue of objects, for example, is built up from images of the objects themselves, a price table and forms describing the objects in the catalogue. Whenever the electronic document is enriched with logical and semantic links between the different items of information contained in it and the different ways of representing the same piece of information, then it may be seen as a hyperdocument.
In a hyperdocument, information my be navigated according to criteria more in conformity with the logic of the information searcher than that of the author.
A hypertext may organise a single document or library of documents and contains the cross-references which constitute the document of library.
In each document, it is possible to isolate parts which, decontextualised, conserve their intrinsic informational value. To refer to these parts, we shall use the term Documentary Units. The documentary unit is a part of a document which is artificially circumscribed with a view to further processing: bibliographic description, the description of contents, storage, information retrieval, the diffusion of the document[3]. A document is effectively a physical unit, but may entail information of a different nature or different subjects which it may be interesting to isolate. Below we use the terms Documentary Unit and Information (Unit indistinctly). The diffusion of word processing and electronic document management software has influenced writing and reading habits. The result is greater inconsistency in style and reasoning than was the case when documents were written by hand. With the advent of hyperdocuments on the other hand. readers have tended to personalise texts, again because of their logical and stylistic unevennes and also because it is now possible to access information at different moments and along different paths (spatial-temporal disorder in information access). The user thus perceives the possibility of running through each document interpreting it according to the objectives of the moment.
A characteristic feature of hypermedia systems is that links exist between the elements of knowledge contained in each document. These links constitute a logical structure which is, in general, different from the one originally conceived by the author. Thus, the user may follow the reading path best suited to his mental aptitudes. his knowledge and his information needs. This approach makes access and information retrieval more flexible and emphasises opportunities personalisation.

**3. Computerised help for information access: the process of breaking down technical reports**

The work we performed under the "Progetto Finalizzato Edilizia" set out to develop a hypermedia system organising and aiding the retrieval of information in the concluding technical reports of the first three years of the 1 RIESCE subproject. The first stage was to define the steps necessary to organise the relevant information. We decided, in particular, to work on the full text, not on synopses. Our first actions were:

- systematic, multiple reading to identify and transpose concepts and information units;
- definition of links both explicit and implicit) between concepts and information units, representing them in text, table and/or graph form;
- definition of a reading pattern different from the one prefigured by the author of the document though a logical path integrating the sequential reading strategy characteristic for a text on paper with typical mental processes whereby thought may follow different, sometimes highly complex paths.

The concepts referred to in the first of the three points identify classes of information by areas of competence of the reader who, on the basis of his specific knowledge, manages to restrict the field of research solely to the documents or portions of documents associated with the chosen class of information. This led to a reduction in quantity of the materials processed (without producing a significant loss of information) and to their schematisation.
In view of the need to think in terms of information organised according to the logical path followed by the user in the retrieval process, we broke the text down into information units, defining the architecture and different keys to the information which the system makes available to the user. The breakdown of the text consisted, basically, of the following stages: reading of the full text, identification of the Information Units (IU), attribution of the title to the Complex Information Units (CIU) identified, identification of concepts inside the single Information Unit, identification of anchor points in each information unit, identification

of links from one anchor point in an information unit of departure to the anchor point in the information unit of arrival, normalisation of the list of concepts in the single technical report, definition of a structure of concepts in view of the peculiar sector of application. Now let us see each of these stages in detail.

*a) Identification of the information units*

The concept of the Information Unit has already been introduced. Here a further distinction is made between Complex Information Units (CIU) and Elementary Information Units (EIU). A CIU is a self-consisting atom of information containing at least one concept. Unlike the CIU, an EIU is not self-consisting, in so far as its existence is bound to the existence of a reference CIU.

*b) attribution of the title to the Complex Information Unit identified*

The title has to identify the CIU univocally. It must be a synthesis of its content. It is not generally identifiable with a concept in so far as a CIU frequently refers to more than once concept. Moreover, for distinct CIUs which refer to the same concept, the information may be processed at different levels of detail and supply different information. The title of the CIU has to take this into account. The title is not necessarily an alphanumeric string retrieved in the text.

*c) identification of concepts in.side fhe CIU or EIU*

The concepts represent an abstraction of elements of reality which may or may not refer to physical entities. They allow us to express, in a simple, concise and structured way, the contents of the documents of the CIUs and E1Us. Since it is So general, the concept does not necessarily coincide with a string of the text in which the concept itself has to be retrieved. Note that the relationship between CIU/EIU and concepts is a many-to-many one.

*d) identification of anchor points in each information unit*

An Anchor Point is an object univocally identified inside a document by the document itself and by the context in which it finds itself. This is the concept which characterises the point of departures or arrival for each step in a process of navigation. Precisely because of its immediate, contextual characterisation, the anchor point represents a sensitive point In the text with which a link from/towards another information unit may be associated. Thus, to identify an anchor point is to identify a character string inside an information unit, and the above-mentioned link may be associate with the string.

*e) identification of links from one anchor point in an information unit of departure to the anchor point in the information unit of arrival*

Once the anchor points have been identified, It will be necessary to establish links with other information units according to criteria of adjacency and analysis.

*f) normalisation of the list of concepts in the single technical report*

After extracting concepts from each technical report, it is necessary to normalise them. This involves reducing redundant concepts and retrieving synonyms.

*g) definition of a structure of concept in view of the particular sector of application*

Once the concepts have been identified, they have to be organised in a network encompassing the concepts retrieved from all the technical reports. The hierarchy must be obtained by taking into account each particular sector of application.

Now let us see the architecture used to organise information in the hypermedia system.

**4. Architecture of the information in the hypertext**

Setting out from the structure of the technical reports of subproject 1 on the first three years of the "Progetto Finalizzato Edilizia", two blocks were identified to define the architecture of the hypertext: the data base and the information access key structure[2].

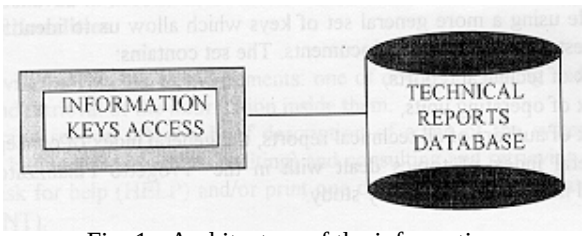


Fig. 1 - Architecture of the information

The data base contains information about the text and the illustrations of each technical report, organised both as they appear in the full text on paper and in information units. a distinction was introduced between Complex Information Units CIU) and Elementary Information Units (EIU).
Attempting a parallel with the relational model, the CIU might be seen as the basic structure of a hypermedia data base, in the same way that a relation is the fundamental logical structure of a relational data base. Hence the user may access the CIU at any moment of the navigation.
The Elementary Information Unit represents one or more properties of the CIU and, as such, exists only in so far as a reference CIU also exists.

The data base, finally, contains master information about the different operating units and the authors of technical reports.
Given the nature of the documents, the information contained in the full text is not generally organised according to a standardised structure. Besides being subject to very high discretionary factor, the organisation also appears closely dependent on the document's domain of application. For example, very specific documents, such as normative documents, seem to have a sufficiently standardised structure to allow organisation of the full text according to univocally identified entities (legal articles, paragraphs etc.). In general, however, the only structure that can be retrieved inside the full text is the hierarchical one, already used for a generic document on paper. according to which the text may be organises into sections, each section into chapters and so on.
The system therefore has to ensure access to information (hence to the full text) through keys already supplied for documents on paper. It must also be possible to access each document through its general index. The documents introduced into the hypertext are technical reports of 150-200 pages containing free text, illustrations, legislation. data collection cards, references to other documents and so on. A certain homogeneity was recorded with the type of publication, but not with the typology of the data involved. These considerations and analysis of the structure built up from the process of breaking documents down suggest that digital processing of information validates access with the following set of keys:

- CIU index.
- index of the authors of the technical report in question
- index of the concepts of the technical report (We think of concepts as classes of objects which represent reality and in which common properties may be identified)

The keys described above refer to single documents. Note, however, that the user who Interacts with the system has to retrieve information in a universe of technical reports. This is why the relevant document has to be chosen in advance. This selection can be made using a more general set of keys which allow us to identify the document which interests us in a library of documents. The set contains:

- the index of technical reports,
- the index of operating units,
- the index of authors of all technical reports, the general index of concepts,
- the general index of topics dealt with in the "Progetto Finalizzato Edilizia" and identified in the initial feasibility study.

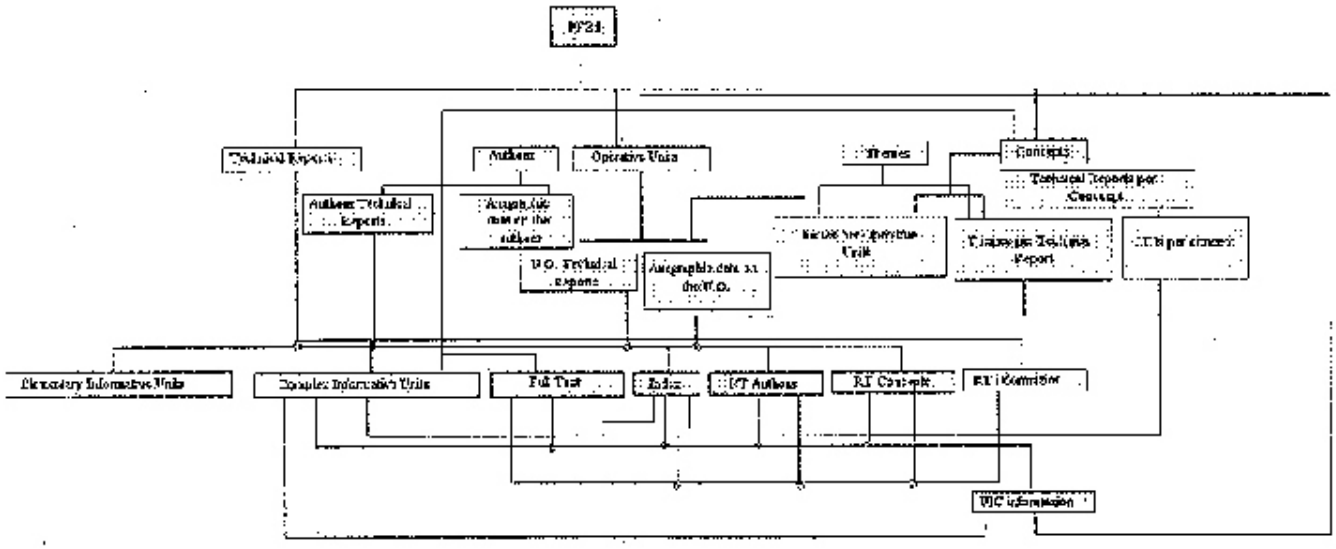


Fig. 2 - The hypertext organisation

**5. User requirements**

The system proposed here aims to render the knowledge and results of the research promoted under the "Progetto Finalizzato Edilizia" interactive and coordinated, highlighting connections, Implicit knowledge, propositions and suggestions. The users it addresses, as we have seen. form a somewhat heterogeneous group in terms of their professional backgrounds. People who may be interested in using the information range from professionals, researchers who intend to obtain the results of the research under the project, to students who use the same results for tutorial purposes. Users thus differ by knowledge of the domain of application, by objectives in the use of the system and by knowledge of the functionalities and interaction primitives which it displays. Within this framework the prerequisites which the system has to meet are:

- the system functionalities and interface must reduce the user's disorientation on access to information; the system must allow the user to access the information at least with the access keys he is accustomed to using with paper documents in relation to his mental model;
- the links between the different information units must be semantic allowing reconstruction of the logical paths followed by the user in navigating information;
- the system must offer the option of interfacing files with the most commonly used word processors and commercial graphics packages.

These prerequisites were the premise for designing functionalities of the system.

**6. System functionalities**

The system envisages two basic environments: one of document editing and another for consultation and retrieval of the information inside them.
This paragraph section offers a brief description of some of the system's practical aspects. While building the hypertext (editing) and consulting and retrieving information, the user may ask for help (HELP) and/or print one or more of the information elements retrieved (PRINT).
To reduce the risk of the user's being disoriented in THE use of the hypermedia system and to reduce his mental load (cognitive overhead) when, faced with a high number of alternative informational paths, he struggles to choose the one which Interest him, the MAP, RECENT and BACK functions have been added to the prototype. MAP function allows the user to visualise a map of types of information present in the system (full text, index, concepts of the technical report etc.), and thus to have a clear picture of the state of the system. With the RECENT function, the system allows the user to return to the nodes of the hypermedia system which he has already visited. The BACK function, finally, allows him to backtrack along the most recent consultation path. Numerous other facilities have been designed into the prototype to help the user.
The system developed contains an interface module friendly for the most heterogeneous range of users. Hence the interface (shown in fig. 3) is based on the use of menus and icons to minimise language barriers and user disorientation in interaction with the system.

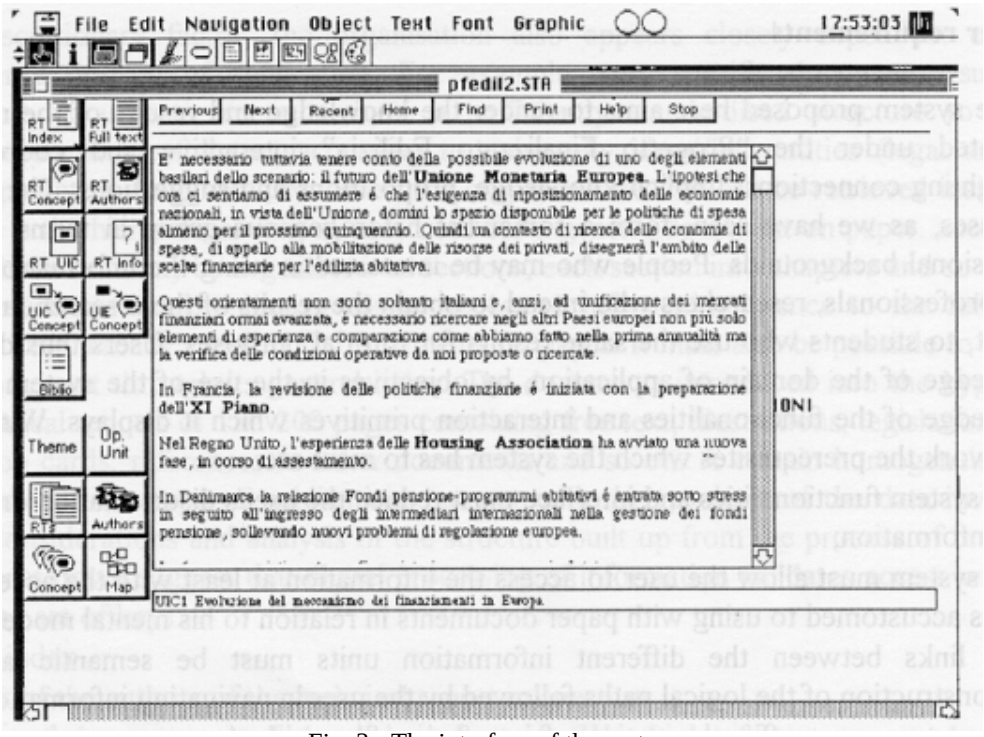

Fig. 3 - The interface of the system

**7. Conclusions**

The prime advantage of the use of a hypermedia system is the flexibility of its information access and non linear document consultation.
The prototype was developed on both an Apple platform and an IBM platform using the PLUS 2.0 tool. The prototype we have developed is a vital step in the study of the documentation produced by the "Progetto Finalizzato Edilizia" and the design of its structure. The next step will be to develop a hypermedia data base organising data on the conclusive results of the Project activities according to a predefined set of fields. In this way it will be easier to check data consistency and to foster an update process with a view to lasting use.

**References**

[1] Agosti Maristella, Crivellari Franco, Melucci Massimo. A Formalization of the Explicit model mechanisms for a development application environment, Atti del II Convegno Nazionale Sistemi Evoluti per Basi di Dati, Rimini, 6-8 Giugno 1994

[2] Baudin Catherine, Pell Barney, Kedar Smadar. Incremental Acquisition of Conceptual Indices for Multimedia Design Documentation, Atti della Giornata di Studio su Accesso all'informazione multimediale distribuita, Frascati, 14 Marzo 1995.

[3] Baudin Catherine, Pell Barney, Kedar Smadar, Using Induction to Refine Information Retrieval Strategies, Proceedings of the AAAI94 conferences, Seattle 1994.

[4} Conklin Jeff. Hypertext: a survey and introduction, IEEE Computer, 20(9), 1987