The browsing interfaces as a mean for an effective access to the data

Oreste Signore

CNUCE, Institute of CNR
via S. Maria, 36
56126 Pisa
Italy

Tel.: +39 50 593201
Fax: +39 50 576751
e-mail: oreste@icnucevm.cnuce.cnr.it

Abstract

In the access to unstructured information, the representation of the relevant concepts is a fundamental issue for the implementation of an effective mean of sharing knowledge between indexers and users. Browsing of the concepts may play an important role in the user interface. Two implementation of this concept are briefly presented.

1. Concept browsing interfaces

In the Information Retrieval approach ([1], [2], [3]), if the user cannot suitably express what he/she is looking for, the system cannot give adequate results. As it was pointed out in [4], some users of the information retrieval systems need often support to define or refine the topics of interest to them, others should improve their knowledge on the subject to decide exactly which are the interesting elements, others have a clear idea about the concepts to be found out, but they are not able to express such concepts. Others, at last, start from clearly defined subjects, but they realize that the starting concept is or too much indefinite (therefore getting too much high number of documents), or too much specific (causing the finding of a too much scanty number of documents). In any case, a refining of the

required concept is necessary. In order to facilitate the user's searches it's useful to represent the specific knowledge of a subject area by the explicit representation of the relationships of synonymy, preference and hierarchy among the terms, that is to create a thesaurus ([5]). A system of this type requires the user to move in the hierarchy of the concepts. Furthermore, the thesauri are considered as integrating part of hypertextual systems, as they form the structure through which the users can move to select the fundamental parts of the documents under investigation ([6], [7]).

The importance of a suitable user interface was emphasized since many years as an effective mean for communicating concepts between the users and the system. In [8] is supported the thesis according to which an approach based upon knowledge assures a consistency in the representation between indexers and users, provided that a semantic system of concepts accessible to both them can be available.

One of the possible approaches is formed by the graphic browsing interfaces[1], which have as prerequisite the availability of a high quality thesaurus. In the hypertexts and in the IRS the browsing has been used both for the assistance to query's formulation and the direct search of the information in the documents. Some remarkable prototypes ([8],[9],[10]) have been developed.

Two applications([11], [12]), developed in different environments, will be described in the following paragraphs.

2. HERMES:hypertext for hypertext

---

[1] The browsing represents an informal or euristic research that, through a collection of well connected documents, allows to find the important information which is needed. A browsing interface has to give suggestions and helps for the data base research when the user has a little knowledge of its content and organization, and to allow the users, particularly when they start from not well defined queries, to prepare or to refine their necessities, giving the possible alternatives of research.

In a specific application environment (the management of the cultural heritage) the usefulness of the implementation of an effective mean for accessing data pertaining coat-of-arms was envisaged. Even if a coat of arms is described by a well defined language (the heraldic grammar) it must be noted that this language is not commonly known. In addition, it is not easy to specify the kind of figure (natural or artificial) present in the coat of arms. Furthermore, considering the kind of potential users, the availability of an user friendly interface, which would also support the learning of the heraldic grammar, appeared essential.

All these considerations, and the presence of interrelationships of many unstructured informations (historical documents, maps, descriptive texts, etc.) lead to the the decision of implementing a hypertyext.

After an examination of the characteristics of escutcheon's descriptions, it was possible to conclude that the correct identification of the fundamental elements of the graphic composition of the coat of arms was an essential feature. It resulted that the more effective representation of the essential concepts was a structured thesaurus, which was implemented as a hypertext, too.
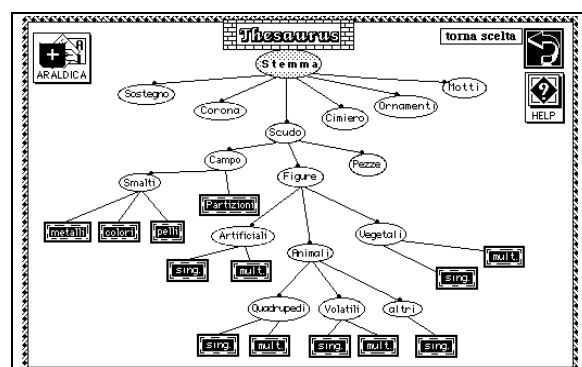

Fig. 1 - Visualization of the thesaurus

The thesaurus structure is visualized to the user under a tree shape (Fig. 1), where the *classification's facets* are represented by the ellipses, and the *classes of descriptor terms* by the rectangles. For each descriptor term, there is a leaf card, where the content of the related concepts can or cannot be visualized, according to user's discretion.

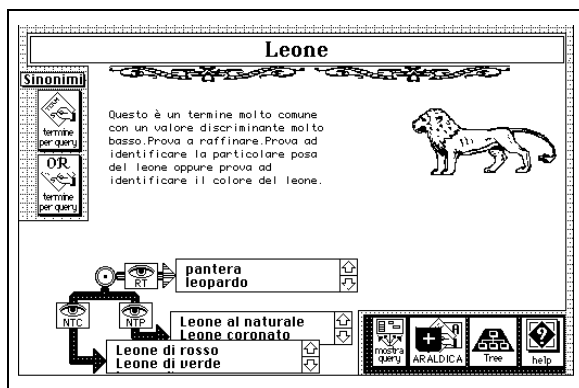The passage from one term to the other happens simply by following the associative links (Fig. 2).

Fig. 2 -A thesaurus leaf card

The thesaurus can be utilized either in *Didactic browsing* or in *Help for query* mode.

In the *didactic browsing* mode, we navigate in the graphic Thesaurus, visiting the cards of the descriptors terms, from which we may learn the heraldic term for a peculiar element, as well as peculiar rules of the nobility and an explanation of an eventual allegoric meaning.

In case we choose the *Help for query*, we may move in different ways:

• *Query by browsing*.
The system will take the job of composing the query according to the indications given from the user navigating in the thesaurus. On the cards of descriptors terms some new buttons are available, which allow to select the term considered more suitable (with or without its eventual synonyms) and put it in boolean AND/OR with the previous ones. The query can be later changed through a further navigating in the thesaurus.

• *Query by expert*
The user may enter the query, which is *filtered* by the system in order to verify the exactness of the terms utilized.All the terms which, even if aren't included in the thesaurus, are anyway present in the escutcheon's descriptions, are saved, to allow an intervention of the thesaurus's

manager (inclusion of the term and its connections with other terms).

The *result* card is common to all the query modes, and shows the identifiers of the retrieved documents pertinent to the query. Starting from the list of the retrieved documents, the user may initiate the navigation in the hypertext. It is worthwhile to note that during the navigation:

• the appearance of a marker indicates that the card under examination is an element of the retrieved documents set;

• the user may move both in the retrieved documents set and in the hypertext;

• at any moment the user may go back to the result card, to select a different entry point.

3. The PROTEUS interface

The growth and the spread of online information retrieval services has been much less then what was expected at the beginning of '80. This has been due in part to technological problems (networks reliability and speed, logon procedures, etc.), and in part to the intrinsic limitations of the Information Retrieval Systems. As a matter of fact, the available systems are substantially functionally equivalent, but differ, often only for minimal details, in the query language.

It is well known that, in order to be able to formulate effective queries, the user must be well aware of the content and the structure of the document. It is also evident that for the user it would be much more easy to interact with the database by mean of a conceptual representation of the document.

These considerations lead to the decision of implementing a general purpose interface, named Proteus because it is adatpable to any information retrieval system.

The architecture of Proteus is depicted in fig. 3. It is evident that the two basic functions (graphic interface and submission of the query to the host) are seen and managed as two separate tasks.

The graphic interface task is constituted by two processes which can be executed in multitasking. Each process, in turn, is a multi thread process: one thread is in charge of accepting the user's inputs, while the other is in charge of executing the queries on the support database. Processes are syncronized by means of semaphores, threads by message passing.
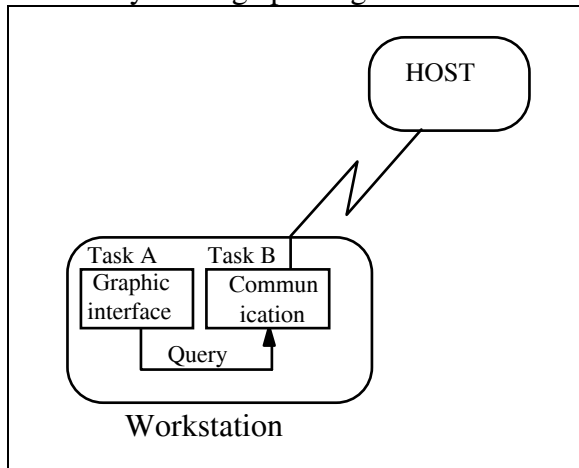

Fig. 3 - The PROTEUS Architecture

The Proteus interface rely on a database which contains all the information describing the available hosts and the characteristics of the accessible databases[2]. As soon as the user gets access to one of the databases he is registered to, he is faced with a conceptual representation of the document, independent on the features of the specific IRS. Needless to say, he may get a description of the content and the meaning of the various fields, or may enter a value that will be saved for the query formulation. If the user wants to enter a value for a controlled dictionary field, he may get a list of allowed terms, and perform a single or multiple choise in a list box, or just enter values, that will be checked against the dictionary.

As it has already been pointed out, the *Thesaurus Manager* has been implemented for tree structured thesauri as well as for multi-tree thesauri[3], and exibits a different behaviour depending on the specific thesaurus, even if the basic functionalities are the same, that is:

- graphical representation of the thesaurus as a graph, where the nodes are the thesaurus terms, and arcs are the connecting relationships;
- "point and click" interaction style.

Once the user has identified the kind of relationships he is interested to, he may select a term, either directly, either from a list box, and will have a display of a neighbour of the term, where only the selected relationships are shown.

Purely hierarchical thesauri are displayed as a tree (fig. 4), while for multi-tree thesauri we have a "butterfly display" (fig. 5).
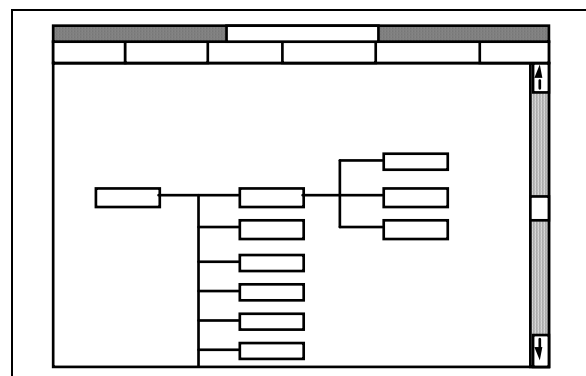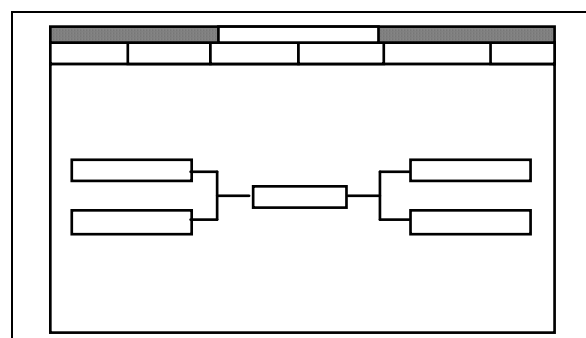

Fig. 4: A tree display.

Fig. 5: A butterfly display.

He may move around the structure (and ask for detailed explanation of the terms) using

---

[2] Each field of the databases may be a *free-text* field, or a *Dict* field (the field may take only the values defined in a list), or a *Thes* field (the allowed valued are organized as a thesaurus), or a *Cod* field (the field has an internal representation, different fron the structure perceived by the user, as discussed in [13]).

[3] As a matter of fact, it has been tested on two different thesauri, both consisting of several thousand words: ICONCLASS GEODOC.

the scroll bar, and may extend the tree toward the root or toward the leaves.

The use of the colors makes immediatly identifiable the currently active node (in *red*) and the origin of the browsing (*dark cyan*), from the others (*green*).

Any term may be selected in order to build a query to the database: all terms selected in a single navigation will be OR'ed.

4. Conclusions

Two different implementations of the concept browsing interfaces have been briefly described. In both cases, the most significant aspect is constituted by the possibility of a navigation on a graphical representation of the hierarchies of the concepts, and by the support given to the user in order to formulate the appropriate query.

REFERENCES

[1] Salton G., McGill M.J., *Introduction to modern Information Retrieval*, McGraw-Hill, New York (1983)

[2] Salton G.**:** *Automatic text processing*, Addison-Wesley (1989), ISBN 0-201-12227-8

[3] Van Rijsbergen C.J.: *Information retrieval, Second edition*, Butterwoths, London(1979)

[4] Smith P.J., Shute S.J., Galdes D., Chignell M.H., *Knowledge-Based Search Tactics for an Intelligent Intermediary System*, ACM TOIS, Vol. 7, N. 3 (July 1989)

[5] ISO 2788: *Documentation Guidelines for the establishement and development of monolingual thesauri,* ISO, Switzerland (1986)

[6] Thompson R.H., Croft W.B., *Support for browsing in an intelligent text retrieval system,* Int. J. Man-Machine Studies, Vol.30, 639-668,1989.

[7] Trigg R.H., Weiser M.: *TEXTNET: A network-based approach to text handling*, ACM TOIS, Vol. 4, N. 1 (January 1986)

[8] Monarch I, Carbonell J., *CoalSORT: A Knowledge-Based Interface*, IEEE Expert (Spring 1987), pp.39-53

[9] Pollitt A.S.: *End user touch searching for cancer therapy literature: a rule based approach*, in Proceedings of the Sixth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Vol.17, N.4, (June 1983)

[10] Frei H.P.,Jauslin J.F., *Graphical presentation of information and services:a user-oriented interface*, Information Technology: Research and Development, N.2

[11] R.Bartoli,A.Garibaldi, M.Greco, R.Potenza, O.Signore, B.Testa: *Una interfaccia per il browsing grafico di thesauri strutturati,* II Workshop Informatica e Scienze della Terra, Sarnano (MC), 22-23 October 1990. (To be published)

[12] Aulisi R.,Ceccanti V.,Signore O: *Hypertext for Hypertext: a Figured Thesaurus*, Database and Expert Systems Application, Proceedings of DEXA91 International Conference in Berlin, Germany, 1991, (D. Karagiannis ed.), Springer-Verlag

[13] Signore O., Bartoli R.: *Implementation of a historical/geographical database with support of imprecise dates*, DEXA 90: Proceedings of the International Conference in Vienna, Austria, 1990 (Tjoa A.M., Wagner R., Eds.), Springer-Verlag