BRAQUE: A Hypertext-based Interface for Accessing Large Text Databases

Pier Giorgio Marchetti

1. Introduction

The amount of literature indexed in bibliographic and text collections continues to grow, is complex and multidisciplinary in nature. The access to such sources of information, their consultation, is required by more and more professionals. Tools like personal computers and high speed networks to connect to on-line catalogues and bibliographic collections are easily available. Nevertheless a professional knowledge of collection contents, indexing methods and query language is normally required to allow a fruitful access to large text collections and retrieval of relevant information. This paper addresses the personal computer version (MS Windows) of the BRAQUE (Browse And Query) interface designed to overcome the above mentioned problems and give significant results in the information gathering process to the professional information hunter as well to the occasional searcher. BRAQUE relies on the implementation of a two level hypertext model to support the hypertext metaphor in a very large document space supported by meta information (concept space). Even thought the underlying model is valid for any kind of information object (including compound documents containing graphs, pictures or sound), previous experimental results and current implementation refer to text information objects. The validity of the BRAQUE approach, therefore includes the general case of compound information objects (compound documents) but there is room for further development and exploitation of the model and of the related implementation.

2. BRAQUE: Project Background

The BRAQUE interface has been designed trying to match two completely different sets of requirements. The first requirement was to satisfy the needs of as many user profiles as possible. In the past this requirement got solutions which were trying to categorise users according to their professio-
nal skills, and then to tailor interface behaviour according to such categories like «expert», «novice», which proved very rigid and unproductive. Following a cognitive approach [Ingwersen 1984] a number of efforts has been made up to an explicit use of Cognitive Task Analysis (CTA) [Roth Woods 1989], with the attempt to identify the actions performed by individuals in information access.

The initial design effort [Belkin Marchetti 1990], used the Cognitive Task Analysis of information retrieval (IR) interaction, based on the distributed expert problem treatment model developed by [Belkin et al. 1983].

In the earlier work the analysis had to rely on the specific situation of user-intermediary interaction in document retrieval.

Some insight was gained, but after some analysis it became clear that user behaviour was influenced and changed according to the information problem or according to feedback from the interaction with the (digital/on-line) information sources. Several attempts to perform synthesis and design starting from user interactions and CTA failed mainly because users of information systems try to satisfy their information requirements with several information seeking strategies. It is therefore extremely difficult to design an user interface pretending to use the insight from CTA or from user interaction records and try to generate a corresponding finite state machine. As a matter of fact, the (users’) information seeking strategies are influenced by the nature of the information problem, by the nature of the information itself (e.g. information, meta-information), by the user’s goal (e.g. learn, select), by the information access method (e.g. scan, search) or by the information access mode (e.g. recognise, specify) [Belkin Marchetti Cool 1993].

So the design specification of BRAQUE was tailored around the aim of providing users with at least some of the information seeking strategies mentioned above at any time.

The second requirement was to exploit the concepts’ categorisation performed during the document indexing process, to provide feedback both in the concept and document browsing, and in the querying of the document collection. In this view the information retrieval process can be represented according to the schema in Fig. 1 which extends (including, in an explicit way, concepts indexing, thesaurus and feedback) the information retrieval schema according to [Belkin Croft 1987].

A conceptual modelling of IR data resources, through a two-level description of IR systems, had been used to describe and model the functionality of an hypertext environment for interacting with large textual databases [Agosti et al. 1991, 1992]. The HYPERLINE prototype of this functional
schema has been under test for a long time in an operational environment [Marchetti-Muehlhauser 1991]. The list and a schema of the atomic functions which it is possible to build upon the two level hypertext model is found in Fig. 2.

3. braque: Design and Function

The braque interface was designed to support at least two basic interaction methods: browsing and searching (BRAQUE = BROWSE And QUERY). As a basic working hypothesis it was assumed that the user had the freedom to modify his interaction method (according to a new information seeking strategy) at any time during the interaction itself. This lead to a design centred around the availability of the complementary functionality (either browsing or searching) to any of the two basic interaction methods (searching and browsing), and the availability of a seamless bridging capability among them.

Browsing can be interpreted either as scanning through information items (specifically text documents or bibliographic references), or as scanning through the concepts. In the two level hypertext model of Fig. 2 the information problem can be solved at any of the two available levels (concept space and document space) with a number of interactions and accesses

Fig. 1
Read Function: 1,4
Associate: 1+2,2
Navigate: 3

Meta-Information

Level of Documents

3

1 2 4

to and through the available spaces. The model of Fig. 1 shows how the two levels of the hypertext model are in practice built. What has to be noted is that the arrows indicating the flow of feedback in Fig. 1 are describing the potential interactions (the actual information seeking strategies), in a very approximate and trivial way.

BRAQUE design was driven by the above mentioned considerations drawn from cognitive task analysis (CTA) and information seeking strategies (ISS), the two level hypertext models, and a model for the transparent construction of Boolean queries, which fits the desired general concepts of query formulation support. This model, as exemplified in the OAK system [Meadow et al. 1989], supports users in constructing faceted query formulations without explicit Boolean operators.

The general interface characteristics of BRAQUE are as follows:

- Integrated as one of two choices for the on-line information system: the native command language and BRAQUE;
- Graphical user interface;
- Direct manipulation for all the operations, including selection of items and functions;
- Window-based, with window sequences corresponding to functions;
- Pull-down menus that are general to all windows;
- Command buttons specific to particular windows or particular functional sequences;
- Icons, which when selected open the windows for specific functions;
- Context sensitive help in hypertext help format;
- Extensive service feature description in hypertext help format.

The overall structure of BRAQUE is organised according to functional sequences of operations or choices. The consolidated design [Belkin Marchetti Cool 1993], did foresee a large number of functional sequences, structured in a quasi hierarchy, with each sequence being organised with a number of choices, as in a hierarchical menu-based system. The resulting implementation lead to a quite complex interface software and to a relatively crowded interface screen populated by too many objects. In the implementation phase it was recognised that some of the functional sequences were designed to support specific information seeking strategies (e.g. ‘Find Known Document’) which in turn could rely on window structures already in place to support other seeking strategies (e.g. ‘Search Strategy Formulation’). Therefore an adequate parametrisation of the input window has been enough to offer a unique solution to different interaction problems.

On the other side, the number of required interaction or functional windows has been created recognising that some of the windows where required in order to support different user profiles. The original idea of a hierarchy of interaction sequences tailored around the requirements of specific user profiles (e.g. expert and aiming at minimising search costs, as opposite to novice and aiming at comprehensive searches and browsing interactions), was then replaced by a more schematic approach (only searching or browsing were allowed). This schematic approach has been mitigated by the possibility to customise the BRAQUE interface behaviour at any time during the session. It is therefore possible to set:

- The communication access mode (e.g. asynchronous or TCP/IP);
- The default operator for proximity searching;
- The automatic or manual access to titles from retrieved searches;
- The format of retrieved references.

Furthermore two other facilities have been integrated in the BRAQUE interface:

- Automatic query interpretation;
- Context check to allow a seamless exchange between the terminal interaction mode and the BRAQUE interaction mode.
For that which concerns the automatic query interpretation, the interaction from users with different degrees of familiarity with the query language is supported by means of a faceted query formulation window (named Document Searcher) which allows either the input of simple search terms as in the original schema suggested by Meadow or the input of the entire (query) search language, allowing complex Boolean queries including Boolean logic, proximity, prefixes, suffixes and truncation in the same search window.

The context check allows an expert user to start a BRAQUE session using the search and the browse windows, to continue his interaction directly entering commands on the terminal window and then to resume the BRAQUE session, with a context check automatically performed and the issue of an error message if the on-line session contains a status (e.g. deleted sets) which does not allow a consistent continuation of the interaction in windows mode.

The available functional sequences are as follows:

- **DATABASE SELECTION**: This window accomplishes choice of database or databases (multiple database searching). In particular, it allows the user to view pre-structured displays of databases grouped on specific topics (e.g. aerospace, environment, ...).

- **DOCUMENT SEARCHER**: This window supports the formulation of a structured query without explicit Boolean logic, through the use of a faceted structure entry box. As in OAK [Meadow et al. 1989], users are asked to consider their queries as a combination of concepts (facets), and to enter terms descriptive of, or related to each concept, in separate regions of the query formulation window. For searching, terms in each region are ORed, and regions are ANDed. Users enter terms by typing, or by copy and paste from the reference display window (BOOK), from the reference title window or from the Term Browser window. Furthermore, the user can reuse and/or modify previous search formulation saved in the Search Term Pool. The original Meadow’s facet concept is enriched allowing for the possibility to qualify any of the facets with the desired field identification (e.g. title, author, etc.). Use by novices is supported with the possibility to set a default proximity operator which is inserted automatically between adjacent terms in the facets. As mentioned above the expert professional can directly type any complex query statement in the facet.

- **TERM BROWSER**: This window offers and supports a variety of term
browsing options, including browsing in the thesaurus (with semantic association), browsing in a dictionary file, or browsing in secondary indices like the author index. All browsing sequences allow direct selection of terms from the display for immediate search and retrieval of documents or inclusion in the facets pertaining to the Document Searcher for more elaborated search formulations.

- **TITLE LIST:** This window allows the user to examine a title list after any search action. Selecting one entry in the list opens an overlapping window for individual document display. The link to the term browser is ensured via a button in the window.

- **DOCUMENT VIEW:** This window allows the user to read any individual document. When the window is open it is possible to copy elements of text into the clipboard for further inclusion in the Term Browser, in the Document Searcher or in any word processor for report generation. The print-out of the retrieved document on any of the printers available is also ensured via the main menu bar.

- **DOCUMENT POOL:** This window allows the user to save retrieved documents in the main storage in order to create local collections of relevant literature. Guarantee of clean export of entire documents to word processors for report generation is ensured by an export function.

- **TERM POOL:** This window allows the user to save complete search profiles for later reuse and/or manipulation.

- **SHELF:** This windows contains all the facilities to allow the user to search set manipulation. It contains a log of the sets created during the session. By point and click it is possible to narrow (ANDing) or broaden (ORing) searches, as well as jump to the Term Browser.

4. **Braque: An Example of Interaction**

Assume that the user (u) works in a technology broker company in charge of technological transfer. His boss has to prepare a report on the commercial exploitation of space.

His boss knows a lot about the technology, but very little about the international agreements ruling the acquisition and distribution of data from space and international law. Our user is then asked to find literature on the topics his boss feels he is not familiar with. He starts selecting the database he thinks may contain relevant technical information: the European Aerospace Database. He knows an existing exploitation issue for space generated
data is related to remote sensing products, so he thinks this is a proper starting point. He tries first to use the Document Searcher entering the three concepts «law, remote sensing, environment» (see Fig. 3). The initial results are rather poor. In fact, only four references have been found out of which three seem relevant to his information problem (see Fig. 4).

The user thinks he has to broaden the view of his search and therefore he tries the «concept» button on the titles that look more promising (see Fig. 5).

As a matter of fact the use of the term «law» is rather ambiguous in a technical database where there are a lot of descriptions of physics laws, and therefore the thesaurus (concept space) browsing proves to be very useful as reported in the following navigation (see Fig. 6).

The user tries a direct search on the thesaurus term getting immediately the view of published material on the topic of interest to his boss (see Fig. 7).

The simple example shows here how an hypertext browsing metaphor can be useful to address some of the problems users have in interacting with large text databases. The example above shows only a search and a concept browse action for the sake of simplicity. As a matter of fact, the two distinct actions (i.e. browse and search) can be iterated a large number.
Fig. 4

Document(s) 4 of 4

1. Conditions of access to Earth observation data: Legal aspects
2. Protection of SPOT data and derived products under private agreements.
3. Environment observation and climate modelling through international
4. Variation of atmospheric effects on measured radiance as a function of

Fig. 5

Search in: Thesaurus

Term: LAW (JURISPRUDENCE)

Proposed terms for set LAW AND ENVIRONMENT AND REMOTE SENSING

1. REMOTE SENSING
2. EARTH OBSERVATIONS (FROM SPACE)
3. ORGANIZATIONS
4. PROTECTION
5. SATELLITE IMAGERY
6. SPOT (FRENCH SATELLITE)
7. AEROSOLS
8. AGREEMENTS
9. ALTITUDE TESTS
10. ATMOSPHERIC EFFECTS
11. ATMOSPHERIC MODELS
12. ATMOSPHERIC MOISTURE
of times allowing a great degree of freedom to the user. In fact, the concept
browsing process can be either very finalised with the purpose of identi-
fying the relevant information elements or can be rather cognitive, letting
the user understand and familiarise himself with language specific problems
(e.g. the ambiguity of the word «law») as well as with topic and database
specific solutions (e.g. the use of the term «space law» or of the term
«international law» to resolve above mentioned ambiguities).

5. Conclusion

This paper describes the personal computer (MS Windows) version of
BRAQUE, a BROWSE And QUERY interface that, relying on a two level hypertext
model of large text collections gives multiple information seeking strategies
to users of a classical on-line information retrieval service (ESA-IRS). The
design process was based on a wide set of techniques ranging from precursor
user interaction studies via intermediaries to the cognitive approach (Co-
gnitive Task Analysis) and previous interface experiences, like the OAK
system. The BRAQUE interface supports a variety of user interaction styles,
addressing the problems that users face in accomplishing their tasks.
Through the two-level hypertext structure, it allows direct movement
between the meta information structure (i.e. the thesaurus) and specific
information elements (i.e. the references or documents). The preliminary
results show that the chosen approach is a viable one to support access to
multidisciplinary information sources from heterogeneous user communities.
Future work will focus on the exhaustiveness of implemented functionality
and on the development of a client structure able to operate in a client-
server environment according to the standard protocol for «search and
retrieve» applications i.e. ANSI Z39.50. Results are also to be expected from
ongoing research activities in related areas like lattice theory applications
for meta information (thesaurus) [Pedersen 1993] as well as from the
exploitation of the rule based Unix version of BRAQUE.

References

[Agosti et al. 1991] Agosti M., Gradenigo G., Marchetti, P.G., Architecture and
Functions for a Conceptual Interface to Very Large On-line Bibliographic Collec-
ment for Interacting with Large Textual Databases, in Information Processing and Management, 28(3), 1992, 371-387.


