Palay, A.J., and Fox, M.S., (1981), "Browsing Through Databases", *Information Retrieval Research*, R.N. Oddy, S.E. Robertson, C.J. van Rijsbergen, and P.W. Williams (Eds.), pp. 310-324, Butterworth & Co. (Publishers) Ltd.

Browsing through databases

Andrew J. Palay and Mark S. Fox

20.1 Introduction

Significant progress has been made over the last decade in the design and construction of database systems that go beyond their specialised applications in providing multi-purpose access to data. Structures such as relational, hierarchic and network databases provide efficient and flexible structuring and storage of data. Concurrently, research has progressed in providing powerful languages for accessing these databases — for example, relational calculus and algebra, QBE (Zloof, 1977) and natural-language interfaces (Mylopolous *et al.*, 1975; Waltz, 1975; Sacerdoti, 1977; Hayes-Roth, Mostow and Fox, 1978).

It is important to note that the majority of the research in the area of database access has focused on the area of *parameterised search* (PS). PS can be characterised as strictly focused, in the sense that the user must specify exactly the set of attributes that the records must contain and/or satisfy. There are a number of problems associated with PS-based access to databases. In particular:

(1) The interface to the database system is often too complex for the untrained user. In order to properly specify the parameters of a search, the user must understand a highly stylised and idiosyncratic interface language. Work in the area of natural-language understanding has been directed towards solving this problem.
(2) The logical structure of the database is normally different from the user's view of the database. If the user does not understand the structure of a database, it is doubtful that he will be able to specify the proper set of parameters and values in order to retrieve the information he desires.
(3) The user is unable to become familiar with the organisational structure of the database. Records are inserted into the database according to some classification schema. Since the user only specifies parameters and never views the system as a whole unit, the user has little chance of learning the classification schema.
(4) There is no quick and easy way to access *related* records. Each access, whether related or independent, requires a new search specification.

The BROWSE system has been developed as an alternative to PS-based database access systems. The BROWSE system relies on *browsing* as the primary mechanism for information retrieval. Browsing can be characterised as a heuristic search in a well-connected space of records. Some of the heuristics most commonly used in browsing library databases are:

(1) *If* book X is interesting, *then* what else has the author of X written?
(2) *If* book X is interesting, *then* what other books are in the same category?
(3) *If* a symposium article is interesting, *then* what else appeared in the same symposium?
(4) *If* the author of an interesting paper is from an institution X, *then* what else has been published at that institution?
(5) *If* there is an interesting paper in a journal, *then* what else appeared in that journal?

The goal of the BROWSE system is to provide *browsing* access to databases by building the search heuristics directly into the database as quick access paths between related records.

20.2 An example

A Browse system database consists of a set of frames. A frame is a single CRT (TV) screen of information. The purpose of the frame is to provide information to the user and to provide quick access to further related information. Options provide links to related frames. A user moves between the frames by selecting an option, which results in the system's displaying a new frame. The database can be viewed as a set of tokens, with the options providing the relations between the tokens.

The following example is taken from a BROWSE database developed for an on-line library catalogue system.

The first frame (*Figure 20.1*) that is displayed welcomes the user to the BROWSE system. In the upper right-hand corner of every frame is the name of the frame (in this case ZOG1). At the bottom of the frame is a set of standard options ('help', 'back', 'next', ..., 'find'). These options (called global pads) will appear in every frame. They provide a set of system functions that are useful throughout the entire network. The first frame consists of the text welcoming the user to the BROWSE system and three options. The first option allows the user to continue receiving instruction on how to use the BROWSE system. The second option allows the user to move directly to the top of the classification hierarchy. The third option allows the user to get news describing changes to the system. To select an option, the user types in the first letter of the option (an I to select the first option, a T for the second or an N for the third). If there is a pointing device (mouse, touch screen, etc.) available, then the user only needs to point to the option in order to select it. The naive user would continue by selecting the I option for more instruction. An experienced user would select the T option, thus displaying the top of the classification hierarchy (*Figure 20.2*). There are 12 options to this frame. Options that contain a dash (-) after the first two characters do not point to any other frames. They are included as pointers to information that one day will be included. An important property of the BROWSE system is that it can support

```
THE BROWSE-NET          6 July 1979                          ZOG1

                    THE CARNEGIE-MELLON UNIVERSITY

                    COMPUTER SCIENCE DEPARTMENT

                    LIBRARY BROWSING SYSTEM

          **** ZOG DIFFERENTIATES BETWEEN UPPER AND LOWER CASE ****


               I. Continue for instruction. (Type CAPITAL I to continue)

                    T. To begin BROWSING.

                                                        N. News

  help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.1

```
Top of the Browse-net.                                          browse5

You are now at the top of the BROWSE-NET.  The following are the
access paths available for browsing.

  1. Computing Review.                    6.-Institution

  2.-CMU Computer Science Dept.           7.-Keyword

  3.-Dewey-Mounts.                        8.-Journal

  4.-Library of Congress                  9.-Symposium

  5.-Author                               0.-Publisher

            N. New Entries

            I. Instruction and help information.


  help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.2

multiple views of the database. At present only the *Computing Reviews* (1976) classification and a new entries list are available.

By selecting option 1, the user moves on to the top of the *Computing Reviews* classification hierarchy (*Figure 20.3*). The frames forming the classification hierarchy each contain a title (0.: Computer Science), a definition section, a list of sub-categories, an entry list option (E), a lost map option (L) and a parameterised search option (S). If there are additional sub-categories, then a more selections option is included (M). Finally, if there is a designated


```
0.: Computer Science                                            CATEG73

        Computing Reviews

                                          The Computing Reviews
1.  General Topics and Education [1.]     classification is published in
                                          Computing Reviews by the
                                          Association of Computing
2.  Computing Milieu [2.]                 Machinery. (Computing Reviews,
                                          May 1976).

3.  Applications. [3.]

4.  Software [4.]

M.  More Selections

    E.-Entry List
    L. Lost: Map.
                                                  S. Parameterized Search


  help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.3

primary supra-category, then an option (P) is included that links the current classification frame to its primary supra-category. The entry list option points to a list of the entries that have been directly classified under the current category. In this case there are no such entries. Finally, the parameterised search option causes the system to begin a parameterised search and would proceed to a set of frames that allows the user to specify the parameters.

The initial goal of this sample session is to browse through the database for information related to learning. By selecting the option for Applications, the user learns that the category includes cognitive processes (*Figure 20.4*). By selecting the more selections option (M), the user discovers that Artificial Intelligence is one of the sub-categories of Applications (*Figure 20.5*). If applications was not a useful category, the P option could be selected to go back up the hierarchy, to allow the user to select another search path.

Now by selecting option 2, the user moves onto the frame describing artificial intelligence (*Figure 20.6*) and discovers that learning is one of the sub-categories (under Learning and Adaptive Systems). Note that an entry frame exists for this category. If the user would want to see entries that are directly classified under Artificial Intelligence, he would select option 'E'. With an interest in learning systems, the user now selects option 2 and the goal category frame has been reached (*Figure 20.7*).

There are no sub-categories to Learning and Adaptive Systems. The user has reached a terminal frame in the classification hierarchy. The entry list (*Figure 20.8*) is displayed by selecting the E option. Along with the list of entries, the entry list frame also contains a pointer back to the category frame that leads to the entry list (option R), and options to move back and forth through the entry list (options M and P, which is not shown). This is just one form of indexing provided by BROWSE. In addition, the autogeneration of hierarchic and alphabetical indices has been added.

```
3.: Applications                                          CATEG109

        Computing Reviews

1. Natural Sciences [3.1]                  This category contains
                                           subcategories concerned with the
2. Engineering [3.2]                       use of computers - where, how,
                                           when, and why they are used.  It
3. Social and Behavioral Sciences[3.3]     also deals with the relationships
                                           between human cognitive and
4. Humanities [3.4]                        perceptual processes and
                                           computing.
M. More Selections

    E.-Entry List
    L. Lost: Map.
    P. <0.: Computer  Science>                S. Parameterized Search

help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.4

```
3.: Applications                                          CATEG108

        Computing Reviews

1.  Management Data Processing [3.5]       This category contains
                                           subcategories concerned with the
2.  Artificial Intelligence [3.6]          use of computers - where, how,
                                           when, and why they are used.  It
3.  Information Retrieval [3.7]            also deals with the relationships
                                           between human cognitive and
4.  Real-Time Systems [3.8]                perceptual processes and
                                           computing.
M.  More Selections

    E.-Entry List
    L. Lost: Map.
    P. <0.: Computing Review>                    S. Parameterized Search

help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.5

The user decides that entry 6 looks interesting. That option having been selected, the frame in *Figure 20.9* is displayed. This frame provides the basic information about the article in question. The user may find additional information about the authors by selecting options A or B. If the user wants information about the Computer Science Department at Carnegie-Mellon University he may do so by selecting option I. In all three cases a list of entries associated with the author or institution will be available. The user can gain


```
3.6: Artificial Intelligence                              CATEG76

        Computing Reviews

1. Induction and Hypothesis-Formation      This category contains
   [3.61].                                 subcategories pertaining to
                                           induction and the formation of
2. Learning and Adaptive Systems [3.62]    hypotheses; learning and
                                           inductive systems; pattern
3. Pattern Recognition [3.63]              recognition; problem solving;
                                           simulation of natural systems;
4. Problem Solving [3.64]                  theory of heuristic methods,
                                           and general and miscellaneous
M. More Selections                         subjects within the broad area
                                           of artificial intelligence, or
    E. Entry List                          the machine simulation and
    L. Lost: Map.                          modeling of human functions,
    P. <3.: Applications:>                 particularly human
                                           intelligence.
                                              S. Parameterized Search

help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.6

```
3.62: Learning and Adaptive Systems                       CATEG158

        Computing Reviews

                                           Applications in which a computer
                                           modifies its programs according to
                                           input and/or memory, including
                                           modification of logical paths,
                                           self-adaptive pattern changes, and
                                           changes in parameter values.

    E. Entry List
    L. Lost Map.
    P. <3.6: Artificial Intelligence>         S. Parameterized Search

help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.7

additional information about the symposium in which this article appeared by selecting the option F. Along with information about the symposium, a list of all articles will be provided. Finally, there are a set of options that provide additional information about the current entry. Option 1 will provide a list of acronyms and keywords. The keywords are organised as a list of options. If the user wants to see what other entries in the database share a keyword, he can select that keyword. The second option provides the abstract to the paper. The fourth option leads to a list of categories under which the current article has

```
3.62: Learning and Adaptive Systems                              ENTRY89

1.  EG -- A case study in problem solving with king and pawn endings;
    Perdue, C.; Symposium or Conference Paper;
2.  Encoding knowledge in partitioned networks; Hendrix, Gary G.; Technical
    Report;
3.  Experiences in evaluation with BKG--A program that plays backgammon;
    Berliner, H.; Symposium or Conference Paper;
4.  Inference in the conceptual dependency paradigm: a personal history;
    Schank, Roger C.;
5.  Knowledge acquisition from structural descriptions; Hayes-Roth, F.;
    Symposium or Conference Paper;
6.  Knowledge-guided learning of structural descriptions; Fox, M. S.;
    Symposium or Conference Paper;
7.  Models of learning systems; Buchanan, Bruce G.;

8.  On fuzzy resolution; Aronson, Alan R.; Technical Report;


                                                    R. Root Category
M. More Entries

help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.8

```
                    Symposium                                SYMPOSIUM31

Knowledge-guided learning of structural descriptions


A. Fox, M.S.                          B. Reddy, D. R.
1. Computer Science Dept.
   Carnegie-Mellon University
   Pittsburg, Pennsylvania

Page number:

F. Proceedings of the Fifth International Joint Conference on Artificial
   Intelligence


                        1. Acronyms and Keywords.
                        2. Abstract.
                        3. Circulation Information.
                        4. Alternate Category.


help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.9

been classified. These categories are options that point back into the classification hierarchy.

By selecting option 2 the user can view the abstract of the paper (*Figure 20.10*). The user decides that he wants to see additional information on M. S. Fox. By selecting option A in *Figure 20.10*, he moves to a frame giving some information about the author (*Figure 20.11*). Options are provided for linearly moving through the author list (<, >), and going to the author index frame (↑). Selecting option 1, the user is led to a frame listing of all M. S. Fox articles that


```
                    Abstract                                 SYMPOSIUM33

Knowledge-guided learning of structural descriptions

                A. Fox, M.S.
                B. Reddy, D. R.

We demonstrate how the use of domain dependent knowledge can reduce the
combinatorics of learning structural descriptions, using as an example the
creation of alternative pronunciations from examples of spoken words.
Briefly, certain learning problems (Winston, 1970; Fox & Hayes-Roth, 1976)
can be solved by presenting to a learning program exemplars (training
data) representative of a class. The program constructs a characteristic
representation (CR) of the class that best fits the training data.
Learning can be viewed as search in the space of representations. Applied
to complex domains the search is highly combinatorial due to the: 1)
Number of alternative CRs. 2) Size of training set. 3) Size of the
exemplars.


help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.10

```
                    Mark S. Fox                                  author1

NAME:  Mark S. Fox

ADDRESS:
  Computer Science Department
  Carnegie-Mellon University
  Pittsburgh, Pennsylvania 15213

DATE OF BIRTH:  9 May 1952


                        1.  Publications.

                        2.  Research interests.


   <. Previous author      ↑. Authors with initial F      >. Next author

help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.11

are currently in the database (*Figure 20.12*). Seeing nothing of additional interest the user now decides to see M. S. Fox's other areas of interest (*Figure 20.13*).

Also interested in Information Retrieval, the user selects option 4. This puts him back into the classification hierarchy (*Figure 20.14*). He may now continue browsing in that area. If the user has managed to get lost, there is the lost map option (L). By selecting that option the user can get a global view of the classification hierarchy that surrounds the current category (*Figure 20.15*). Surrounding categories can be reached by selecting any of the options.

```
Publications: Mark S. Fox                                        author2

  1. Knowledge-Guided Learning of Structural Descriptions; Symposium paper.

  2. Maximal Consistent Interpretations of Errorful Data in Hierarchically
     Modelled Domains; Symposium paper.

                                                          ↑. Root Frame
help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.12

```
Research interests: Mark S. Fox                                  author3
CATEGORIES:                          KEYWORDS:
  1. Artificial Intelligence.          5.-Artificial Intelligence.
  2. Learning and Adaptive Systems.    6.-Learning.
  3. Database Systems.                 7.-Knowledge Representation.
  4. Information Retrieval.            8.-Discovery.
                                       9.-Man-Machine Communication.
                                       A.-Databases.
                                       B.-Speech Understanding.
                                       C.-Software Design.
                                                          ↑. Root Author
help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.13

20.3 Browsing and searching

The BROWSE system relies on browsing as its primary method of database access. However, there are many times when searching is desired. The user may already know exactly for what he is looking and should not have to move through the network of frames to get there.

The BROWSE system includes the ability to specify searches. As was seen in the example (*Figure 20.2*), each category frame in the current system has an


```
3.7: Information Retrieval                                       CATEG164
        Computing Reviews
                                        This category embraces
1. Content Analysis [3.71]              subcategories concerned with the
                                        systematic computer analysis,
                                        organization, storage, recovery,
2. Evaluation of Systems [3.72]         and dissemination of data.

3. File maintenance [3.73]

4. Searching [3.74]

M. More Selections

    E. Entry list
    L. Lost: Map.
    P. <3.: Applications;>                    S. Parameterized Search

help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.14

```
Lost Map For: Information Retrieval (Select * to Return)          Map2

                  |- 1. General Topics     |- A. Natural Science
                  |- 2. Computing Milieu   |- B. Engineering
                  |- 3. Applications-------|- C. Social Science
                  |- 4. Software           |- D. Humanities      |- I. Content Analysis
0. Computer Science |- 5. Mathematics of Computation  |- E. Management  |- J. Evaluation
                  |  6. Hardware           |- F. Artificial Intelligence  |- K. File Maintenance
                  |  7. Analog Computers   |- * Information Retrieval-----|- L. Searching
                  |  8. Functions          |- H. Real Time                |  M. Vocabulary

help   back   next   mark   return   top   display   comment   goto   find
```

Figure 20.15

option that allows the user to specify a parameterised search. The parameterised search differs from searches in normal PS-based systems in the following:

(1) The search that is initiated is context dependent. The user may browse through the classification hierarchy until he finds the area in which he is interested. When he initiates a search at that point, only entries that are