Search Techniques for Accessing CD-Rom Databases
Anjana Bhatnagar

Abstract
It is an indisputable fact that the computer information revolution has fundamentally changed library services across the nation for the past decades. Two recent advancements of information technologies, in particular, had significant impact on Library services. First, the introduction of CD-ROM technology has provided library users powerful access to large databases. Second, the creation of electronic networks has made online remote information access a reality. With the introduction of CD-ROM had a powerful impact on the storage and retrieval of information. CD-ROMs have proved particularly effective for storing bibliographical references. What are the CD-ROM Databases, why we use them and why we don’t want to use them in academic libraries are discussed briefly. Types of databases are explained by illustration. Search features commonly available in four selected CD-ROM databases such as keyword search, phrase search, Boolean search etc are discussed briefly. The selected CD-ROM databases are Ulrich’s on disc from Directory, Ei Compendex Bibliographic database, and Emerald full-text database of journal article. Some of the general issues related to the searching and retrieval of information from CD-ROM databases are briefly discussed. Some common points have been identified that the user must consider in order to quickly accessing information on CD-ROM databases in an effective way.

Keywords: Information Storage and Retrieval, Databases, Search Techniques, LISA, Engineering Index, Ulrich’s on disc, Emerald

0. Introduction
Due to the development of information technology, information is increasingly recorded in digital format and electronic databases which are continually becoming more complex, information storage and retrieval processes have profoundly changed. Over the past decades three major migrations have occurred in library academic services: from printed resources to online electronic databases; from online databases to bibliographic and full-text/full-image databases; and from CD-ROM databases to online Internet access to bibliographic and full-text/full-image databases.
1. **What are CD-ROM Databases**

CD-ROM is abbreviation for ‘Compact Disc-Read-Only Memory’ a type of optical disk capable of storing large amounts of data up to 1GB, although the most common size is 650 MB (megabytes).

The term ‘database’ has been used in the literature in two different connotations. In Computer Science this is referred to as a database management system. In this sense a database has a logically consistent structure in which records are actually linked. In library and Information science, these are referred to as bibliographic or inventory in nature and comprise a collection of related logical records here a database is defined as an organized and generally unlinked set of machine, or it can be said, these databases are the electronic version of published literature traditionally appearing in printed form in various field.

Databases are produced by different agencies in a computer readable form, the contents of which are accessible in various ways according to requirements with the help of suitable software. Various agencies (governmental and non-governmental) are involved in developing databases by collecting information from varieties of sources and packaging the information in the form of databases. Databases can be stored on CD-ROM, on the Internet, or on commercial sites that are accessible only if user’s library has a subscription. CD-Rom in the library contains periodical indexes, abstracts, statistics, directories, and other complete texts. Some databases provide information that is primarily numeric, such as data from a recent government census and some contain information that is primarily textual. Database some time also called data bank.

2. **Why We Use CD-ROM Databases in Libraries?**

Access to CD-ROM databases is fast, accurate and thorough. A clever search by a users may get them in seconds most of the information they need. In contrast, traditional print services may take months for same task, with still great deal of information left out. As a matter of fact, libraries can not possibly continue with the sluggish manual information service methods and hope to remain in business tomorrow. Libraries can have economic access to information, space saving and users friendly Libraries and information centers used these databases in providing various library services as mentioned below:

- Literature search services, current and retrospective literature searches
- Current awareness services in various forms including SDI
- Full text databases may be utilized for search services
- Document delivery services
- For producing digest trend reports etc.
- Compilations of bibliographies
- Abstracting and Indexing Services

3. **Why We Don’t Want to Use CD-ROM Databases**

It is always beneficial to know about the pros and cons of any information so that one can proceed systematically.

3.1 CD-ROM databases are not updated in time, some web journals or indexing services are available online even before their print editions roll out of the press. For example, Engineering Index on disc is updated quarterly, and Dissertation Abstracts on disc only once in a year.

3.2 Retrieval from CD-ROM databases take seconds, as against online databases, which take only milliseconds, however complex the command search statement may be (of courses with higher bandwidth telecom pipes provided)
3.3 If not acquired from a single producer, CD/DVD Rom databases require skills in using different search engines. Making menu-driven index searching or single keyword searching at command prompt may be easy to use in each system. But, users must make command, or Boolean searches, to get at exacter or less abstract hits. Sometimes users have been found key boarding DIALOG syntax instead of SPIRS one to search into MathSci Disc or PsycLIT from SilverPlatter, and vice versa. Both the search engines are so dissimilar in construct. Again, generally, users keyboard a collection of key words at the command prompt (including particles and other abstract terms, more often than not). They are flummoxed to discover that one access engine parses the search statement with “or” operator between words, and the other with ‘and’ operator. The former results come in too many hits and the latter in too few. Therefore, all new uses require basic education in the techniques of bibliographic database searching.

3.4 If adequate jukeboxes are not provided for permanent mounting of optical databases, it is a great hassle for users to switch from one disc to another for either a different file or part of the same file. By comparison, on online services, like DIALOG’s the process is just a click away.

Yet another disadvantage of CD-ROM, they are vulnerable to damage by dust, fingerprints and scratches, etc. Therefore they need very careful handing. Libraries will have to make backup copy of each CDS disc to be kept in safe custody, so fresh copies can be made.

4. Searching CD-ROM Database

Conducting effective searches requires knowledge of a number of techniques. Users should not only have been an understanding of the nature, content and structure of the database, but also be familiar with the various option available in the search screen, the tools and techniques for searching, and the techniques and formats for display of the search output. The search features commonly available are Keyword Search, Phrase Search, Boolean Search, Truncation, Index and /or Thesaurus Support, Proximity Search, Field-Specific Searches, Free-Text Search, Combing Search Sets and Search Refinement, Limiting or Range Search, Searching through the Retrieved Records and they may vary from one CD-ROM to another.

5. Selection of Databases for Comparison

Four CD-ROM Databases have been selected for a discussion on the various search and retrieval features available, ‘Ulrich’s on disc’, ‘Books in Print’ from the category of general or ready reference sources, ‘Ei Compendex’, ‘LISA plus’ are Bibliographic databases containing details of documents including abstracts and Emerald full-text database of journal articles.

5.1 Ulrich’s on Disc

Ulrich’s on Disc from R. R. Bowker provides the ability to search and browse the entire Ulrich’s family of databases. They included nearly 250,000 titles from over 200 countries, Information on about 11,000 new titles per year, annotations for almost 90,000 titles, full text reviews from Magazines for Libraries and Library Journal for over 8,700 publications, indicators for over 21,000 refereed publications and complete names and addresses for 80,000 serials publishers and distributors. The main screen of Ulrich’s is shown in Figure 1.
5.2 Ei Compendex Bibliographic Database

DIALOG OnDisc® Compendex, produced by (Ei) Engineering Information Inc., provides coverage of the world's significant engineering and technical literature. Subject coverage includes but is not limited to the various disciplines of engineering, applied physics, electronics and instrumentation, light and optical technologies, and other areas of significant technology. Compendex contains references to and abstracts from journals, technical reports, books, proceedings and conference papers, and more. Author-prepared abstracts are used when available. Publications from around the world are indexed, including approximately 4,500 journals and 2,000 conference proceedings per year. Approximately 10 percent of the documents indexed are in a language other than English.

Opening main Screen is like as Figure 2.

FIG. 1 ‘ULRICH’S on disc’ Main Screen

Fig. 2 ‘Ei Compendex’ Main Screen
5.3 Library & Information Science Abstracts (LISA plus)

LISA plus is the world’s best known resource for the coverage of ongoing research in all aspects of library and information studies. Over 245,000 abstracts keep user well informed about such topics as artificial intelligence, information and knowledge management, publishing and copyright, World Wide Web resources and much more since 1969. Lisa plus main screen is as Figure 3.

5.4 Emerald Full-Text Database of Journal Articles

Emerald is a full text database from MCB, covers more than 82 top journal across ten broad subject areas. It provides access to the full text of articles in PDF and HTML format published from 1994 to date. The subjects covered include Marketing, Human Resources, Quality Management, Information Management, Library and Information Services, Training and Education, General Management Property, Operations and Production Management etc. Users can choose any subject and get a list of all the journals covered in that subject. First screen of Emerald is as Figure 4.
6. Accessing Different Databases

Four databases mentioned above may be accessed by the following ways

6.1 Searching in Specific Field of Selected CD-ROM Databases

There are two forms of searching for information in CD-ROM databases – field-specific and free-text search. A user may know which field to search, or might like to restrict a given search to one or more fields. This is a field specific search. If the user are not sure of the field to be searched, they can conduct a free-text search, which means that the search is not restricted to any particular field. While the free-text search is not universally available, field specific search is the simplest form of search, and CD-ROM databases offer various options to conduct searches on one or more specific fields. The simplest option is the form search, where the user can select a specific box for a particular field as shown in figure no. 1 and can key in the search term/phrase. Which search field can be searched in a database depends on the content and structure of the database concerned, and there fore they differ from one database to another. Various search fields of selected databases are given below

Ulrich's on disc: Search fields are Abstracting and Indexing Service, Area Code, Circulation, CODEN Number, Country, Dewey Number, Document Availability, Electronic Vendor, ISSN, Keyword, Keyword In Title, LC Class, Media Type, Personnel Name, Price (US Dollar), Publication Code, Publisher, Special Features, Special Index, Status Code, Subjects, Title, US State/Zip, Year First Published. Search screen is shown by Figure 5

![ULRICH’S on disc Search Fields Screen](image)

Ei Compendex: Under option ‘Search/Modify’ search fields are Word/Phrase Index, EI Subject Headings, Author Name, Author Affiliation, Title Words, Journal Name, Conference Search Options (Conference Title, Conference Location, Conference Sponsor, Conference Year), Limit Options ‘English…’, (English only, Journal Articles Only, Conference Papers Only, Latest On Disc Records Only), Additional Search Options (Words/Phrases, EI Classification Codes, Major Subject Headings, Treatment Codes, Year of Publications, Language) different fields and Search results screen shown in Figure 6
Keyword search, author search subject search are most common search features. Search terms or phrases for the same or different fields can be combined using Boolean operators. Following table no.1 shows how searches on specific fields containing keywords and phrases can be conducted and what options are available in selected CD-Rom databases for searches on keywords or phrase

| CD-ROM Option For Conducting Search Method |
|--------------------------------------------|------------------------------------------|
| Ulrich's on disc Browse Index Select a field, enter the search term in the appropriate box, and search. Click on Find after entering the term After entering and selecting the chosen term, click on view titles to see the results |
| LISA plus Browse Select the particular field, enter the same term and search user can combine search terms, from the same or different fields by Boolean operators. Form Search Screen is shown in Figure 7 Select any particular field and choose to view the index. Select the index and search |
| Ulrich's on disc Form Search Enter the search term preceded by a field code (chosen from a list shown in the search fields box) and an equals sign, and then search |
| LISA plus Easy Search Select any particular field code and choose to view the index. Select the term from the index to search |
| Ulrich's on disc Search Search/Modify Click on ‘search/modify’ and then click on ‘word/phrases’ index. Enter the term in the small box after ‘enter’ and user will see the corresponding portion of the index. Select one or more terms and click OK. User will then see a search set with the given |
terms in the 'search history' window and the output (the complete record) in the output window in the output window Enter the keyword in the box for search terms and click on the search, the system will then conduct the search and show the number of hits and brief information about each record (document number, title, author, document type, and quality indicator) will appear, double click on any record to see the full text. Multiple keywords can be joined by Boolean operator. Click on browse and click on any field (author, keyword, etc) and the corresponding index will be displayed, browse through the index and double click on any terms to select and search. Note that here keywords can be single words or phrases.

Emerald Advanced Search Select a field (eg. full text, author, and keyword) enter search terms and select search. User can combine search terms with any other terms from the same or different field combined with Boolean operation (and, or, not)

Table No. 1. Options available for Searches on Keywords and Phrases

| FIG. 7. ULRICH’S on disc’ Form Search Screen |

6.2 Boolean Search

Combine terms to make the search broader or narrower. How broad or narrow the search is depends on the search logic used to combine search terms. Three logical connectors, or Boolean operators, are used to define the relationship between search terms: “and”, “or”, and “not”. The Boolean operators may be used to join search terms within search fields in all search modes and to join search criteria between search fields in the Advanced and Expert search modes only. (Novice mode always uses a Boolean AND between search fields). Boolean searches cannot be conducted in the 'browse index' mode of CD-ROM.

Ulrich’s on disc: Operators are AND, OR, NOT/AND NOT and Options are. In the Form Search mode user can combine terms/phrases with Boolean operators in a chosen field box, or can type one term/phrase in each box and click on a Boolean operator in the box next to the box next to the term on the right. In the search mode user have to type the Boolean operator to combine terms/phrases from the same or different fields.
Ei Compendex: Operators are AND, OR, NOT and Options are: use these operators to combine search terms in the ‘Modify Search Option’

**LISA Plus:** Operators are AND, OR, ANDNOT (not can also be used in free-text search mode) Options: In the Easy Search mode, user can combine terms/phrases with Boolean operators in a chosen field box, or can type one term/phrase in each box and click on a Boolean operators in the box next to the term on the right. In the Expert Search Mode, user has to type the Boolean operator to combine terms/phases from the same or different fields. No nesting is possible.

Emerald: Operators are And or, not and Options enter these operators in the search in the search as well as Advanced Search Mode

### 6.3 Truncation Search

Truncation of the three kinds of truncation (left, right and middle truncation), left truncation is not commonly available in CD-ROM databases. The symbols for truncation and their syntax and effect vary, and the user has to learn these to conduct an effective search.

**Ulrich’s on disc:** * or $ substitute for any number of characters and can be used for right truncation only. ‘?’ Substitutes for one character, can be used anywhere in a word, or user can use multiple ‘?’ Symbols.

LISA: Truncation is the substitution of a wildcard symbol for any portion of a word to retrieve a group of words. Most often, truncation is used to abbreviate a word — to trim a term back to its stem, or root word. Truncating lets user search for word variants thereby broadening the search. The asterisk * is a multi-character wildcard and the question mark ? is a single character wildcard. Wildcards are typically used at the end of a word but may be embedded within a word as well. Right and middle or internal truncation is allowed. Both the symbols can be used for right as well as middle truncation. User can enter the symbol as appropriate in the Easy Search as well as the Expert Search mode. Multiple ? can be used. Wildcards are typically used at the end of a word but may be embedded within a word as well. Avoid truncating words to less than five characters or user may retrieve unwanted results.

**Ei Compendex:** ‘?’ for any no. of characters. A combination of ‘?’ and space can be used to conduct different types of truncation. To search on words that start with a word stem and that have no more than one character after the stem, enter ‘?’ (Question mark, space, question mark). For example, ROBOT? ? Retrieves robot and robots. To search on words that starts with a word stem and have no more than two characters after the stem. Enter ‘??’ (two question marks) for example, Robot?? Retrieves: robot, robots, robotic.

**Emerald:** Use a ? to find any single character in the position. Use an * to find any one or more characters in the position.

### 6.4 Index Search and Thesaurus Search

The indexes allow users to select the search terms/phrases from the term index, and a thesaurus allows users to consult a map of available terms to widen or narrow down a given search, as required. CD-ROM retrieval software provides index search facilities, ie users can select an index to browse and select from the index for searching. There may be an index file or a separate index for each searchable field so that the user can choose a field and then browse the corresponding index file. Index search facilities are more commonly available than thesaurus facilities. Thesauri are hierarchically structured. This means that they go from broad terms to narrower, more specific, terms.

**Ulrich’s on disc:** User can choose a field from either the search or the From Search mode. In each case user can display and browse the corresponding index and select a term/phrase from the index.
Ei Compendex: In order to conduct a search, users need to choose an index: author, subject heading, keywords, etc., and then the corresponding index will be open for users to browse and to select the search term.

LISA plus: The thesaurus may be accessed via the Browse window by clicking on the Index drop-down list and selecting Thesaurus Term. Highlight the required term and click View Titles. In the Expert Search mode, enter the thesaurus term in the Search Query Box, click on Ok or press Enter. Press the 'View Brief' button to display the results of the search, i.e., the thesaurus block. A thesaurus term can be selected to conduct a search in the database as shown in Figure 8.

![Fig. 8 'LISA plus' Thesaurus Search](image)

Emerald: In the Advanced Search mode, users can choose one or more fields to search, and once users have entered a search term corresponding to a field, they can click on the 'Thesaurus' button to get other words corresponding to the search term.

6.5 Proximity Search

With a proximity search, users can specify the number of words allowed between the search terms. In a general sense, there is a correlation between the number of intervening words (the words creating the "distance" between search terms) and the topical relevance of users' search results. The closer the distance between users' search words, e.g., 5 words, the more relevant but fewer the results may be. Likewise, the wider the distance between the search words, e.g., 20 words, the less relevant but larger the CD-ROM results may be. The type of proximity search operator to use depends upon whether users need search terms to be found in the same order as listed in the search statement (adjX) or found in the same or reverse order of what users originally specified (nearX).

Ulrich's on disc: Not available.

Lisa plus: ANDx or NEARx locates both words, in any order within x words in the same field. NOTx locates the first word, but not the second within x word in the same field. WITH: locates both words in adjacency, forward order only, in the same field. WITHx: locates both words, forward order only with x words in the same field. Option in proximity search is only allowed for the free text search. It can be used in both the Easy Search and Expert Search modes. Expert Search Screen of IISA is shown in figure 9.
El Compendex: Write two or more words consecutively to search as a phrase. Use periods (.) for one or more intervening words. Option for search expression ‘artificial intelligence’ will be searched as phrase, i.e., the word next to each other in the same order. User can specify the maximum no. of intervening words by one or more periods, e.g., information; management will retrieve both information management and information resource management.

Emerald: Same document, same paragraph, word apart, exact order. In the Advanced Search mode, user can choose any of the four options to conduct a proximity search.

6.6 Free-Text Search

Free-Text Search enables the users to specify one or more search or phrases that are not limited to any particle field. There may also be index for free-text search terms. Instead of searching field, in some cases the user might like to conduct a search in all or many different fields. This is possible through what is known as the free search facility. In some cases, the default setting is the free text search, i.e., whenever a user enters a search term it is searched in all or a number of selected fields as shown in figure 1. Free text option in selected databases as

Ulrich’s on disk: No option is available for free-text searching

El Compendex: No free text search is possible; user have to choose an index (i.e., a field) to conduct a search

LISA plus: Here a free-text term is any single word taken from any field in the record, with the exception of stop words (). Free text differs from keyword searching in that a space between two terms is treated as meaning true proximity in relative position and order. For example: ft = catalog$ will retrieve all references that have ‘catalog’ as the first seven characters of the term; ‘ft= line’, will retrieve all references that contain the terms ‘line’ including ‘line management’, author/research worker ‘Line Maurice B’ ft=public libraries will retrieve all references where the term ‘public libraries’ occurs, but it will not retrieve records with the combinations, ‘Libraries public’ or with the phrase ‘public and academic libraries’. If a term is entered in the Search mode’ with out an operators a free text (ft) will be automatically carried out.
Emerald: In the search option, or using the ‘Full-text’ option in the Advanced Search mode, user can conduct a full-text search, i.e. a search on any part the document

7. Results

After the comparison we come to know that all databases have almost same basic accessing techniques as keyword search, Phrase Search, Boolean Search, Truncation, Index and/or Thesaurus Support, Proximity Search, Field-Specific Searches, Free-Text Search, Combing Search Sets and Search Refinement, Limiting or Range Search, Searching through the Retrieved Records and they may vary from one CD-ROM to another with a slight variation as given below

- Boolean searching is possible in all the CD-ROM database, through how it is carried out varies. Boolean searches cannot be conducted in the ‘browse index’ mode of a CD-ROM database for the obvious reason that only one term/phrase can be keyed in to display the corresponding portion of the index.
- None of the sample databases provides left-truncations facilities. The symbols for truncation vary from one database to another.
- Search refinement of a given search can be refined by adding or dropping one or more search terms or criteria, such as limiting the output to a given language, year of publication, price range, and so on.
- Some CD-ROM products have more than one search interface, for example, Ulrich’s on disc, while others have on search interface giving options for both - simple, novice and other advanced/ expert search.
- Some retrieval system such as EI Compedex show the result instantly, while in others the user has to chose the view/display option, as in LISA plus.
- Proximity search in CD-ROM databases allows users to conduct proximity searches in order to specify how closely the terms prescribed in a search expression.
- Two types of indexing practice are followed word index and phrase index. in the word index, each individual word in a field is indexed, while in the phrase index, a given phrase that may comprise several keyboards, is indexed. The user may need to check whether a field is a word- or phrase-indexed, by browsing through the index for the particular field, before entering a search term/phrase.
- CD-ROM databases do not usually provide facilities for ranking the search results according to relevance.

8. Conclusion

The mission of library services in the age of computer and electronic technologies is to provide information at the point and the moment of users’ needs. Unquestionably, content is still the key to the success of a library electronic database. The medium of information delivery systems has evolved over the years from printed indexes to magnetic tapes, disks and compact discs. CD-ROM technology seems to be increasingly popular among library users. Users are not sophisticated searchers; their acceptance of the CD-ROM technology indicates that they are willing to learn. With initial training users may become less reliant on library staff, at least for routine database operations. Improvements in the user-friendliness of software interfaces and standardization of navigational functions should make CD-ROM databases easier to use and allow users to concentrate on developing good search strategies rather than such mundane activities as how to print.

As the expertise of users increases, so do the expectations. With more databases containing abstracts and full text becoming available, user demand for immediate access to documents will grow. The
availability and preference for the electronic version of documents will have an effect on print subscriptions and inter-library loan. It is important that CD-ROM database developers and libraries keep abreast of the changing attitudes and use of CD-ROM by patrons in order to offer the services which will meet both present and future user information needs.

The promise of new technology is for libraries to extend the reach of research and education, improve the quality of learning, and reshape scholarly communication. This is not an extravagant claim for the technology, but rather a declaration of an ambition shared by many who are developing and managing the technology. And the key to fulfilling that promise lies within the communities of higher education, science, and public policy responsible for applying the technology to those ends.

9. References

4. www.royalroads.ca/coppul/glossary.html accessed on June 21, 2004
5. www.cert.org/homeusers/homecomputersecurity/glossary.html
10. Ulrich’s on disc summer 2004 RR Bowker (CD-ROM)

About Author

Dr Anjana Bhatnagar has been working at PK Kelkar Library at Indian Institute of Technology, Kanpur for the last twenty years and presently working in Computer Aided Reference Services of Digital Division of its library. She has been consistently involved in assisting the users by guiding them to use the databases according to the latest integrated technology. With the motive to expand and share her working knowledge and practical experiences about how to access the databases quickly and wisely, she wishes to provide information about searching retrieval, and user interfaces with the help of comparative study of databases commonly used in academic libraries.