Module VI

CHAPTER 6

User Interface and Visualization

Syllabus

Human Computer interaction, the information access process, starting points, query specifications, context, using relevance judgments, interface support for the search process.

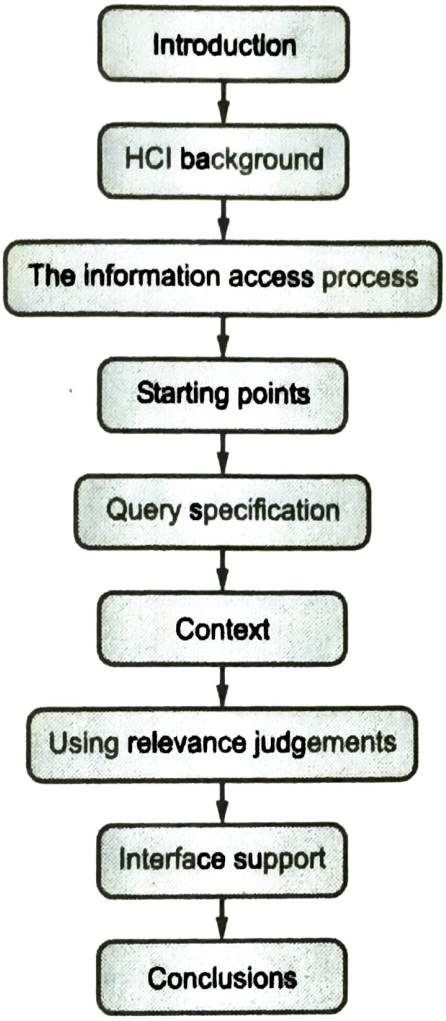
Self-learning Topics : SeeSoft

GQ. What are main components of user interface?

GQ. What are the main methods of retrieving data and information?

GQ. What are the main challenges in information retrieval?

- Here, we'll talk about user interfaces for interacting with information retrieval systems on a human information seekers.
- Users frequently approach information retrieval systems with only a vague idea of how they might accomplish their objectives.
- As a result, the user interface should facilitate the expression and interpretation of information needs. Additionally, it ought to assist users with query formulation, choice of information sources, understanding of search results, and tracking of search activity.



(1F1)**Fig. 6.1 : User Interfaces and Visualization Flow**

6.1 HUMAN COMPUTER INTERACTION

GQ. What is human computer interface?
GQ. Discuss the design principles of human computer interface?
GQ. What is the role of visualization in human computer interface?
GQ. What is GUI in visualization?

- Ben Shneiderman, an ex- pert in the field of Human Computer Interaction defines HCI as,

  "Well designed, effective computer systems generate positive feelings of success, competence, mastery, and clarity in the user community. When an interactive system is well-designed, the interface almost dis- appears, enabling users to concentrate on their work, exploration, or pleasure".

- Shneiderman provides a list of guidelines for user interface design as a

means of achieving these objectives.

(1) Provide informative feedback,

(2) Enable easy action reversal,

(3) Support an internal locus of control,

(4) Lessen working memory load, and

(5) Offer alternate interfaces for novice and expert users are some of the factors that are particularly crucial for information access. Depending on the specific application of the interface, each principle should be instantiated in a different way.

The principles that are particularly pertinent to information access systems are covered below.

6.1.1 Design Principles

(1) Offer informative feedback

- This idea is important for information access interfaces in particular.
- We'll look at the most recent approaches to informing users about the connections between their query definition and the documents that were returned, as well as the connections between the documents themselves and the metadata defining collections.
- The system offers an internal locus of control if the user may decide how and when feedback is sent.

(2) Reduce working memory load

- The process of gaining access to information is iterative, and as new information is discovered, its objectives vary and change.
- Information access interfaces that offer mechanisms for remembering choices made during the search process enable users to resume temporarily abandoned strategies, switch between strategies, and retain information and context over the course of multiple search sessions.
- This is one important way that information access interfaces can reduce memory usage.

- Providing browsable material that is pertinent to the present step of the information access process is another memory-helping tool.
- This provides search beginning points like lists of sources and topic lists, as well as suggestions for related terms or information.

(3) Provide alternative interfaces for novice and expert users

- Simplicity versus power is a key trade-off in every user interface design.
- Less flexibility and occasionally less effective use come at the tradeoff of simpler interfaces being easier to master.
- Powerful interfaces can be time-consuming to learn and place a memory burden on users who only occasionally use the system, but they also give knowledgeable users additional options and control over how the interface functions.
- Using a "scaffolding" approach is a common solution. The application is offered to the novice user with a straightforward interface that is easy to learn, delivers the essential functionality, but is limited in power and versatility.
- For more seasoned users, alternate interfaces are provided, providing them more power, more functionality, and possibly even whole distinct interaction models.
- A well-designed user interface offers logical links between basic and sophisticated interfaces.

(4) Information access

- Tradeoffs between simplicity and power are unique to interfaces.
- The quantity of information displayed regarding how the search engine itself operates is one such trade-off.
- A crucial design decision in information access interfaces is how much information the user should see.

6.1.2 The Role of Visualization

- Most computer users nowadays are familiar with windows, menus, icons, dialogue boxes, and other interface design features.

- These offer a more user-friendly interface than command line -based displays because to the usage of bit mapped displays and computer graphics.
- Information visualization, a less well-known but expanding field, aims to offer visual representations of very large information spaces.
- Humans are very sensitive to visual information and visuals. Some types of information can be conveyed through a visual representation far more quickly and efficiently than through any other means. Think about the distinction between a written description of a person's face and a picture of him.
- A fast developing area of this discipline is **scientific visualization,** which maps physical phenomena onto two- or three-dimensional images.
- A vivid representation of the pattern of peaks and valleys on the ocean floor is an illustration of scientific visualization; it offers a picture of physical events that cannot be captured on camera.
- The image is instead created using data that represents the underlying phenomena.
- The main information visualization techniques, in addition to the use of icons and colour highlighting, including brushing and linking, panning and zooming, focus-plus-context, magic lenses, and the use of animation to maintain context and make obscured information accessible. These methods enable interactive, dynamic use.
- **Brushing and linking** is the joining of two or more views of the same data so that any changes to one view's representation also affect the other views' representations. Consider a display that has a histogram and a list of titles as two examples.
- The histogram displays the number of documents published annually for a given set of documents.
- The titles of the associated documents are displayed in the title list. By employing brushing and linking, a user may choose a colour, like red, for one bar of the histogram, which would highlight all the titles in the list display that were released in that year.
- The operations of a movie camera to scan sideways over a scene (panning), move in for a close-up(zooming), or back away to acquire a

wider view are referred to as **panning and zooming** .

- Text clustering, for instance, can be used to display a top-level perspective of the primary themes in a collection of documents. You can zoom in closer still to see the text linked with individual papers after moving closer to reveal individual documents as icons.
- **Focus-plus-context** is a technique for enlarging the focus of attention while simultaneously making the surrounding items smaller.
- Like the effect seen in a fisheye camera lens, an object appears smaller the further it is from the focus of attention.
- When placed on top of another data type, **magic lenses**, which are directly manipulable transparent windows, apply a transformation to the underlying data, altering the underlying data's appearance. Magic lenses are most commonly used for sketching jobs, and they are especially helpful when used as a two-handed interface.

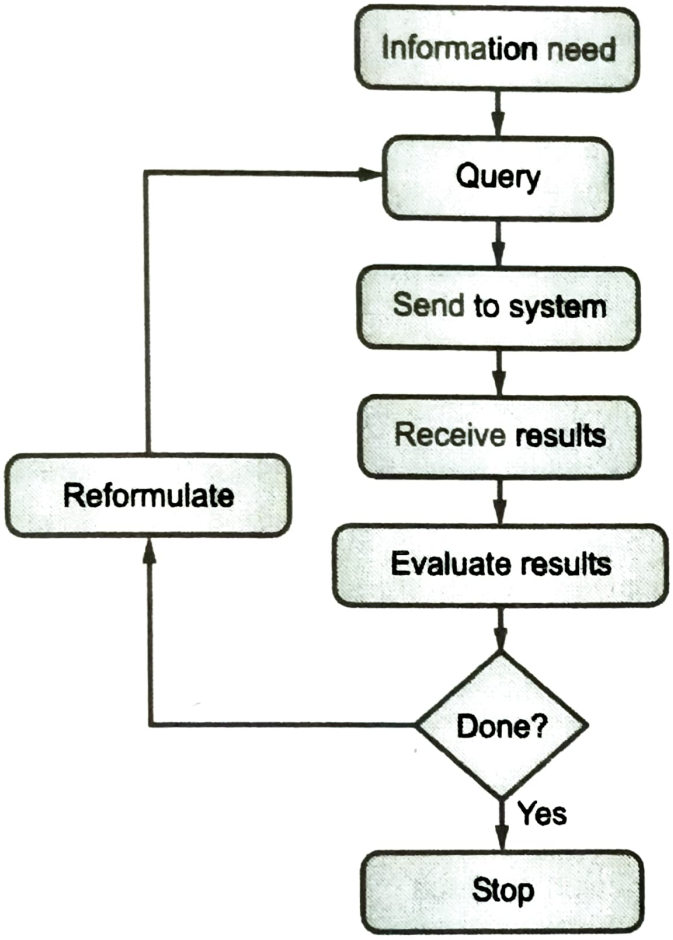
☞ Evaluating Interactive Systems

- The approach for evaluating user interface strategies is an important aspect of human-computer interaction.
- Although they are frequently used to compare the rankings produced by non-interactive systems, precision and recall metrics are less suitable for evaluating interactive systems.
- In the TREC tasks, systems are compared based on how successfully they return the top 1000 documents, unlike to the typical evaluations that place an emphasis on high recall levels.
- To evaluate highly interactive information access systems, useful metrics outside of precision and recall include: **time required to learn** the system, **time required to achieve goals** on benchmark tasks, **error rates**, and **retention of the use of the interface** over time.
- However, in many interactive settings, users only need a few relevant documents and do not care about high recall.

⏭ 6.2 THE INFORMATION ACCESS PROCESS

GQ. Explain the information access process in detail?
GQ. What are the different types of Overviews?

- Finding a plumber, keeping up with a business rival, drafting a publishable scholarly piece, or even looking into a fraud charge are all goals that require access to information.
- To achieve a specified goal, information access tasks are used. These activities include everything from asking targeted questions to conducting in-depth research on a subject. Between these two extremes are other tasks.
- Observing a well-known topic over time, following a plan or predefined set of searches to reach a certain goal, and researching a topic in an undirected way were identified to be the three main types of information seeking tasks, according to a study of business analysts.
- The Standard Information Access Process are described according to the following sequence of steps :
  (1) Start with an information need.
  (2) Select a system and collections to search on.
  (3) Formulate a query.
  (4) Send the **query** to the system.
  (5) Receive the results in the form of information items.
  (6) Scan, evaluate, and interpret the results.
  (7) Either stop, or,
  (8) Reformulate the query and go to step 4.
- Today, the majority of information searchers only encounter this simple interaction model.
- The fact that many users dislike being presented with a lengthy, jumbled list of retrieval results that do not specifically answer their information demands is not included in this paradigm.
- Additionally, it incorporates the underlying presumption that the user's information needs are static and that the process of finding information consists of iteratively fine-tuning a query until it returns all documents that are pertinent to the initial information demand.



(1F2)**Fig. 6.2.1 : A simplified diagram of the standard model of the information access**

- It is used by Web search engines.
- It ignores the fact that most users find lengthy lists of documents that don't directly address their query disturbing.
- It makes the assumption that a user's information needs are static and that the best way to find information is to successfully refine a query until it returns all of the documents that are pertinent to the initial information need.

Non-Search Parts of the Information Access Process

- It makes the assumption that a user's information needs are static and that the best way to find information is to successfully refine a query until it returns all of the documents that are pertinent to the initial information need.

- When the analytical procedures were more closely studied, it was discovered that there were six main kinds of work that made up 80% of the work: finding trends, drawing comparisons, aggregating information, identifying a critical subset, assessing, and interpreting.
- Cross-referencing, summarising, finding dynamic visuals for reports, and other miscellaneous tasks made up the remaining 20% of the work.
- These findings suggest that search/retrieval and analysis/synthesis of results represent the two primary parts of the information access process. User interfaces should support a seamless integration of both types of activities.
- However, analysis and synthesis are tasks that can be completed independently of information seeking, therefore it is important to distinguish between the two types of activities for our purposes.

☞ Starting Points

- Users must have easy entry points in search interfaces. A user cannot select how to begin a search by looking at an empty screen or a blank entry form.
- Users typically don't start by expressing their information needs in great detail. According to studies, users frequently begin their searches with relatively simple queries, review the results, and then make little adjustments as a result of the feedback.
- The assistance of users in choosing the sources and collections to search on is one of the functions of an information access interface.
- The user's first problem formulation can be aided by directing them to the appropriate collection of beginning points.
- Traditional bibliographic searches presume that the user starts by perusing a list of source names and selecting which collections to search on, whereas Web search engines erase source distinctions and thrust the user in the middle of a Web site with little information about the relationship of the search hit to the rest of the collection.
- The interface to the sources that are available is not especially useful in either situation.
- Let's discuss four main types of starting points: *lists, overviews, examples,* and *automated source selection.*

☞ Lists of Collections

- The user must renew the query on a different collection if they are dissatisfied with the results on one collection.
- Through experience, formal education, or suggestions from friends and colleagues, frequent searchers gradually learn a collection of sources that are helpful for their areas of interest.
- A "favorites" list, sometimes referred to as a bookmark list or a hotlist on the Web, can be used to store frequently used sources. Recent research investigates how to keep track of a user's or work group's tailored information profile based on the information sources they've previously accessed.
- However, a list of well-known sources is insufficient when users want to do searches in areas outside of their areas of expertise.
- Professional information seekers, like librarians, get knowledge of the best sources through practice and years of training.
- Exploration and the finding of new beneficial sources are discouraged by the constrained nature of traditional interfaces to information collections.
- To aid users in understanding the contents of collections as a method to start their search, researchers have lately developed a number of approaches.

☞ Overviews

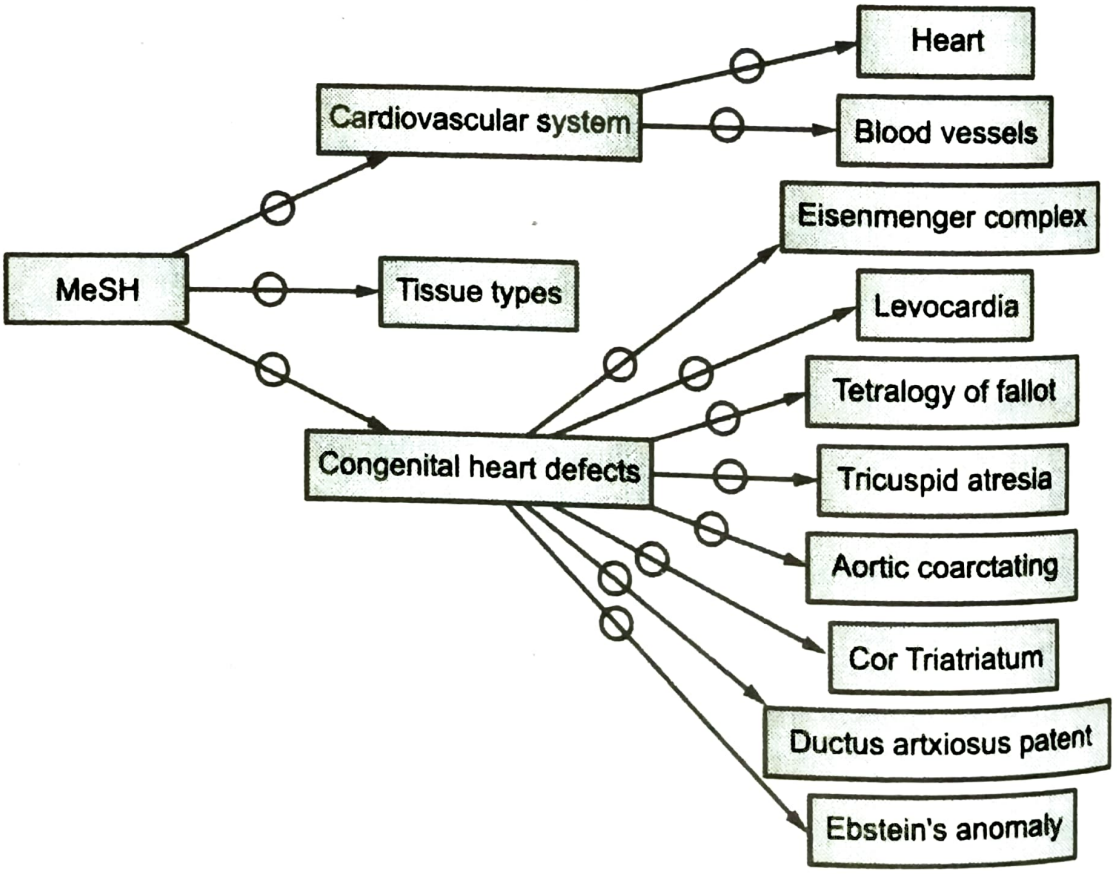
- An overview can assist users choose or rule out sources by displaying the topic domains that are covered by the collection.
- An overview can guide people into generic neighborhoods to get them started, and they can then navigate using more specific descriptions.
- The researcher suggests a method of engagement in which the user first views an overview of the data they will be working with, pairs and zooms to identify prospective interest areas, and then views details.

☞ Category or Directory Overviews

- In some collections, category labels have different meanings. Most are made to aid in document organization and query specification.
- A document is typically associated directly with the node of the category hierarchy to which it has been allocated in interfaces that graphically represent category hierarchies. For instance, in Yahoo!, selecting a

category link displays a list of the documents that belong to that category.

- The document is conceptually kept inside the category label. In order to navigate the search results in Yahoo!, the user must scan a list of category labels and determine which one is most likely to include references to the desired topic.
- If you choose the wrong turn, go back and try again while keeping track of which pages include certain pieces of information.
- This can be a time-consuming and frustrating process if the needed information is hidden deep in the hierarchy or not at all. Users cannot build searches based on combinations of categories using this interface because documents are theoretically stored "within" of categories.
- Designing an effective interface to incorporate category selection into query specification is challenging, in part because displaying category hierarchies requires a significant amount of screen space.
- Lets consider One web-based service that enables the integration of search with display and MeSH category label selection is Internet Grateful Med.



(1F3)**Fig 6.2.2 : The MeSHBrowse interface for viewing category labels hierarchically**

- The MeSHBrowse system allows users to interactively browse a subset of semantically associated links in the MeSH hierarchy.
- Clicking on a category from a particular beginning point causes the connected categories to be shown in a two-dimensional tree form.
- As a result, only the pertinent subset of the hierarchy is displayed at a time, making reading through this enormous hierarchy a more manageable task.
- The interface lacks tools for searching over an underlying document collection and suffers from the space restrictions associated with a two-dimensional hierarchical presentation. In Fig. 6.3, it is depicted.

☞ Automatically Derived Collection Overviews

- The most prevalent general topics that appear within the collection have been automatically extracted in numerous efforts to show summary information. Unsupervised analytic techniques, typically variations of document clustering, are used to derive these themes.
- The centroids of the clusters determine the themes in the collections. Clustering classifies documents according to how similar they are to one another.
- The Scatter/Gather browsing paradigm aggregates documents into thematically related groupings and shows the user descriptive text summaries.
- According to a user research, Scatter/Gather can successfully transmit some of the information and organisation of a huge text collection. However, that study also shown that a regular similarity search was not as good as Scatter/Gather for discovering documents that were pertinent to a query.
- In other words, participants who were only allowed to look through retrieval results and not construct their own questions performed worse than subjects who were allowed to make their own queries and browse the hierarchical structure of clusters covering the whole collection.
- By using clustering to arrange the retrieved documents after a query, it is possible to connect Scatter/Gather with traditional search technologies.
- The study discovered that, if the clusters are created from the top-ranked documents returned in response to the query, papers relevant to the question tend to fall predominantly into one or two out of five clusters.

- The study also demonstrated that the best cluster's precision and recall were higher than the retrieval results as a whole.
- The consequence is that a user may save time by focusing on the cluster's contents that has the greatest number of relevant documents while avoiding clusters that contain mostly irrelevant documents.
- As a result, clustering of retrieval results may be helpful in guiding users to a segment of the results that contain a significant number of pertinent documents.
- Using Kohonen's feature map algorithm, maps that graphically depict the overall content of a document collection or sub collection have been produced (See Fig. 6.2.3).
- The size and form of the 2D map's regions fluctuate according to how frequently the collection's documents assigned to each region's related region occur.
- The adjacency of regions is intended to reflect the semantic relatedness of the themes within the collection. Regions are described by a single word or phrase.
- The titles of the documents that are most closely related to a document region are displayed in a pop-up window when the cursor is moved over that region.

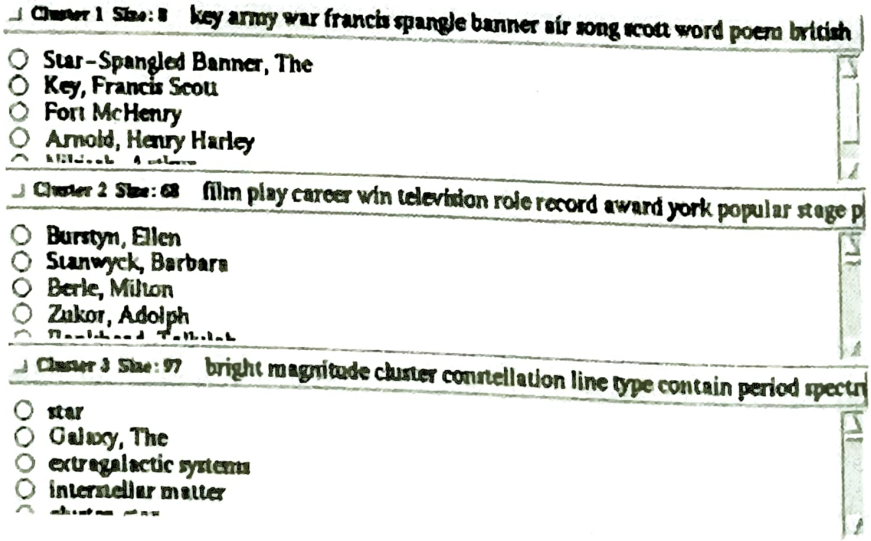


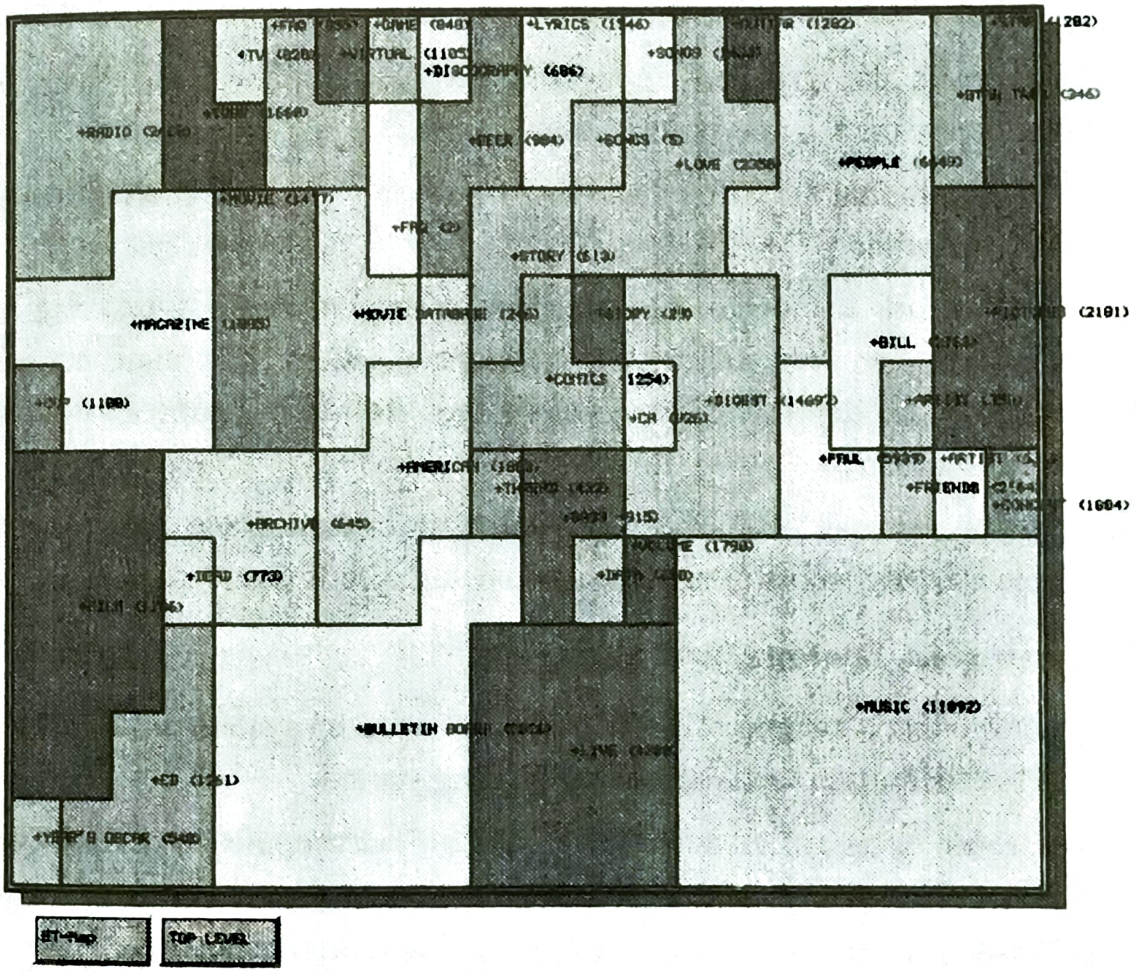
Fig. 6.2.3 : Display of Scatter/ Gather clustering Retrieval Results

☞ Evaluations of Graphical Overviews

- The assessments that have been made so far offer no positive support for their usefulness.
- According to one study, the results of clustering were going to be difficult to use for non-expert users, and graphical representations (such as depicting clusters with circles and lines connecting documents) were much more challenging to use than textual representations because it can

be challenging to determine a document's contents without actually reading some text.

- A different recent study contrasted Yahoo! with the Kohonen feature map overview representation on a browsing task. One of the tasks required the subjects to locate a "interesting" Web page in the Yahoo! entertainment category and to arrange the same Web pages into a Kohonen map style.
- Whether participants began in Yahoo! or the graphical map affected the experiment. After completing the browsing activity, participants were instructed to try to repeat it using the alternative tool.
- Within 10 minutes, 11 out of 15 of the subjects who started with the Kohonen map visualization found an intriguing page. Eight of them were successful in using Yahoo! to locate the same page.14 out of 16 people were successful in finding engaging home pages.
- Only two out of the 14 were successful in locating the page in the graphical map display, though.



(1F4)**Fig 6.2.4 : A two-dimensional overview created using a Kohonen feature map learning algorithm on Web page having to do with the topic Entertainment**

- The author came to the conclusion that this interface is more suited for browsing than search.
- Unsupervised thematic overviews have not generally been demonstrated to be useful for usage in the information access process, but they may be most helpful for giving users a "gist" of the type of content that can be found inside the document collection.

☞ Co-citation Clustering for Overviews

- Long known as a means to provide a summary of a collection's contents, citation analysis. Based on co-citation patterns, the fundamental concept is to identify publications that are "centrally located."
- The frequency with which two sources are cited by a third source is one approach to gauge citation patterns. Pairing articles that quote the same third article is another option. In both situations, it is assumed that the paired articles have some things in common.
- Documents are grouped depending on how similar their co-citation patterns are once a matrix of co-citations has been constructed.
- The clusters that are formed are understood to represent recurrent themes in the material.
- In an effort to locate key authors within a topic, clustering can concentrate on the authors of the papers rather than their contents. Recently, the Referral Web initiative used web-based papers to put this idea into practice.
- The concept has also been applied to websites, where major topical themes among websites are identified using web link structure.

☞ Examples, Dialogs, and Wizards

- Giving users a sample of how to engage with the system at the outset is another technique to assist them in getting started.
- The term "retrieval through reformulation" also applies to this method. The Rabbit system, which shows graphical representations of sample database queries, is an early implementation of this concept.
- The user is given a broad outline for a query, which they can subsequently alter to create a partially comprehensive explanation of what they want.

- The system then displays an illustration of the kind of information that corresponds to this sketchy description.
- If a user searches a database of computer items and shows a preference for discs, for instance, an example item is retrieved with its disc descriptors filled in.
- The user can iterate the process and utilize or alter the shown descriptors.
- The concept of retrieval by reformulation has been explored further and applied to the fields of software engineering and user interface design.
- The interactive dialogue is a more exciting version on this concept. Since the beginning of information retrieval research, dialog-based interfaces have been investigated in an effort to imitate the interaction offered by a human search intermediary.
- An animated focus-plus-context dialogue is used in the DLITE system interface to walk users through typical system operating sequences.
- The dialog's initial phases are listed out in outline form at the top. Any individual step's explanation can be expanded by the user by clicking on its description.
- When the user wants to concentrate on the current tactic, they can close the dialogue again after expanding it to see what questions will be asked next.
- Wizard dialogue is a more constrained style of dialogue that has gained widespread use in commercial systems.
- This application aids users in completing time-sensitive tasks, but it makes no attempt to explicitly teach users how to carry out the tasks.
- The wizard reduces user input to just a few options with default settings, providing a step-by-step shortcut through the series of menu selections (or strategies) that a user would typically carry out in order to complete a task.
- According to a recent study, wizards are helpful for tasks requiring numerous steps, for users lacking the requisite domain knowledge (such as a restaurant owner installing accounting software), and when steps must be carried out in a specific order (for example, a procedure for hiring personnel).

- Successful wizards had features including the ability to be run again and modified by the user, an overview of the supported functions, and clear descriptions and clear results for choices.
- When the user interface failed to solve an issue successfully, wizards were proven to be ineffective (for example, a commercial wizard for setting up a desktop search index requests users to specify how large to make the index, but supplies no information about how to make this decision).
- Wizards were also discovered to be useless when intended to instruct the user on how to utilise the interface and when not put through a user testing process.

☞ Automated Source Selection

- There is a lot of room for research in the domain of human-computer interfaces that assist in directing users to the right sources.
- It calls for both identifying consumers' information needs and determining which needs may be met by which sources.
- Building a model of the source and the user's information needs, then attempting to identify which best fits the two, is an ambitious approach.
- Making a representation of the information sources' contents and comparing it to the query definition is a simpler solution.
- This strategy is used by GIOSS, a system that seeks to predict beforehand which bibliographic database to send a search request to in light of the query's terms.
- The system uses a simple analysis of the combined frequencies of the query words within the individual collections.
- Automatically sending a query to many sources and then combining the responses from different systems is the opposite of automatically choosing the best source for a query.
- There are a tonne of metasearch engines online. Research is ongoing in the domain of effective result fusion, also known as collection fusion.

6.3 QUERY SPECIFICATION

GQ. Define property/attribute specification and explain its sequence.

- In order to create a query, a user must first choose the collections, metadata descriptions, or information sets that the query will be compared to and then they must provide the words, phrases, descriptors, or other types of data that will be compared to or matched against the collections' data.
- In order to display the results to the user, the system generates a set of documents, metadata, or other information types that in some way fit the query definition.
- Five main human-computer interaction styles are noted in the study. These include direct manipulation, natural language, menu selection, form filling, command language, and form filling.
- Every method has been applied in query specification interfaces, and each one has benefits and drawbacks.

☞ Boolean Queries

- The matching procedure in modern information access systems often uses a statistical ranking algorithm.
- However, until recently, the majority of bibliographic systems and commercial full text systems only enabled Boolean queries.
- Therefore, the difficulties users face while specifying Boolean questions have been the topic of several information access research.
- Unfortunately, studies have repeatedly demonstrated that most users struggle to formulate queries in Boolean style and frequently predict the wrong outcomes.
- Multiple factors make boolean queries challenging. The primary one is that most people find the fundamental syntax to be counterintuitive.
- When Boolean operators are expressed using the English words AND and OR rather than their logical equivalents, many English-speaking users presume that everyday semantics are also linked with Boolean operators.

- Using AND implies an expansion of the query's scope to inexperienced users because more types of information are being asked.
- Additionally, the majority of query languages that include Boolean operators demand that the user supply complicated syntax for additional connector types as well as for descriptive information.
- The majority of users are not familiar with the concepts of operator precedence or the use of parentheses for nested evaluation.
- The creators of World Wide Web search engines had to come up with more user-friendly methods of query specification because they were catering to a large audience with little experience in query specification.
- They give users a variety of standard, straightforward ways to combine query terms, such as all the words (put all phrases in a conjunction) and any of the words (place all phrases in a disjunction).
- Allowing syntax-based query formulation while offering a simpler or more user-friendly syntax is another web-based alternative.
- The fact that pure Boolean systems do not rank the documents returned based on how closely they fit the query is another drawback. A document must either satisfy the query or not under the pure Boolean framework.
- Web- based systems usually rank order the results of Boolean queries using statistical algorithms and Web-specific heuristics.

☞ From Command Lines to Forms and Menus

- Another issue with pure Boolean query definition in online bibliographic systems, aside from conceptual misunderstandings of the logical meaning of AND and OR, is the arbitraryness of the syntax and the contextlessness of the TTY-based interface in which they are mostly offered.
- Typically, input is written in response to a prompt and has the following form:

  COMMAND ATTRIBUTE value (BOOLEAN-OPERATOR ATTRIBUTE value}* e.g.,

  FIND PA darwin AND TW species OR TW descent FIND TW Mt St. Helens AND DATE 1981

- The names of the commands and attributes must be recalled by the user because they are easily forgotten in between system usages.
- This issue is made worse by the fact that, after more than ten years, one of the two main online bibliographic systems at UC Berkeley still generates an error if the author field is specified as PA rather than PN, as is done in the other system, despite the fact that the command languages for the two systems have different but very similar syntaxes.
- Interfaces made to fit the system rather than the users are characterised by a lack of flexibility in the syntax.
- The user no longer needs to recall the names and categories of accessible qualities thanks to the new Web-based version of Melvyl, which offers form completion and menu selection.
- Users choose the types of metadata they want to use from list boxes, and attributes are clearly displayed so that selection can take the place of definition.
- The Melvyl application for the Web enables the retention of context between searches, the storage of previous results in tables, and the hyperlinking of these results to lists that contain the obtained bibliographic data.
- By checking a box next to the query record, users can also edit any of the previously submitted questions.
- Many of the system's sophisticated choices are made apparent and immediately available by the graphical interface, which most users would not learn by using the command-line interface.
- Compared to command-line interface, bit-mapped displays are an improvement, but they do not completely fix the issues.

☞ Faceted Queries

- Another issue with Boolean searches is that their rigid interpretation frequently results in result sets that are either empty or excessively large depending on whether the user conjoins terms in an attempt to minimise the result set or includes several items in a disjunct.
- The user's ignorance about the collection's contents and the functions of the terms inside it is a major contributing factor to this issue.

- In systems with command-line-based interfaces like DIALOG, a typical approach to solving this issue is to create a series of brief queries, check the amount of documents returned for each, and then aggregate those questions that yield a reasonable number of results.
- For instance, in DIALOG, each query returns a group of documents to which an identifying name is given. DIALOG displays the set number with a lifting of the number of matched documents rather than returning a list of titles.
- By specifying the specified number and giving the command to display the titles, titles can be displayed. Document sets that are not empty can be referred to by a set name and new sets can be created by combining them using AND operations.
- The user can go back and try a different combination of sets if this set in turn proves to be too tiny, and this process is continued until a set of documents that is a reasonable size is produced.
- This type of query formulation is frequently referred to as a "faceted query" to denote that the user's search is broken down into themes or facets, each of which should be included in the pages that are returned [553, 348]. For instance, a query on medicines for the treatment of osteoporosis can have three facets, denoted by the disjuncts (drugs OR pharmaceuticals OR osteoporosis OR "bone loss") (prevention OR cure)
- This request suggests that the user wants to see documents that cover all three subject areas.
- Post-coordinate or quorum-level ranking is a method to impose an ordering on the outcomes of Boolean queries. According on the size of the subset of the query phrases each document contains, this method ranks documents.
- A situation that bridges full Boolean syntax and free-form natural language inquiries is produced when faceted queries and quorum ranking are combined.
- A list of entry lines could be the format of an interface for expressing this kind of interaction. Each subject is entered by the user as one topic per entry line, and each topic is composed of a list of phrases that are disjunctly concatenated and have semantic relationships.
- Writings that include at least one term from each facet are given a higher ranking than documents that only use terms from a few or one facet.

- As a result, documents that discuss several of the user's themes will be ranked higher than those that only mention one.

☞ Graphical Approaches to Query Specification

- Many different types of graphical interfaces, both dynamic and directly manipulable, have been created to make specifying Boolean grammar easier.
- User surveys typically show that these graphical interfaces outperform their command-language counterparts in terms of accuracy and speed.
- Below are three such methods that are described.
- It has been suggested numerous times that graphical representations of **Venn diagrams** be used to enhance Boolean query formulation.
- A ring or circle represents a search phrase, while the intersection of rings denotes a conjunction of terms.
- Typically, the applicable portions of the diagram display the number of documents that satisfy the various conjuncts.

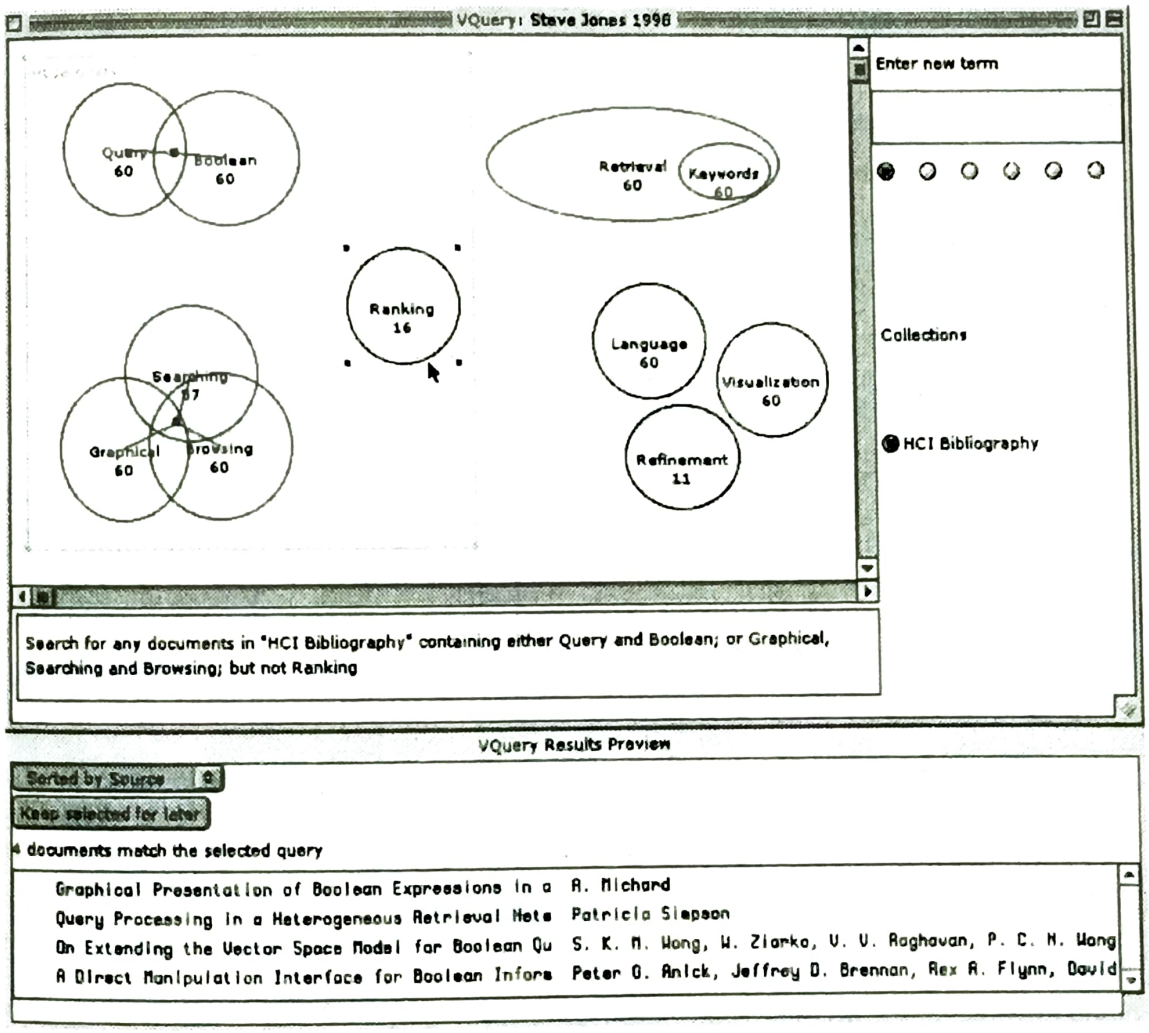


Fig. 6.3.1 : The VQuery Venn diagram visualization for Boolean query specification

- According to the study, a straightforward Venn diagram representation surpassed a Boolean query syntax in terms of speed and accuracy.
- But this format's restrictions on the expression's complexity are a drawback. For instance, in a typical Venn diagram, only three question terms can be ANDed together.
- This issue has been addressed by innovations, such as the VQuery system (see Fig. 6.3.1). A direct manipulation interface in VQuery lets users give ovals any number of search terms.
- An AND is suggested between those terms if two or more ovals are positioned so that they overlap, and if the user chooses the location of their junction.
- An OR is indicated between the respective phrases if the user selects anywhere inside the ovals but outside the area of intersection.
- Any term whose oval shows in the display's active area but is still selected is connected to a NOT operation.
- A conjunction is made up of all groupings of ovals that are contained within an active region, which represents the current query. You can move search phrase ovals out of the active area for future use.
- The user picks attributes from a second list of attribute types that is displayed across the top of the screen after being presented a scrollable list of attribute types on the left side of the screen.
- A list box containing the values for each attribute is presented in the screen's main area after clicking on its name.
- The user then decides which attribute values to allow the flow through. Sequencing two or more of these qualities produces a conjunct meaning across the chosen values.
- A disjunct has the semantics of a pair or more of these when they are arranged in parallel.
- The width of the "water" flowing from one attribute to the next indicates the number of pages that match the query at each location. (See Fig. 6.3.2).
- The amount of flow can be decreased by a conjunct. The far right side of the screen displays the items that match the entire query.

- According to a user research, the filter flow paradigm results in fewer errors than a typical SQL database query.

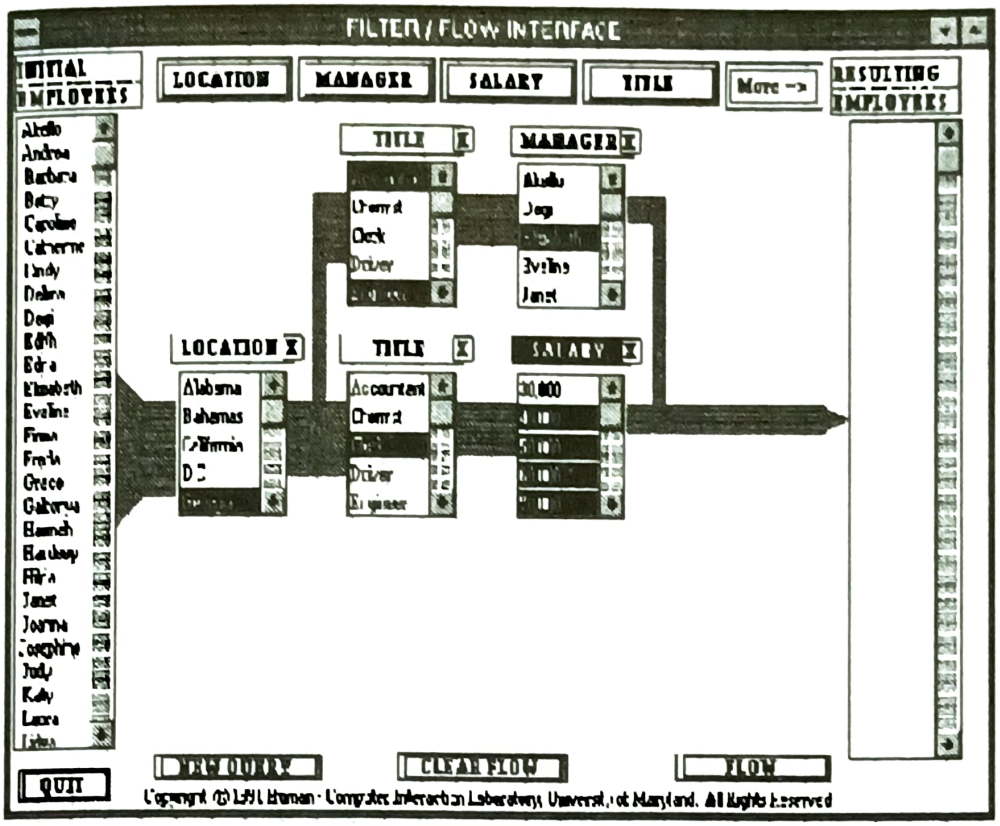


Fig. 6.3.2 : The filter flow visualization for Booleñn query specification

- Magic lenses are a last illustration of a graphical way to query specification.
- The study has recommended extending the use of this visualization tool to Boolean query formulation.
- Lists or icons are used to represent information in a 2D environment. Filters are used on the document set by lenses. In Fig. 6.3.3, For instance, a translucent Iens can be related to a word.
- This lens can make all of the documents in a set of documents disappear when it is put over an icon of the set of documents.
- If a second lens representing a different word is placed on top of the first, the two lenses work together to conceal any documents that don't contain both words in the collection of documents.
- Additional information, such as a minimum threshold for the term's frequency in the texts or an on/off switch for word stemming, can be changed dynamically. For instance, Fig. 6.3.3 depicts a disjunctive

search that looks for places with high yearly wages or cheap housing costs.

- One lens "calls out" and labels a group of cities in southern California.

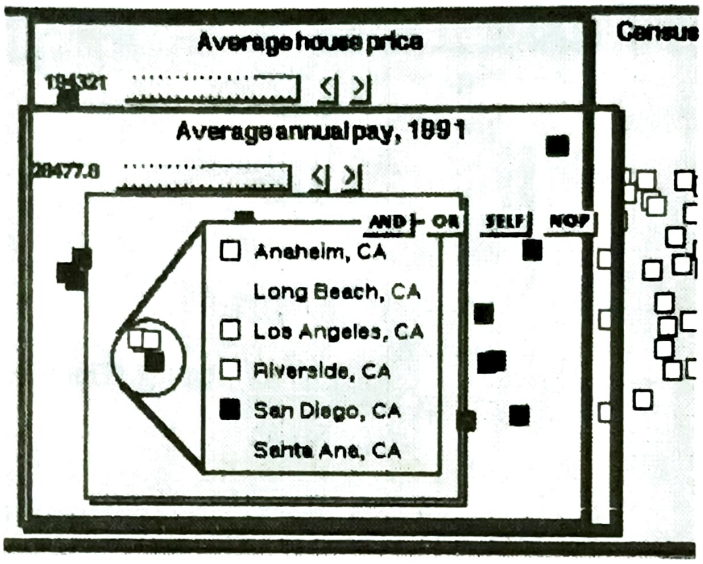


Fig. 6.3.3 : A magic lens interface for query specification

Phrases and Proximity

- In general, proximity information can significantly increase search accuracy.
- On the Internet, the distinction between a single-word query and a two-word exact phrase match might be the difference between a long list of mostly relevant items and an overwhelming mass of documents that are returned.
- There have been a lot of different ways to specify phrases developed. The proximity range must be given with an infix operator in accordance with LEXIS-NEXIS syntax.
- White w/3 house, for instance, denotes "white within 3 words of home, regardless of order." Simply placing two words side by side, separated by a space, allows you to specify the precise closeness of phrases.
- The placement of the terms within quotation marks is a common technique utilized by web search engines. In order to specify facets, Shneiderman advises offering a list of entry labels.
- The distinction is that each line's terms are considered as a phrase rather than a disjunction. The idea is to use this to direct people toward more accurate query specification.

- These approaches have the drawback of requiring an exact match of phrases, even though it is frequently the case that there are one or two words in between the keywords of interest.

Natural Language and Free Text Queries

- The benefit of using statistical ranking algorithms is that users can specify queries naturally without having to consider Boolean or other operators.
- However, they have the problem of offering the user less control and input regarding the outcomes.
- The output of a statistical ranking often consists of a list of documents with a score, likelihood, or percentage next to each title.
- Users receive limited information regarding the ranking of the content and the functions of the search phrases. If the user is particularly interested in one of the query phrases being present, this may be problematic.
- The declaration of "mandatory" phrases within the natural language query is one search approach that can help with this specific issue with statistical ranking algorithms.
- Instead than depending on the ranking algorithm to accurately weight the query phrases, this enables the user control which terms are thought to be essential.
- However, in order to include a mandatory specification, the user must be aware of a specific command and how it operates.
- You can try to respond to a question by using its natural language grammar.
- Since the necessary information is embedded within the text of documents rather than being described by the database design, information access question answering differs from that of database management systems.
- The Murax system can recognize if a user is asking for a person, place, or date based on the grammar of the enquiry.
- The next step is to search encyclopedia articles for sentences that contain the noun phrases that appear in the query, as these sentences are likely to have the solution.
- The FAQ finding system, which matches question-style queries against question-answer pairings on various themes, is another kind of automated question answering.

- After locating the most plausible FAQ files for the given question using a typical IR search, the system compares the phrases in the question to the question section of the question-answer pairings.

6.4 CONTEXT

GQ. Discuss the Query Term Hits Within Document Content?

GQ. Write a note on Query Term Hits between Document Content?

- In order to make the current document set more intelligible, this section describes interface strategies for integrating the document set into the content of other information kinds.
- This includes demonstrating how the document set relates to other documents in the set as well as to query words, collection overviews, descriptive metadata, hyperlink structure, and document structure.

Document Surrogates

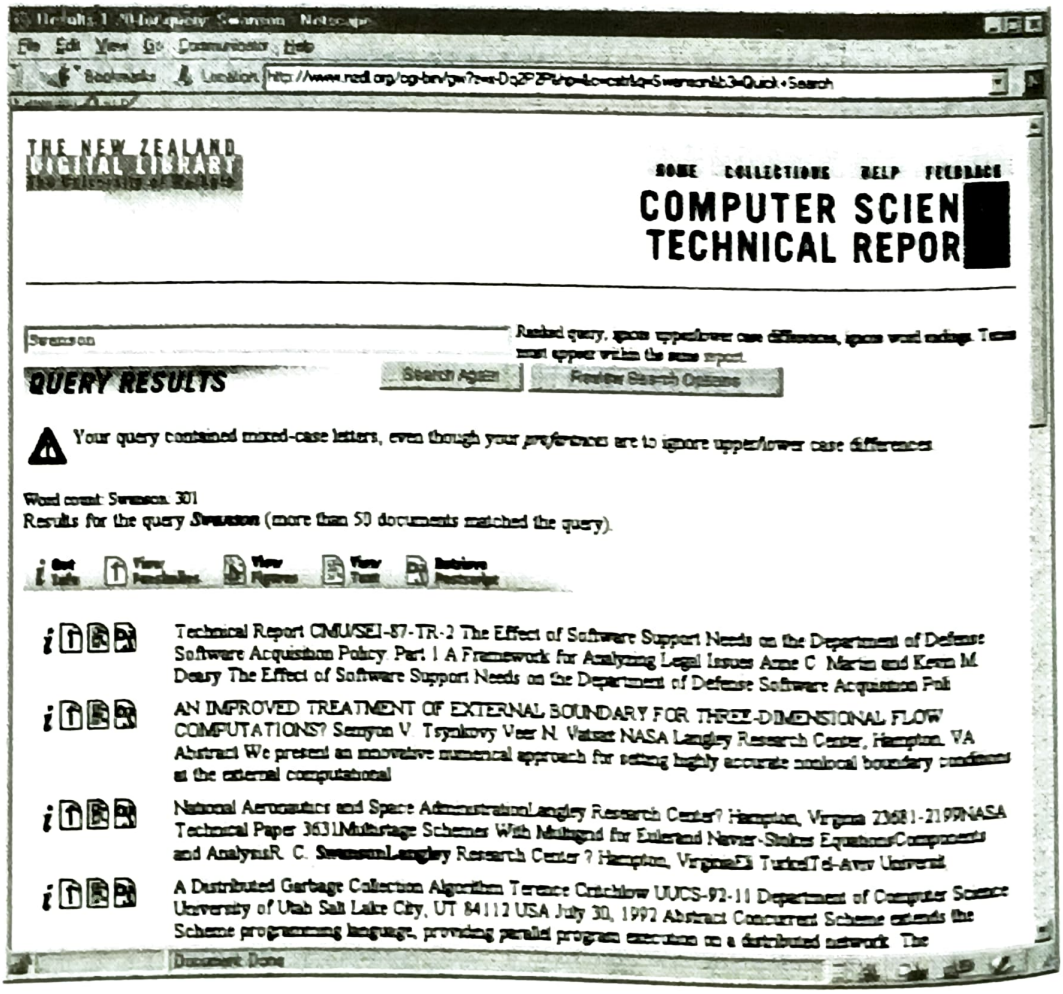


Fig. 6.4.1 : An example of a ranked list of titles and other document surrogate information

- The most typical method of displaying results for a query is to list document details in the order of their computed relevance to the query.
- As an alternative, documents are ranked according to a metadata attribute, such date, for pure Boolean ranking.
- The document title and a selection of significant metadata, such the date, the author, and the length of the piece, are typically included in the document list.
- In systems that use statistical ranking, the title is frequently accompanied by a numerical score or percentage that represents the degree of similarity or likelihood of relevance as determined by the system.
- A document surrogate is another name for this kind of information. Look at Fig. 6.4.1.
- Users can choose between a brief and a detailed look in some systems. An abstract or summary often appears in the detailed view.
- In most user interfaces, selecting a document's title or an icon next to its name will open a view of the actual document, either in a new window on the screen or in place of the search results list.

Query Term Hits Within Document Content

- It is frequently helpful to highlight the instances of the phrases or descriptors that match those of the user's query in systems where the user can read the full text of a retrieved document.
- Additionally, it may be helpful for the system to scroll the document view to the first passage that includes one or more of the search terms, then highlight the matched terms with a different colour or reverse video.
- The goal of this display is to help the user focus on the areas of the document that are most likely to be pertinent to their query. It has consistently been discovered that highlighting of search phrases is a valuable feature for information access interfaces.
- Recently, it has been discovered that colour highlighting is helpful for skimming lists of bibliographic records.

KWIC

- The Keyword-in-content (KWIC) document surrogate is a feature associated with highlighting. Sentence fragments, complete sentences, or collections of sentences containing search phrases are taken from the entire text and shown for viewing with other types of Surrogate data (such as document title and abstract).

- A KWIC listing differs from an abstract, it should be noted. Although it might not include references to the terms in the query, an abstract summarizes the key points of the content.
- Sentences that succinctly describe how the search phrases are used in the document are displayed in a KWIC extract.
- This display can include both the subsets of query terms that are included in the retrieved documents as well as their relationships to one another.
- There must be trade-offs made between the number of text lines to display and the lines themselves.
- The best contexts to show are those that appear close to the beginning of the page and that contain the biggest subset of query terms, according to results from text summarization research, although it is unknown which contexts are best chosen for viewing.
- Because the system needs access to a copy of the original content in order to extract the sentences containing the search terms, the KWIC function is typically not displayed in Web search result displays.
- Typically, only the index sans term location data is saved by web search engines. Systems that index certain websites are able to display KWIC data in the document list display.

TileBars

- Through the TileBars interface, a more condensed version of the query term hit display is made available. One topic is entered per line as the user types a query in a faceted format.
- A graphic bar is shown next to each document's title after the system has retrieved the documents to indicate the degree of match for each facet.
- Thus, TileBars quickly show which passages in each article contain which themes as well as how often each topic is brought up. The image of each document is a rectangular bar.
- A case is shown in Fig. 6.4.2. The facets of the query are represented by rows within the bar. Each TileBar has a top row for "osteoporosis," a second row for "prevention," and a third row for "research."
- The bar is further divided into columns, each of which points to a different portion of the document. Overlapped hits inside a single section are more likely to point to a relevant document than scattered hits across the page.

- The patterns are intended to show if a facet's terms are used frequently throughout the document as a main issue, occasionally as a subtopic, or only in passing.
- The degree of darkness in each square indicates how many times the search term appears in that passage of text; the darker the square, the more hits. No hits on the search keyword are indicated by white. The user can rapidly determine whether a subset of the phrases appear in the same section of the page as a result.
- The first document's middle section, but neither the beginning or the finish, can be shown to contain considerable overlap. Thus, it is likely to cover subjects other than osteoporosis research.
- The user will likely be interested in the second through fourth documents as well, even though they are much shorter and contain overlap with all terms of interest.
- The next three publications are all lengthy, and based on the TileBars, we may infer that they talk about research and prevention but barely mention osteoporosis. As a result, they are most likely of no interest.

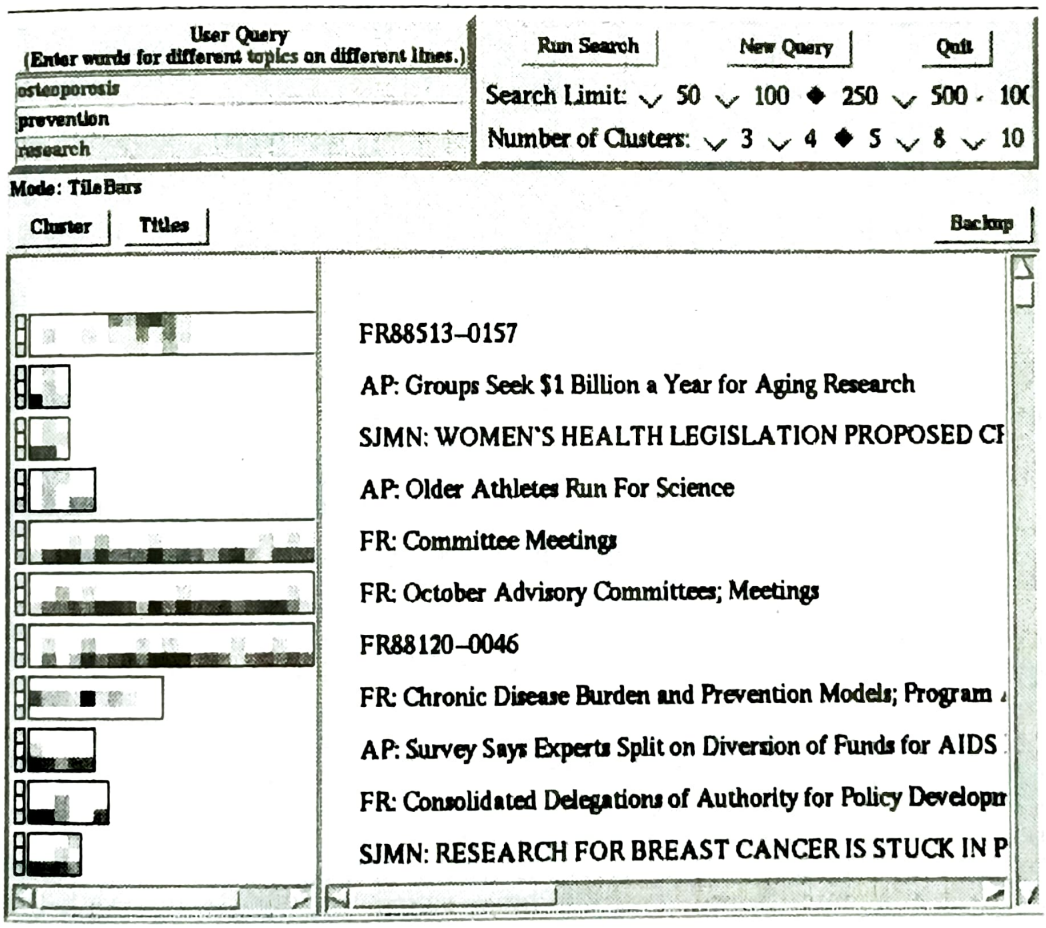


Fig. 6.4.2 : An example of the TileBars retrieval results visualization

SeeSoft

- The SeeSoft visualization portrays text as one "line" of text on each horizontal line of the strip, imitating newspaper text columns.
- In Fig. 6.4.3 the illustration is neat and appealing in appearance. A summary illustrating the volume and shape of the text is provided through graphics, which abstract away the specifics.
- Color highlighting is used to draw attention to a variety of details, such as the location of a certain word within the text.
- A pop-up window allows users to study the details of a smaller area of the display, whereas the overview displays more text but with less detail.

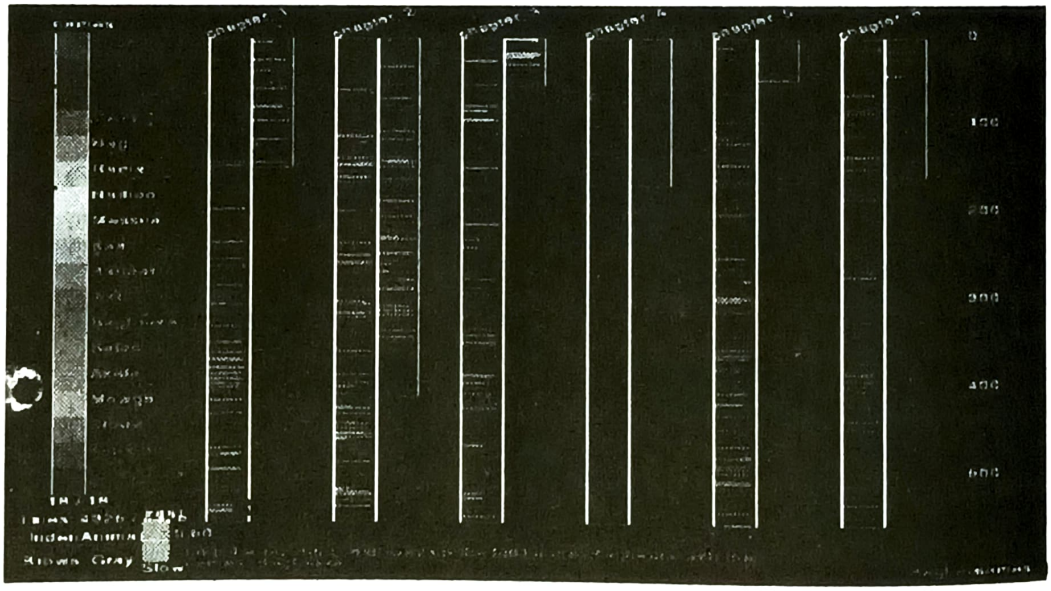


Fig 6.4.3 : An example of the SeeSoft visualization for showing locations of characters within a text

- A line of text is a significant unit of information when developing software, which is how SeeSoft was first intended to be used.
- As a result, SeeSoft displays information pertinent to the programming area, such as which "line of code" was edited by which programmer, how frequently specific lines have been modified, and the number of days since the lines were last modified.

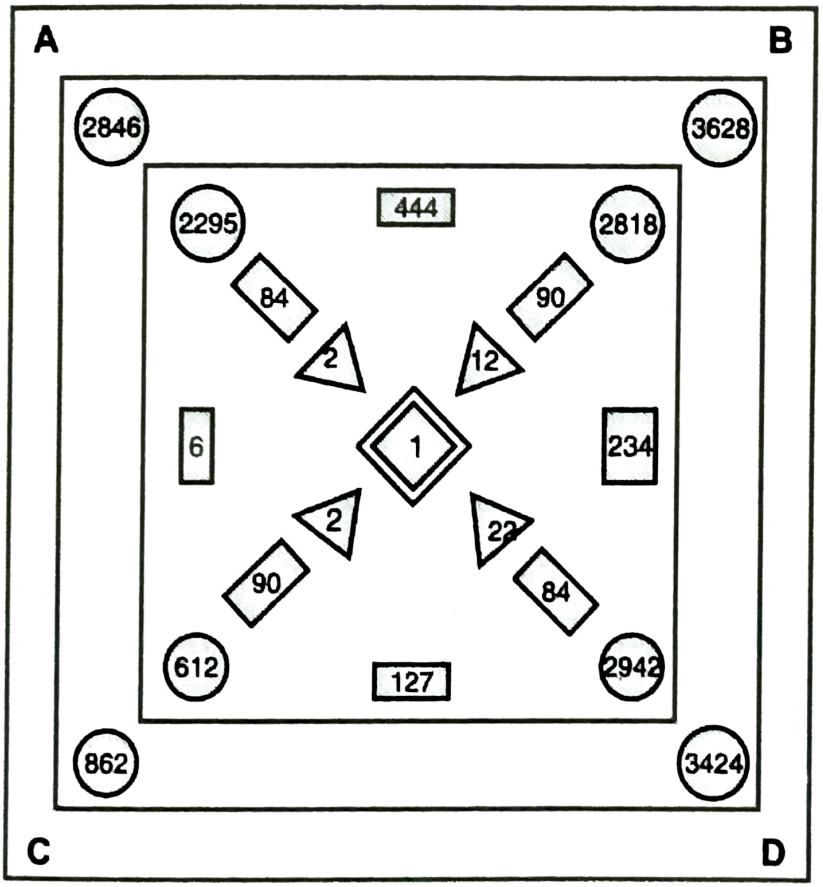
Query Term Hits between documents

- To display a different type of information regarding the relationship between search terms and retrieved documents, other visualization concepts have been developed.

- These systems give an overview or summary of the retrieved documents according to which subset of query terms they contain, unlike KWIC interfaces and TileBars, which illustrate how query phrases appear within individual documents.
- Variations on this concept are described in the following subsections.

InfoCrystal

- The InfoCrystal displays the number of documents that contain a given subset of search criteria. By doing this, the user is spared from having to define Boolean ANDs and ORs in their query while still being able to see which term combinations are truly present in papers that were ranked statistically.
- The InfoCrystal enables visualization of every relationship between N user-specified "concepts" conceivable. The InfoCrystal shows the number of papers that contain each conceivable subset of the N concepts, in a brilliant expansion of the Venn diagram model.
- Fig. 6.4.4 is an outline of what the InfoCrystal might show in response to a search against four keywords or Boolean expressions, denoted by the letters A, B, C, and D.

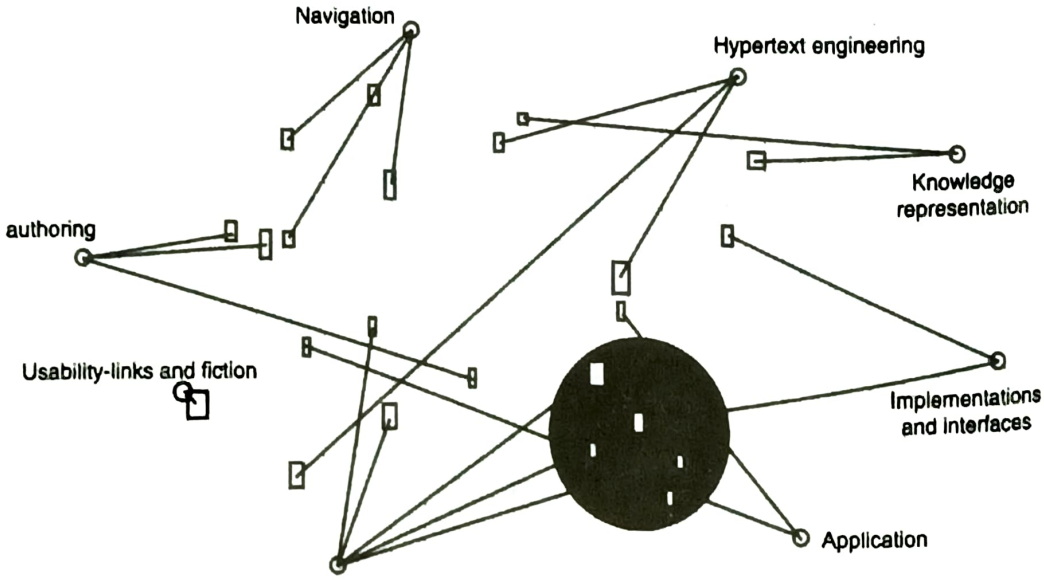


(1F5)**Fig 6.4.4 : A sketch of the InfoCrystal retrieval results display**

- One document was found that had all four keywords, as indicated by the diamond in the center. The triangle marked "12" denotes that 12 documents containing the characteristics A, B, and D, and so forth, were discovered.

VIBE and Lyberworld

- Queries are displayed in an abstract graphical space in VIBE and Lyberworld displays.
- After the search, icons are made to show how many documents each subset of the search phrases was found in. The location of the icon denotes the subset status of each group of documents.
- For instance, in VIBE, a group of documents that contain three of the five query terms are displayed at a place positioned halfway between the representations of the three query terms in question on an axis connecting these three terms.
- In Fig. 6.4.5 This concept is presented in 3D on Lyberworld.



(1F6)**Fig. 6.4.5 : An example of the VIBE retrieval results display**

Lattices

- For the purpose of formulating queries, many academics have used a graphical representation of a mathematical lattice, where the query consists of a set of restrictions on a hierarchy of categories.
- A document containing the terms A, B, C, and D might be positioned at a point in the lattice with these four categories as parents. This is one approach to the issue of showing documents in terms of numerous attributes.

- The lattice arrangement would typically be too complicated to allow for reading if such a representation were used for retrieval results rather than query formulation.

SuperBook : Context via Table of Contents

- To present query word hits in context, the SuperBook system uses the structure of a huge page. On the left side of the display, a table of contents (TOC) for a book or manual is shown in a hierarchy, and on the right, the full text of a page or section is displayed.
- The user can enlarge or reduce the display of the sections and subsections by adjusting the table of contents. The currently seen portions are compressed and their viewing areas are enlarged using a focus-plus-context method.
- The display changes dynamically when the user moves the cursor to a different area of the TOC, enlarging the new focus and contracting the previously viewed sections.
- The search results are displayed in the context of the hierarchy of the table of contents once the user sets a query for the book.
- In Fig. 6.4.6 the portions containing search results are expanded, while the others are compressed.

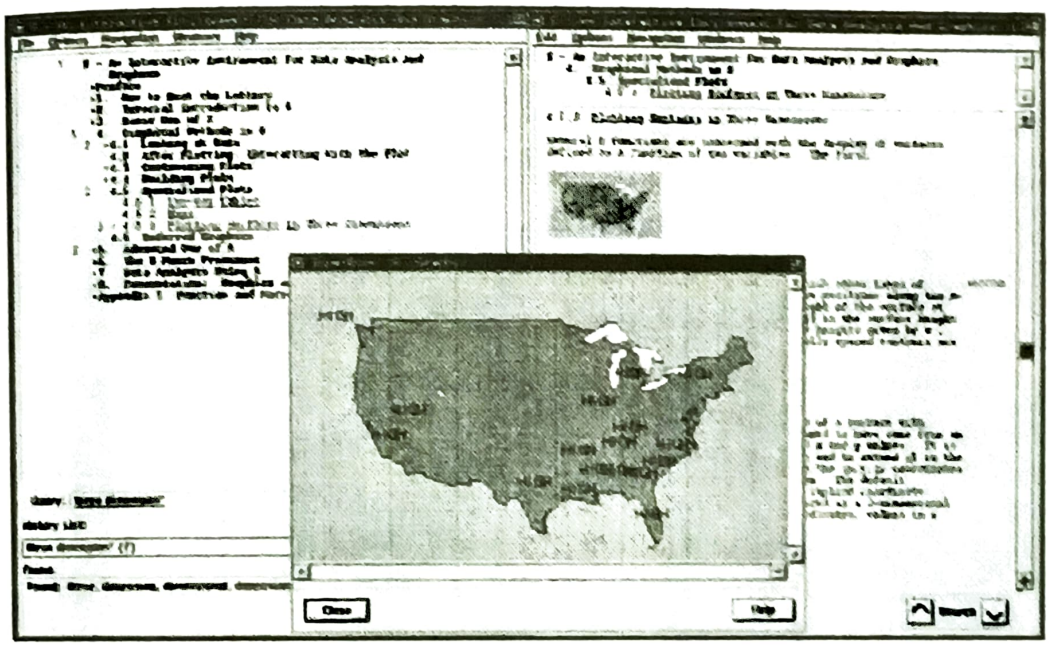


Fig 6.4.6 : The SuperBook interface for showing retrieval results on a large manual in context

- In reverse video, the search terms that appear in chapter or section names are highlighted.
- The page itself is presented on the right side of the screen when a user

selects a page from the table of contents view, and the page's search terms are highlighted in reverse video.

- In comparison to a regular system, SuperBook subjects were either faster and more accurate or both.
- The investigators thoroughly reviewed the records and made plausible hypotheses when disparities between SuperBook and the standard system appeared.
- Following the initial trials, they changed SuperBook in accordance with these theories, and the results were typically better.

Categories for Results Set Context

- The context of a query's results can also be provided by category metadata. For instance, SuperBook's original design allowed for navigating within a computer manual, a highly structured document.
- The central concept was expanded upon by the CORE project to encompass approximately 1000 full-text chemical papers.
- A study of this representation showed that it performed a number of tasks better than a typical search system. Since this collection lacks a table of contents, context is created by grouping papers into a hierarchy of categories that include chemistry-related terminology.
- A hierarchical view of the collection is provided by the categories themselves being structured into a hierarchy, which lists the documents associated to each category when that category is chosen for further in-depth viewing.

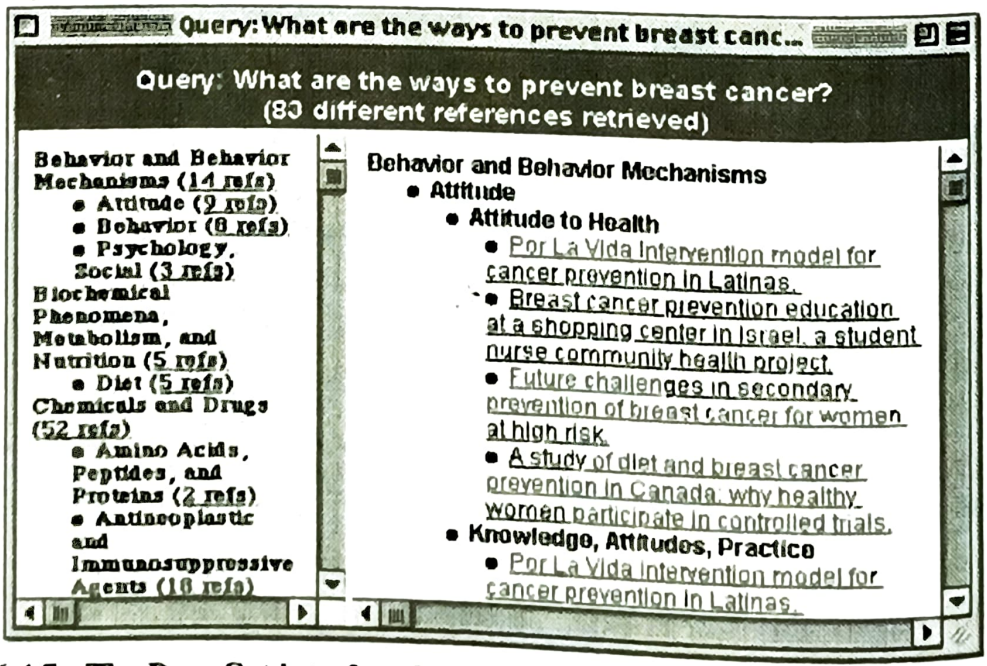


Fig 6.4.7 : The DynaCat interface for viewing category labels that correspond to query types

- Results for a search on breast cancer prevention are displayed in Fig. 6.4.7. Three windows make up the interface.
- The user's search term and the quantity of papers found are displayed in the top window.
- The left window provides a table of contents view of how the search results are organized by displaying the categories in the top two levels of the hierarchy.
- The titles of all the papers that fall under each category are shown in the right pane along with the hierarchy's categories.

Using Hyperlinks to Organize Retrieval Results

- Although the authors of the SuperBook refer to it as a hypertext system, it is actually more accurately described as a way of presenting search results inside the framework of a structure that users can comprehend and examine all at once.
- Although the hypertext element was not examined independently to determine its significance, the authors typically do not refer to it when discussing the design's strengths.
- In reality, it appears to be to blame for one of the primary issues with the system's updated version, which is that users have a propensity to leave the pages they are reading, sometimes extending SuperBook's time spent on a subject.
- SuperBook uses a non-standard type of hypertext, in which any word is instantly linked to instances of the same word in other areas of the document, which may contribute to this wandering.
- Contrary to expectations, this is not how hypertext linkages are actually made in actuality.
- Today, hyperlink connections are used much more selectively in hyperlinked help systems and online content.
- These links frequently have labels that are somewhat intelligible when viewed in the context of their surroundings.
- Back-of-the-book indexes frequently only identify the most significant uses of a word or the first in a sequence of usage rather than all of its occurrences. Perhaps the same ideas should guide automated hypertext connecting.
- Additionally, at least one study revealed that users built more accurate mental models of a limited, hierarchically arranged hypertext system than they did of one that permitted more flexible access.

Cha-Cha: SuperBook on the Web

- The Cha-Cha intranet search system expands the SuperBook concept to a sizable heterogeneous Web site, like that which might be found in an organization's intranet.
- Illustration provided in Fig. 6.4.8. Several factors set this system apart from SuperBook.

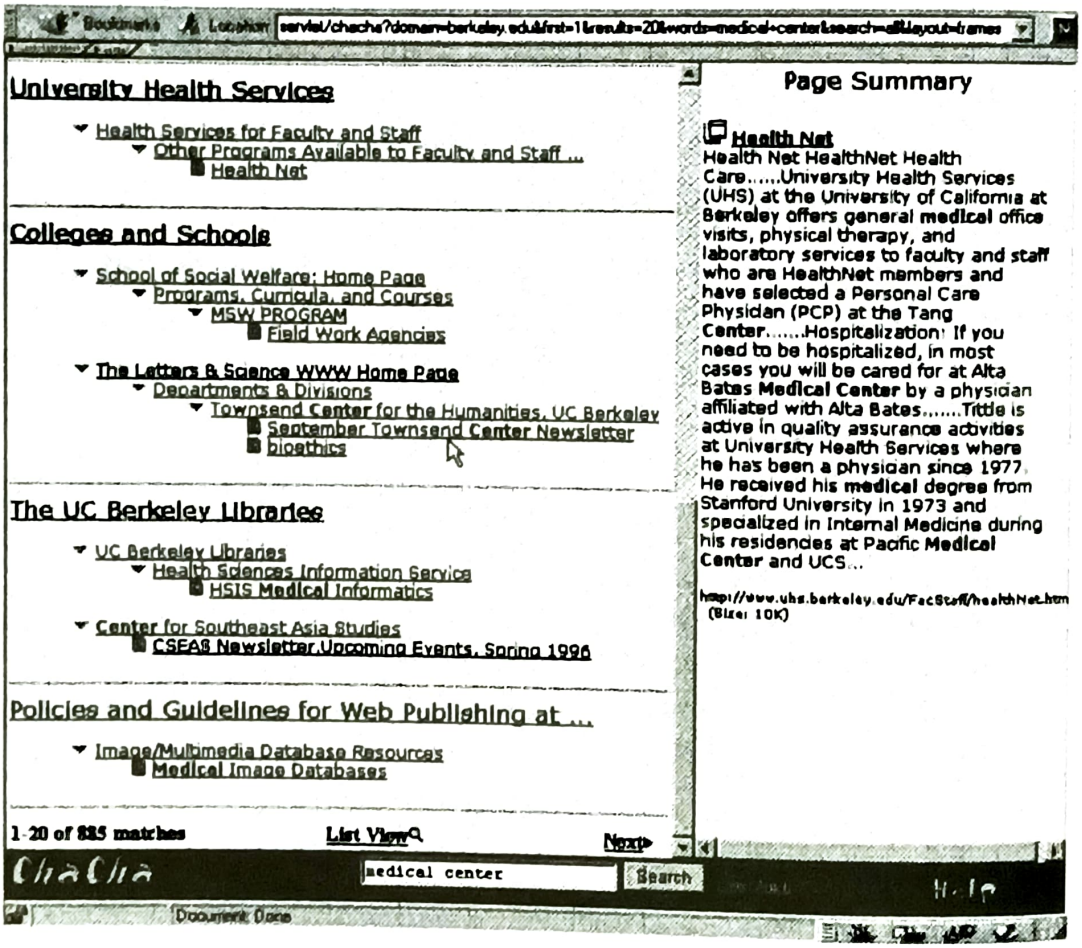


Fig. 6.4.8 : The Cha-Cha interface for showing Web intranet search results in context displaying results on the query 'medical center'

- The majority of websites lack a true table of contents or category structure, and huge universities' or enterprises' intranets aren't typically set up according to a single central hub.
- Cha-Cha overlays what is meant to be a meaningful organization on top of the underlying chaos using the link structure already available on the website.
- If 'medical' appears as a term in a document within the health center section of the Web, the home page of this center will be presented together with the more specific hits if a user does not know to use the term 'health center' but instead queries on 'medical center.'

Mapuccino : Graphical Depiction of Link Structure

- The Mapuccino system enables users to request information from a specific website.
- The algorithm continuously searches the website, evaluating each page it encounters for relevance to the search.
- The weights of a page's outlinks are raised when a relevant page is discovered. As a result, a portion of the search is based on the presumption that relevant pages would be located close to one another on the website.
- In a nodes-and-links view, the portion of the website that has been crawled is graphically shown (see Fig. 6.4.9). Instead of telling the user what the contents of the pages are, this type of presentation only reveals their connection structure.

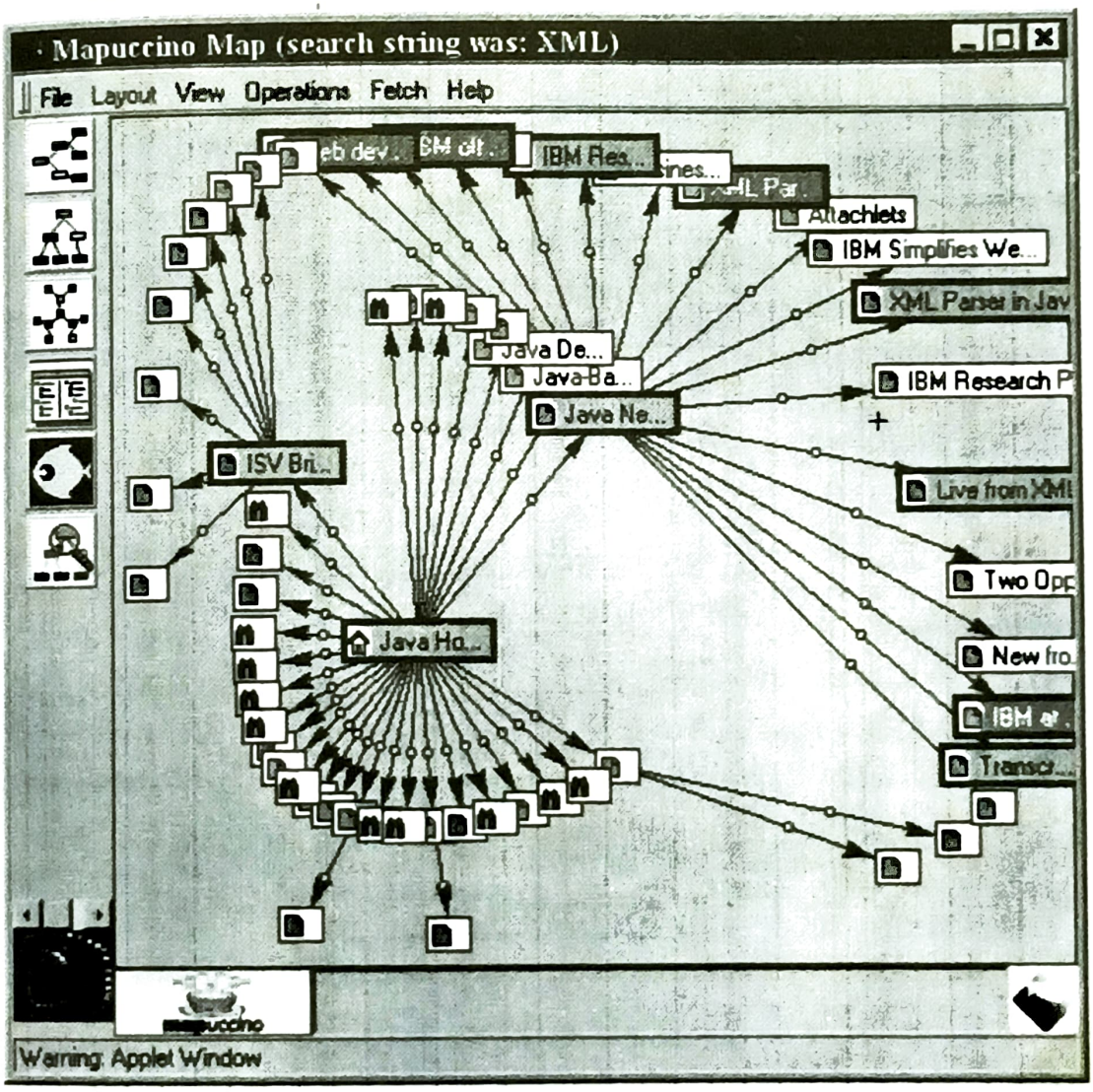
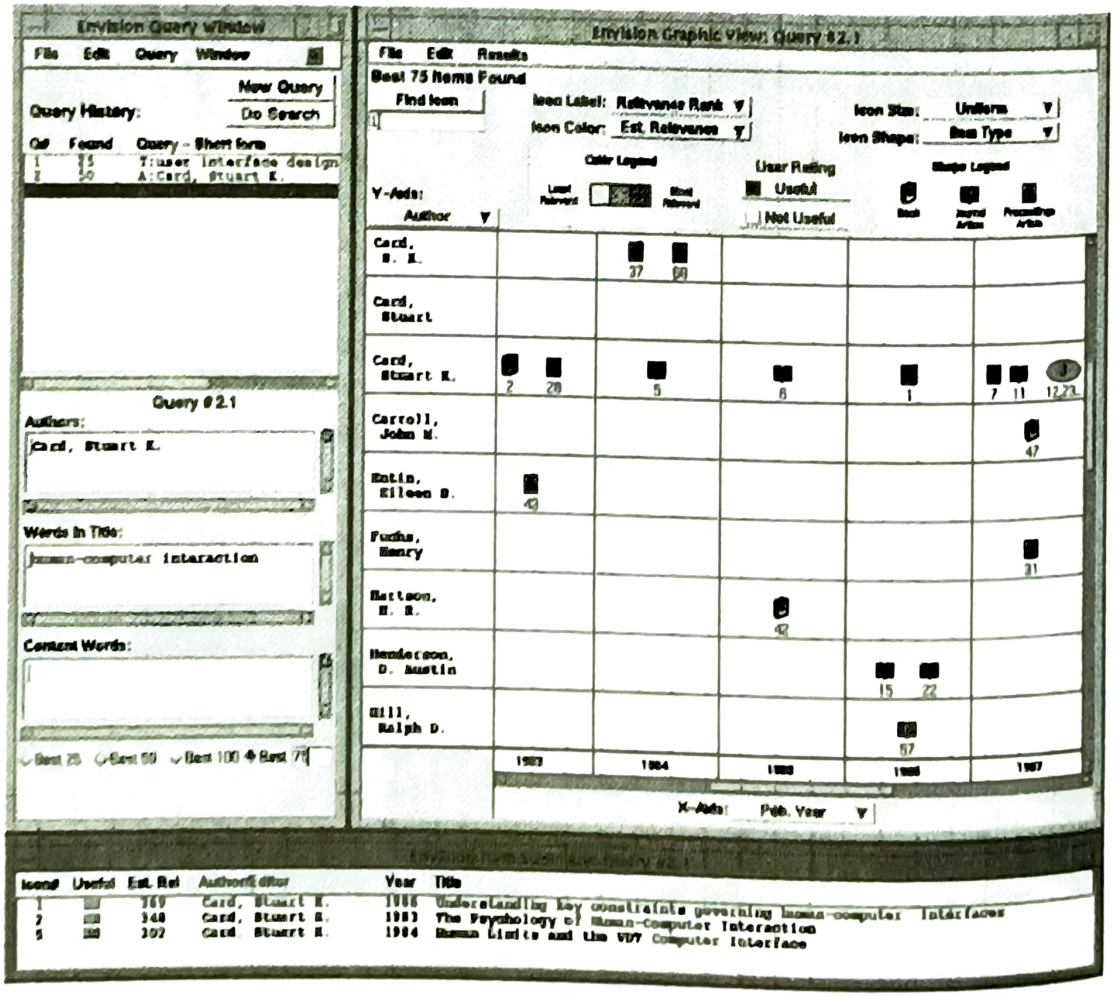


Fig. 6.4.9 : Example of a Web subset visualized by Mapuccino

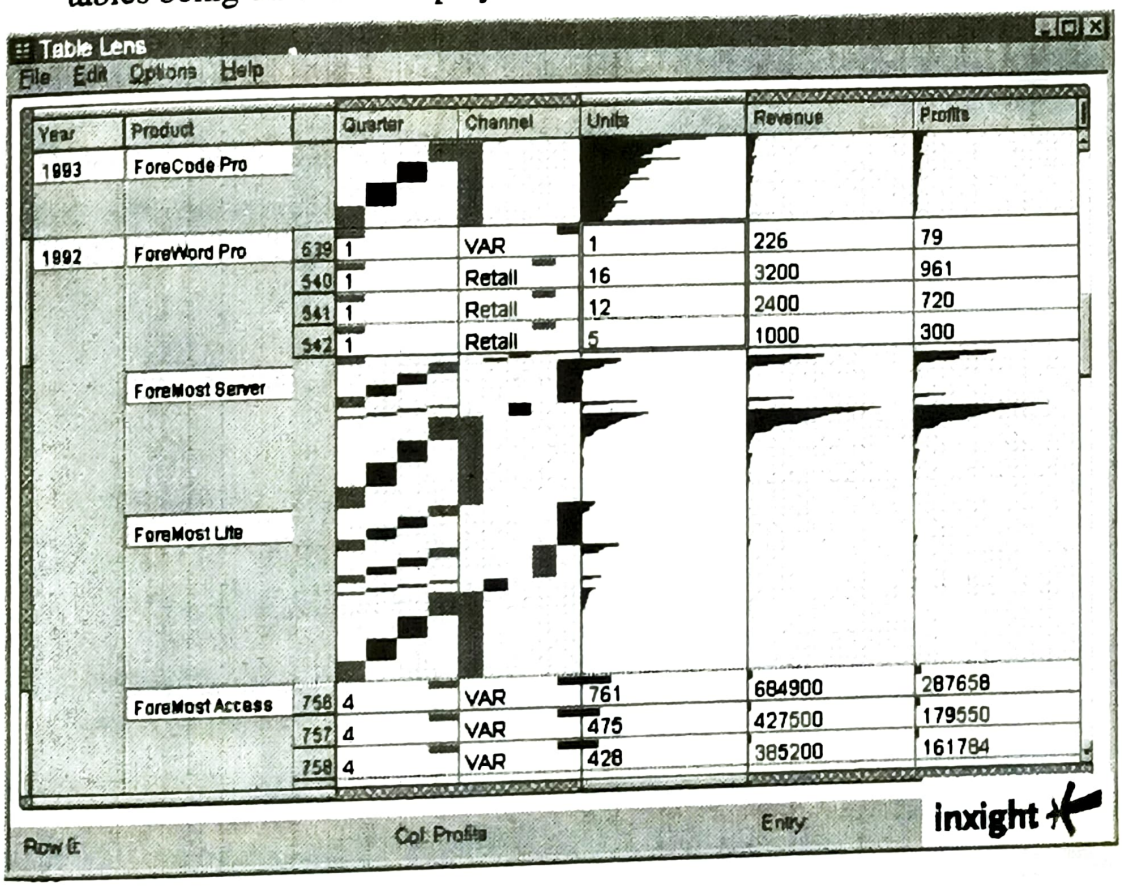
Tables

- Another method for displaying the connections between retrieved documents is using tabular displays.
- The Envision system lets users arrange results along X and Y axes according to metadata like author or date, and employs graphics to display values for properties related to retrieved documents inside each cell (see Fig. 6.4.10).
- An iconographic representation of a document's colour, shape, and size are used to display the document's type, computed relevance, and other characteristics.
- When you click on an icon, a new window with further details about the document appears.



(1F7)**Fig 6.4.10 : The Envision tabular display for graphically organizing retrieved documents**

- An innovative interface for viewing and dynamically rearranging very big informational tables is called the Table Lens (see Fig. 6.4.11).
- It uses focus-plus-context to cram hundreds of rows of data into a place where conventional spreadsheets could only hold a maximum of two dozen rows.
- Users can quickly switch from a view centered around one type of metadata to another since it enables quick reorganization via column sorting.
- The relative ranks of various articles by the same author, for instance, can be displayed by first sorting documents by rank and then by author name.
- A resort by date might reveal trends in the relevance ratings in relation to the publication.
- This quick resorting capability aids in getting over issues related to tables being unable to display numerous simultaneous intersections.



(1F8)**Fig 6.4.11 : The Table Lens visualization**

6.5 USING RELEVANCE JUDGEMENTS

GQ. Write the challenges of relevance feedback

GQ. Describe the difference between relevance feedback and query expansion in terms of user interaction?

- Query reformulation is a crucial step in the information access process, and reference feedback is a tried-and-true method for query reformulation.
- In its original form, relevance feedback referred to a cycle of user-system interaction in which the user chose a select group of pages that were pertinent to the query, and the system subsequently used features obtained from these chosen documents to revise the original question.
- A fresh set of documents is then returned after this updated query has been completed. Although they are likely to appear in a different rank order, documents from the initial batch can nevertheless appear in the new results list.
- Relevance feedback introduces crucial design decisions, such as whether actions should be initiated and controlled by the user against those that should be carried out automatically by the system.
- The researcher goes into great length about this matter and claims that, despite modern systems' emphasis on trying to automate the entire process, a middle approach in which the system assists in automating search at a strategic level is preferable.
- The researcher proposes comparing an automatic camera to one with lens- and shutter-speed adjustments.

Interfaces for Standard Relevance Feedback

- A list of titles with check boxes next to them that allow the user to mark items as relevant makes up the basic relevance feedback interface.
- Depending on the system, this could mean that unmarked documents are irrelevant or that no judgment has been made regarding unmarked documents. Another approach is to offer a selection of checkboxes that say if something is relevant or not.
- Users may be permitted to select a value on a relevance scale in particular circumstances. conventional relevance feedback algorithms

Negative relevance judgment evidence typically does not result in higher performance, but machine learning algorithms can benefit from negative feedback..

- The system can either automatically reweight the question and re-run the search once the user has made a set of relevance judgments and sent a search command, or it can produce a list of terms the user can choose from in order to supplement the original query. Refer to Figure 6.19. Systems typically don't offer suggestions for terms to omit from a query.
- A fresh list of titles appears upon a new execution of the query. Keeping a marker, such as a checked box next to the papers that the user has already evaluated, can be useful.

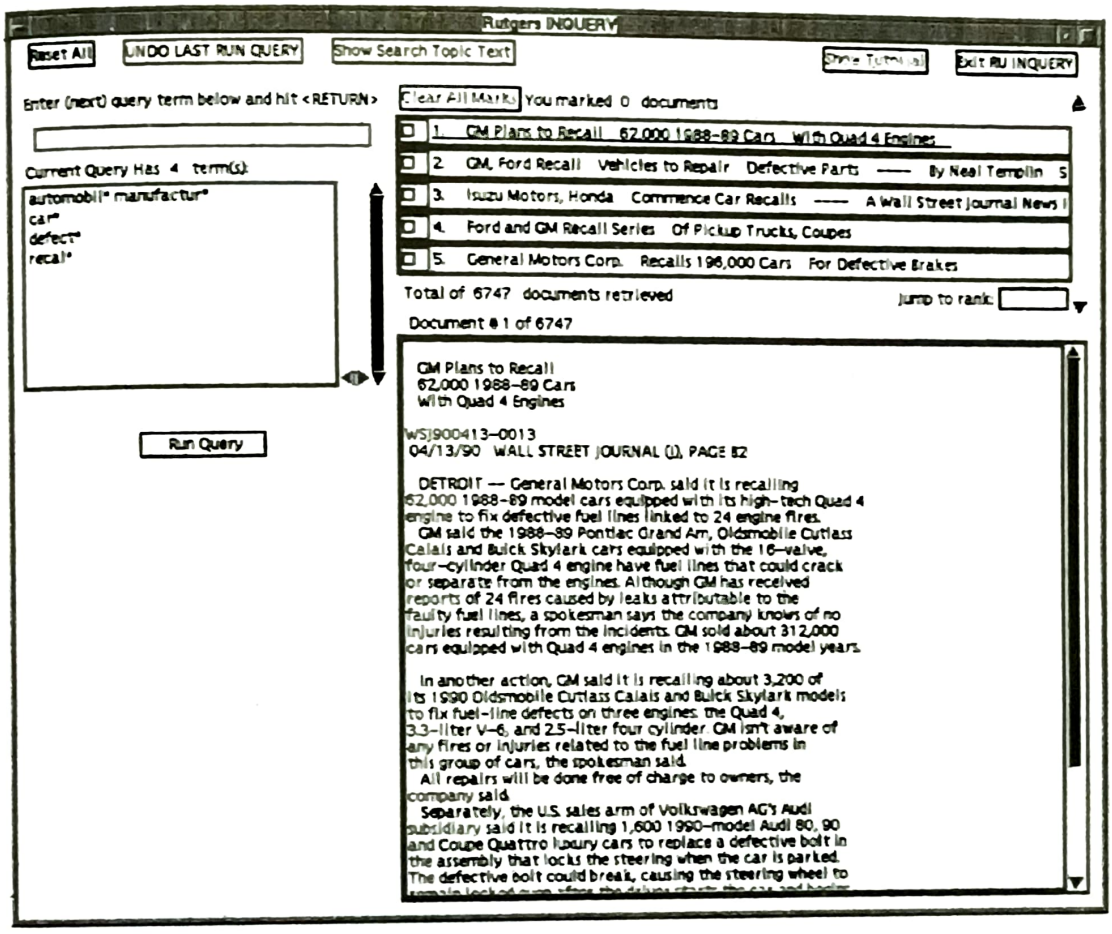


Fig. 6.5.1 : An example of an interface for relevance feedback

Studies of User Interaction with Relevance Feedback Systems

- Standard relevance feedback presupposes user involvement through the specification of pertinent documents.
- Users may also choose which terms to include to the query in some user interfaces.

- However, because most ranking and reweighting algorithms are challenging to comprehend or forecast, it's possible that consumers would find it challenging to directly regulate a relevance feedback system.
- A recent study was carried out to directly examine the extent to which user management of the feedback process is advantageous.
- The advantages of letting consumers peek "under the hood" during relevant feedback were quantified by the researchers.
- They put the Inquery system to the test in four cases :
  (1) **Control :** The subjects could only manually reformulate the question because there was no relevance feedback.
  (2) **Opaque :** The subjects simply chose relevant documents, and the rankings were updated.
  (3) **Transparent :** The system's reformulated queries and updated rankings were shown to the subjects.
  (4) **Penetrable :** In the middle of the reranking process, the system is paused. The terms that the system would have used for opaque and transparent query reformulation are displayed to the subjects. After then, the subjects decide which, if any, of the new terms to include in the search. The updated rankings are then shown by the system.

Fetching Relevant Information in the Background

- The objective of a standard relevance feedback is to enhance an ad hoc inquiry or create a profile for a routing query.
- Recently, academics have started creating systems that track users' behavior and progress across protracted periods of contact in an effort to foretell which documents or actions the user is likely to seek in the future.
- These programmes, often known as semi-automated assistants or recommender "agents," frequently employ machine learning strategies.
- Some of these systems ask the user for explicit input in the form of a goal statement or relevance assessments, while others covertly monitor user behavior and attempt to draw conclusions from it.
- Researchers developed a method that predicts how users would manage email messages and organize meetings in calendar manager software.

- The software "looks over the shoulders" of the users, logging each action into a database that is relevant.
- The system predicts a user's action based on the resemblance of the current scenario to situations previously encountered once sufficient data has been gathered.
- For instance, the system might offer to automatically save email messages from a specific sender into a specific file the next time one of their messages arrives if the user almost always does so. This system incorporates user feedback, both implicit and explicit.
- The system interprets a user's disregard for the system's recommendation as unfavorable feedback and adds the user's overriding action to the action database as a result.
- When a user makes a specific kind of inaccurate prediction, the system prompts them with questions that allow them to change the feature's weight. Lastly, the system can be explicitly trained by the user by giving it fictitious examples of input-action pairs.

Group Relevance Judgements

- Recently, there has been a lot of interest in leveraging relevance judgments from a wide variety of users to rank or score general interest information.
- Some forms of this social recommendation approach ignore the representation of the material being rated entirely, relying instead on similarities among relevance assessments made by persons with similar tastes.
- For rating information where taste plays a big part, like movie and music suggestions, this has been found to be very effective. More recent research has coupled content information with group relevance judgments.

Pseudo-Relevance Feedback

- What is referred to as pseudo-relevance feedback is at the extreme end of the system versus user feedback spectrum.
- In this strategy, the system simply assumes that its top-ranked articles are relevant and uses these documents to augment the query with a relevance feedback rating mechanism instead of relying on the user to select the top k relevant documents.

- In some circumstances, perhaps those where the initial query statement is detailed and lengthy, it has been observed that this method is very effective.
- Using the output of retrieval result clustering as the input to a relevance feedback mechanism, either by having the user or the system select the cluster to be used, is an intriguing expansion to this idea, but it has not yet been tested.

6.6 INTERFACE SUPPORT FOR THE SEARCH PROCESS

GQ. Describe the interface support for the search process?
GQ. What is the purpose of query interface in information retrieval?
GQ. What are the 3 main user interface types used to interact with a computer?

- The user interface designer must decide how to organize different types of information on the computer screen and how to organize potential interaction sequences. Particularly challenging is this design problem for a complicated activity like information access.
- In this section, we talk about the arrangement of information in complex information systems and use examples of existing interfaces to illustrate our points.
- We start out by talking about relatively basic search interfaces, like those used for string searches in "find" operations, before moving on to multi window interfaces and complex workspaces.
- The integration of scanning, choosing, and querying inside information access interfaces is then discussed. Finally, interface support for keeping track of the search history is included.

Interfaces for String Matching

- The 'find' action, which is often performed over the contents of a document that is being examined, is a frequent simple search need.
- Typically, this function only performs a straightforward string match with no ranking output or Boolean term combinations (without regular expression capabilities).
- Usually, a dedicated purpose search window with a few basic controls is

built (e.g., case-sensitivity, search forward or backward). The user enters the query string into a form, and the target document highlights any matches (see Fig. 6.6.1).

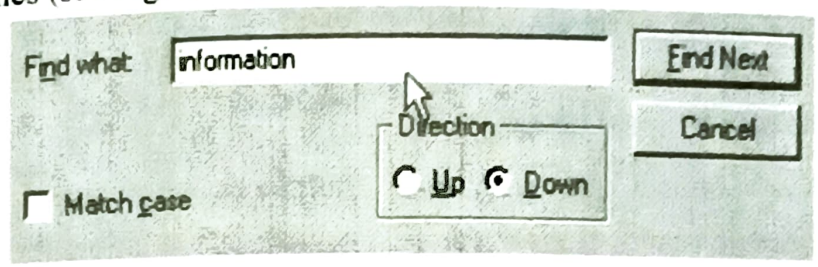


Fig. 6.6.1

Fig 6.6.1 An example of a simple interface for string matching, from Netscape Communicator.

- The 'find' function for searching across small collections, like the files on a computer's hard drive, or the history list of a web browser, is the next level of sophistication.
- Another common implementation of this kind of function is a straightforward string match. Once more, the controls and parameter settings are displayed at the top of a special purpose search window, where checkboxes and entry forms are used to specify the various options.
- As opposed to the prior illustration, this one shows a results list directly on the search interface (see Fig. 6.6.2).

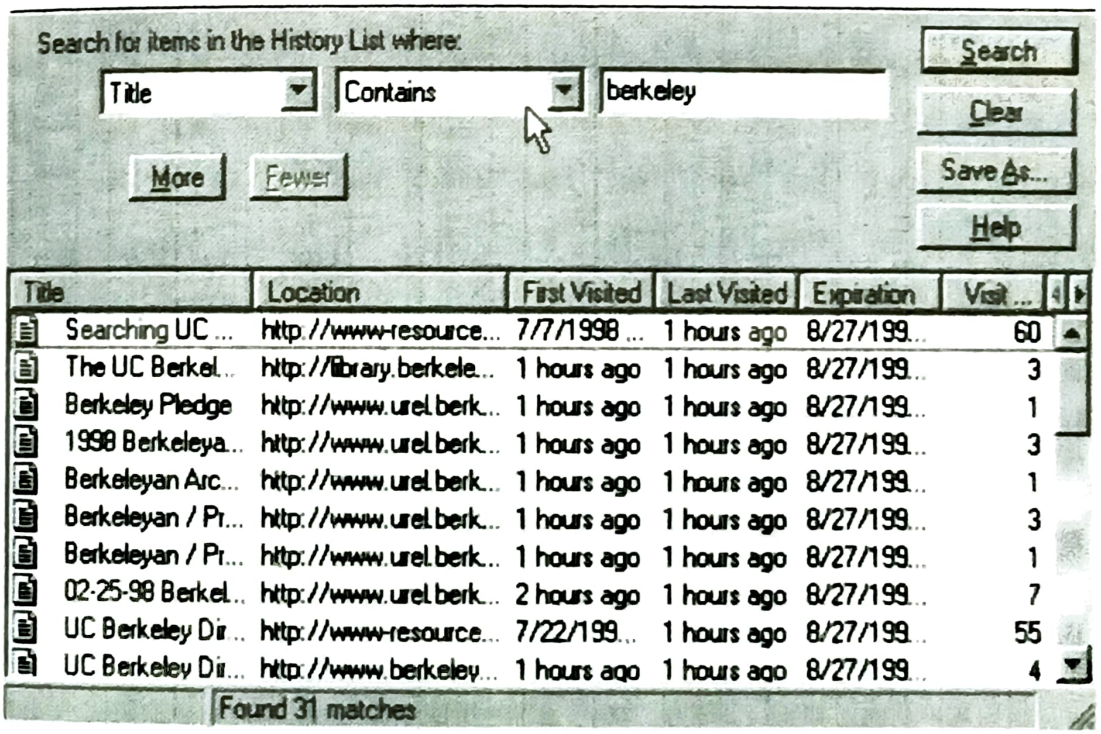


Fig 6.6.2 : An example of an string matching over a list.

Window Management

- The interface designer must choose how to arrange the numerous choices and information displays within the interface for search jobs that are more complicated than the straightforward string matching find operations mentioned above.
- The conventional command-line interfaces or menus used by bibliographic search engines are TTY-based. When the system replies to a command, the previous results screen's contents are erased, forcing the user to recall the previous one's information.
- The primary design decisions in such a system relate to the command or menu structure and the sequence in which the various alternatives are presented.
- The windowing concept can be utilized in contemporary graphical interfaces to split functionality into various, concurrently visible windows.
- Linking information from one window to another in information access systems is frequently helpful. In SuperBook, for instance, this is done by connecting documents to their positions in a table of contents.
- The option allows users to copy and paste data from one window into another.
- The designer must decide between a monolithic display, in which all the windows are laid out in predefined positions and are all simultaneously viewable, tiled windows, and overlapping windows, while arranging information within windows.
- User studies comparing these choices when used for various tasks have been done.
- There are currently no established standards for information access interfaces, and the outcomes of these research typically rely on the domain in which the interface is utilized.
- The inherent limitation on the number of information types that can be exhibited at once is a challenge for any information access interface. Information access systems must always provide space for a text display area.

- For the text to be readable, this area must occupy a sizable amount of the screen. For instance, a tool within a paint software can be made fairly small while still being recognisable and functional.
- It is challenging to compress many of the information displays required for an information access system in this way due to concerns about legibility.
- A good layout, graphics, and font design can help; for instance, spacing, typeface, and other minor details can have a huge impact on how Web search results appear.
- Although they offer versatility in design, overlapping windows can soon result in a crowded, disorderly display.
- Elastic windows are an expansion of the idea of a workplace or room in terms of how 2D tiled windows are organized. By altering how much screen space is taken up by the current role, the main goal is to make switching between roles or tasks easier.

Example Systems

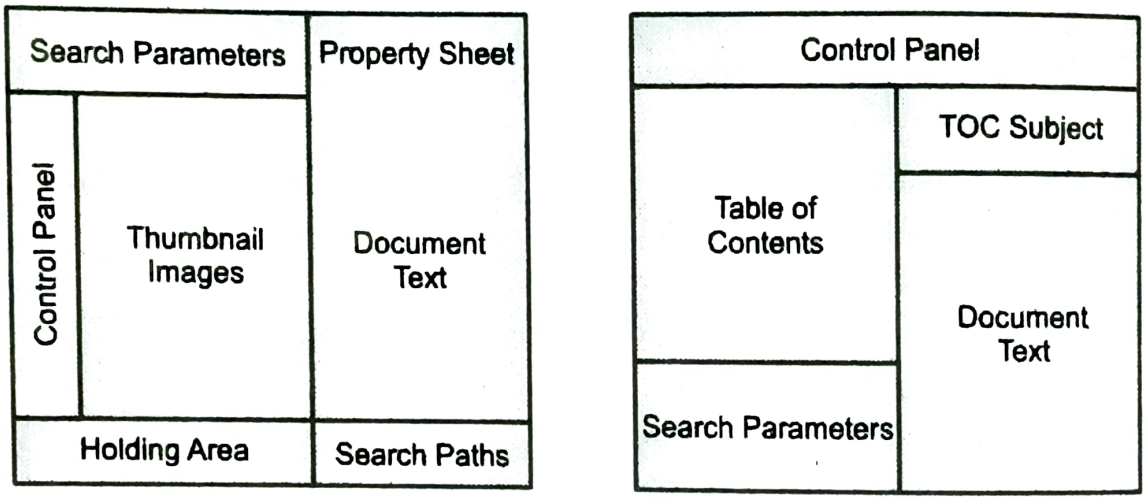
The information layout and management strategies used by numerous contemporary information access interfaces are described in the following sections.

The InfoGrid Layout

- A typical illustration of a monolithic design for an information access interface is the InfoGrid system.
- The layout, which is divided into a left and right side, is predicated on the availability of a huge display (see Fig. 6.6.3).
- The left side is further separated into a top section with structured entry forms for specifying a query's characteristics, a column of recognisable controls lining the left side, and a bottom space for storing important documents.
- The primary central section is where retrieval results are viewed. These results may be displayed as thumbnails of the original documents or as derived organizations of the documents, such as Scatter/Gather-style cluster findings.
- Users can choose papers from this area and either see them in the right-hand side or store them in the holding area below. The right-hand side of the display is used to see chosen documents for the most part, with

metadata for those papers displayed in the upper part.

- A graphic history of previous interactions is supposed to be displayed in the space below the document display.



(1F9)**Fig 6.6.3 : Diagrams of monolithic layouts for information access interfaces**

The SuperBook Layout

- The InfoGrid's design is fairly reminiscent of SuperBook's layout. The primary distinction is that SuperBook keeps the table of contents-like display in the main left-hand pane, along with counters that show the proportion of documents with search hits at each level of the outline.
- Selected documents are shown in the main pane on the right side, much like InfoGrid.
- Just below the table of contents display, queries are created. Also displayed in this box are terms associated with the user's search. Pop-up overlapping windows show large images.
- The Super Book's layout came about after multiple iterative design cycles. In earlier iterations, overlapping windows were utilized in place of a monolithic layout, allowing users to sweep out a new text box by selecting a rectangle on the screen.
- The users of this new text box might jump to the table of contents or to instances of highlighted words in other papers using its own set of buttons.
- After taking into account the findings of experimental studies demonstrating that users can be more productive if provided fewer, well-chosen interaction paths rather than being given a wide latitude, SuperBook was rebuilt.

The DLITE Interface

- A variety of intriguing design decisions are made by the DLITE system. Functionality is divided into two sections: control of the search process and result display. An animated graphic direct manipulation display serves as the control part (see Fig. 6.6.4).
- Query objects, sources, documents, and collections of retrieved documents are all graphical objects. An editable field in a query function Object() { [native code] } object is filled out by the user to create a query.
- The system creates a query object, which is symbolized by a tiny icon and can be put onto icons for collections or search services. In response, if a service is running, it creates an empty results set object and attaches the query to it.
- The retrieval results are shown as a circular pool, and the papers that make up the pool are shown as symbols placed around the edge. Documents can be put into other services, such a document summarizer or a language translator, after being pulled out of the results set pool.
- The user can copy the query icon and drop it onto another search engine in the interim.
- The contents of the underlying query are displayed in a "tool-tips" box that appears when the mouse is placed on the query's icon. Queries can be saved and used again later, making it easier to keep track of tried-and-true search tactics.

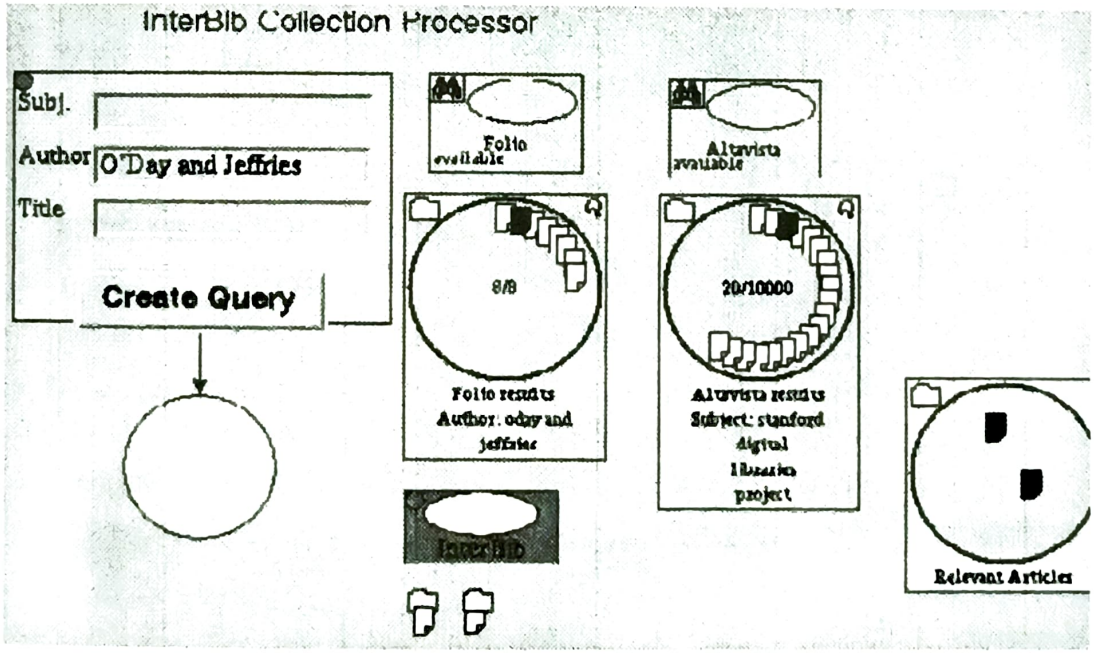


Fig 6.6.4 : The DLITE interface

The SketchTrieve Interface

- The SketchTrieve interface's guiding principle is the representation of information access as an informal process, in which partially completed concepts and explored paths can be saved and returned to compare to subsequent interactions.
- The results can then be combined via operations on graphical objects and connectors between them. Users have been seen to organize information using the physical arrangement of cells in a spreadsheet.
- This concept served as the inspiration for the creation of SketchTrieve, a tool that enables users to arrange retrieval results side by side to promote comparison and recombination (see Fig. 6.6.5).
- In the future, there should be a wider adoption of the idea of a canvas or workspace for the preservation of the previous context.
- Numerous problems are difficult to resolve, such as how to display the outcomes of a collection of linked searches after making tiny alterations based on query expansion, relevance feedback, and other types of modification.

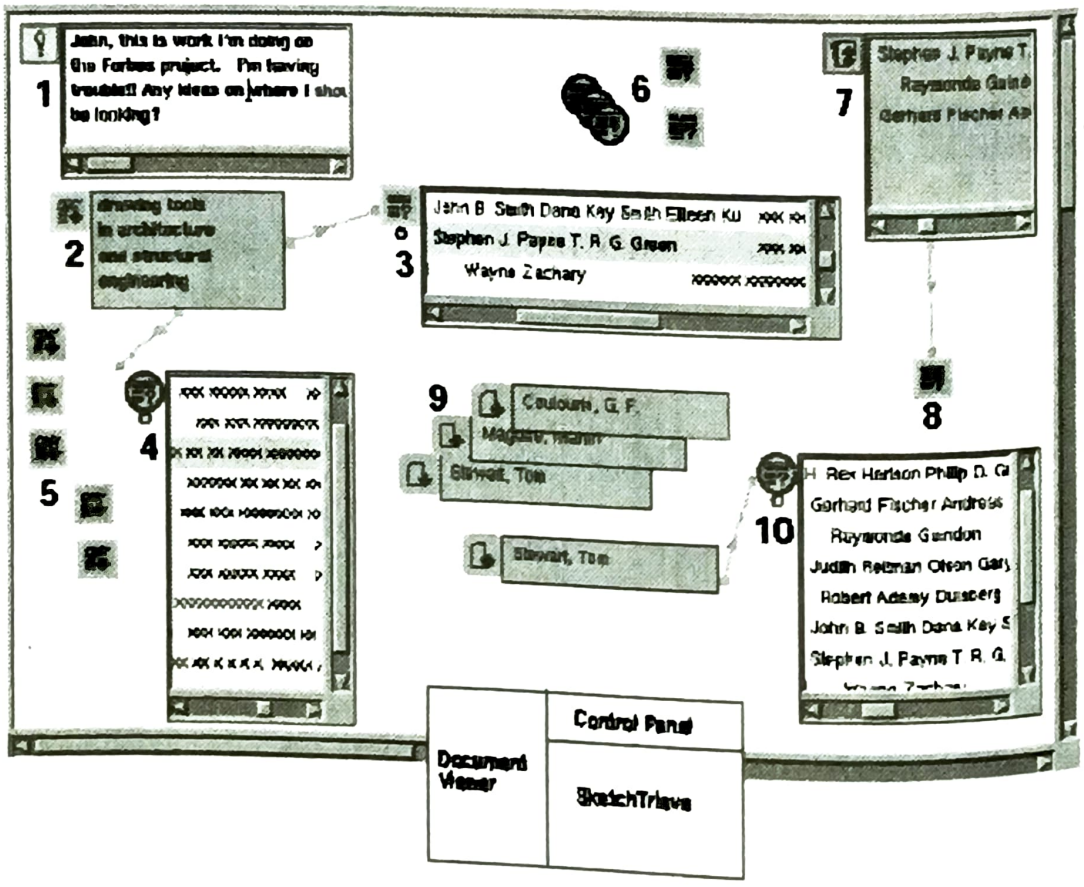


Fig. 6.6.5 : The Sketch Trieve interface

Examples of Poor Use of Overlapping Windows

- Problems can occasionally arise when switching from a command-line interface to a graphical display.
- The researchers discuss bad design choices made in a display for a bibliographic system that had overlapping windows. Another scenario where problems can arise is when designers "literally" convert a TTY interface to a graphical one.
- The existing LEXIS-NEXIS interface, which does not take advantage of window systems' ability to let users display multiple types of information at once, is a good example of the implications. Instead, the interface does not maintain window context when the user goes from one function to another, despite taking up the entire screen.
- For instance, instead of overlaying the information with a pop-up window or rearranging the available space with resizable tiles, viewing a small amount of metadata about a list of retrieved titles causes the list of results to vanish. Furthermore, rather than using the bit-map capabilities of a graphical interface, this metadata is shown in poor ASCII.
- When a user chooses to view a document's complete text, it is displayed in a small area, a few paragraphs at a time, as opposed to growing to take up the entire available space.

Retaining Search History

- According to the debate, the user interface should allow the user to annotate the decisions they make and the information they discover along the route, as well as display the options that are accessible at any given time as well as past moves and both short- and long-term strategies.
- Users should be able to save specific chunks of search sessions as well as the entire search session, and have flexible access to and modification of each. Additionally, there is growing interest in the incorporation of user preferences and usage data into the creation of search queries as well as the use of search results.
- The majority of these tactics are not well supported by the user interfaces that are now available, although several mechanisms have been established that start to satisfy these needs. For these jobs, systems that

save the search's past history are especially helpful.

- Most search systems have historically offered some sort of history function. These often provide a list of the commands that were previously executed. The introduction of graphical history in more recent times has made it possible to trace both commands and results.
- The VISAGE system for information visualization serves as an example of an innovation of particular interest for information access interfaces by saving state in small form in a "slide sorter" view (see Fig. 6.6.6).
- The additional benefit of the VISAGE application's visual design makes it simpler to identify individual states.
- This interface should be helpful for keeping track of search activity history despite being designed to be used as a presentation creation tool.

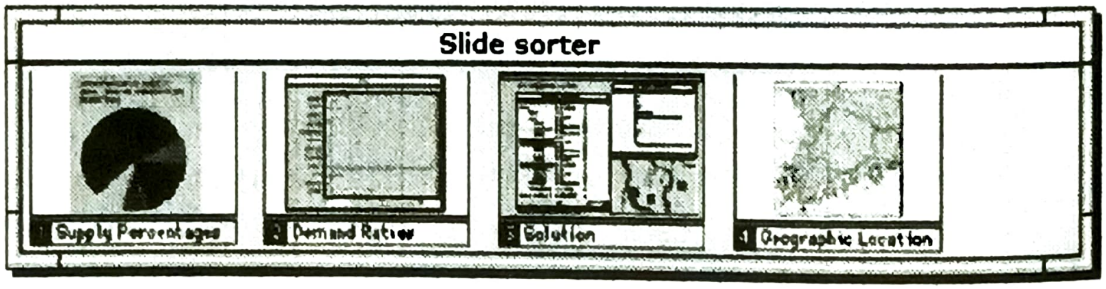


Fig. 6.6.6 : The VISAGE interaction history visualization

Integrating Scanning. Selection, and Querying

- In general, user interfaces for information access do a poor job of supporting strategies or even moves from one operation to the next.
- In most interfaces, even something as straightforward as leveraging the output of retrieval results from one query as input to another query conducted in a later search session is not properly supported.
- Users preferred user interfaces that integrated scanning and query specification, according to the researchers. They did not, however, notice improved outcomes from such contacts.
- They postulated that interaction between two different modes needs greater guidance since, if interactions are too unrestrained, they may result in wasteful or mistaken behavior.
- This implies that additional adaptability is required, albeit within limitations.

- The new Melvyl system for the Web offers means to edit a query's result after it has been run in order to run it again (see Fig. 6.6.7).
- Systems based on workspaces, like DLITE and Rooms, allow for the saving and reuse of past states.
- These systems do not, however, effectively combine the general search process with scanning and information selection from auxiliary structures. In general, there has to be better integration between scanning, selection, and querying.

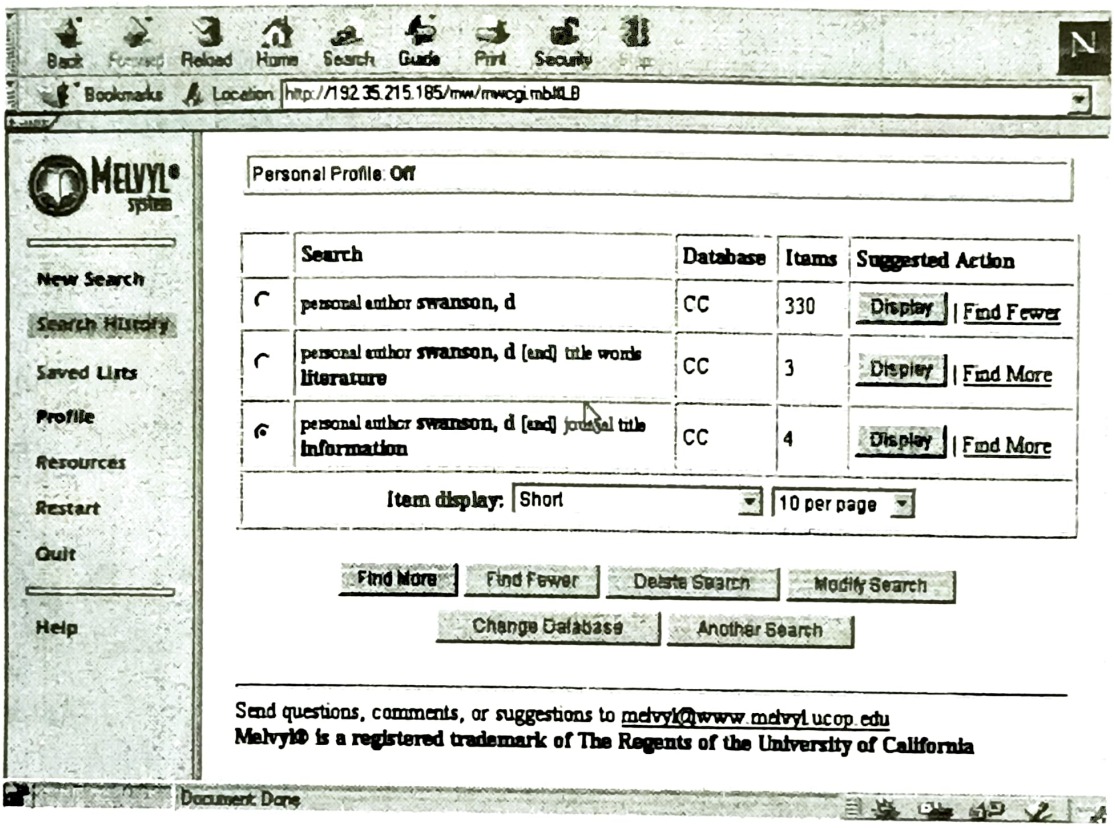


Fig. 6.6.7 : A view of query history revision in the Web-based version of the Melvyl bibliographic catalog. Copyright @, The Regents of the University of California

General Questions

Q. 1 Write the challenges of relevance feedback.

Q. 2 Write about the importance of relevance feed back in probabillstic model.

Q. 3 What are the 3 main user interface types used to interact with a computer?

Q. 4 What is the purpose of query interface in information retrieval?

Q. 5 What are the major components in information retrieval system?

Q. 6 What are the main challenges in information retrieval?

Q. 7 What are main components of user interface?

Q. 8 What are the main methods of retrieving data and information?

Q. 9 Describe the difference between relevance feedback and query expansion in terms of user interaction.

Q. 10 What is GUI in visualization?

Q. 11 What is human computer interface?

Q. 12 Define visualization?

Q. 13 Define property/attribute specification and explain its sequence.

Chapter Ends...

❑❑❑