Moves in online searching

Raya Fidel

Abstract: Moves, or changes in query formulation, are made to resolve three problem situations: (1) when retrieved sets are too large; (2) when they are too small; or (3) when retrieved sets are off-target. Observation and analysis of about ninety searches resulted in a list of eighteen operational moves, or modifications of query formulation, that keep the meaning of query components unchanged, and twelve conceptual moves which change the meaning of query components. All these moves are explained and then related to search tactics and strategies.

Introduction

From the first encounter with an information need to the moment when a final answer set is printed and delivered, the process of an online search is usually described to include several steps. After an information need is registered, searchers develop a search plan. During this step they give an explicit verbal form to the request, decide which search systems and what databases to use and in what sequence, and they develop query formulations for the databases to be searched. A query formulation consists of a list of query components and Boolean operators. Each query component (or part, facet, concept group) is represented by search terms. The Boolean operators relate these search terms to one another either by intersecting retrieved sets with the AND operator or by grouping them together with the OR operator. The next step is to execute the search plan and print an answer set that will satisfy the information need.

The most important advantage of online searching is that searchers can revise their search plan throughout the entire search process, and more specifically, they can modify their query formulations at any step of the search. Any such change in a query formulation will be called a move. The purpose of this presentation is to describe various modifications of query formulations, or moves, made by searchers to improve search results.

The moves described here were made by seven experienced online searchers who were observed while performing approximately ninety searches as part of

The author is at Graduate School of Library and Information Science, University of Washington, Seattle, Washington, USA.
their regular workday. Their search processes were recorded and then systematically analyzed to identify moves that they made. These moves are presented here as a list structured according to two styles of searching and by the three problem situations the moves are meant to resolve.

It was found that the observed searchers employ one of two distinct searching styles: operationalist or conceptualist [1]. Searchers were classified as using one or the other of these styles by characteristics such as: the method they used to analyze requests, the type of search terms they selected, and their attitudes toward new developments in search systems. Interestingly, it was also found that moves most frequently made by searchers of one style are different in nature from moves made by searchers of the other style. Operationalist searchers habitually select moves which rely on the trade-offs between free-text and controlled vocabulary searching and which use search system capabilities in order to modify retrieved sets without changing the specific meaning that those sets represent. Such moves are called here operational moves. In contrast, conceptual moves modify retrieved sets by modifying the meaning of the concepts the sets represent and are often supported by subject-related analyses of requests and by the structure of index languages. Although each style has its typical moves, searchers of one style sometimes use moves of the other style.

It was also observed that the searchers made moves to improve query formulations in three problem situations: when (1) they thought that retrieved sets were too large; (2) they thought that retrieved sets were too small; or when (3) they considered a retrieved set to be off-target. For that reason, the moves within each style are presented by the problem they are expected to eliminate.

A note about precision and recall is in order here. Moves to reduce or to enlarge
"The distinction, therefore, between moves to reduce or enlarge retrieved sets and moves to improve precision or recall is blurred"

the size of a retrieved set relate to attempts to improve precision or recall, respectively. However, the correspondence is not clear-cut. Any move to increase precision alone (i.e., eliminating only non-relevant documents) automatically reduces the size of a set, but not every move to reduce the size of a set is aimed at improving precision. On the other hand, all moves to increase recall alone enlarge the size of a set, and moves to enlarge the size of a set improve recall because searchers are attempting to retrieve additional relevant documents. However, searchers sometimes try to enlarge a set because they think that they have not retrieved enough documents, as opposed to a concern that some documents have been missed. Given a certain situation, searchers make one move if they want to enlarge a set, but may make another if they think, say, that some highly relevant documents are missed. The distinction, therefore, between moves to reduce or enlarge retrieved sets and moves to improve precision or recall is blurred. The difference lies in the searchers’ intentions when they make a certain move in the context of a particular search.

Although this list is restricted to moves that were actually made by the observed searchers, readers may keep in mind that each move listed here suggests another one that is theoretically conceivable. That is, each move to reduce the size of a retrieved set has its counterpart move to enlarge the size of a set, and vice versa. For example, if to group together synonyms with a descriptor (a thesaurus term assigned by indexers) is a move to enlarge the size of a set, to eliminate synonyms as representatives in a query component is a move to reduce its size. But while the observed searchers often added synonyms to enlarge a set, they never eliminated then a later move. Thus, the moves in this list are described as they were observed and counterpart moves not selected are not listed.

Experienced online searchers are probably familiar with most of the moves presented here. Moreover, most of the moves are scattered in the textbook literature (e.g., [2], [3], [4]), and in articles published in the online journal literature. It is useful, though, to present them in a problem-oriented fashion: when faced with a particular difficulty, searchers can screen the moves pertinent to their problem and select the most appropriate ones. The list provides novice searchers with ready access to a variety of moves that may help them to master efficient search techniques. Experienced searchers can use this presentation, and particularly the table, as a check list of options. This may also prove of particular benefit to searchers who search in a specific style: searchers who tend to select operational moves, for example, can be reminded by the list of conceptual moves of moves to which they are not accustomed and thus expand their repertoire. The moves are also related to
<table>
<thead>
<tr>
<th><strong>Operational moves</strong></th>
<th><strong>Conceptual moves</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Moves to reduce the size of a set</strong></td>
<td></td>
</tr>
<tr>
<td>Weight 1</td>
<td>Limit a descriptor to be a major descriptor.</td>
</tr>
<tr>
<td>Weight 2</td>
<td>Intersect free-text set with a broader descriptor.</td>
</tr>
<tr>
<td>Weight 3</td>
<td>Limit free-text terms to occur in a predetermined field.</td>
</tr>
<tr>
<td>Weight 4</td>
<td>Require that free-text terms occur closer to one another in the searched text.</td>
</tr>
<tr>
<td>Weight 5</td>
<td>Limit to documents of a certain form.</td>
</tr>
<tr>
<td>Negate</td>
<td>Eliminate unwanted elements by using the AND NOT operator.</td>
</tr>
<tr>
<td>Limit 1</td>
<td>Limit to documents written in a particular language.</td>
</tr>
<tr>
<td>Limit 2</td>
<td>Limit to documents published, or indexed, in a particular period of time.</td>
</tr>
<tr>
<td>Limit 3</td>
<td>Limit to documents retrieved from a specific portion of the database.</td>
</tr>
<tr>
<td>Limit 4</td>
<td>Limit to sources that have, or do not have, a certain term in their titles.</td>
</tr>
<tr>
<td>Cut</td>
<td>Submit only part of the retrieved answer set, arbitrarily selected.</td>
</tr>
<tr>
<td></td>
<td>Moves to enlarge the size of a set</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Add 1</td>
<td>Add synonyms and variant spellings.</td>
</tr>
<tr>
<td>Add 2</td>
<td>Add descriptors as free-text terms.</td>
</tr>
<tr>
<td>Add 3</td>
<td>Add descriptors assigned to relevant citations retrieved.</td>
</tr>
<tr>
<td>Add 4</td>
<td>Add terms from database's index that have a high number of postings.</td>
</tr>
<tr>
<td>Include</td>
<td>Group together a descriptor with all the descriptors that are its narrower terms.</td>
</tr>
<tr>
<td>Cancel</td>
<td>Eliminate restrictions previously imposed.</td>
</tr>
<tr>
<td>Refine</td>
<td>Find a &quot;better&quot; descriptor.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
search tactics and strategies and these relationships are discussed after the list of moves is presented.

Operational moves

Moves to reduce the size of a retrieved set

Operational moves to reduce the size of a retrieved set primarily involve the use of various methods for weighting search terms or operators and limiting sets to include documents that have specific features only.

The use of weights enables searchers to incorporate into query formulations the 'importance' of search terms or of their relations. Therefore, using weights is usually expected to improve the precision of retrieved sets. In most databases that employ weighted indexing, topics that are main points of a document are represented by major descriptors. The most direct way of using weights, then is to require that a descriptor be a major descriptor (Weight 1).

Weighting can also be used in free-text searching in various ways, although in indirect manner. When the precision of a set created by free-text terms needs to be improved, searchers intersect this set with a set which is created by entering a broader descriptor (Weight 2). This move ensures that a specific concept (expressed in free-text terms) is not merely mentioned but is actually discussed in retrieved documents. Another way to weight free-text terms is to require that they occur in a predetermined field (Weight 3). Entering the restriction that free-text terms, representing a certain concept, occur in titles, for instance, retrieves a set of documents for which the concept is important enough to be mentioned in their titles.

Some search systems facilitate another mechanism to increase precision by providing word-proximity operators. The hierarchy in word-proximity operators can be used in reduction/precision-oriented moves. Searchers can require that free-text terms occur closer to one another in the searched text (Weight 4). For example, if a retrieved set is an intersection of two sets, each created by free-text terms occurring anywhere in the text, searchers can impose the restriction that terms in the two sets must appear in the same field. In this move it is the relations between free-text terms that are weighted.

Controlled vocabulary can provide an additional means for weighting if it includes form descriptors, e.g., review, research, that describe the type of a document. That is, when searchers cannot identify ways to reduce the size of a retrieved set by eliminating documents that are not highly relevant, they may require that all documents in an answer set are also assigned a selected form descriptor, if suitable (Weight 5). This move enables searchers to retrieve documents that are expected to carry more information than others. While precision might be improved if properly done, this move is usually employed when a particular type of document is required, or when the precision of an original answer set is reasonable but its size too far exceeds request requirements.

As a last resort to increase precision, searchers use the negation operator. If they can identify an unwanted element that consistently appears in non-relevant
"When everything else fails, and searchers do not manage to reduce retrieval to a satisfactory size by precision-oriented moves, they submit only part of an answer set to users by randomly selecting, say, the first twenty citations."

citations, searchers try to eliminate all documents that are indexed under the undesirable term (descriptor or free-text) by using the AND NOT operator (Negate). This move is made with caution and searchers frequently review the citations for rejected items to make sure that no relevant documents were included among them.

To this point, the moves described always increase precision. The following moves in this section are always aimed at reducing the size of a retrieved set but their effect on precision depends on the specific request and on the particular move made. Limiting moves are primarily facilitated by search system capabilities and by database structure. For example, limiting a retrieved set to include documents published in a certain language (Limit 1) may indeed improve precision if a user does not read any other language but for another user who reads several languages, it may eliminate some relevant documents rather than weeding only irrelevant ones.

When search system capabilities and database structure permit, searchers limit a retrieved set to documents published or added to the database during a certain period of time (Limit 2). Or searchers may limit a set to be retrieved from a particular portion of a database specifically designated by a database producer (Limit 3). Such sub-databases may cover a distinct area within the subject field that is covered by a database, or they may cover selected sources, e.g., the core journals in the field. Unlike the other limiting moves, the effect here depends on the criteria by which the sub-database is defined.

For databases that do not have designated sub-databases, searchers can still create an on-the-spot sub-database by retrieving documents that are published in sources that have, or do not have, certain terms in their titles (Limit 4). For instance, if in searching a medical database only nursing literature is desirable, searchers enter the requirement that only documents published in sources that have the stem "nurs" in their titles will be retrieved. By the same token, nursing literature can be excluded, if unwanted, by requiring that the stem does not occur in sources' titles.

When everything else fails, and searchers do not manage to reduce retrieval to a satisfactory size by precision-oriented moves, they submit only part of an answer set to users by arbitrarily selecting, say, the first twenty citations (cut). Just as with Weight 3 and Negate, searchers try to avoid making this move and refer to it only in exceptional cases, e.g., when an answer set is exceedingly large, its precision is acceptable, and a user needs to see just a few documents and wants to get them right away.
Moves to enlarge the size of a retrieved set

Operational moves to enlarge the size of a retrieved set primarily involve the addition of search terms to represent a concept. Therefore, all enlargement moves also improve recall of retrieved sets.

The most frequently used operational moves to enlarge sets rely upon the capability of free-text searching. A set can be enlarged by:

- adding synonyms and variant spellings found in various sources such as in retrieved citations (Add 1)
- searching a descriptor also as a free-text term, or the reverse (Add 2)
- adding descriptors assigned to relevant documents retrieved by free-text terms, if they represent the same concept (Add 3).

The display of entries in a database’s index may suggest additional search terms to represent a query component. When displayed, searchers examine the number of postings for a term (how many citations would be retrieved by a search term) and select terms with the largest number of postings (Add 4). Consider, for example, a request about the effect on humans of a certain chemical which is used for grain fumigation (in a hypothetical database). Suppose searchers decide to establish a set to represent the concept grain fumigation. Displaying entries in a database’s index, they discover that the term fumigation has a low number of postings, while the term fumigated will retrieve a large number of citations. They add then the set retrieved by fumigated grain to the previous formulation and thus substantially increase the number of retrieved documents.

A frequently used mechanism to enlarge the size of a retrieved set is to use a command that groups together a descriptor with all the descriptors that are its narrower terms according to the hierarchy displayed in an index language (Include). The inclusive search [5] is a move heavily used by searchers in databases which provide such system command, but it is quite time-consuming in databases lacking this command because the narrower descriptors must be entered one by one.

One operational move that may improve recall but is bound to decrease precision is the one that eliminates restrictions previously imposed (Cancel). These restrictions are usually introduced to secure precision before searchers find out that they actually need to enlarge the size of a retrieved set. This move is selected at a point where searchers are willing to sacrifice precision in order to retrieve more documents.

Moves to improve both precision and recall

One operational move that attempts to improve a retrieved set, and which is not size-related, is the move to find a ‘better’ descriptor (Refine). Consider again the request about the effect of a certain chemical on humans. A strictly direct translation of the request may result in the combination of two query components: the name of the chemical, and the descriptor human. For a database in which indexing is specific, a better descriptor might be human exposure, since documents about this subject relate to human exposure and are likely to be indexed under it. Searchers entering the first descriptor may discover that the second one was assigned to retrieved documents and choose to use it instead. Precision and recall are probably improved and changes in set size are immaterial at this point.
Conceptual moves

Moves to reduce the size of a retrieved set

Conceptual moves to reduce the size of a retrieved set are aimed at discovering more specific representations of concepts. To achieve this end, searchers intersect a set with other sets and thus combine concepts, or they may narrow the meaning of the concepts within a query component itself. These moves always increase precision.

The most straightforward such move to increase precision is to intersect a set with another set representing an additional query component (Intersect 1). A set of documents dealing with the effect of a chemical on humans, for example, may be intersected with a set of documents about grain fumigation. However, if such a move is impossible or undesirable, searchers make moves that result in a more specific representation for individual query components rather than introducing additional components.

When using descriptors to represent concepts in query components, searchers usually try to select the most specific descriptors. If they cannot identify a specific descriptor, they often enter a broader descriptor to represent a concept. In such cases, a move to increase precision is to intersect a set retrieved by a broader descriptor with a set created by an extensive list of free-text terms representing the specific concept (Narrow 1). Suppose that the chemical in our example is tri-chloroethanol, and that the most specific descriptor in a database vocabulary is ethanol. If they think that precision needs to be improved, searchers intersect a set formed by this descriptor with a set created by grouping together free-text terms such as trichloroethanol, chloroethanol, tri, and chloro. Keeping the descriptor ethanol increases the probability that retrieved documents are of sufficient relevance for trichloroethanol.

Role indicators facilitate moves to increase precision. Index languages that provide role indicators enable indexers (and searchers) to express the role in which a concept, represented by a descriptor, relates to the subject of a document. For example, documents about the use of a chemical for grain fumigation can be indexed by a descriptor for the chemical, and then be qualified with a role indicator such as agricultural use. These documents are thus differentiated from others that deal with the effect of grain fumigation on the concentration of the same chemical in the environment. A move to improve precision, then, is a move to qualify descriptors with role indicators, when available (Narrow 2).

In indexing, a role indicator is never assigned without an accompanying descriptor. However, some search systems facilitate searching by allowing role indicators to be entered ‘detached’ or ‘free-floating.’ Searching such systems, searchers can intersect a set for a query component with role indicators (Intersect 2). If, for instance, a chemical set is intersected with a set of documents whose indexing includes the role indicator agricultural use, the documents that are retrieved will deal with the original chemical and the agricultural use of at least one chemical, which may or may not be the original one.

Searchers who use role indicators frequently in their searching attempt to improvise when they must search databases that provide no role indicators. To
"Searchers often consult thesauri of other databases as sources for broader, related, or synonymous search terms to represent a query component"
gation, for instance, misses some relevant documents, they may instead retrieve a set created by all the descriptors that represent the chemical qualified with the role indicator agricultural use. For some requests this move may not improve recall, and for some requests it clearly reduces precision. However, it may be worth a trial for a request where, say, the use of the chemical is of primary importance and particular applications are only a secondary issue.

Another conceptual move to increase recall is to eliminate from a query formulation a component which is present by definition in all documents indexed in a database (Exclude). For example, when searching an agricultural database, the component grain fumigation is eliminated from a query formulation since the chemical is likely to be mentioned in documents only when it is used in agriculture, in which case it is used for grain fumigation.

When searchers review a potential answer set and suspect that some relevant documents are still missing, they supplement this set with sets of documents dealing with broader subjects (Expand 5). A set of documents about the effect on humans of a chemical used for grain fumigation can be supplemented with a set of documents about the effect of the chemical on humans, and/or by another one that deals with the use of the chemical in grain fumigation.

**Moves to improve both precision and recall**

Searchers sometimes improve a retrieved set by changing their strategy altogether. A conceptual move to support such a strategy shift is to construct an indexing-probe set (Probe 1). Such a set is retrieved strictly to get more information about indexing procedures and policies; it is a probing set in the sense that its output is not considered to be part of an answer set — in fact, documents in the retrieved probing set may not be relevant to the request at all. This kind of set is usually formed when the correspondence between a concept and a descriptor is ambiguous, or when no matching descriptor could be found, because a thesaurus is not available to searchers or because they could not locate an appropriate descriptor. An indexing-probe set is normally constructed at the beginning of the terminal session for a search and its structure is kept narrow so as to retrieve only specific documents.

Various strategies may be used to form indexing-probe sets:
- display the indexing of relevant documents suggested by users
- display the indexing of documents written by a known author in the subject area
- combine related and/or broader terms into a narrow representation. While terms might be broad, the combination is supposed to result in a narrow set.

Suppose searchers want to explore the way in which indexing expresses the fact that a chemical used for grain fumigation actually affects humans. For this purpose, they intersect sets for three concepts: the class of chemical compounds, occupational diseases, and agricultural workers. Reviewing descriptors assigned to documents dealing with affected workers is likely to inform searchers about the way in which the effect of the chemical is expressed in indexing.

Another procedure used in attempts to improve precision and recall is to
examine the number of postings of a search term in several databases. The differences among the number of postings in the various databases then guide the decision about which query components to restrict and which to expand in which database (Probe 2). For example, if a search term for a chemical would retrieve many citations in a chemical database, but very few in an agricultural database, searchers conclude that many of the references for the chemical in the chemical database would not be relevant for a request about agricultural uses of the chemical. They learn, then, that the agricultural aspect should be carefully worked out when they search the chemical database.

**Tactics, strategies and moves**

Information search tactics are described and named by Bates [6], and search strategies have been suggested by Bourne [7].* The moves presented here are complementary to tactics and strategies: when searchers pick up an information search tactic that seems promising — or when they decide to follow a selected search strategy — they can readily identify the optional moves that would enhance the performance of their plan.

Bates defines a search tactic as a move made to further a search regardless of the medium on which information is stored. As such, the moves presented here are moves within her search tactics because they are discussed in terms of online conventions only and they require a change in query formulation.

For some of the moves, the correspondence with Bates' tactics is straightforward. Negate, for example, is an example of the Block tactic, i.e., to reject, in the search formulation, items containing or indexed by certain terms. Intersect 2, Expand 3, or Expand 4 are examples of the Vary tactic, i.e., to alter or substitute one's search terms in any of several ways. Probe 1, on the other hand, fits into the Scaffold tactic, i.e., to design an auxiliary, indirect route through the information files and resources to reach the desired information, but the correspondence is not immediately obvious and additional moves may possibly fit into this tactic.

For some tactics, this list of moves provides no examples. The Sub tactic of moving downward hierarchically to a more specific term, for instance, is theoretically possible but the corresponding move was not observed to have been made by the examined searchers because they always selected the most specific descriptors to begin with. The move equivalent to the Sub tactic is therefore not listed here. However, the discrepancy between the list of theoretically possible tactics delineated by Bates, and the list of moves actually made by professional searchers which is presented here can be partially resolved if one remembers that each move listed here suggests a counterpart move that is theoretically conceivable. For instance, the counter move of Expand 1 is an example of the tactic Sub.

Search strategies describe optional plans for a search after a request is analyzed, its conceptual components are identified, and databases are selected. Determined by the particular requirements of a search, searchers decide whether to proceed with, say, the Citation Pearl Growing strategy, according to which searchers start

---

*As reported by Markey and Atherton.*
"... the list of moves presented here enables searchers to review possible options for changes in query formulations when faced with a particular problem. It provides concrete suggestions for techniques to be used in search tactics and supports decisions made during the planning of search strategies.

with the most specific query formulation and then incorporate new terms derived from displayed sets into the query formulation, or whether the Building Block approach is more suitable [7]. Once a strategy is determined, searchers can review the moves which support its implementation. For example, when searchers adopt the Building Block strategy, they first try to represent each query component comprehensively and only then to combine them. In their attempts at extensive representation for a query component, searchers can select from the list moves to enlarge a set that involve only one query component (i.e., Add 1 through Add 4, Include, Expand 1, Expand 2).

Searchers can fine-tune their decisions about how to proceed with a specific search strategy by determining which type of move, operational or conceptual, best serves their purposes. Suppose they plan a search to retrieve the Most Specific Facet First [7]. When reviewing the set retrieved for the most specific query component, they find that the number of citations is too large. Rather than to review all the possible moves that will reduce the size of a set, they can decide at this point whether a change in the meaning of the most specific query component is desirable. If a change in meaning is the choice, searchers can review the list of conceptual moves which reduce the size of a set and select the most appropriate ones. If, instead, they choose to preserve the meaning of the most specific set, they can refer to the list of operational moves to find suitable moves.

In summary, the list of moves presented here enables searchers to review possible options for changes in query formulations when faced with a particular problem. It provides concrete suggestions for techniques to be used in search tactics and supports decisions made during the planning of search strategies.

References


