in $h(t_j)$'s hierarchy tree. The variable f is in the interval **(0,1)** and H is the height of $h(t_j)$'s hierarchy tree. In practical terms, it is desirable to generate a small but non-trivial number of related queries, say 5–10. Depending on the length and complexity of the original query and the value s, f-family can produce quite a substantial number of queries. In such a case, only a reasonable number are forwarded to the search engine. The queries are chosen by ordering the items to be modified from the outermost Boolean operator, to the innermost until the goal number is reached. The choice of parameter f is based on experience of the particular user, as well as the query specification. On the other extreme, if the query is very basic, such as a single word, all the possible related queries may be generated.

Although the invention is illustrated in terms of broadening and restricting one item at a time and determining the replacement by using the same variable fraction f for each item, it is contemplated that any number of items can be simultaneously changed and that the replacement for each can be determined independently by traversing any number of edges in the item's hierarchy tree.

The results returned by the search engine contain the complete answer to the user's query, together with the number of matches that each of the related queries would elicit. The relative number of matches that the related queries produce is useful in providing the user with a measure of the relative restrictiveness of the different items in the query. Based on this extra information, the user may be able to reformulate a query in a more intelligent way than he could otherwise.

An example implementation of the user interface is shown in FIG. **11**. In addition to the query, the user can specify the number of matches she would be most interested in seeing. Out of all the related queries generated by the query tuner, only those that yield close to the desired number of results are displayed to the user.

Although illustrated and described herein with reference to certain specific embodiments, the present invention is nevertheless not intended to be limited to the details shown. Rather, various modifications may be made in the details within the scope and range of equivalents of the claims and departing from the scope of the invention.

The invention claimed is:

1. A method for generating search queries to be sent to a search engine for searching a information management system, comprising the steps of:
   a) receiving an initial search query;
   b) converting the initial search query to general boolean language;
   c) identifying a level in a respective hierarchy tree for each search query item in the initial search query;
   d) formulating additional related search queries by substituting items from the respective hierarchy tree for selected items in the query, the substituted item having a level in the hierarchy tree that is greater than or less than the level of the query item in the initial query; and
   e) forwarding the initial search query and the additional search queries in parallel to the search engine.

2. A method for generating search queries according to claim **1** wherein the information management system to be searched is a global information network.

3. A method for generating search queries according to claim **1** wherein the search query items are selected from a group consisting of: a meta-datum, a keyword and a Boolean operator.

4. A method for generating search queries according to claim **1** wherein the additional related queries consist of a set of queries generated by selecting each query item in the initial search query for substitution to form a respectively different one of the related queries.

5. A method for generating search queries according to claim **4** wherein the initial search query is modified by restricting and broadening the search query by traversing one edge up and one edge down on the hierarchy tree for the selected query item.

6. A method for generating search queries according to claim **5** wherein the number of edges of the hierarchy tree to be traversed is increased to generate additional queries.

7. A method for generating search queries according to claim number **1** wherein the step of receiving an initial search query includes retreiving keywords from a prior search and adding them to the search query.

8. A method for organizing a set of records into clusters comprising the steps of:
   a) receiving the set of records;
   b) analyzing at least one text field from each record of the set of records to determine any patterns;
   c) partitioning each record by the text field analyzed in step (b) into clusters based on a shared pattern, wherein the step of partitioning a set of records consists of the steps of:
       c1) hashing all the single words, all the pairs of consecutive words and all long sequences of words for each record in the set of records; and
       c2) partitioning a set of records by the text field hashed in step (c1) into clusters
   d) analyzing the partitions to identify the clusters for display; and
   e) displaying the clusters to the user.

9. A method for organizing a set of records into clusters, comprising the steps of:
   a) receiving the set of records;
   b) analyzing at least one text field from each record of the set of records to determine any patterns;
   c) partitioning each record by the text field analyzed in step (b) into clusters based on a shared pattern;
   d) analyzing the partitions to identify the clusters for display; and
   e) displaying the clusters to the user;
   wherein the step of analyzing the partitions to identify the clusters for display includes analyzing a best cluster for display by weighing both the number of documents that contain the shared pattern and the length of the shared pattern.

10. A method for organizing a set of records into clusters according to claim **9**, wherein each record is a document having a uniform resource locator (URL), a title, a document excerpt and date information, and the fields to be analyzed are selected from a group consisting of: the title, the URL, the document excerpt and the date information.

11. A system for generating search queries to be sent to a search engine for searching a information management system, comprising:
   a) means for receiving an initial search query;
   b) means for converting the initial search query to general Boolean language;
   c) means for placing each search query item in the initial search query into a hierarchy tree;
   d) means for formulating additional related search queries by replacing at least one search query item with an item either above or below the search query item in its hierarchy tree; and

e) means for forwarding the initial search query and the additional search queries in parallel to the search engine.

12. A system for organizing a search engine's results including a set of documents each document including a plurality of fields, comprising:

a) means for receiving the set of documents;

b) means for analyzing several fields from each document to determine patterns

c) means for partitioning each document by the fields analyzed in step (b) into clusters based on a shared pattern;

d) means for analyzing the partitions to identify clusters for display;

e) means for displaying the clusters of documents to the user; and

f) means for user selection of the clusters for viewing;

wherein the means for analyzing the partitions to identify the clusters for display includes means for analyzing a best cluster for display by weighing both a number of documents that contain the shared pattern and a length of the shared pattern.

13. A carrier including a computer program which, when executed by a processor, causes the processor to generate search queries to be sent to a search engine for searching a information management system, the computer program causing the computer to perform the steps of:

a) receiving an initial search query;

b) converting the initial search query to general Boolean language;

c) placing each search query item in the initial search query into a hierarchy tree;

d) means for formulating additional related search queries by replacing at least one search query item with an item either above or below the search query item in its hierarchy tree; and

e) forwarding the initial search query and the additional search queries in parallel to the search engine.

14. A carrier including a computer program which, when executed by a processor, causes the processor to organize a set of documents into clusters, by causing the computer to perform the steps of:

a) receiving the set of documents;

b) analyzing at least one field from each document to determine a pattern;

c) partitioning each document by the fields analyzed in step (b) into clusters based on a shared pattern;

d) analyzing the partitions to identify the clusters for display;

e) displaying the clusters to the user; and

f) allowing the user to select one or more of the clusters for viewing;

wherein the step of analyzing the partitions to identify the clusters for display includes the step of analyzing a best cluster for display by weighing both a number of documents that contain the shared pattern and a length of the shared pattern.

* * * * *