(iv) *"automatically identifying a set of entities in the selected document content for searching information related thereto using the information retrieval system;"*

See discussion of the corresponding element of claim 1, where the instructions are stored in a memory of the match server 14.

(v) *"automatically categorizing the selected document content using the organized classification of document content for assigning the selected document content a classification label from the organized classification of content; and"*

See discussion of the corresponding element of claim 1, where the instructions are stored in a memory of the match server 14.

(vi) *"automatically formulating the query to restrict a search at the information retrieval system for information concerning the set of entities to the category of information in the information retrieval system identified by the assigned classification label."*

See discussion of the corresponding element of claim 1, where the instructions are stored in a memory of the match server 14.

4. <u>Claims 2, 7, 15 and 19 are unpatentable over Wieser taken in view of Stibel under 35 U.S.C. § 103(a)</u>

Requestor respectfully submits that claims 2, 7, 15 and 19 are unpatentable over Wieser taken in view of Stibel under 35 U.S.C. § 103(a). A claim chart applying Wieser and Stibel to these claims is submitted herewith as Appendix P.

(a) Wieser and Stibel render obvious claim 2

Claim 2 depends from independent claim 1, and further requires *"limiting the query by adding terms relating to context information surrounding the set of entities in the selected document content."* As stated above, Wieser discloses forming a keyword query by identifying keywords in the content of a selected document, such as a Web page. As described in more detail below, Stibel teaches a system that processes a keyword query to generate an enhanced keyword query that includes all of the terms of the original query along with one or more additional terms. The one or more additional terms are meaning terms that define a term in the original query and/or are terms that are expected to be found in Web pages directed to the same meaning as that of a term in the original query. (*Appendix E at 11/36 to 12/24, and 9/21-43, all of which is shown in claim chart at Appendix P*) One of ordinary skill in the art would have easily recognized that the system of Wieser is ready for improvement through use of Stibel's techniques. The resulting enhanced query would include additional terms that relate to context information surrounding the keywords in the selected document content, rather than to context information surrounding other portions of the selected document content or surrounding a different document entirely. That is, these additional terms, in being directed to the meaning of the keyword terms in the selected document content and in corresponding to terms that are expected to be found in Web pages directed to the same meaning, provide context information surrounding the keyword terms in order to achieve a more precise search.

Specifically, Stibel teaches a system that receives a search query having one or more query terms and processes the search query to generate a new search query that will more effectively retrieve information that is relevant to the original search query. *(Appendix E at 2/42-47, and shown in claim chart at Appendix P*) The processing includes identifying a meaning or Sense of one or more query terms, identifying related terms that are likely to appear on a Web page when a user is querying for information associated with that identified meaning, and then modifying the original query to form the new search query by appending the meaning and the one or more related terms to the original query. *(Appendix E at 10/30-36 and 11/56-66, all of which is shown in the claim chart at Appendix P*) For example, if the original query includes only the term "java" and is determined to correspond to the meaning "coffee" and have related terms "espresso" and "beverage," the system may modify the search query to be the new search query "java + coffee + espresso + beverage." (*Appendix E at 12/16-19)* Stibel asserts that this new modified query will result in a more effective search in that the Internet search engine is more likely to return a meaningful hit list in response to the new modified query as compared to the initial query. Stibel teaches that "[a]s expanded, the query now includes terms that are selected to increase the likelihood that an Internet search engine will return a meaningful hit list." *(Appendix E at 12/18-21)*

As mentioned previously, the system of Wieser generates keyword queries just like those that are processed by Stibel's system. Therefore, one of ordinary skill in the art would have easily recognized that the system of Wieser is ready for improvement through use of Stibel's techniques. Specifically, a person of ordinary skill in the art would have readily recognized that Stibel's techniques for processing a search query could be applied to Wieser's query to make that query more effective at finding desired information by further limiting the query through the addition of meaning terms and related terms. Furthermore, the combination of Stibel's teachings with those of Wieser would involve merely combining and/or substituting known prior art elements to yield predictable results. One skilled in the art could have easily combined the known elements as claimed by known methods, with each element in the combination performing the same function as it does separately, and the combination yielding nothing more than predictable results to one skilled in the relevant art.

(b) Wieser and Stibel render obvious claim 7

Claim 7 depends from claim 6, and further requires *"extracting with the text categorizer a set of terms relating to the document content; and appending to the query ones of the set of terms extracted by the text categorizer to contextualize the query."* Wieser describes extracting keywords from the selected document content and appending the keywords to the query to contextualize the query. Stibel teaches that meaning terms and related terms can also be added to contextualize the query. The keywords, meaning, and/or related terms contextualize the query by reflecting the environment from which the query was generated – i.e., the textual contents of the selected document.

In particular, Wieser describes extracting keywords from selected document content and appending the keywords to a query via construction of a query context vector. (*Appendix C at page 14, line 19 to page 15, line 5, and shown in claim chart at Appendix P*) Stibel describes that further contextual information may be added to the query by identifying one or more meanings of the terms in the initial query and appending the one or more meanings and related words to the initial query. The related words represent a word or a meaning that is likely to appear on a web page when a user is querying for information associated with the one or more meanings of the terms in the initial query. *(Appendix E at 9/21-43, 10/30-36, 11/56-66 and 12/13-18, all of which is shown in the claim chart at Appendix P*)

(c) Wieser and Stibel render obvious claim 15

Claim 15 depends from independent claim 14, and further requires *"a short length aspect vector generator for generating terms relating to context information surrounding the set of entities in the selected document content; wherein the query generator adds the terms relating to the context information to limit the query."* The match server 14 of Wieser, when modified in accordance with

Stibel's teachings, includes electronic components that perform this function, as set forth in the discussion of the corresponding element of claim 2.

(d) Wieser and Stibel render obvious claim 19

Claim 19 depends from independent claim 18, and further requires *"wherein the instructions stored in the memory further comprise limiting the query by adding terms relating to context information surrounding the set of entities in the selected document content."* Instructions for performing this function are stored in a memory of the match server 14 of the system taught by Wieser, as set forth in the discussion of the corresponding element of claim 2.

5. <u>Claims 3 and 8 are unpatentable over Wieser taken in view of Stibel and Syskill under 35 U.S.C. § 103(a)</u>

Requestor respectfully submits that claims 3 and 8 are unpatentable over Wieser taken in view of Stibel and Syskill under 35 U.S.C. § 103(a). A claim chart applying Wieser, Stibel, and Syskill to these claims is submitted herewith as Appendix Q.

(a) Wieser, Stibel and Syskill render obvious claim 3

Claim 3 depends from claim 2, and further requires *"wherein the number of terms added is limited to a predefined number."* Syskill describes a system that develops a user profile for a user that browses the World Wide Web by receiving a rating from the user for each visited Web page and analyzing the information on each Web page. The system is then able to generate a query based on the user profile that is submitted to a search engine to find Web pages that may be of interest to the user. Syskill's system limits the number of terms in the search query submitted to the search engine to a predetermined number (e.g., 14 terms) in recognition of the fact that search engines cannot accept very long queries.

Specifically, Syskill teaches that a system that provides queries to a search engine may limit a query to a predetermined number of terms in recognition of the fact that the search engine may not be able to process very long queries. (*Appendix F at Page 56, Second Column, Lines 3-6, and shown in claim chart at Appendix Q*) Syskill operates on keyword queries just like those generated by the system of Wieser, as modified based on the teachings of Stibel. As noted previously, Syskill expressly teaches what a person of ordinary skill in the art would have already known – i.e., that any realistic implementation of an electronic search engine, like that of Wieser, as modified based on Stibel's teachings, has a limited amount of computing resources and, therefore, would limit the number of terms/keywords in a keyword query inputted into the search engine for processing. Accordingly, a person of ordinary skill in the art would have readily recognized that Syskill's limitation of terms in a query to a predetermined number could be easily applied to the search queries of Wieser, as modified based on Stibel's teachings, in recognition of the limited processing capabilities of search engines, as expressly disclosed by Syskill. Furthermore, the combination of Syskill's teachings with those of Wieser and Stibel would involve merely combining and/or substituting known prior art elements to yield predictable results. One skilled in the art could have easily combined the known elements as claimed by known methods, with each element in the combination performing the same function as it does separately, and the combination yielding nothing more than predictable results to one skilled in the relevant art.

(b) Wieser, Stibel and Syskill render obvious claim 8

Claim 8 depends from claim 7, and further requires *"abbreviating the set of terms extracted by the text categorizer to a predefined number of terms."* As discussed above with reference to claim 3, Syskill describes limiting the number of terms in a query to a predefined number. Applying Syskill's teachings to the Wieser system, as modified based on Stibel's teachings, would result in abbreviating the set of keywords extracted by Wieser's contextual matching server 300 to a predefined number. As described above, a person of ordinary skill in the art would have readily recognized that Syskill's limitation of terms in a query to a predetermined number could be easily applied to the search queries of Wieser, as modified based on Stibel's teachings, in recognition of the limited processing capabilities of search engines, as expressly disclosed by Syskill.

6. Claims 1, 2, 5-7, 10-15 and 18-19 are Anticipated by Rhodes under 35 U.S.C. § 102(a) and 35 U.S.C. § 102(e)

Requestor respectfully submits that claims 1, 2, 5-7, 10-15 and 18-19 are anticipated by Rhodes under 35 U.S.C. § 102(a) and 35 U.S.C. § 102(e). A claim chart applying Rhodes to these claims is submitted herewith as Appendix R.

(a) Rhodes anticipates independent claim 1

(i) *"A method for automatically generating a query from selected document content, comprising:"*

The Rhodes reference teaches an electronic searching system that automatically analyzes a current document with which the user is interacting (e.g., an e-mail displayed in an Emacs window) and automatically generates a query from the current document. The query may be generated both from text information of the current document and/or meta-information of the current document, and the results provided in response to the query are documents deemed most similar to the current document. Specifically, Rhodes teaches taking the body text from a current document and finding documents "similar" to the current document based on word similarities.

> The RA works in two stages. First, the user's collection of text documents is indexed into a database saved in a vector format. These form the reservoir of documents from which later suggestions of relevance are drawn; that is, stored documents will later be "suggested" as being relevant to a document currently being edited or read. The store documents can be any sort of text document (notes, Usenet entries, webpages, e-mail, etc.). This indexing is usually performed automatically every night, and the index files are stored in a database. After the database is created, the other stage of the RA is run from Emacs, periodically taking a sample of text from the working buffer. The RA finds documents "similar" to the current sample according to word similarities; that is, the more times a word in the current sample is duplicated in a candidate database document, the greater will be assumed the relevance of that database document. The RA displays one-line summaries of the best few documents at the bottom of the Emacs window.

(Appendix D at 1/56-2/6)

Rhodes also teaches that meta-information for documents can also be represented by vectors in a similar manner as the body text of the documents.

> Analysis module 133 first indexes all the documents in a corpus of data (which, again, are stored as files mass storage device 106, which is assumed for explanatory purposes to be a hard disk), and writes indices to disk. Unlike the RA, the invention preferably keeps several vectors for each document. These include not only the wordvec vector for text (if any) in the document but also vectors for meta-information, e.g., subject, people, time, date, day of week, location, etc.

(id. at 10/42-51)

Rhodes discloses that the Remembrance Agent (RA) similarity algorithm used to identify similar documents based on the body text of a selected document (i.e., the current document) can also be used for identifying similar documents based on the meta-information of the selected document.

> 4. Determination of relevance
> For each element of each discrete [meta-information] vector in a query – the generation and vectorization of which is described below – the algorithm used by the RA may be used to determine relevance to documents in the corpus.

(id. at 12/53-57)

Rhodes indicates that the RA similarity algorithm will generate a single similarity value for each type of meta-information (which, as stated above, can include the body text information itself). The different similarity values are then weighted and aggregated to generate a single overall similarity value for each document that is being compared to the selected document.

> 5. Weighted addition of vectors
> The result of the foregoing operations is a single similarity value for each type of meta-information. These values are associated with each document in the indexed corpus, and are used to compute the overall similarity using bias values for query and document types, by the following formula:
>
> Query biases = bq pq sq lq dq etc. (i.e., body_query_bias, person_query_bias, etc.)

(id. at 13/1-8)

> Each vector similarity is multiplied by its respective bias and the resulting biased similarity is summed , to produce an overall similarity between zero and one.

(id. at 13/15-19) After determining a single overall similarity value for each document that is being compared to the selected document, Rhodes describes supplying a ranked list of the documents having the highest overall similarity values. Rhodes also describes that documents having overall similarity values below a predetermined threshold may be omitted from the result set.

> Analysis module 133 supplies a ranked list of the most relevant documents, which may be continually, intermittently, or upon request

presented to the user over display 126. If desired, or upon user command, the list may be pruned to include only documents whose relevance level exceeds a predetermined threshold.

(id. at 13/42-47)

(ii) *"defining an organized classification of document content with each class in the organized classification of document content having associated therewith a classification label; each classification label corresponding to a category of information in an information retrieval system;"*

Rhodes performs a faster similarity analysis by using an organized classification of the documents in which the documents are grouped by wordcodes set forth in a wordvec file having the format shown in table 2 below.

(int)	(width*uns int)	(int)	(uns int)	(uns int)		(uns int)
NUM_WORDS,	WORDCODE-1,	NUM_DOCS=N1,	DOC-1,	DOC-2,	. . . ,	DOC-N1,
	WORDCODE-2,	NUM_DOCS=N2,	DOC-1,	DOC-2,	. . . ,	DOC-N2,
	etc.					

As shown in the above table, a wordcode "WORDCODE-1" is a category/class of documents that contains "N1" documents listed as "DOC-1, DOC-2, … , DOC-N1." A query having a word corresponding to "WORDCODE-1" will be restricted to searching the corresponding documents listed in the same row (i.e., in the same word vector) of the wordvec file as ""DOC-1, DOC-2, … , DOC-N1."

Accordingly, Rhodes describes a process that indexes a corpus of documents by assigning the documents into different groups, each group being labeled by a unique wordcode. Each of the wordcode groups is made up of only those documents in the corpus that include at least one occurrence of a particular word that uniquely corresponds to the wordcode. For example, a wordcode "12" may uniquely correspond to the word "pickle." The indexing of the corpus of documents performed by Rhodes may classify Documents 12, 201, and 2102 of the corpus as including at least one occurrence of the word "pickle" and thus as being within the wordcode class "12." The wordcode class "12" and its associated documents (i.e., Documents 12, 201 and 2102) may then be set forth as a word vector in the wordvec file, which may be subsequently leveraged to enable more efficient searching for similar documents than that achieved through the use of document vector dot products. Accordingly, each wordcode is a classification label and includes within its class all documents within the corpus that include at least one occurrence of the word that uniquely corresponds to the wordcode.

Specifically, Rhodes describes a similarity algorithm that represents documents as document vectors and that is able to estimate a similarity value between two documents by performing a dot product of the two corresponding document vectors.

> Briefly, the concept behind the indexing scheme used in RA is that any given document may be represented by a multidimensional vector, each dimension or entry of which corresponds to a single word and is equal in magnitude to the number of times that word appears in the document. … The advantages gained by this representation are relatively speedy disk retrieval, and an easily computed quantity indicating similarity between two documents: the dot product of their (normalized) vectors.

(Appendix D at 2/15-24)

Rhodes, however, acknowledges that such a dot product calculation becomes too slow for large databases (i.e., a large corpus of documents). Therefore, Rhodes provides an alternative similarity algorithm that groups documents into word vectors labeled by wordcodes and stores the word vectors in a wordvec file.

> Experience with the RA has shown that actually performing a dot product with each indexed document is prohibitively slow for large databases. In preferred implementations, therefore, document vectors are not stored; instead, word vectors are stored. The "wordvec" file contains each word appearing in the entire indexed corpus of documents followed by a list of each document that contains that particular word.

(id. at 4/20-27)

Accordingly, each wordcode is a label for a group of documents that include at least one occurrence of the word uniquely associated with the wordcode. The wordcode and the list of document identifiers (and the frequency of the uniquely associated word in each corresponding document) are included in the wordvec file.

> Each word in the wordvec is represented by a unique numerical code, the "width" indicating the number of integers in the code (the RA uses two integers per code). The NUM_DOCS field indicates the number of documents containing the word specified by the associated wordcode. The word-count variables DOC-1, DOC-2, … , DOC-N1 each correspond

> to a document containing the word, and reflect the number of occurrences of the word divided by the total number of words in the document.

(id. at 4/45-55)

(iii) *"automatically identifying a set of entities in the selected document content for searching additional information related thereto using the information retrieval system;"*

The Rhodes reference describes automatically extracting words from a text document (e.g., the body text of an e-mail) for searching additional information related thereto. The words are formed into a query vector for searching.

In particular, Rhodes describes an analysis module 133 that generates queries automatically from a current document in a document buffer 140. The analysis module 133 may extract both meta-information for the document and text information from the document, and it may use the extracted information to find additional information related to the current document.

> Analysis module 133 preferably generates queries autonomously from the current document in document buffer 140 or by reference to a current context. In the former case, analysis module 133 classifies the document either by its header or by reference to a template, and extracts the appropriate meta-information. In the latter case, the user's physical or interpersonal surroundings furnish the meta-information upon which the query is based. It is not necessary for the documents searched or identified to correspond in type to a current document. Furthermore, the query may not be limited to meta-information. Instead, the invention may utilize both a meta-information component (with relevance to candidate documents determined as discussed above) and a text component (with relevance determined in accordance with RA).

(Appendix D at 13/19-34)

Specifically, Rhodes describes extracting words from the text of the document, and then using the words to find additional documents related to the words through a word similarity analysis. *(id. at 1/56-2/6)* The words that are extracted from the text of the selected document are words with a substantive meaning. In other words, the words are not stop words. Moreover, the words are stemmed prior to being used in the word similarity analysis.

> Briefly, the concept behind the indexing scheme used in RA is that any given document may be represented by a multidimensional vector, each dimension or entry of which corresponds to a single word and is equal in magnitude to the number of times that word appears in the document. … RA creates vectors in three steps: … Step 1: Remove stop words … Step 2: Stem words … Step 3: Make the document vector .

(id. at 2/15-54)

(iv) *"automatically categorizing the selected document content using the organized classification of document content for assigning the selected document content a classification label from the organized classification of content;"*

Rhodes teaches automatically categorizing a selected text document (e.g., the body text of an e-mail) using the organized classification of document content set forth in its wordvec file by identifying one or more wordcodes corresponding to the query vector generated from the selected text document. Stated differently, in developing the query vector, words appearing in the selected document are identified and used to determine one or more wordcodes using the wordvec file format. For example, a selected document including the word "pickle" is matched against and produces the numeric wordcode "12". The wordcode "12" then yields a vector, revealing all documents that include at least one occurrence of the corresponding word "pickle" against which more complex processing can be performed. In this process, derivation of the wordcode "12" for the word "pickle," is done for the selected document, such that the selected document is automatically categorized.

The grouping of documents by wordcodes in a wordvec file format is shown by Rhodes in a table, reproduced below:

(int)	(width*uns int)	(int)	(uns int)	(uns int)		(uns int)
NUM_WORDS,	WORDCODE-1, WORDCODE-2, etc.	NUM_DOCS=N1, NUM_DOCS=N2,	DOC-1, DOC-1,	DOC-2, DOC-2,	. . . , . . . ,	DOC-N1, DOC-N2,

As shown in the above table, a wordcode "WORDCODE-1" is a category/class of documents that contains "N1" documents listed as "DOC-1, DOC-2, … , DOC-N1." A query having a word corresponding to "WORDCODE-1" will be restricted to searching the corresponding documents listed in the same row (i.e., in the same word vector) of the wordvec file as ""DOC-1, DOC-2, … , DOC-N1."

In particular, Rhodes describes looking at a word in the query vector, which, consequently, appears in the selected document, and identifying the corresponding wordcode for purposes of performing the similarity analysis in response to the query. The existence of wordcodes demonstrates teaching of an organized classification of document content by Rhodes, and association of wordcodes represents teaching by Rhodes of an assignment of classification labels from the organized classification to categorize the document and enable identification of a group of documents that will be analyzed by the similarity analysis.

> Accordingly, for each word in the query vector, the RA first looks up the word in the word offset file, and from that the word's entry is looked up in the wordvec file. An array of document similarities is used to maintain a running tally of documents and their similarities, in terms of numbers of word matches, to the query vector. The array is sorted by similarity, with the most similar documents at the top of the list. Similarity is computed for each word in the query vector by taking the product of the query-vector entry and the weight of each document in the corresponding wordvec file. To normalize this product, it is then divided by the query-vector magnitude (computed in the same manner as the document magnitude) and also by the document magnitude. The final value is added to the current running-total similarity for that document, and the process is repeated for the next word in the query. In summary, the query vector is analyzed wordcode by wordcode, with the similarities array indicating the relevance to the query of each document."

(*Appendix D at 5/12-28*)

(v) *"automatically formulating the query to restrict a search at the information retrieval system for information concerning the set of entities to the category of information in the information retrieval system identified by the assigned classification label."*

Rhodes teaches automatically formulating the query to restrict the search for information concerning the word(s) identified from the selected text document to the group(s) of documents in the corpus identified by the wordcode(s) assigned to the selected text document. Specifically, Rhodes describes automatically identifying one or more wordcodes for the selected document and performing a similarity analysis that restricts the documents in the corpus that are analyzed to only those documents that are within the word vector group or groups corresponding to the wordcode(s). (*Appendix D at 5/12-28, reproduced above*)

(b) Rhodes anticipates claim 2

Claim 2 depends from independent claim 1, and further requires *"limiting the query by adding terms relating to context information surrounding the set of entities in the selected document content."* The Rhodes system teaches further limiting the query by adding to the query meta-information terms that relate to context information surrounding the set of keywords identified in the selected text document. For example, if the selected text document is the body text of an e-mail, the query can be further limited by using the subject of the e-mail specified in the "Subject" field of the e-mail and the name of the person specified in the "From" field of the e-mail to generate a subject query vector and a person query vector, respectively, both of which are then used in conjunction with the body query vector to find similar documents.

In this example, the name of the person specified in the "From" field and the subject specified in the "Subject" field of the e-mail relate to context information surrounding the set of keywords in the body text of the e-mail in that they place the keywords of the body of text of the e-mail in their proper context – i.e., that they were sent by the sender and are directed to the specified subject. Furthermore, these additional terms relate to context information surrounding the keywords in the e-mail body text, rather than to context information surrounding other portions of the e-mail or surrounding a different e-mail entirely.

Moreover, the subject and the name of the person that sent the e-mail further limits the query in that a similarity value is determined for each document with respect to the body query vector, another similarity value is determined for each document with respect to the person query vector, and yet another similarity value is determined for each document with respect to the subject query vector. To identify documents that are responsive to the query, an overall similarity value for each document is then determined by calculating biased similarity values using bias values for each query type (i.e., a body_query_bias, a person_query_bias, and a subject query bias) and adding together the biased similarity value for the body query, the biased similarity value for the person query, and the biased similarity value for the subject query. Those documents with the highest overall similarity values are produced in response to the query.

In particular, Rhode's analysis module 133 may generate queries automatically from a current document by analyzing not only the text of the current document but also meta-information associated with the current document. In the above example, both the meta-information in the "Subject" field and the meta-information in the "From" field relate to context information surrounding the terms in the body text of the e-mail and, moreover, are displayed in the header of the e-mail that precedes and, therefore, surrounds the terms in the body text of the e-mail. (*Appendix D at 13/19-34 and 11/10-31, all of which is shown in claim chart at Appendix R*)

Rhodes teaches that the meta-information and the text information of the current document can be set forth in vectors In a manner similar to that described for text information (*Appendix D at 10/42-51 and 12/41-49, all of which is shown in the claim chart at Appendix R*) A similarity value can then be determined for each type of meta-information and text information that represents the degree of similarity between the corresponding meta-information and text information of a given document and that of the selected document. (*id. at 12/53-57, and shown in claim chart at Appendix R)* A single overall similarity value can then be calculated to determine the overall similarity between the given document and the selected document by performing a weighted addition of similarity values. *(id. at 13/15-19, and shown in claim chart at Appendix R)* The results of the query are presented to the user as a ranked list of most similar documents based on the calculated overall similarity value. Documents below a predetermined threshold similarity may be excluded from the result set. (*id. at 13/42-47, and shown in claim chart at Appendix R*)

(c) Rhodes anticipates claim 5

Claim 5 depends from independent claim 1, and further requires *"wherein the organized classification of document content is defined using a hierarchical organization."* As stated above with respect to claim 1, the organized classification of document content is the wordcodes in the wordvec file. This wordvec file can be representative of words in text information in documents, such as the body text of an e-mail. Rhodes, however, also contemplates using a similar wordvec file format for meta-information that is represented as text, such as, for example, names in the person field of an e-mail. Accordingly, the organized classification of document content described by Rhodes includes groups of documents organized by text wordcodes and groups of documents organized by meta-information wordcodes, such as person_name wordcodes. Moreover, this organized classification of document content is hierarchical in that Rhodes describes that the body text of an e-mail may be deemed more important with respect to similarity calculations than the person names in the "From" field of the e-mail. In asserting this, the Rhodes system contemplates a higher ranking for the text wordcode groups of documents than that of the person_name wordcode groups of documents.

In particular, Rhodes describes that discrete meta-information can be captured in vectors in a file that has the same file format as the wordvec file for text information. Accordingly, Rhodes contemplates wordcodes being used for meta-information and, thus, contemplates grouping documents in a corpus by meta-information wordcodes for performing meta information similarity analyses. *(Appendix D at 12/41-49, and shown in claim chart at Appendix R)*

Rhodes describes biasing the different similarity values generated by the similarity analysis for a selected document such that the text and the various different types of meta-information are thereby ranked relative to each other in importance. Therefore, in ranking the text and meta-information relative to each other in importance, Rhodes affects a hierarchical ranking of the corresponding text wordcode groups of documents in a corpus relative to the corresponding meta-information wordcode groups of documents in the corpus. For example, if the document is an e-mail, the Rhodes system may rank the text wordcode groups of documents in the corpus higher than the sender name wordcode groups of documents by setting a ratio of two-to-one for the bias of the text similarity value relative to the person name meta-information similarity value. (*Appendix D at 11/62-12/4, 13/1-8 and 13/15-19, all of which is shown in the claim chart at Appendix R*)

(d) Rhodes anticipates claim 6

Claim 6 depends from independent claim 1, and further requires *"using a text categorizer to assign the classification label assigned from the organized classification of content."* Rhodes teaches using an analysis module 133 that analyzes the text of the selected text document to identify one or more important words and then uses the important words to assign one or more wordcodes to the selected text document.

In particular, Rhodes describes an analysis module 133 that generates query vectors autonomously from a current document in a document buffer 140. The analysis module 133 may extract text information for the document to form the query vector. *(Appendix D at 13/19-34, and shown in claim chart at Appendix R*) Rhodes teaches automatically categorizing a selected text document (e.g., the body text of an e-mail) using the organized classification of document content set forth in the wordvec file by identifying one or more wordcodes corresponding to the query vector generated from the selected text document. *(Appendix D at 5/12-28, and shown in claim chart at Appendix R*)

(e) Rhodes anticipates claim 7

Claim 7 depends from claim 6, and further requires *"extracting with the text categorizer a set of terms relating to the document content; and appending to the query ones of the set of terms extracted by the text categorizer to contextualize the query."* In extracting meta-information and using the meta-information for the similarity analysis, Rhodes describes extracting with the analyzer 133 a set of terms relating to the selected document content and appending to the query ones of the set of terms extracted by the analyzer 133 to contextualize the query.

Specifically, the analyzer 133 can extract meta-information, such as the person names in the "From" field of the e-mail. This meta-information is then appended to the query by forming and using an additional meta-information query vector that results in similarities that are summed up, after biasing, with the similarities calculated based on the text information of the selected text document.

Put another way, Rhodes generates a query based on a current document to find documents that are similar to the current document. The documents deemed similar to the current document are those documents that have the highest overall similarity scores. Rhodes describes that the overall similarity score can be calculated by adding a text similarity score to one or more meta-information similarity scores. Accordingly, by summing the text similarity score with the one or more meta-information scores for each document and only producing those documents having the highest resulting aggregate scores in response to the query, Rhodes is appending to the query meta-information terms to contextualize the query. (*Appendix D at 11/62-12/4 and 13/15-19, all of which is shown in claim chart at Appendix R*)

(f) Rhodes anticipates claim 10

Claim 10 depends from independent claim 1, and further requires *"wherein each class in the organized classification of document content has associated therewith a characteristic vocabulary."* Each wordcode within the system of Rhodes is associated with a word, which describes the category represented by the wordcode in that it sets forth the characteristic, i.e., the word, that must be found in all documents within the wordcode category. *(Appendix D at 4/45-55, and shown in claim chart at Appendix R*)

(g) Rhodes anticipates claim 11

Claim 11 depends from claim 10, and further requires *"ranking results from the query performed at the information retrieval system in accordance with one of the assigned classification label and the* *characteristic vocabulary."* Rhodes ranks search results in accordance with the one or more wordcodes assigned to the selected text document in that the one or more assigned wordcodes are used in the calculation of the similarity values. Specifically, Rhodes describes using wordcodes to determine similarity values for documents. *(Appendix D at 5/12-28, and shown in claim chart at Appendix R)* Rhodes also describes using the similarity values to rank the set of searched documents. *(Appendix D at 13/42-47, and shown in claim chart at Appendix R)*

(h) Rhodes anticipates claim 12

Claim 12 depends from claim 11, and further requires *"using the method in a system for enriching selected content of a document with personalities that identify enrichment themes."* Rhodes teaches enriching the selected document content in that the additional information produced in response to the query is used to annotate the display of the selected document content. For example, if the selected document content is an e-mail displayed in an Emacs interface window, the Rhodes system will display one line summaries of the most similar documents to the displayed e-mail at the bottom of the Emacs window. *(Appendix D at 1/65-2/6, and shown in claim chart at Appendix R)*

Requestor notes that the '979 patent states that to "enrich" a document means merely to "annotate a document in accordance with a predefined personality." (*Appendix A at col. 6, lines 63 and 64*)

(i) Rhodes anticipates claim 13

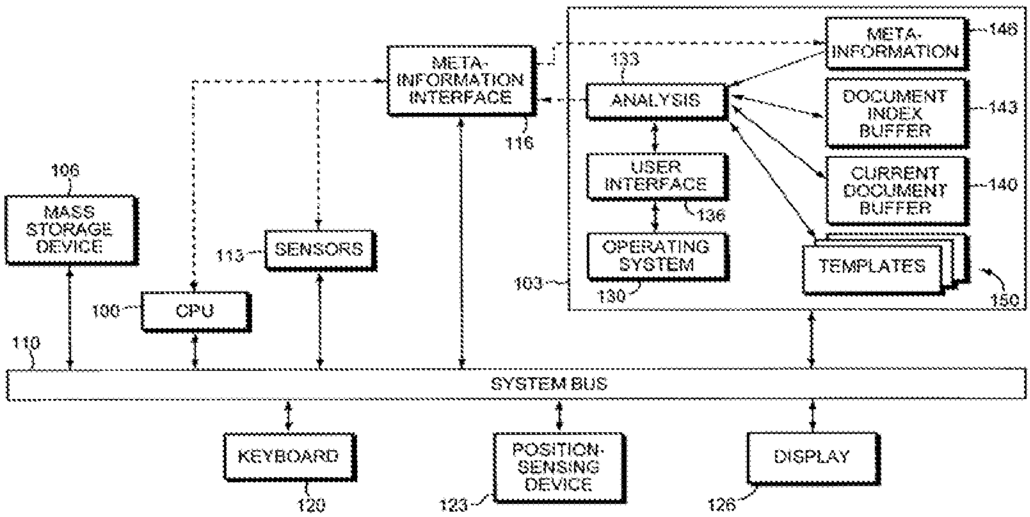
Claim 13 depends from independent claim 1, and further requires *"automatically identifying the set of entities using a service that recognizes entities of a predefined type."* The Rhodes system automatically identifies the words in the selected text document using a service that recognizes entities of a predefined type – i.e., words in the selected text document that are not stop words. (*Appendix D at 2/15-54, and shown in claim chart at Appendix R*)

(j) Rhodes anticipates independent claim 14

(i) *"A system for automatically generating a query from selected document content, comprising:"*

Rhodes describes a computer system shown in Fig. 1 (reproduced below) that includes a central-processing unit ("CPU") 100 and additionally includes a main system memory 103 and a mass storage device 106 that store instructions that are executed by the CPU 100. The computer system shown in

Fig. 1 performs the operations set forth above with reference to claim 1 to automatically generate a query.



> Refer now to FIG. 1, which illustrates, in block-diagram form, a hardware platform incorporating a representative, generalized embodiment of the invention. As indicated therein, the system includes a central-processing unit ("CPU") 100, which perform operations on and interacts with a main system memory 103 and components thereof. System memory 103 typically includes volatile or random-access memory ("RAM") for temporary storage of information, including buffers, executing programs, and portions of the computer's basic operating system.

(Appendix D at 9/18-34)

(ii) *"an entity extractor for automatically identifying a set of entities in the selected document content for searching information related thereto using an information retrieval system;"*

As described above with reference to claim 1, the analyzer module 133 of the main system memory 103 of the Rhodes system executes the operations corresponding to this feature.

> The main memory 103 contains a group of modules that control the operation of CPU 100 and its interaction with the other hardware components. These modules are implemented as executable machine

> instructions, running (by means of CPU 100) as active processes effectively capable of interacting (i.e., exchanging data and control commands) as illustrated. An operating system 130 directs the execution of low-level, basic system functions such as memory allocation, file management, and operation of mass storage devices 106. At a higher level, an analyzer module 133 directs execution of the primary functions performed by the invention, as discussed below; and instructions defining a user interface 136 allow straightforward interaction over display 126. User interface 136 generates words or graphical images on display 126 to facilitate user action and examination of documents, and accepts user commands from keyboard 120 and/or position-sensing device 123.

(Appendix D at 10/1-17)

(iii) *"a categorizer for defining an organized classification of document content with each class in the organization of content having associated therewith a classification label; each classification label corresponding to a category of information in the information retrieval system; the categorizer automatically assigning the selected document content a classification label from the organized classification of content; "*

As described above with reference to claim 1, the analyzer module 133 of the main system memory 103 of the Rhodes system executes (by means of the CPU 100) the operations corresponding to this feature.

(iv) *"a query generator for automatically formulating the query to restrict a search at the information retrieval system for information concerning the set of entities to the category of information in the information retrieval system identified by the assigned classification label."*

As described above with reference to claim 1, the analyzer module 133 of the main system memory 103 of the Rhodes system executes (by means of the CPU 100) the operations corresponding to this feature.

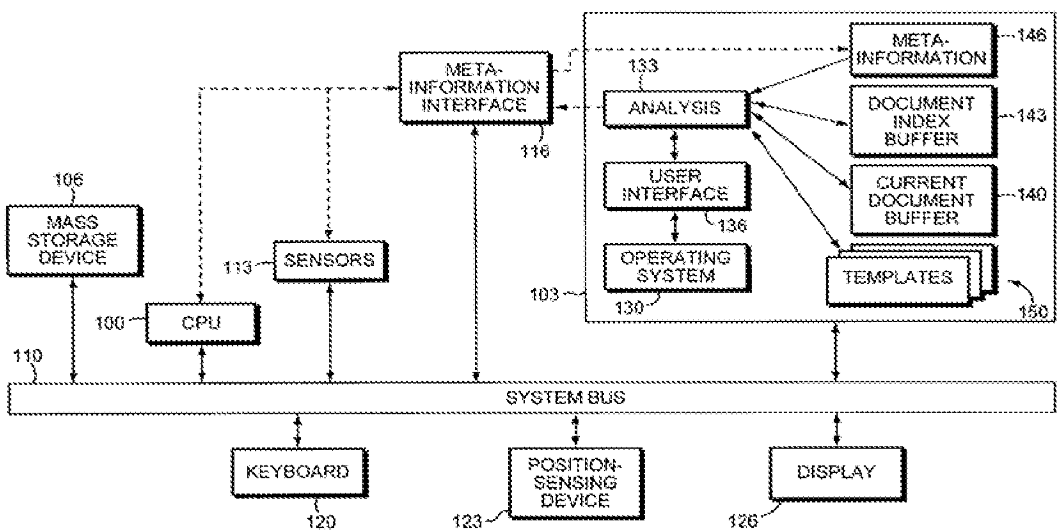
(k) Rhodes anticipates claim 15

Claim 15 depends from independent claim 14, and further requires *"a short length aspect vector generator for generating terms relating to context information surrounding the set of entities in the selected document content; wherein the query generator adds the terms relating to the context information to limit the query."* See discussion of corresponding element of claim 2, where the analyzer module 133 of the main system memory 103 of the Rhodes system executes (by means of the CPU 100) operations corresponding to this element.

(l) Rhodes anticipates independent claim 18

(i) *"An article of manufacture for use in a computer system, comprising: a memory;"*

Rhodes describes a main system memory 103 and a mass storage device 106 shown in Fig. 1 (reproduced below) that store instructions that are executed by the CPU 100.



Refer now to FIG. 1, which illustrates, in block-diagram form, a hardware platform incorporating a representative, generalized embodiment of the invention. As indicated therein, the system includes a central-processing unit ("CPU") 100, which perform operations on and interacts with a main system memory 103 and components thereof. System memory 103 typically includes volatile or random-access memory ("RAM") for temporary storage of information, including

> buffers, executing programs, and portions of the computer's basic operating system. The platform typically also includes read-only memory ("ROM") for permanent storage of the computer's configuration and additional portions of the basic operating system, and at least one mass storage device 106, such a hard disk and/or CD-ROM drive. All components of the platform are interconnected by and communicate over, a bidirectional system bus 110.

(*Appendix D at 9/18-34*)

(ii) *"instructions stored in the memory for operating a method for automatically generating a query from selected document content, comprising:"*

The main system memory 103 stores instructions for operating a method for automatically generating a query from selected document content. (*Appendix D at 9/18-34*). Analysis module 133 of the main system memory 103 directs execution of the instructions to automatically generate a query from selected document content.

> The main memory 103 contains a group of modules that control the operation of CPU 100 and its interaction with the other hardware components. These modules are implemented as executable machine instructions, running (by means of CPU 100) as active processes effectively capable of interacting (i.e., exchanging data and control commands) as illustrated. An operating system 130 directs the execution of low-level, basic system functions such as memory allocation, file management, and operation of mass storage devices 106. At a higher level, an analyzer module 133 directs execution of the primary functions performed by the invention, as discussed below; and instructions defining a user interface 136 allow straightforward interaction over display 126. User interface 136 generates words or graphical images on display 126 to facilitate user action and examination of documents, and accepts user commands from keyboard 120 and/or position-sensing device 123.

(*Appendix D at 10/1-17*) See also discussion of corresponding element of claim 1.

(iii) *"defining an organized classification of document content with each class in the organized classification of document content having associated therewith a classification label; each classification label corresponding to a category of information in an information retrieval system;"*

See discussion of corresponding element of claim 1, where the main memory 103 stores the instructions. (*Appendix D at 10/1-17*)

(iv) *"automatically identifying a set of entities in the selected document content for searching information related thereto using the information retrieval system;"*

See discussion of corresponding element of claim 1, where the main memory 103 stores the instructions. (*Appendix D at 10/1-17*)

(v) *"automatically categorizing the selected document content using the organized classification of document content for assigning the selected document content a classification label from the organized classification of content; and"*

See discussion of corresponding element of claim 1, where the main memory 103 stores the instructions. (*Appendix D at 10/1-17*)

(vi) *"automatically formulating the query to restrict a search at the information retrieval system for information concerning the set of entities to the category of information in the information retrieval system identified by the assigned classification label."*

See discussion of corresponding element of claim 1, where the main memory 103 stores the instructions. (*Appendix D at 10/1-17*)

(m) Rhodes anticipates claim 19

Claim 19 depends from independent claim 18, and further requires *"wherein the instructions stored in the memory further comprise limiting the query by adding terms relating to context information surrounding the set of entities in the selected document content."* As described above, the main system memory 103 of Rhodes stores instructions for operating a method for performing this function. (*Appendix D at 9/18-34*). Analysis module 133 of the main system memory 103 directs execution of the instructions to limit the query by adding terms relating to context information surrounding the set of entities in the selected document content. See discussion of corresponding element of claim 2.

7. Claims 4, 16-17, and 20 are unpatentable over Rhodes taken in view of Stibel under 35 U.S.C. § 103(a)

Requestor respectfully submits that claims 4, 16-17, and 20 are unpatentable over Rhodes taken in view of Stibel under 35 U.S.C. § 103(a). A claim chart applying Rhodes and Stibel to these claims is submitted herewith as Appendix S.

(a) Rhodes and Stibel render obvious claim 4

Claim 4 depends from claim 2, and further requires *"limiting the query by adding terms defining the assigned classification label."* Each wordcode category of Rhodes shares a one-to-one correspondence with a word in the query. (*Appendix D at 4/45-49*) Accordingly, a wordcode category is defined by a word in the query and, therefore, may be further defined by a term that corresponds to the meaning of the word in the query. Stibel teaches adding meaning terms to a query to improve its precision. Therefore, in including meaning terms of words corresponding to wordcodes in the query, the Rhodes system, as modified based on Stibel's teachings, further limits the query by adding terms defining the assigned wordcode category or categories.

In particular, Stibel teaches a system that receives a search query having one or more query terms and processes the search query to generate a new search query that will more effectively retrieve information that is relevant to the original search query. (*Appendix E at 2/42-47, and shown in claim chart at Appendix S*) The processing includes identifying a meaning or Sense of one or more query terms, identifying related terms that are likely to appear on a web page when a user is querying for information associated with that identified meaning, and then modifying the original query to form the new search query by appending the meaning and the one or more related terms to the original query. (*Appendix E at 10/30-36 and 11/56-66, all of which is shown in the claim chart at Appendix S*) For example, if the original query includes only the term "java" and is determined to correspond to the meaning "coffee" and have related terms "espresso" and "beverage," the system may modify the search query to be the new search query "java + coffee + espresso + beverage." (*Appendix E at 12/16-19*) Stibel asserts that this new modified query will result in a more effective search in that the Internet search engine is more likely to return a meaningful hit list in response to the new modified query as compared to the initial query. Stibel teaches that "[a]s expanded, the query now includes terms that are selected to increase the likelihood that an Internet search engine will return a meaningful hit list." *(Appendix E at 12/18-21)*

As noted previously, the system of Rhodes generates keyword queries just like those that are processed by Stibel's system. That is, one of ordinary skill in the art would have easily recognized that the system of Rhodes is ready for improvement through use of Stibel's techniques. Specifically, a person of ordinary skill in the art would have readily recognized that Stibel's techniques for processing a search query could be applied to the Rhode's query to make it more effective at finding desired information by further limiting the query through the addition of meaning terms. Furthermore, the combination of Stibel's teachings with those of Rhodes would involve merely combining and/or substituting known prior art elements to yield predictable results. One skilled in the art could have easily combined the known elements as claimed by known methods, with each element in the combination performing the same function as it does separately, and the combination yielding nothing more than predictable results to one skilled in the relevant art.

(b) Rhodes and Stibel render obvious claim 16

Claim 16 depends from claim 15, and further requires *"wherein the query generator further limits the query by adding terms defining the selected classification label provided by the categorizer."* The analyzer module 133 of the main system memory 103 of the Rhodes system executes (by means of the CPU 100) the operations corresponding to this element. (*Appendix D at 10/1-17).* Additionally, see the discussion of the corresponding element of claim 4, above.

(c) Rhodes and Stibel render obvious claim 17

Claim 17 depends from claim 16, and further requires *"wherein the instructions stored in the memory further comprise further limiting the query by adding terms defining the assigned classification label."* The analyzer module 133 of the main system memory 103 of the Rhodes system executes (by means of the CPU 100) the operations corresponding to this element. (*Appendix D at 10/1-17)* Rhodes teaches enriching the selected document content in that the additional information produced in response to the query is used to annotate the display of the selected document content. For example, if the selected document content is an e-mail displayed in an Emacs interface window, the Rhodes system will display one-line summaries of the most similar documents to the displayed e-mail at the bottom of the Emacs window. *(Appendix D at 1/65-2/6, and shown in claim chart at Appendix S)*

Requestor notes that the '979 patent states that to "enrich" a document means merely to "annotate a document in accordance with a predefined personality." (*Appendix A at col. 6, lines 63 and 64*)

(d) Rhodes and Stibel render obvious claim 20

Claim 20 depends from claim 19, and further requires *"wherein the query generator further limits the query by adding terms defining the selected classification label provided by the categorizer."* The main memory 103 stores the instructions for performing this function. (*Appendix D at 10/1-17)*. Additionally, see the discussion of the corresponding element of claim 4, above.

8. Claims 3 and 8 are unpatentable over Rhodes taken in view of Syskill under 35 U.S.C. § 103(a)

Requestor respectfully submits that claims 3 and 8 are unpatentable over Rhodes taken in view of Syskill under 35 U.S.C. § 103(a). A claim chart applying Rhodes and Syskill to these claims is submitted herewith as Appendix T.

(a) Rhodes and Syskill render obvious claim 3

Claim 3 depends from claim 2, and further requires *"wherein the number of terms added is limited to a predefined number."* Syskill describes a system that develops a user profile for a user that browses the World Wide Web by receiving a rating from the user for each visited Web page and analyzing the information on each Web page. The system is then able to generate a query based on the user profile that is submitted to a search engine to find Web pages that may be of interest to the user. Syskill's system limits the number of terms in the search query submitted to the search engine to a predetermined number (e.g., 14 terms) in recognition of the fact that search engines cannot accept very long queries.

Specifically, Syskill teaches that a system that provides queries to a search engine may limit a query to a predetermined number of terms in recognition of the fact that the search engine may not be able to process very long queries. (*Appendix F at Page 56, Second Column, Lines 3-6, and shown in claim chart at Appendix T*). Syskill operates on keyword queries just like those generated by the system of Rhodes. Syskill expressly teaches what a person of ordinary skill in the art would have already known – i.e., that any realistic implementation of an electronic search engine, like that of Rhodes, has a limited amount of computing resources and, therefore, would limit the number of terms/keywords in a keyword query inputted into the search engine for processing. Accordingly, a person of ordinary skill in the art would have readily recognized that Syskill's limitation of terms in a query to a predetermined number could be easily applied to the search queries of Rhodes in recognition of the limited processing capabilities of search engines, as expressly disclosed by Syskill. Furthermore, the combination of Syskill's teachings with those of Rhodes would involve merely combining and/or substituting known prior art elements to yield predictable results. One skilled in the art could have easily combined the known elements as claimed by known methods, with each element in the combination performing the same function as it does separately, and the combination yielding nothing more than predictable results to one skilled in the relevant art.

(b) Rhodes and Syskill render obvious claim 8

Claim 8 depends from claim 7, and further requires *"abbreviating the set of terms extracted by the text categorizer to a predefined number of terms."* As discussed above, Syskill describes limiting the number of terms in a query to a predefined number. Applying Syskill's teachings to the Rhodes system would result in abbreviating the set of keywords extracted by Rhodes from the selected document content to a predefined number. As described above, a person of ordinary skill in the art would have readily recognized that Syskill's limitation of terms in a query to a predetermined number could be easily applied to the search queries of Rhodes, in recognition of the limited processing capabilities of search engines, as expressly disclosed by Syskill.

9. <u>Claim 9 is unpatentable over Rhodes taken in view of Black, Syskill, and Mauldin under 35 U.S.C. § 103(a)</u>

Requestor respectfully submits that claim 9 is unpatentable over Rhodes taken in view of Black, Syskill, and Mauldin under 35 U.S.C. § 103(a). A claim chart applying Rhodes, Black, Syskill, and Mauldin to this claim is submitted herewith as Appendix U.

(a) Rhodes, Black, Syskill, and Mauldin render obvious claim 9

(i) *"The method according to claim 8, wherein said abbreviating comprises:"*

As described above, claim 8 is rendered obvious by Rhodes taken in view of Syskill.

(ii) *"extracting noun phrases from the selected document content; ranking the noun phrases by those that occur most frequently in the document content; defining a subset of noun phrases by identifying those ranked noun phrases that occur more frequently than a first predefined frequency; ranking those words in the subset of noun phrases by their frequency of occurrence to define an ordered list of words; defining a subset of the ordered list of words by identifying those ranked words that occur more frequently than a second predefined frequency;"*

Black discloses extracting words (which may be noun phrases, since a noun phrase can be a single word) from the selected document content and performing a frequency analysis of the words in the selected document content. The words are then ranked by order of frequency of occurrence in the selected document content. *(Appendix G at 3/29-32, and shown in claim chart at Appendix U)*

A subset of the words is defined based upon a predefined desired frequency such that only a subset of the words in the selected document content is thereby included as the keywords in the query. Specifically, Black teaches: "[t]hose [words] that match [the words in the pre-selected keyword list], based upon a desired frequency become keywords to be combined with the hook to form focused, optimal queries."*(Appendix G at 3/33-35) (emphasis added)*

As evidenced by Rhodes, a person of ordinary skill in the art would have recognized that the importance of a keyword in a selected document content is directly proportional to the frequency of occurrence of the word in the document. (*Appendix D at 3/6-15*, where the document term weight DTW is shown to be directly proportional to the term frequency "tf").

In view of this, one skilled in the art would have understood that Black's indication that words are identified for inclusion in a query "based upon a desired frequency" indicates that only those words having frequencies above a predefined desired frequency are chosen for inclusion in the subset of words selected by Black for use in the query.

Accordingly, applying Black's teachings to the query generated by Rhodes results in Rhodes abbreviating the terms in the query to a predefined number by ranking words extracted from the selected document content based on their frequency of occurrence in the selected document and then reducing the number of words to a subset of words by selecting only those words that occur more frequently than a predefined desired frequency.

A person of ordinary skill in the art would have readily appreciated that Black's teachings could be applied to the system of Rhodes to decrease the number of keywords in the query of Rhodes by ranking the words of the selected document in accordance with their frequency of occurrence in the selected document and then using only the subset of the words that satisfy a desired frequency (i.e., a predefined frequency threshold) as the keywords in the query. The person of ordinary skill in the art would have readily recognized, based on the teachings of Syskill, that such a reduction in the number of keywords in the query would be desirable to allow the query to be processed by search engines having limited query processing capabilities.

(iii) *"re-ranking the subset of words in inverse frequency to their use in the category of information in the information retrieval system identified by the assigned classification label; using only those highest ranked words in the re-ranked subset of words to define the set of terms appended to the query."*

Mauldin teaches that keywords extracted from a document should be weighted by both their frequency in the document (i.e., term frequency) and the inverse frequency to their use in the corpus of documents being searched (i.e., inverse document frequency). Mauldin describes ranking the keywords based on their term frequency-inverse document frequency and selecting only the highest ranks words (e.g., the top 100) for searching. *(Appendix I at 6/64-67, 7/1-13, and shown in claim chart at Appendix U)*

Applying Mauldin's teachings to the Rhodes/Black query, which includes the subset of words, would result in a re-ranking of the subset of words in inverse frequency to their use in the corpus being searched and a selection of only the highest ranked words in the re-ranked subset (e.g., the top 100 words) for inclusion the query.

One skilled in the art would have appreciated that the subset of words identified for the query based on the Rhodes/Black teachings could be re-ranked in inverse frequency to their use in the category of information to be searched (i.e., the corpus) and only the highest ranked words would be allowed to remain as keywords in the query. The person of ordinary skill in the art would have readily recognized, based on the teachings of Syskill, that such a further reduction in the number of keywords in the query would be desirable to allow the query to be processed by search engines having limited query processing capabilities.

10. <u>Claims 1, 2, 4-7 and 10-20 are unpatentable over Black taken in view of HyPursuit under 35 U.S.C. § 103(a)</u>

Requestor respectfully submits that claims 1, 2, 4-7 and 10-20 are unpatentable over Black taken in view of HyPursuit under 35 U.S.C. § 103(a). A claim chart applying Black and HyPursuit to these claims is submitted herewith as Appendix V.

(a) Black and HyPursuit render obvious independent claim 1

(i) *"A method for automatically generating a query from selected document content, comprising:"*

Black teaches automatically generating a query by analyzing a body of selected content, such as an article on the SR-71 Blackbird Airplane. The automatic generation of the query involves an automatic generation of a hook, i.e., a concept, primary subject matter or main topic of the body of selected content, and the automatic generation of keywords from the body of selected content. A query is then formed by "AND"ing a keyword with the hook. For example, the automatic query that may be generated for the SR-71 BlackBird Airplane article may be "BLACKBIRD and AVIATION", where Blackbird is the main topic automatically determined for the article and aviation is a keyword identified in the article. (*Appendix G at 1/64-67*)

Thus, the Black reference teaches "[a] system for conducting queries from any document displayed on any computer device." (*Appendix G at Abstract*) Specifically, the Black reference teaches "automatic … statistical and empirical analysis" (*Appendix G at 2/43*) of a body of selected content that may be "text, such as magazine articles, news stories or any other text" (*Appendix G at 1/47-48*) where the text can be an online article (*Appendix G at FIG. 1, 4/18-20*).

The Black reference uses its analysis results to automatically generate a group of suggested queries for finding additional information about the document. Black teaches that "[b]rilliant queries require a preparation process that analyzes any text to enhance and generate a set of suggested searches based on that analysis and certain pre-set user parameters." (*Appendix G at 1/50-54*)

The automatic generation of the query involves analyzing the text of the selected document content to identify a hook term that reflects the primary subject matter of the document.

> The hook is the concept, primary subject matter or main topic for a body of text. The hook is used to define a query as narrowly as possible on a particular topic for a selected information source.

(*Appendix G at 2/10-14*)

> Automatic Generation of the Hook – One embodiment of the brilliant query to enable an automatic process for generating brilliant queries for a body of text, is to determine the hook by extracting the highest frequency proper names from the text body.

(*Appendix G at 2/57-61*).

The automatic generation of the query also involves analyzing the text of the selected document content to identify one or more keywords that are reflective of additional topics for the search.

> A brilliant query requires a list of keywords that are generated by automatic … statistical and empirical analysis of the body of content to be enhanced or a comparable body of content."

(*Appendix G at 2/42-45*)

> Keywords are simply a collection of words, generated automatically … , that are deemed to be indicative of the topic matter or one of the topics for a given content selection.

(*Appendix G at 2/26-29*)

> Automatic Generation of Keywords – A word frequency analysis is done on all of the text, with stopwords excluded, and the resulting words, by order of frequency are compared to a pre-selected keyword list. Those that match, based upon a desired frequency become keywords to be combined with the hook to form focused, optimal queries.

(*Appendix G at 3/29-35*)

A suggested query for the document is generated automatically by ANDing the hook and a keyword.

> Once the keywords have been selected and the hook for a body of text has been determined or automatically generated, the searches are created by generating a link for every keyword extracted from the body of text and combining it with the hook in a search that results in a result set that is the logical intersection of the results generated by the hook and the keyword. Basically, each entry in the list of search results must contain both the hook and the keyword and not just one or the other.

(Appendix G at 3/37-45)

> After the hook and the keywords have been established the query is conducted selecting one of the hook-keyword sets.

(Appendix G at 3/55-57)

(ii) *"defining an organized classification of document content with each class in the organized classification of document content having associated therewith a classification label; each classification label corresponding to a category of information in an information retrieval system;"*

HyPursuit teaches categorizing hypertext documents using a cluster heirarchy shown in Fig. 3 of HyPursuit, reproduced below:

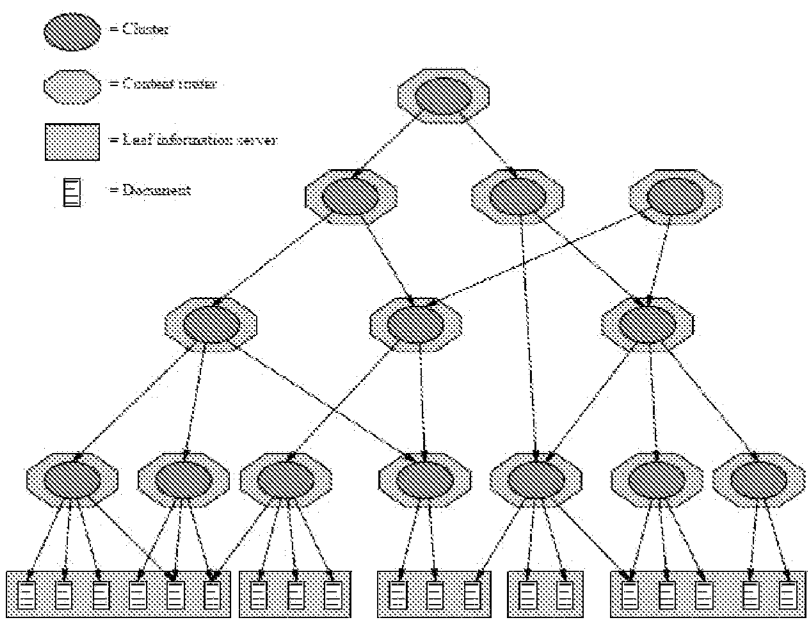


Figure 3: Content Routing as Cluster Hierarchies

As shown in Fig. 3, documents are organized into different clusters that are each handled by a different content router. If, under a broadest reasonable construction standard, a cluster is considered a class and a set of clusters is considered an organized classification of document content (the assumption), the HyPursuit system defines an organized classification of document content by grouping document content into different clusters, each cluster having an assigned content label. Given this assumption, the content label of a cluster corresponds to a category of information in HyPursuit's information retrieval system.

Specifically, HyPursuit teaches categorizing documents into a hierarchy of clusters of related documents.

> The HyPursuit prototype is a scalable system that uses *content-link* hypertext clustering, based on document contents and link information, to structure the information space and to support the entire range of search activities. Content-link clustering automatically computes sets of related documents called *clusters*. HyPursuit admits multiple coexisting cluster hierarchies based on different principles of grouping documents, such as the Library of Congress catalog scheme and institutional structures. These hierarchies may be constructed automatically or manually.

(*Appendix J at page 180, column 2*) (emphasis included)

> For example, documents can be clustered based on institutional boundaries or based on Library of Congress catalog subjects.

(*Appendix J at page 184, column 1*)

A separate information server may represent each cluster and perform operations on its local data.

> Clusters also provide convenient units for the partitioning of work and resource allocation among the distributed components of the system. For example, a separate information server on a separate host may represent each individual cluster, performing operations on its local data.

(*Appendix J at page 181, column 1 and Fig. 3*, which shows one content router server per cluster)

Each cluster has associated with it a content label that summarizes it.

> Each content router users its abstraction functions to compute a content label that summarizes its associated cluster.

(*Appendix J at page 184*)

The query taught by Black is a keyword query, just like those processed by HyPursuit's search engine. (*Appendix J at Fig. 6*). Accordingly, one of ordinary skill in the art would have easily recognized that Black's system is ready for improvement through use of HyPursuit's techniques. Specifically, a person of ordinary skill in the art would have readily appreciated that HyPursuit's techniques could be used to process Black's keyword query to enable scalable and efficient query processing. In applying HyPursuit's techniques to Black's query, Black's document would be automatically categorized via identification of one or more clusters as corresponding to Black's query, and the corresponding category or categories would be used to restrict application of Black's query to only those information servers corresponding to the identified cluster(s).[9] Furthermore, the combination of HyPursuit's teachings with those of Black would involve merely combining and/or substituting known prior art elements to yield predictable results. One skilled in the art could have easily combined the known elements as claimed by known methods, with each element in the combination performing the same function as it does separately, and the combination yielding nothing more than predictable results to one skilled in the relevant art. Among other features, the disclosure of the Black and HyPursuit references teach items deemed missing from prior art during original prosecution, and responsible for Examiner allowance of the claims.

(iii) *"automatically identifying a set of entities in the selected document content for searching additional information related thereto using the information retrieval system;"*

Black teaches automatically identifying keywords from the selected body of content by performing a word frequency analysis on all of the text, with stop words excluded, and the resulting words, by order of frequency, are compared to a pre-selected keyword list. Those that match are deemed the keywords for which additional information will be sought through inclusion of the keywords in automatically generated search queries.

Specifically, Black teaches automatic identification of keywords from the selected body of content that are indicative of a topic matter or one of the topics for the body of content. Black teaches that "[k]eywords are simply a collection of words, generated automatically … , that are deemed to be indicative of the topic matter or one of the topics for a given content selection." *(Appendix G at 2/26-29*)

---

[9] Requestor applies the Patent Owner's apparent claim construction in the related litigation without waiver. See *n.* 1.

The list of keywords can be generated by an automatic statistical and empirical analysis of the selected body of content to be enhanced. As taught by Black, "[a] brilliant query requires a list of keywords that are generated by automatic … statistical and empirical analysis of the body of content to be enhanced or a comparable body of content." (*Appendix G at 2/42-45*)

The analysis to determine the list of keywords is a word frequency analysis that is done on all of the text of the selected body of content, with stopwords excluded, and the resulting words are then compared to a pre-selected keyword list. Words that match the pre-selected keyword list are retained as the keywords for the selected body of content based upon a desired frequency.

> Automatic Generation of Keywords – A word frequency analysis is done on all of the text, with stopwords excluded, and the resulting words, by order of frequency are compared to a pre-selected keyword list. Those that match, based upon a desired frequency become keywords to be combined with the hook to form focused, optimal queries.

(*Appendix G at 3/29-35*).

(iv) *"automatically categorizing the selected document content using the organized classification of document content for assigning the selected document content a classification label from the organized classification of content;"*

Black teaches automatically identifying a query that includes a hook search term that is reflective of the main topic or primary subject matter of the selected body of content. HyPursuit teaches receiving a query and automatically classifying the query into one or more clusters/categories, thereby assigning the query the corresponding content label(s).[10] HyPursuit's assignment of one or more content labels to the query is used to enable more efficient and scalable query processing by leveraging a cluster hierarchy to restrict the processing of the query to only those information servers corresponding to the assigned content label(s).

Black's query is a keyword query, just like those processed by HyPursuit's search engine. (*Appendix J at Fig. 6*). Accordingly, one of ordinary skill in the art would have easily recognized that Black's system is ready for improvement through use of HyPursuit's techniques. Specifically, a person of ordinary skill in the art would have readily appreciated that HyPursuit's techniques could be used to

---

[10] As stated previously, Requestor applies the Patent Owner's apparent claim construction in the related litigation without waiver. See *n*.1.

process Black's keyword query to enable scalable and efficient query processing. In applying HyPursuit's techniques to Black's query, Black's document would be automatically categorized via identification of one or more clusters as corresponding to Black's query, and the corresponding category or categories would be used to restrict application of Black's query to only those information servers corresponding to the identified cluster(s).

Specifically, Black teaches automatically identifying a query that includes a hook search term that is reflective of the primary subject matter of a document.

> The hook is the concept, primary subject matter or main topic for a body of text. The hook is used to define a query as narrowly as possible on a particular topic for a selected information source. To determine a "hook", a content layer must exist for which a context can be determined. There must be a perceivable structure to the information source and each content entry must have an associated context or place or places within the structure of the information source.

(*Appendix G at 2/10-18*)

> Automatic Generation of the Hook – One embodiment of the brilliant query to enable an automatic process for generating brilliant queries for a body of text, is to determine the hook by extracting the highest frequency proper names from the text body.

(Appendix G at 2/57-61) HyPursuit teaches receiving a query and automatically identifying one or more content labels of clusters relevant to the query and then forwarding the query to only those information servers corresponding to the identified content labels.

> To support a variety of query processing operations, HyPursuit uses query routing to identify relevant clusters, forward queries to the information servers for those clusters, and merge the results.

(*Appendix J at page 182, column 1*)

> HyPursuit uses query routing to support the search operations. Query routing uses the content labels stored in the content router to determine which of the child servers are likely to contain documents related to the user query. The query is then forwarded to these servers, and the results from each server are merged into a single result set. Documents returned by more than one child server are displayed only once.

(*Appendix J at page 186, column 2*)

> The abstraction function for query routing, on the other hand, computes a manageable set of terms that are used for identifying portions of the information space relevant to particular queries.

(*Appendix J at page 185, column 1*) HyPursuit teaches that such a restriction is desirable in that it provides scalable query processing.

> To support operations like query processing in a scalable way, HyPursuit uses manageable summaries of cluster contents, called *content labels*, to approximate complete knowledge of the information space

(*Appendix J at page 184, column 1*) (emphasis included)

> To support scalable query processing, HyPursuit uses manageable summaries of cluster contents, called *content labels*, to approximate complete knowledge of the information space

(*Appendix J at page 181, column2*) (emphasis included)

(v) *"automatically formulating the query to restrict a search at the information retrieval system for information concerning the set of entities to the category of information in the information retrieval system identified by the assigned classification label."*

Black teaches automatically formulating a query through the automatic generation of a hook for the selected body of content, the automatic generation of keywords from the selected body of content, and then ANDing the hook and one of the keywords to automatically form the query.

For the reasons stated above, a person of ordinary skill in the art would have readily recognized that Black's selected body of content could be assigned one or more content labels through the categorization of Black's query in accordance with HyPursuit's teachings and that Black's search could then be restricted to the category of information identified by the assigned content labels in order to provide scalable and efficient query processing as taught by HyPursuit.

Specifically, Black teaches automatically formulating a query by ANDing an automatically generated hook and an automatically generated keyword determined from selected document content.

> Once the keywords have been selected and the hook for a body of text has been determined or automatically generated, the searches are created by generating a link for every keyword extracted from the body

> of text and combining it with the hook in a search that results in a result set that is the logical intersection of the results generated by the hook and the keyword. Basically, each entry in the list of search results must contain both the hook and the keyword and not just one or the other.

(*Appendix G at 3/37-45*)

> After the hook and the keywords have been established the query is conducted selecting one of the hook-keyword sets.

(*Appendix G at 3/55-57*) HyPursuit teaches receiving a query and automatically identifying one or more content labels of clusters relevant to the query and then forwarding the query to only those information servers corresponding to the identified content labels.

> To support a variety of query processing operations, HyPursuit uses query routing to identify relevant clusters, forward queries to the information servers for those clusters, and merge the results.

(*Appendix J at page 182, column 1*)

> HyPursuit uses query routing to support the search operations. Query routing uses the content labels stored in the content router to determine which of the child servers are likely to contain documents related to the user query. The query is then forwarded to these servers, and the results from each server are merged into a single result set. Documents returned by more than one child server are displayed only once.

(*Appendix J at page 186, column 2*)

> The abstraction function for query routing, on the other hand, computes a manageable set of terms that are used for identifying portions of the information space relevant to particular queries.

(*Appendix J at page 185, column 1)* HyPursuit teaches that such a restriction is desirable in that it provides scalable query processing.

> To support operations like query processing in a scalable way, HyPursuit uses manageable summaries of cluster contents, called *content labels*, to approximate complete knowledge of the information space

(*Appendix J at page 184, column 1*) (emphasis included)

> To support scalable query processing, HyPursuit uses manageable summaries of cluster contents, called *content labels*, to approximate complete knowledge of the information space

(*Appendix J at page 181, column2*) (emphasis included)

(b) Black and HyPursuit render obvious claim 2

Claim 2 depends from independent claim 1, and further requires *"limiting the query by adding terms relating to context information surrounding the set of entities in the selected document content."* HyPursuit describes further limiting a query by adding refinement terms to the query that correspond to the highest weighted terms in the clusters (specifically, the sub-clusters) identified as matching the query. These terms are referred to as "collocated terms" in that they are terms that are located in documents within the same sub-cluster that matches the terms of the original query. (*Appendix J at page 186, column 2*) HyPursuit teaches that adding these terms to the original query can improve both the recall and the precision of the searching. (*Appendix J at page 186, column 2*).

Black's query is a keyword query, just like those processed by HyPursuit's search engine. (*Appendix J at Fig. 6*). Accordingly, one of ordinary skill in the art would have easily recognized that Black's system is ready for improvement through use of HyPursuit's techniques. That is, a person of ordinary skill in the art would have readily appreciated that HyPursuit's techniques to improve a query could be used to process Black's keyword query to thereby improve the recall and the precision of the resulting search. In applying HyPursuit's techniques to Black's query, Black's query would be modified to further include terms that are located in documents within the same sub-cluster that matches the terms of Black's query, as taught by HyPursuit. (*Appendix J at page 186, column 2*)

In particular, in applying HyPursuit's teachings to Black, the terms of the original query correspond to keywords extracted from Black's selected body of content. The additional collocated terms added to these keywords in accordance with Hypursuit's teachings, therefore, relate to context information surrounding the keywords of Black's selected body of content in that they relate to words that are commonly found in documents within the clusters/sub-clusters that match the Black keywords.

Furthermore, these additional terms relate to context information surrounding the keywords in the selected body of content, rather than to context information surrounding other portions of the body of content or surrounding a different body of content entirely. For example, if the selected body of content is an article on the SR-71 Blackbird and Black's query is "Blackbird AND aviation," the collocated terms relate to context information surrounding the terms "Blackbird" and "aviation," rather than to context information surrounding other terms in the article, such as "Elint," "reconnaissance" or "transport," or surrounding a different article entirely. (*Appendix G at 1/64-2/9*)

Specifically, Hypursuit teaches further limiting a query through a query refinement process that adds collocated terms to the query. The collocated terms are the highest weighted terms in the clusters (specifically, the sub-clusters) identified as matching the query. Hypursuit teaches that adding these terms to the original query can improve both the recall and the precision of the searching. (*Appendix J at page 186, column 2, and shown in claim chart at Appendix V*)

(c) Black and HyPursuit render obvious claim 4

Claim 4 depends from claim 2, and further requires *"limiting the query by adding terms defining the assigned classification label."* As stated above, HyPursuit teaches adding collocated terms to the query to further limit the query. The collocated terms define the content label(s) assigned to the query in that, as shown in Fig. 5 of HyPursuit (reproduced below), the content label is defined as a set of attributes that include the collocated terms. The example content label shown in Fig. 5 shows a cluster that has three sub-clusters, with the first sub-cluster including the terms "committee" and "eecs" (among others) as potential collocated terms that may be added to a query. *(Appendix J at page 185, column 1 and column 2 to page 186, column 1, all of which is shown in claim chart at Appendix V*)

```
({version: 1}
 {url: "http://www.psrg.../cgi-bin/cra/www-eecs.mit.edu"}

 {cluster
  ({cluster-num: 1) (size: 6763) (num-docs 5) (num-terms: 485)
  (url: "http://www-eecs  /committees.html")
  (url  "http://www-eecs.../hq/index.html")
  (url: "http://www-eecs.../staff.html")
  (url. "http://www-eecs.../stucrg.html")
  (url. "http://www-eecs.../test pictures html")
  (term: ((attribute: header) (value: administrative) (tf: 1) (df: 1)))
  (term: ((attribute. header) (value  committees) (tf: 2) (df  1)))
  (term: ((attribute: header) (value  computer) (tf: 5) (df  5)))
  (term: ((attribute: header) (value: eecs) (tf: 5) (df: 5)))
  (term: ((attribute: header) (value: department) (tf  11) (df: 5)))
  ...
  (term: ((value: committee) (tf: 13) (df: 1)))
  (term: ((value: eecs) (tf: 27) (df: 5)))
  (term: ((value: officer) (tf: 4) (df. 1)))
  (term: ((value: organizations) (tf: 2) (df: 1)))
  (term: ((value: personal) (tf: 3) (df: 1)))
  ..  ))

 {cluster:
  ((cluster-num: 2) (size: 17385) (num-docs: 8) (num-terms: 1080)
  (url: "http://www-eecs.../AY94-95/announcements/index.html")
  (url: "http://www-eecs.../AY95-96/announcements/1.html")
  (url: "http://www-eecs  ./AY95-96/announcements/2.html")
  (url: "http://www-eecs.  /AY95-96/announcements/3.html")
  (url. "http://www-eecs   /AY95-96/announcements/index.html")
  (url: "http //www-eecs.. /current/announcements/index.html")
  (term: ((attribute: header) (value: ay95) (tf: 1) (df: 1)))
  (term  ((attribute: header) (value: computer) (tf: 8) (df: 8)))
  (term. ((attribute: header) (value: current) (tf: 1) (df: 1)))
  (term: ((attribute: header) (value: department) (tf: 17) (df: 8)))
  ...
  (term: ((value: announcements) (tf: 22) (df: 8)))
  (term  ((value: department) (tf  32) (df  8)))
  (term: ((value: eecs) (tf: 38) (df. 8)))
  ...
  (term. ((attribuse: title) (value: announcements) (tf: 8) (df: 8)))
  (term: ((attribute: title) (value: eecs) (tf: 8) (df: 8)))
  ))

 {cluster:
  ((cluster-num. 3) (size. 6706) (num-docs  3) (num-terms: 1108)
  (url: "http://www-eecs.../class-materials.html")
  (url. "http://www-eecs   /comment-form.html")
  (url: "http://www-eecs.../web-pages.html")
  (term: ((attribute: header) (value: comment) (tf: 1) (df: 1)))
  (term: ((attribute: header) (value: course) (tf. 1) (df. 1)))
  (term: ((attribute: header) (value: csars) (tf: 1) (df: 1)))
  (term: ((attribute: keyword) (value: mail) (tf: 1) (df: 1)))
  (term: ((attribute: keyword) (value: name) (tf: 1) (df: 1)))
  ...
  (term: ((attribute: title) (value: form) (tf: 1) (df: 1)))
  (term: ((attribute. title) (value: home) (tf: 1) (df: 1)))
  (term: ((attribute: title) (value: pages) (tf: 3) (df: 3)))
  ))
 ...
)
```

Figure 5: Sample Clustered Content Label

(d) Black and HyPursuit render obvious claim 5

Claim 5 depends from independent claim 1, and further requires *"wherein the organized classification of document content is defined using a hierarchical organization."* If, under a broadest reasonable construction standard, a cluster is considered a class and a set of clusters is considered an organized classification of document content (the assumption), HyPursuit teaches that the organized classification of document content is defined using a hierarchical organization as shown in Fig. 3 of

HyPursuit, reproduced above. That is, as shown in Fig. 3, documents are categorized into different clusters in a hierarchy of clusters and each cluster is handled by a particular content router. The documents themselves are handled by leaf information servers, which are located at the lowest level of the heirarchy. *(Appendix J at page 180, column 2 and at page 184, column 1, all of which is shown in the claim chart at Appendix V)*

(e) Black and HyPursuit render obvious claim 6

Claim 6 depends from independent claim 1, and further requires *"using a text categorizer to assign the classification label assigned from the organized classification of content."* The Black system, as modified based on HyPursuit's teachings, uses a text categorizer to assign one or more content labels corresponding to clusters of documents in accordance with HyPursuit's teachings, where the query is derived from a document and reflects the primary subject matter of the document in accordance with Black's teachings. The one or more content labels are assigned from an organized classification of content. (*Appendix G at 2/10-18 and 2/10-18, 2/57-61, Appendix J at page 182, column 1, Appendix J at page 186, column 2 and Appendix J at page 185, column 1, all of which is shown in Appendix V*)

(f) Black and HyPursuit render obvious claim 7

Claim 7 depends from claim 6, and further requires *"extracting with the text categorizer a set of terms relating to the document content; and appending to the query ones of the set of terms extracted by the text categorizer to contextualize the query."* Black extracts the hook and keywords from the selected document content and appends the hook and/or a keyword to the query to contextualize the query. As stated previously, HyPursuit teaches that collocated terms can be extracted from sub-clusters related to the query and also can be added to the query to contextualize the query. The hook, keyword, and/or collocated terms contextualize the query by specifying the environment from which the query was generated – i.e., the textual contents of the selected document and its matching sub-clusters.

Specifically, Black describes automatically analyzing the textual contents of the selected document to determine the primary subject matter or topic of the document. Black teaches that "[t]he hook is the concept, primary subject matter or main topic for a body of text. The hook is used to define a query as narrowly as possible on a particular topic for a selected information source." *(Appendix G at 2/10-14)* Black also describes automatically analyzing the textual contents of the selected document to determine keywords that correspond to one or more other topics of the document. Black teaches that "[k]eywords are simply a collection of words, generated automatically … , that are deemed to be indicative of the topic matter or one of the topics for a given content selection." (*Appendix G at 2/26-29*) Black describes adding the hook and/or keyword terms to queries to provide additional context information for the queries that reflect the document content from which the queries are derived. (*Appendix G at 3/37-45, and shown in claim chart at Appendix V)*

HyPursuit describes further limiting a query through a query refinement process that adds collocated terms to the query. The collocated terms are the highest weighted terms in the clusters (specifically, the sub-clusters) identified as matching the query. HyPursuit teaches that adding these terms to the original query can improve both the recall and the precision of the searching. (*Appendix J at page 186, column 2, and shown in claim chart at Appendix V*)

(g) Black and HyPursuit render obvious claim 10

Claim 10 depends from independent claim 1, and further requires *"wherein each class in the organized classification of document content has associated therewith a characteristic vocabulary."* HyPursuit describes that each cluster in HyPursuit's organized classification of document content[11] has associated with it a set of terms that characterize the cluster for purposes of query routing. Specifically, HyPursuit teaches that each content router uses an abstraction function to compute a manageable set of terms used to identify the router's cluster for purposes of identifying portions of the information space relevant to particular queries. (*Appendix J at page 185, column 1, and shown in claim chart at Appendix V*)

(h) Black and HyPursuit render obvious claim 11

Claim 11 depends from claim 10, and further requires *"ranking results from the query performed at the information retrieval system in accordance with one of the assigned classification label and the characteristic vocabulary."* HyPursuit teaches that a result set of documents produced in response to a query can be ranked in accordance with the assigned content label(s) in that the documents may be grouped according to the sub-clusters specified in the assigned content label(s). In other words, the ranking of the search result documents is in accordance with the assigned content label(s) in that the ranking is consecutive for search result documents within the same assigned sub-cluster. An example of this is shown in Fig. 7 of HyPursuit (reproduced below), which shows one document in a first sub-cluster (i.e., "James William O'Toole Jr"), three documents ranked consecutively in a second sub-cluster

---

[11] As stated previously, Requestor applies the Patent Owner's apparent claim construction in the related litigation without waiver. See *n.* 1.

(i.e., "MIT Programming Systems Research Group", "MIT Programming Systems Research Group" and "Programming Systems Research Group Publications") , and one document in a third sub-cluster (i.e., "Mark Sheldon's Publications"). (*Appendix J at page 186, column 2, and shown in claim chart at Appendix V*)

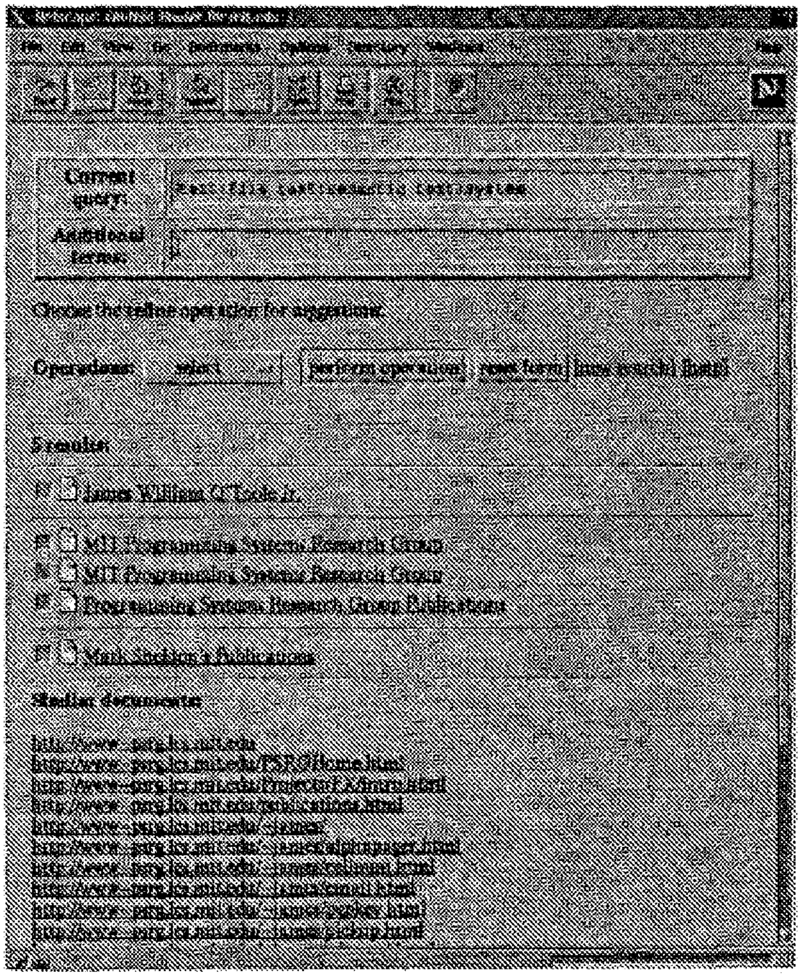
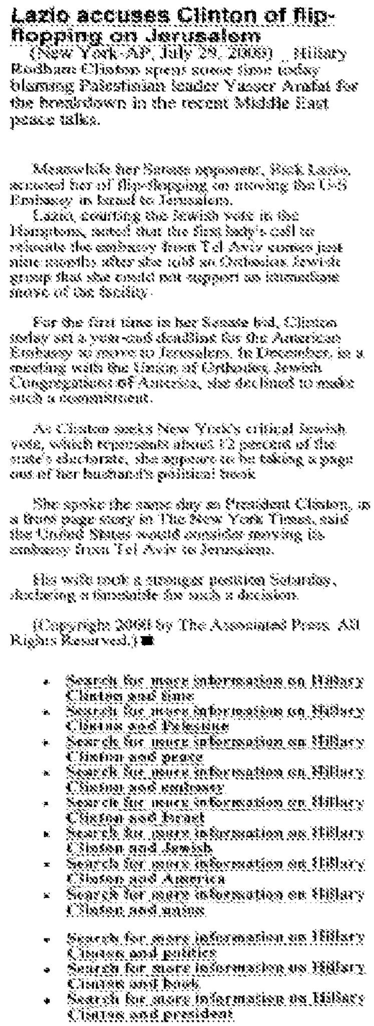


Figure 7: Suggesting Additional Documents

(i) Black and HyPursuit render obvious claim 12

Claim 12 depends from claim 11, and further requires *"using the method in a system for enriching selected content of a document with personalities that identify enrichment themes."* The Black reference teaches enriching the selected document by annotating the selected document to include the automatically generated queries as hyper-links, as shown in FIG. 5 of Black (reproduced below).



Lazio accuses Clinton of flip-flopping on Jerusalem

(New York-AP, July 29, 2000) _ Hillary Rodham Clinton spent some time today blaming Palestinian leader Yasser Arafat for the breakdown in the recent Middle East peace talks.

Meanwhile her Senate opponent, Rick Lazio, accused her of flip-flopping on moving the US Embassy in Israel to Jerusalem.

Lazio, courting the Jewish vote in the Hamptons, noted that the first lady's call to relocate the embassy from Tel Aviv comes just nine months after she told an Orthodox Jewish group that she could not support an immediate move of the facility.

For the first time in her Senate bid, Clinton today set a year-end deadline for the American Embassy to move to Jerusalem. In December, in a meeting with the Union of Orthodox Jewish Congregations of America, she declined to make such a commitment.

As Clinton seeks New York's critical Jewish vote, which represents about 12 percent of the state's electorate, she appears to be taking a page out of her husband's political book.

She spoke the same day as President Clinton, in a front page story in The New York Times, said the United States would consider moving its embassy from Tel Aviv to Jerusalem.

His wife took a stronger position Saturday, declaring a timetable for such a decision.

- Search for more information on Hillary Clinton and time
- Search for more information on Hillary Clinton and Palestine
- Search for more information on Hillary Clinton and peace
- Search for more information on Hillary Clinton and embassy
- Search for more information on Hillary Clinton and Israel
- Search for more information on Hillary Clinton and Jewish
- Search for more information on Hillary Clinton and America
- Search for more information on Hillary Clinton and [illegible]
- Search for more information on Hillary Clinton and politics
- Search for more information on Hillary Clinton and book
- Search for more information on Hillary Clinton and president

Fig. 5

As shown in Fig. 5, the article titled "Lazio accuses Clinton of flip-flopping on Jerusalem" has been annotated to include various hyperlinks including the hyperlink "Search for more information on Hillary Clinton and time." *(Appendix G at FIG. 5 and 4/30-34, and shown in claim chart at Appendix V)* Requestor notes that the '979 patent states that to "enrich" a document means merely to "annotate a document in accordance with a predefined personality." (*Appendix A at col. 6, lines 63 and 64*)

(j) Black and HyPursuit render obvious claim 13

Claim 13 depends from claim 1, and further requires *"automatically identifying the set of entities using a service that recognizes entities of a predefined type."* The Black reference teaches automatically identifying the keywords using a service that recognizes words of a predefined type, where the predefined type of words corresponds to words that are not stop words. (*Appendix G at 2/45-49 and 3/30/31, and shown in claim chart at Appendix V*)

(k) Black and HyPursuit render obvious independent claim 14

(i) *"A system for automatically generating a query from selected document content, comprising:"*

Black discloses a system that generates a query from selected document content.

> 1. A system for submitting to an information source, a query based upon any document displayed on any computer device comprising, conducting a statistical frequency analysis of the word occurrence in the document to determine the subject matter of the document, the subject matter being the most frequently occurring proper name or non-stopword, and the keywords in the document, the keywords being determined by the frequency analysis or by comparison to a pre-selected keyword list and creating queries by combining the subject matter with each of the keywords.

(*Appendix G at 4/58-67)* The system enables display of an HTML page of a document in which hyperlinks are automatically added by the system.

> Brilliant queries require a preparation process that analyzes any text to enhance and generate a set of suggested searches based on that analysis and certain pre-set user parameters. The output of this preparation process can be used to add links to an HTML page of a document either automatically or through manual insertion of the resulting analysis.

(Appendix G at 1/50-57) In describing that the system includes a computer device that displays an HTML page that has been automatically annotated to include links corresponding to queries, Black contemplates performing the query generation and invocation processes through use of a computer device having a corresponding memory and processor. See also discussion of the corresponding element of claim 1.

(ii) *"an entity extractor for automatically identifying a set of entities in the selected document content for searching information related thereto using an information retrieval system;"*

See discussion of the corresponding element of claim 1, where the computer device of Black performs the recited function and, thereby, includes the recited entity extractor. (*Appendix G at 4/58-67 and 1/50-57*).

(iii) *"a categorizer for defining an organized classification of document content with each class in the organization of content having associated therewith a classification label; each classification label corresponding to a category of information in the information retrieval system; the categorizer automatically assigning the selected document content a classification label from the organized classification of content;"*

See discussion of corresponding element of claim 1. Each content router/information server of HyPursuit is implemented as a CGI script that is run on a Sun SparcStation. (*Appendix J at page 189, column 2*). Accordingly, the categorizer is the electrical components within the Sun SparcStations that run the CGI scripts corresponding to the content routers/information servers that perform the categorization operations set forth previously with respect to claim 1. (*Appendix J at page 184 to page 186 and page 189, column 2).*

(iv) *"a query generator for automatically formulating the query to restrict a search at the information retrieval system for information concerning the set of entities to the category of information in the information retrieval system identified by the assigned classification label."*

See discussion of corresponding element of claim 1. Black's computer device automatically generates a query from a selected document that is processed by HyPursuit's content routers/information servers to produce search results. (*Appendix G at 4/58-67 and 1/50-57*) Each content router/information server is implemented as a CGI script that is run on a Sun SparcStation. (*Appendix J at page 189, column 2*). Accordingly, the query generator corresponds to Black's computer device in combination with components of HyPursuit's Sun SparcStations that together automatically formulate a search query that is restricted to only those information servers corresponding to clusters identified by the assigned content labels. (*Appendix J at page 182, column 1; page 185, column1; page 186, column 2)*

(l) Black and HyPursuit render obvious claim 15

Claim 15 depends from independent claim 14, and further requires *"a short length aspect vector generator for generating terms relating to context information surrounding the set of entities in the selected document content; wherein the query generator adds the terms relating to the context information to limit the query."* See discussion of corresponding element of claim 2 above, where Black's computer device in combination with HyPursuit's Sun SparcStations corresponds to a generator that adds collocated terms to the automatically generated query to further refine the query.

(m) Black and HyPursuit render obvious claim 16

Claim 16 depends from claim 15, and further requires *"wherein the query generator further limits the query by adding terms defining the selected classification label provided by the categorizer."* Black's computer device in combination with HyPursuit's Sun SparcStations further refine the query by adding collocated terms that define the content label(s) used to categorize the selected document.

As stated above, HyPursuit teaches adding collocated terms to the query to further limit the query. The collocated terms define the content label(s) assigned to the query in that, as shown in Fig. 5 of HyPursuit (reproduced above), the content label is defined as a set of attributes that include the collocated terms. The example content label shown in Fig. 5 shows a cluster that has three sub-clusters, with the first sub-cluster including the terms "committee" and "eecs" (among others) as potential collocated terms that may be added to a query. *(Appendix J at page 185, column 1 and Appendix J at page 2 to page 186, column 1*)

(n) Black and HyPursuit render obvious claim 17

Claim 17 depends from claim 16, and further requires *"a content manager for enriching the selected document content with results provided from the information retrieval system using the query."* Black teaches a computer device that automatically generates an initial query from selected document content, the query being composed of a hook and a keyword. HyPursuit teaches using this query to identify additional collocated terms that can then be appended to the initial query to improve the precision and the recall of the search. Black teaches that the computer device enriches the selected document content by annotating it to include hyperlinks corresponding to a query for display to a user.

Accordingly, the recited results provided by the information retrieval system resulting from the combination of Black's and HyPursuit's teachings using the query (i.e., the initial query) are the extra terms added to the query to enhance the query. These results are added to the selected document content via inclusion of them in the enhanced query hyperlinks in accordance with Black's teachings. Therefore, the computer device of Black in combination with HyPursuit's Sun Sparcstations are a content manager that enriches the selected document content (e.g., an HTML page) with results (i.e., the links to the enhanced query) provided from the information retrieval system using the query (i.e., using the initial query).

(o) Black and HyPursuit render obvious independent claim 18

(i) *"An article of manufacture for use in a computer system, comprising: a memory;"*

See discussion of corresponding element of claim 1, where Black discloses a computer device that generates a query from selected document content. (*Appendix G at 4/58-67 and 1/50-57).* In performing the automatic query generation and displaying the automatically generated queries on a computer screen like that shown in FIG. 5, the computer device would necessarily have to store program instructions in a memory for execution. (*Appendix G at 4/30-43*) Moreover, the Sun SparcStations of HyPursuit's network include a memory that stores the content router CGI scripts used to further process the queries. (*Appendix J at page 189, column 2*)

(ii) *"instructions stored in the memory for operating a method for automatically generating a query from selected document content, comprising:"*

See discussion of corresponding element of claim 1, where, as detailed below, the system of Black, as modified by HyPursuit's teachings, stores the instructions corresponding to this element in a memory of Black's computer device and a memory of HyPursuit's Sun SparcStations. *(Appendix G at 1/50-57 and 4/58-67; and Appendix J at page 189, column 2)*

(iii) *"defining an organized classification of document content with each class in the organized classification of document content having associated therewith a classification label; each classification label corresponding to a category of information in an information retrieval system;"*

See discussion of corresponding element of claim 1, where the instructions are stored in a memory of HyPursuit's Sun SparcStations. (*Appendix J at page 189, column 2*)

(iv) *"automatically identifying a set of entities in the selected document content for searching information related thereto using the information retrieval system;"*

See discussion of corresponding element of claim 1, where the instructions are stored in a memory of Black's computer device. (*Appendix G at 1/50-57 and 4/58-67*)

(v) *"automatically categorizing the selected document content using the organized classification of document content for assigning the selected document content a classification label from the organized classification of content; and"*

See discussion of corresponding element of claim 1, where the instructions are stored in a memory of HyPursuit's Sun SparcStations. (*Appendix J at page 189, column 2)*

(vi) *"automatically formulating the query to restrict a search at the information retrieval system for information concerning the set of entities to the category of information in the information retrieval system identified by the assigned classification label."*

See discussion of corresponding element of claim 1, where the instructions are stored both in a memory of Black's computer device and a memory of HyPursuit's Sun SparcStations. (*Appendix G at 1/50-57 and 4/58-67; and Appendix J at page 189, column 2)*

(p) Black and HyPursuit render obvious claim 19

Claim 19 depends from independent claim 18, and further requires *"wherein the instructions stored in the memory further comprise limiting the query by adding terms relating to context information surrounding the set of entities in the selected document content."* See discussion of corresponding element of claim 2, where the instructions are stored both in a memory of Black's computer device and in a memory of HyPursuit's Sun SparcStations. (*Appendix G at 1/50-57 and 4/58-67; and Appendix J at page 189, column 2*)

(q) Black and HyPursuit render obvious claim 20

Claim 20 depends from claim 19, and further requires *"wherein the instructions stored in the memory further comprise further limiting the query by adding terms defining the assigned classification label."* See discussion of corresponding element of claim 4, where the instructions are stored both in a memory of Black's computer device and in a memory of HyPursuit's Sun SparcStations. (*Appendix G at 1/50-57 and 4/58-67; and Appendix J at page 189, column 2*)

11. Claims 3 and 8 are unpatentable over Black taken in view of HyPursuit and Syskill under 35 U.S.C. § 103(a)

Requestor respectfully submits that 3 and 8 are unpatentable over Black taken in view of HyPursuit and Syskill under 35 U.S.C. § 103(a). A claim chart applying Black, HyPursuit, and Syskill to these claims is submitted herewith as Appendix W.

(a) Black, HyPursuit, and Syskill render obvious claim 3

Claim 3 depends from claim 2, and further requires *"wherein the number of terms added is limited to a predefined number."* Syskill describes a system that develops a user profile for a user that browses the World Wide Web by receiving a rating from the user for each visited Web page and analyzing the information on each Web page. The system is then able to generate a query based on the user profile that is submitted to a search engine to find Web pages that may be of interest to the user. Syskill's system limits the number of terms in the search query submitted to the search engine to a predetermined number (e.g., 14 terms) in recognition of the fact that search engines cannot accept very long queries.

Specifically, Syskill teaches that a system that provides queries to a search engine may limit a query to a predetermined number of terms in recognition of the fact that the search engine may not be able to process very long queries. (*Appendix F at Page 56, Second Column, Lines 3-6, and shown in claim chart at Appendix W*). Syskill operates on keyword queries just like those generated by the system of Black, as modified by HyPursuit's teachings. As also noted previously, Syskill expressly teaches what a person of ordinary skill in the art would have already known – i.e., that any realistic implementation of an electronic search engine, like that of the system resulting from the combination of the teachings of Black and those of HyPursuit, has a limited amount of computing resources and, therefore, would limit the number of terms/keywords in a keyword query inputted into the search engine for processing.

Accordingly, a person of ordinary skill in the art would have readily recognized that Syskill's limitation of terms in a query to a predetermined number could be easily applied to the search queries of the system resulting from the combination of the teachings of Black and HyPursuit in recognition of the limited processing capabilities of search engines, as expressly disclosed by Syskill. Furthermore, the combination of Syskill's teachings with those of Black and HyPursuit would involve merely combining and/or substituting known prior art elements to yield predictable results. One skilled in the art could have easily combined the known elements as claimed by known methods, with each element in the combination performing the same function as it does separately, and the combination yielding nothing more than predictable results to one skilled in the relevant art.

(b) Black, HyPursuit, and Syskill render obvious claim 8

Claim 8 depends from claim 7, and further requires *"abbreviating the set of terms extracted by the text categorizer to a predefined number of terms."* As stated above, Syskill describes a system that limits the number of terms in the search query submitted to the search engine to a predetermined number (e.g., 14 terms) in recognition of the fact that search engines cannot accept very long queries. A person of ordinary skill in the art would have readily recognized that Syskill's limitation of terms in a query to a predetermined number could be easily applied to the query generated by the system of Black, as modified based on the teachings of HyPursuit, to ensure that the enhanced query can be handled by the search engine that receives it. Accordingly, a person of ordinary skill in the art would, therefore, have abbreviated the set of terms extracted and appended to the query by the Black system, as modified by the teachings of HyPursuit, to a predefined number of terms in accordance with Syskill's teachings to ensure that the enhanced query can be processed by the search engine that receives it.

12. <u>Claims 1, 5, 6, 10-14 and 18 are unpatentable over Black taken in view of Donaldson under 35 U.S.C. § 103(a)</u>

Requestor respectfully submits that claims 1, 5, 6, 10-14 and 18 are unpatentable over Black taken in view of Donaldson under 35 U.S.C. § 103(a). A claim chart applying Black and Donaldson to these claims is submitted herewith as Appendix X.

(a) Black and Donaldson render obvious independent claim 1

(i) *"A method for automatically generating a query from selected document content, comprising:"*

Black teaches automatically generating a query by analyzing a body of selected content, such as an article on the SR-71 Blackbird Airplane. The automatic generation of the query involves an automatic generation of a hook, i.e., a concept, primary subject matter or main topic of the body of selected content, and the automatic generation of keywords from the body of selected content. A query is then formed by "AND"ing a keyword with the hook. For example, the automatic query that may be generated for the SR-71 BlackBird Airplane article may be "BLACKBIRD and AVIATION", where Blackbird is the main topic automatically determined for the article and aviation is a keyword identified in the article. (*Appendix G at 1/64-67*)

Thus, the Black reference teaches "[a] system for conducting queries from any document displayed on any computer device." (*Appendix G at Abstract*) Specifically, the Black reference teaches "automatic … statistical and empirical analysis" (*Appendix G at 2/43*) of a body of selected content that may be "text, such as magazine articles, news stories or any other text" (*Appendix G at 1/47-48*) where the text can be an online article (*Appendix G at FIG. 1, 4/18-20*).

The Black reference uses its analysis results to automatically generate a group of suggested queries for finding additional information about the document. Black teaches that "[b]rilliant queries require a preparation process that analyzes any text to enhance and generate a set of suggested searches based on that analysis and certain pre-set user parameters." (*Appendix G at 1/50-54*)

The automatic generation of the query involves analyzing the text of the selected document content to identify a hook term that reflects the primary subject matter of the document.

> The hook is the concept, primary subject matter or main topic for a body of text. The hook is used to define a query as narrowly as possible on a particular topic for a selected information source.

(Appendix G at 2/10-14)

> Automatic Generation of the Hook – One embodiment of the brilliant query to enable an automatic process for generating brilliant queries for a body of text, is to determine the hook by extracting the highest frequency proper names from the text body.

(Appendix G at 2/57-61).

The automatic generation of the query also involves analyzing the text of the selected document content to identify one or more keywords that are reflective of additional topics for the search.

> A brilliant query requires a list of keywords that are generated by automatic … statistical and empirical analysis of the body of content to be enhanced or a comparable body of content."

(*Appendix G at 2/42-45*)

> Keywords are simply a collection of words, generated automatically … , that are deemed to be indicative of the topic matter or one of the topics for a given content selection.

(*Appendix G at 2/26-29*)

> Automatic Generation of Keywords – A word frequency analysis is done on all of the text, with stopwords excluded, and the resulting words, by order of frequency are compared to a pre-selected keyword list. Those that match, based upon a desired frequency become keywords to be combined with the hook to form focused, optimal queries.

(*Appendix G at 3/29-35*)

A suggested query for the document is generated automatically by ANDing the hook and a keyword.

> Once the keywords have been selected and the hook for a body of text has been determined or automatically generated, the searches are created by generating a link for every keyword extracted from the body of text and combining it with the hook in a search that results in a result set that is the logical intersection of the results generated by the hook and the keyword. Basically, each entry in the list of search results must contain both the hook and the keyword and not just one or the other.

(Appendix G at 3/37-45)

> After the hook and the keywords have been established the query is conducted selecting one of the hook-keyword sets.

(Appendix G at 3/55-57)

(ii) *"defining an organized classification of document content with each class in the organized classification of document content having associated therewith a classification label; each classification label corresponding to a category of information in an information retrieval system;"*

Donaldson teaches categorizing Web pages/Web sites using a categorization scheme, where each category in the categorization scheme has associated with it a category identifier (e.g., a category name). Each category identifier corresponds to a category of Web pages/Web sites. Example categories may include "offensive," "medical," "legal" and "sports." (*Appendix H at 15/38-46, 17/17-19, and 19/51-56*). Specifically, Donaldson teaches categorizing documents using a hierarchical categorization scheme as follows: "Each category may include a listing of sub-categories 865 and web sites 875 within those categories." *(Appendix H at 17/17-19 and Fig. 8c)*

Each category has a category identifier or name that corresponds to a category of Web pages/Web sites in the information retrieval system.

> For example, in one implementation, the hierarchy of category identifiers may include a hierarchy of category names, where groups of the category names are linked together in a hierarchical relationship. In this instance, names in the hierarchy represent categories, the names of which are linked together using sub-categories. The hierarchy of category identifiers also may include other related information, such as a list of web sites that are related to the category by name, description, or otherwise.

(Appendix H at 15/38-46) Example categories may include "offensive," "medical," "legal," and "sports."

> Each electronic information store may contain content that has been classified and stored based on a specified type or types of classification criteria. For instance, the first electronic information store 992 may include content classified as non-offensive and the second electronic information store 994 may include content classified as offensive. Other types of content classification criteria may be implemented in addition to or separate from criteria based on offensive and non-offensive classifications. Other criteria that may be used, for example, include medical and non-medical, legal and non-legal, and sports and non-sports. In one implementation, the first electronic information includes contents relating to non-offensive web sites, and the second electronic information includes contents relating to offensive web sites.

(Appendix H at 19/51-65) Black's query is a keyword query, just like those processed by Donaldson's search engine. (*Appendix H at 14/7-42*). Accordingly, one of ordinary skill in the art would have easily recognized that Black's system is ready for improvement through use of Donaldson's techniques. Specifically, a person of ordinary skill in the art would have readily appreciated that

Donaldson's techniques could be used to process Black's keyword query to improve the quality of search results and/or effect parental controls. In applying Donaldson's techniques to Black's query, Black's document would be automatically categorized via categorization of the term corresponding to the main topic or subject matter of Black's document and then the application of Black's query would be restricted to a data source corresponding to that category to ensure that only desired results are provided. Furthermore, the combination of Donaldson's teachings with those of Black would involve merely combining and/or substituting known prior art elements to yield predictable results. One skilled in the art could have easily combined the known elements as claimed by known methods, with each element in the combination performing the same function as it does separately, and the combination yielding nothing more than predictable results to one skilled in the relevant art.

(iii) *"automatically identifying a set of entities in the selected document content for searching additional information related thereto using the information retrieval system;"*

Black teaches automatically identifying keywords from the selected body of content by performing a word frequency analysis on all of the text, with stop words excluded, and the resulting words, by order of frequency, are compared to a pre-selected keyword list. Those that match are deemed the keywords for which additional information will be sought through inclusion of the keywords in automatically generated search queries.

Specifically, Black teaches automatic identification of keywords from the selected body of content that are indicative of a topic matter or one of the topics for the body of content. Black teaches that "[k]eywords are simply a collection of words, generated automatically … , that are deemed to be indicative of the topic matter or one of the topics for a given content selection." *(Appendix G at 2/26-29*)

The list of keywords can be generated by an automatic statistical and empirical analysis of the selected body of content to be enhanced. As taught by Black, "[a] brilliant query requires a list of keywords that are generated by automatic … statistical and empirical analysis of the body of content to be enhanced or a comparable body of content." (*Appendix G at 2/42-45*)

The analysis to determine the list of keywords is a word frequency analysis that is done on all of the text of the selected body of content, with stopwords excluded, and the resulting words are then compared to a pre-selected keyword list. Words that match the pre-selected keyword list are retained as the keywords for the selected body of content based upon a desired frequency.

> Automatic Generation of Keywords – A word frequency analysis is done on all of the text, with stopwords excluded, and the resulting words, by order of frequency are compared to a pre-selected keyword list. Those that match, based upon a desired frequency become keywords to be combined with the hook to form focused, optimal queries.

(*Appendix G at 3/29-35*).

(iv) *"automatically categorizing the selected document content using the organized classification of document content for assigning the selected document content a classification label from the organized classification of content;"*

Black teaches automatically identifying a hook search term that is reflective of the main topic or primary subject matter of the selected body of content. Donaldson teaches receiving one or more search query terms and automatically classifying the one or more terms into a category, thereby assigning the one or more terms the corresponding category name. As shown in Figs. 9b and 9c of Donaldson (reproduced below), Donaldson's assignment of a category name to the one or more terms is used to improve the search results generated by the query by restricting the application of the search query to an electronic information store classified under that category.

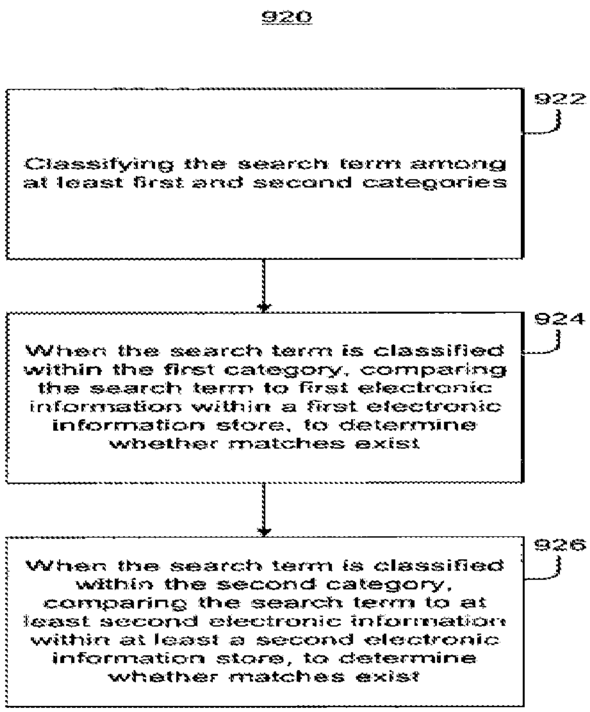


Fig. 9b

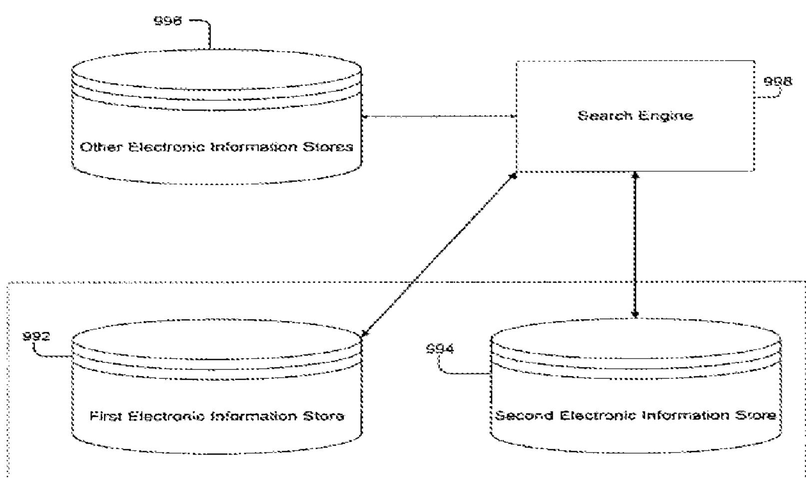


Fig. 9c

Specifically, as shown in Figs. 9b and 9c, a search term is categorized as either being in a first category or a second category. If the search term is categorized as being in the first category, the search is restricted to searching data stored within the first electronic information store 992. In contrast, if the search term is categorized as being in the second category, the search is restricted to searching data stored within the second electronic information store 994.

Black's query is a keyword query, just like those processed by Donaldson's search engine. (*Appendix H at 14/7-42*). Accordingly, one of ordinary skill in the art would have easily recognized that Black's system is ready for improvement through use of Donaldson's techniques. Specifically, a person of ordinary skill in the art would have readily appreciated that Donaldson's techniques could be used to process Black's keyword query to improve the quality of search results and/or effect parental controls. In applying Donaldson's techniques to Black's query, Black's document would be automatically categorized via categorization of the term corresponding to the main topic or subject matter of Black's document and then the application of Black's query would be restricted to a data source corresponding to that category to ensure that only desired results are provided.

Specifically, Black teaches automatically identifying a hook search term that is reflective of the primary subject matter of a document.

> The hook is the concept, primary subject matter or main topic for a body of text. The hook is used to define a query as narrowly as possible on a particular topic for a selected information source. To determine a "hook", a content layer must exist for which a context can be determined. There must be a perceivable structure to the information source and each content entry must have an associated context or place or places within the structure of the information source.

(*Appendix G at 2/10-18*)

> Automatic Generation of the Hook – One embodiment of the brilliant query to enable an automatic process for generating brilliant queries for a body of text, is to determine the hook by extracting the highest frequency proper names from the text body.

(Appendix G at 2/57-61) As shown in Figs. 9b and 9c reproduced above, Donaldson teaches receiving one or more search terms and automatically classifying them into a category

> Classifying the search term (922) generally includes classifying the received search term among one or more categories, with a first category and a second category being described and shown for illustrative purposes. If several search terms are grouped as a single string, the search terms may be collectively classified as a single string based on the grouping of search terms, or they may be classified individually based on each individual search term.

(Appendix H at 18/47-54) Donaldson teaches that these categories are used to restrict the application of the search query to information within a particular electronic information store that corresponds to the identified category.

> Comparing the search terms (step 924) generally includes comparing the search term to first electronic information within a first electronic information store when the search term is classified within the first category. By contrast, comparing the search term (step 926) generally includes comparing the search term to the second electronic information within the second electronic information store to determine whether matches exist when the search term is classified within the second category.

(Appendix H at 18/55-63) Donaldson teaches that such a restriction is desirable in that it can remove undesirable documents from the pool of information that is being searched to ensure that these undesirable documents are not produced as search results in response to the query. In one example described by Donaldson, a search query having a search term or a search string classified in the "non-offensive" category is only applied to an information source corresponding to the "non-offensive" category, thereby ensuring that "offensive" Web pages/Web sites are not provided as search results for the query. Accordingly, a search system modified in accordance with Donaldson's teachings could implement parental controls.

> The following describes an example applying the described search methods of FIG. 9b to this implementation. A user of a client system enters a search term (step 910). The search term is classified as either being offensive or non-offensive (step 922). If the term is classified as being non-offensive, then only the contents of the first electronic information store are searched (924) and results from the search are communicated for display to the user (step 930). In this example, the first electronic information store only contains contents that previously have been classified as non-offensive. If the search term entered by the user is classified as being offensive, the contents of either the second electronic information store or both the first and second electronic information stores are searched (step 926) and the results are communicated for display to the user (step 930).

(*Appendix H at 20/3-18*)

> The described filtering of results between offensive content and non-offensive content based on the classification of the search term may allow a web host to implement a parental type of control in determining

> what search results are displayed to the user. Because the offensive and non-offensive contents are stored in different electronic information stores, the ability to restrict access is enhanced. For instance, parental control can be exercised by blocking the access of a user to one or more electronic information stores. Other forms of data filtering also are enabled through this process and related techniques.

(*Appendix H at 20/19-29*)

(v) *"automatically formulating the query to restrict a search at the information retrieval system for information concerning the set of entities to the category of information in the information retrieval system identified by the assigned classification label."*

Black teaches automatically formulating a query through the automatic generation of a hook for the selected body of content, the automatic generation of keywords from the selected body of content, and then ANDing the hook and one of the keywords to automatically form the query.

For the reasons stated above, a person of ordinary skill in the art would have readily recognized that Black's selected body of content could be assigned a category name through the categorization of Black's hook term in accordance with Donaldson's teachings and that Black's search could then be restricted to the category of information identified by the assigned category name in order to ensure that undesirable documents are not produced as search results in response to the query, as taught by Donaldson.

Specifically, Black teaches automatically formulating a query by ANDing an automatically generated hook and an automatically generated keyword determined from selected document content.

> Once the keywords have been selected and the hook for a body of text has been determined or automatically generated, the searches are created by generating a link for every keyword extracted from the body of text and combining it with the hook in a search that results in a result set that is the logical intersection of the results generated by the hook and the keyword. Basically, each entry in the list of search results must contain both the hook and the keyword and not just one or the other.

(*Appendix G at 3/37-45*)

> After the hook and the keywords have been established the query is conducted selecting one of the hook-keyword sets.

(*Appendix G at 3/55-57*) Donaldson teaches receiving one or more search terms and automatically classifying them into a category

> Classifying the search term (922) generally includes classifying the received search term among one or more categories, with a first category and a second category being described and shown for illustrative purposes. If several search terms are grouped as a single string, the search terms may be collectively classified as a single string based on the grouping of search terms, or they may be classified individually based on each individual search term.

(Appendix H at 18/47-54) As shown in Figs. 9b and 9c of Donaldson (reproduced above), Donaldson teaches that these categories are used to restrict the application of the search query to information within a particular electronic information store that corresponds to the identified category. Specifically, a search term is categorized as either being in a first category or a second category. If the search term is categorized as being in the first category, the search is restricted to searching data stored within the first electronic information store 992. In contrast, if the search term is categorized as being in the second category, the search is restricted to searching data stored within the second electronic information store 994.

> Comparing the search terms (step 924) generally includes comparing the search term to first electronic information within a first electronic information store when the search term is classified within the first category. By contrast, comparing the search term (step 926) generally includes comparing the search term to the second electronic information within the second electronic information store to determine whether matches exist when the search term is classified within the second category.

(Appendix H at 18/55-63) Donaldson teaches that such a restriction is desirable in that it can remove undesirable documents from the pool of information that is being searched to ensure that these undesirable documents are not produced as search results in response to the query. In one example described by Donaldson, a search query having a search term or a search string classified in the “non-offensive” category is only applied to an information source corresponding to the “non-offensive” category, thereby ensuring that “offensive” Web pages/Web sites are not provided as search results for the query. Accordingly, a search system modified in accordance with Donaldson’s teachings could implement parental controls.

> The following describes an example applying the described search methods of FIG. 9b to this implementation. A user of a client system enters a search term (step 910). The search term is classified as either being offensive or non-offensive (step 922). If the term is classified as being non-offensive, then only the contents of the first electronic information store are searched (924) and results from the search are communicated for display to the user (step 930). In this example, the first electronic information store only contains contents that previously have been classified as non-offensive. If the search term entered by the user is classified as being offensive, the contents of either the second electronic information store or both the first and second electronic information stores are searched (step 926) and the results are communicated for display to the user (step 930).

(*Appendix H at 20/3-18*)

> The described filtering of results between offensive content and non-offensive content based on the classification of the search term may allow a web host to implement a parental type of control in determining what search results are displayed to the user. Because the offensive and non-offensive contents are stored in different electronic information stores, the ability to restrict access is enhanced. For instance, parental control can be exercised by blocking the access of a user to one or more electronic information stores. Other forms of data filtering also are enabled through this process and related techniques.

(*Appendix H at 20/19-29*)

(b) Black and Donaldson render obvious claim 5

Claim 5 depends from independent claim 1, and further requires *"wherein the organized classification of document content is defined using a hierarchical organization."* Donaldson teaches an organized classification of document content that is defined using a hierarchical organization. Specifically, Donaldson teaches "[e]ach category may include a listing of sub-categories 865 and web sites 875 within those categories." (*Appendix H at 17/17-19, 15/38-46 and Fig. 8c, all of which is shown in the claim chart at Appendix X*)

(c) Black and Donaldson render obvious claim 6

Claim 6 depends from independent claim 1, and further requires *"using a text categorizer to assign the classification label assigned from the organized classification of content."* The system resulting from the combination of Black's teachings and Donaldson's teachings uses a text categorizer to assign a category name to a search term in accordance with Donaldson's teachings, where the search term is derived from a document and reflects the primary subject matter of the document in accordance with Black's teachings. The category name is assigned from an organized classification of content.

Specifically, Black teaches automatically identifying a hook search term that is reflective of the primary subject matter of a document. (*Appendix G at 2/10-18 and at 2/57-61, all of which is shown in the claim chart at Appendix X*) Donaldson teaches receiving one or more search terms and automatically assigning them a category name by analyzing their text. (*Appendix H at 15/64 to 16/4, and shown in claim chart at Appendix X*)

(d) Black and Donaldson render obvious claim 10

Claim 10 depends from independent claim 1, and further requires *"wherein each class in the organized classification of document content has associated therewith a characteristic vocabulary."* Donaldson discloses that each category in the hierarchy of categories includes a characteristic vocabulary in that Donaldson discloses that each category can be characterized by information related to the category, such as Web site descriptions for Web sites within the category, for purposes of categorizing search terms.

In particular, Donaldson discloses that information related to the categories can be used in addition to the category identifiers to categorize search terms. The information related to the categories may include a name of a Web site corresponding to a category, a description of the Web site, or other related terms. *(Appendix H at 16/5-9, and shown in claim chart at Appendix X)*

Accordingly, for a given category, this information is used to characterize the category for purposes of determining assignment of the category to a search term and, therefore, is a characteristic vocabulary associated with the category.

(e) Black and Donaldson render obvious claim 11

Claim 11 depends from claim 10, and further requires *"ranking results from the query performed at the information retrieval system in accordance with one of the assigned classification label and the characteristic vocabulary."* Black teaches placing the hook as the first term in the query to indicate its greater importance in the query string. Black also teaches that, by placing the hook first in the query, the search results provided by most search engines will rank Web pages having more occurrences of the hook in their contents first. As stated above with respect to claim 4, Donaldson's teachings indicate that the hook can be the same as the assigned category name for Black's selected content and, therefore, the ranking taught by Black of search results in which results are ranked higher if they correspond to document content having more occurrences of the hook corresponds to the recited ranking of results from the query in accordance with the assigned classification label.

In particular, Black describes that the search results generated in response to the query will rank results having more occurrences of the hook in their corresponding contents above other results. *(Appendix G at 1/29-35, and shown in claim chart at Appendix X)* Donaldson contemplates that a search term can be the same as (i.e., match) a category identifier/name to which the search term is assigned. (*Appendix H at 15/64-67 and at 15/38-43, all of which is shown in the claim chart at Appendix X*)

Accordingly, the category name assigned to Black's hook term in accordance with Donaldson's teachings can be the hook term itself and, based on Black's teachings, the ranking of search results can be in accordance with the category name insofar that search results having more occurrences of the category name are ranked higher.

(f) Black and Donaldson render obvious claim 12

Claim 12 depends from claim 11, and further requires *"using the method in a system for enriching selected content of a document with personalities that identify enrichment themes."* The Black reference teaches enriching the selected document by annotating the selected document to include the automatically generated queries as hyper-links, as shown in FIG. 5 of Black (reproduced previously). As shown in Fig. 5, the article titled "Lazio accuses Clinton of flip-flopping on Jerusalem" has been annotated to include various hyperlinks including the hyperlink "Search for more information on Hillary Clinton and time." *(Appendix G at FIG. 5 and 4/30-34, and shown in claim chart at Appendix X)*

Requestor notes that the '979 patent states that to "enrich" a document means merely to "annotate a document in accordance with a predefined personality." (*Appendix A at col. 6, lines 63 and 64*)

(g) Black and Donaldson render obvious claim 13

Claim 13 depends from claim 1, and further requires *"automatically identifying the set of entities using a service that recognizes entities of a predefined type."* The Black reference teaches automatically identifying the keywords using a service that recognizes words of a predefined type, where the predefined type of words corresponds to words that are not stop words. (*Appendix G at 2/45-49 and at 3/30/31, all of which is shown in the claim chart at Appendix X*)

(h) Black and Donaldson render obvious independent claim 14

(i) *"A system for automatically generating a query from selected document content, comprising:"*

Black discloses a system that generates a query from selected document content.

> 1. A system for submitting to an information source, a query based upon any document displayed on any computer device comprising, conducting a statistical frequency analysis of the word occurrence in the document to determine the subject matter of the document, the subject matter being the most frequently occurring proper name or non-stopword, and the keywords in the document, the keywords being determined by the frequency analysis or by comparison to a pre-selected keyword list and creating queries by combining the subject matter with each of the keywords.

(*Appendix G at 4/58-67)* The system enables display of an HTML page of a document in which hyperlinks are automatically added by the system.

> Brilliant queries require a preparation process that analyzes any text to enhance and generate a set of suggested searches based on that analysis and certain pre-set user parameters. The output of this preparation process can be used to add links to an HTML page of a document either automatically or through manual insertion of the resulting analysis.

(Appendix G at 1/50-57) In describing that the system includes a computer device that displays an HTML page that has been automatically annotated to include links corresponding to queries, Black contemplates performing the query generation and invocation processes through use of a computer device having a corresponding memory and processor. See also discussion of corresponding element of claim 1.

(ii) *"an entity extractor for automatically identifying a set of entities in the selected document content for searching information related thereto using an information retrieval system;"*

See discussion of corresponding element of claim 1, where the computer device of Black performs the recited function and, thereby, includes the recited entity extractor. (*Appendix G at 4/58-67 and 1/50-57*).

(iii) *"a categorizer for defining an organized classification of document content with each class in the organization of content having associated therewith a classification label; each classification label corresponding to a category of information in the information retrieval system; the categorizer automatically assigning the selected document content a classification label from the organized classification of content;"*

See discussion of corresponding element of claim 1, where Donaldson's host device 435 (which may include pods 4804, servers 4806 and/or farms 4802 shown in Fig. 4 of Donaldson, reproduced below) executes the operations corresponding to this element. (*Appendix H at Fig. 4 and 13/62-14/6)*

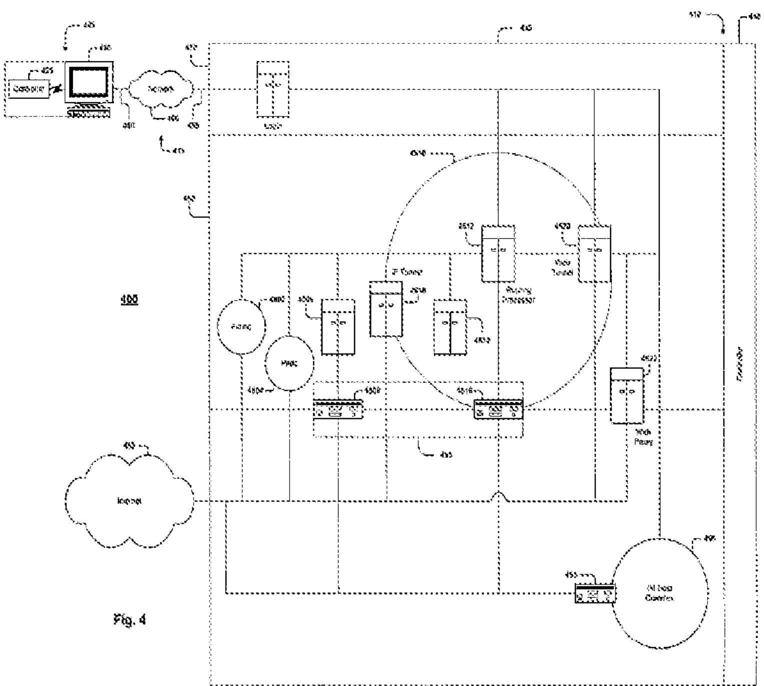


Fig. 4

Donaldson's host device 435 may be a general or a special purpose computer. (*Appendix H at 6/53 to 7/10 and 3/34-44*).

(iv) *"a query generator for automatically formulating the query to restrict a search at the information retrieval system for information concerning the set of entities to* *the category of information in the information retrieval system identified by the assigned classification label."*

See discussion of corresponding element of claim 1, where Black's computer device automatically generates a query from a selected document that is processed by Donaldson's host device 435 to produce search results. *(Appendix G at 4/58-67 and 1/50-57; and Appendix H at Fig. 4 and 13/62-14/6).* Donaldson's host device 435 may be a general or special purpose computer. *(Appendix H at 6/53-7/10 and 3/34-44)* Accordingly, the query generator corresponds to Black's computer device in combination with components of Donaldson's host device 435 that together automatically formulate a search query that is restricted to only searching a data store associated with the assigned category.

(i) Black and Donaldson render obvious independent claim 18

(i) *"An article of manufacture for use in a computer system, comprising: a memory;"*

See discussion of corresponding element of claim 1, where Black discloses a computer device that generates a query from selected document content. (*Appendix G at 4/58-67 and 1/50-57).* In performing the automatic query generation and providing automatically generated queries in a computer screen like that shown in FIG. 5, the computer device would necessarily have to store program instructions in a memory for execution. (*Appendix G at 4/30-43*) Moreover, the host device 435 of Donaldson includes a memory/storage medium on which is loaded a software application that includes instructions for commanding and directing the host device 435 to further process the queries. (*Appendix H at Fig. 4 and 13/62-14/6, 6/53-7/10, 3/45-58)*

(ii) *"instructions stored in the memory for operating a method for automatically generating a query from selected document content, comprising:"*

See discussion of corresponding element of claim 1, where, as detailed below, the system resulting from a combination of Black's teachings and Donaldson's teachings stores the instructions corresponding to this element in a memory of Black's computer device and a memory of Donaldson's host device 435. (*Appendix G at 4/58-67 and 1/50-57; and Appendix H at Fig. 4 and 13/62-14/6, 6/53-7/10, 3/45-58*)

(iii) *"defining an organized classification of document content with each class in the organized classification of document content having associated therewith a classification label; each classification label corresponding to a category of information in an information retrieval system;"*

See discussion of corresponding element of claim 1, where the instructions are stored in a memory of Donaldson's host device 435. (*Appendix H at Fig. 4 and 13/62-14/6, 6/53-7/10, 3/45-58)*

(iv) *"automatically identifying a set of entities in the selected document content for searching information related thereto using the information retrieval system;"*

See discussion of corresponding element of claim 1, where the instructions are stored in a memory of Black's computer device (*Appendix G at 1/50-57 and 4/58-67*)

(v) *"automatically categorizing the selected document content using the organized classification of document content for assigning the selected document content a classification label from the organized classification of content; and"*

See discussion of corresponding element of claim 1, where the instructions are stored in a memory of Donaldson's host device 435. (*Appendix H at 7/22-32 and 34-36, and 13/62-14/6)*

(vi) *"automatically formulating the query to restrict a search at the information retrieval system for information concerning the set of entities to the category of information in the information retrieval system identified by the assigned classification label."*

See discussion of corresponding element of claim 1, where the instructions are stored in a memory of Black's computer device and in a memory of Donaldson's host device 435. (*Appendix G at 1/50-57 and 4/58-67; and Appendix H at 7/22-32 and 34-36, and 13/62-14/6*).

13. Claims 2, 4, 7, 15-17, 19 and 20 are unpatentable over Black taken in view of Donaldson and Stibel under 35 U.S.C. § 103(a)

Requestor respectfully submits that claims 2, 4, 7, 15-17, 19 and 20 are unpatentable over Black taken in view of Donaldson and Stibel under 35 U.S.C. § 103(a). A claim chart applying Black, Donaldson, and Stibel to these claims is submitted herewith as Appendix Y.

(a) Black, Donaldson, and Stibel render obvious claim 2

Claim 2 depends from independent claim 1, and further requires *"limiting the query by adding terms relating to context information surrounding the set of entities in the selected document content."* As stated above, the system resulting from the combination of Black's teachings and Donaldson's teachings forms a keyword query by identifying keywords in the content of a selected document. As described in more detail below, Stibel teaches a system that processes a keyword query to generate an enhanced keyword query that includes all of the terms of the original query along with one or more additional terms. The one or more additional terms are meaning terms that define a term in the original query and/or are terms that are expected to be found in Web pages directed to the same meaning as that of a term in the original query. (*Appendix E at 11/36 to 12/24, and 9/21-43, all of which is shown in claim chart at Appendix Y*) One of ordinary skill in the art would have easily recognized that the system resulting from the combination of Black's teachings and Donaldson's teachings is ready for improvement through use of Stibel's techniques. The resulting enhanced query would include additional terms that relate to context information surrounding the keywords in the selected document content, rather than to context information surrounding other portions of the selected document content or surrounding a different document entirely.

Specifically, Stibel teaches a system that receives a search query having one or more query terms and processes the search query to generate a new search query that will more effectively retrieve information that is relevant to the original search query. (*Appendix E at 2/42-47, and shown in claim chart at Appendix Y*) The processing includes identifying a meaning or Sense of one or more query terms, identifying related terms that are likely to appear on a web page when a user is querying for information associated with that identified meaning, and then expanding the original query to form the new search query by appending the meaning and the one or more related terms to the original query. (*Appendix E at 10/30-36 and 11/56-66, all of which is shown in the claim chart at Appendix Y*). For example, if the original query includes only the term "java" and is determined to correspond to the meaning "coffee" and have related terms "espresso" and "beverage," the system may modify the search query to be the new search query “java + coffee + espresso + beverage.” (*Appendix E at 12/16-19*) Stibel asserts that this new enhanced query will result in a more effective search in that the Internet search engine is more likely to return a meaningful hit list in response to the new enhanced query as compared to the initial query. (*Appendix E at 12/18-21, and shown in claim chart at Appendix Y*)

As mentioned previously, the system resulting from the combination of Black’s teachings and Donaldson’s teachings generates keyword queries just like those that are processed by Stibel’s system. That is, one of ordinary skill in the art would have easily recognized that such a combined system is ready for improvement through use of Stibel’s techniques. Specifically, a person of ordinary skill in the art would have readily recognized that Stibel’s techniques for processing a search query could be applied to the query generated by the system resulting from the combination of Black’s teachings and Donaldson’s teachings to make it more effective at finding desired information by further limiting the query through the addition of meaning terms.

Furthermore, these additional terms relate to context information surrounding the keywords in the selected document content, rather than to context information surrounding other portions of the selected document content or surrounding a different document entirely. That is, these additional terms, in being directed to the meaning of the hook and/or keyword terms in the selected document content and in corresponding to terms that are expected to be found in Web pages directed to the same meaning, provide context information surrounding the hook and/or keyword terms in order to achieve a more precise search.

For example, if the selected document content is an article on the SR-71 Blackbird and Black’s query is “Blackbird AND aviation,” the meaning and related terms correspond to context information surrounding the terms “Blackbird” and “aviation,” rather than to context information surrounding other terms in the article, such as “Elint,” “reconnaissance” or “transport,” or surrounding a different article entirely. (*Appendix G at 1/64-2/9*)

(b) Black, Donaldson, and Stibel render obvious claim 4

Claim 4 depends from claim 2, and further requires *“limiting the query by adding terms defining the assigned classification label.”* Donaldson contemplates that a search term can be the same as (i.e., match) a category identifier/name to which the search term is assigned. (*Appendix H at 15/64-67 and at 15/38-43, all of which is shown in the claim chart at Appendix Y*) Accordingly, the category name assigned to Black’s hook term in accordance with Donaldson’s teachings can be the hook term itself. When the category name assigned to Black’s hook term and, hence, to Black’s selected document content is the hook term itself, the further limiting of the query in accordance with Stibel's teachings by adding terms that correspond to the meaning of Black's hook term is a further limiting of the query by adding terms that define the assigned classification name.

(c) Black, Donaldson, and Stibel render obvious claim 7

Claim 7 depends from claim 6, and further requires *"extracting with the text categorizer a set of terms relating to the document content; and appending to the query ones of the set of terms extracted by the text categorizer to contextualize the query."*

Black extracts the hook and keywords from the selected document content and appends the hook and/or a keyword to the query to contextualize the query. Stibel teaches that meaning terms and related terms can also be added to contextualize the query. The hook, keyword, meaning, and/or related terms contextualize the query by reflecting the environment from which the query was generated – i.e., the textual contents of the selected document.

Specifically, Black describes automatically analyzing the textual contents of the selected document to determine the primary subject matter or topic of the document. *(Appendix G at 2/10-14, and shown in claim chart at Appendix Y)* Black also describes automatically analyzing the textual contents of the selected document to determine keywords that correspond to one or more other topics of the document. Black teaches that "[k]eywords are simply a collection of words, generated automatically … , that are deemed to be indicative of the topic matter or one of the topics for a given content selection." (*Appendix G at 2/26-29*) Black describes adding the hook and/or keyword terms to queries to provide additional context information for the queries that reflect the document content from which the queries are derived. *(Appendix G at 3/37-45, and shown in claim chart at Appendix Y)*

Stibel describes that further contextual information may be added to the query by identifying one or more meanings of the terms in the initial query and appending the one or more meanings and related words to the initial query. The related words represent a word or a meaning that is likely to appear on a web page when a user is querying for information associated with the one or more meanings of the terms in the initial query. (*Appendix E at 10/30-36, 9/21-43, 11/56-66 and 12/13-18, all of which is shown in the claim chart at Appendix Y*)

(d) Black, Donaldson, and Stibel render obvious claim 15

Claim 15 depends from independent claim 14, and further requires *"a short length aspect vector generator for generating terms relating to context information surrounding the set of entities in the* *selected document content; wherein the query generator adds the terms relating to the context information to limit the query."* See discussion of corresponding element of claim 2, where Black's computer device in combination with Donaldson's host device 435, modified in accordance with the teachings of Stibel, correspond to a generator that adds meanings and related terms to the automatically generated query to improve the effectiveness of the query.

(e) Black, Donaldson, and Stibel render obvious claim 16

Claim 16 depends from claim 15, and further requires *"wherein the query generator further limits the query by adding terms defining the selected classification label provided by the categorizer."* Donaldson contemplates that a search term can be the same as (i.e., match) a category identifier/name to which the search term is assigned. (*Appendix H at 15/64-67 and at 15/38-43, all of which is shown in the claim chart at Appendix Y*) Accordingly, the category name assigned to Black's hook term in accordance with Donaldson's teachings can be the hook term itself. When the category name assigned to Black's hook term and, hence, to Black's selected document content is the hook term itself, the further limiting of the query in accordance with Stibel's teachings by adding terms that correspond to the meaning of Black's hook term is a further limiting of the query by adding terms that define the assigned classification name.

(f) Black, Donaldson, and Stibel render obvious claim 17

Claim 17 depends from claim 16, and further requires *"a content manager for enriching the selected document content with results provided from the information retrieval system using the query."*

Black teaches a computer device that automatically generates an initial query from selected document content, the query being composed of a hook and a keyword. Stibel teaches that this query may be used to identify additional meaning and related terms that can then be appended to the initial query. Black teaches that the computer device enriches the selected document content by annotating it to include hyperlinks corresponding to a query.

Accordingly, the recited results provided by the information retrieval system resulting from the combination of Black's teachings, Donaldson's teachings, and Stibel's teachings using the query (i.e., the initial query) are the extra terms added to the query to enhance the query. These results are added to the selected document content via inclusion of them in the enhanced query hyperlinks in accordance with Black's teachings. Therefore, the computer device of Black in combination with Donaldson's host device 435, as modified based on Stibel's teachings, are a content manager that enriches/annotates the selected document content (e.g., an HTML page) with results (i.e., the links to the enhanced query) provided from the information retrieval system using the query (i.e., using the initial query).

Requestor notes that the '979 patent states that to "enrich" a document means merely to "annotate a document in accordance with a predefined personality." (*Appendix A at col. 6, lines 63 and 64*)

(g) Black, Donaldson, and Stibel render obvious claim 19

Claim 19 depends from independent claim 18, and further requires *"wherein the instructions stored in the memory further comprise limiting the query by adding terms relating to context information surrounding the set of entities in the selected document content."*

See discussion of corresponding element of claim 2, where the instructions are stored in a memory of Black's computer device and in a memory of Donaldson's host device 435, and are modified in accordance with Stibel's teachings. (*Appendix G at 1/50-57 and 4/58-67; and Appendix H at 7/22-32 and 34-36, and 13/62-14/6)*.

(h) Black, Donaldson, and Stibel render obvious claim 20

Claim 20 depends from claim 19, and further requires *"wherein the instructions stored in the memory further comprise further limiting the query by adding terms defining the assigned classification label."*

See discussion of corresponding element of claim 4, where the instructions are stored in a memory of Black's computer device and in a memory of Donaldson's host device 435, and are modified in accordance with Stibel's teachings. (*Appendix G at 1/50-57 and 4/58-67; and Appendix H at 7/22-32 and 34-36, and 13/62-14/6)*.

14. <u>Claims 3 and 8 are unpatentable over Black taken in view of Donaldson, Stibel and Syskill under 35 U.S.C. § 103(a)</u>

Requestor respectfully submits that 3 and 8 are unpatentable over Black taken in view of Donaldson, Stibel and Syskill under 35 U.S.C. § 103(a). A claim chart applying Black, Donaldson, Stibel and Syskill to these claims is submitted herewith as Appendix Z.

(a) Black, Donaldson, Stibel and Syskill render obvious claim 3

Claim 3 depends from claim 2, and further requires *"wherein the number of terms added is limited to a predefined number."* Syskill describes a system that develops a user profile for a user that browses the World Wide Web by receiving a rating from the user for each visited Web page and analyzing the information on each Web page. The system is then able to generate a query based on the user profile that is submitted to a search engine to find Web pages that may be of interest to the user. Syskill's system limits the number of terms in the search query submitted to the search engine to a predetermined number (e.g., 14 terms) in recognition of the fact that search engines cannot accept very long queries.

Specifically, Syskill teaches that a system that provides queries to a search engine may limit a query to a predetermined number of terms in recognition of the fact that the search engine may not be able to process very long queries. (*Appendix F at Page 56, Second Column, Lines 3-6, and shown in claim chart at Appendix Z*). Syskill operates on keyword queries just like those generated by the system resulting from the combination of Black's teachings, Donaldson's teachings, and Stibel's teachings. As noted previously, Syskill expressly teaches what a person of ordinary skill in the art would have already known – i.e., that any realistic implementation of an electronic search engine, like that of such a system, has a limited amount of computing resources and, therefore, would limit the number of terms/keywords in a keyword query inputted into the search engine for processing. Accordingly, a person of ordinary skill in the art would have readily recognized that Syskill's limitation of terms in a query to a predetermined number could be easily applied to the search queries of the system resulting from the combination of Black's teachings, Donaldson's teachings, and Stibel's teachings in recognition of the limited processing capabilities of search engines, as expressly disclosed by Syskill. Furthermore, the combination of Syskill's teachings with those of Black, Donaldson and Stibel would involve merely combining and/or substituting known prior art elements to yield predictable results. One skilled in the art could have easily combined the known elements as claimed by known methods, with each element in the combination performing the same function as it does separately, and the combination yielding nothing more than predictable results to one skilled in the relevant art.

(b) Black, Donaldson, Stibel and Syskill render obvious claim 8

Claim 8 depends from claim 7, and further requires *"abbreviating the set of terms extracted by the text categorizer to a predefined number of terms."* As stated above, Syskill describes a system that limits the number of terms in the search query submitted to the search engine to a predetermined number (e.g., 14 terms) in recognition of the fact that search engines cannot accept very long queries. A person of ordinary skill in the art would have readily recognized that Syskill's limitation of terms in a query to a predetermined number could be easily applied to the query generated by the system resulting from the combination of Black's teachings, Donaldson's teachings, and Stibel's teachings to ensure that the modified query can be handled by the search engine that receives it. Accordingly, a person of ordinary skill in the art would, therefore, have abbreviated the set of terms extracted and appended to the query by the combined system to a predefined number of terms in accordance with Syskill's teachings to ensure that the expanded query can be processed by the search engine that receives it.

VI. CONCLUSION

For the foregoing reasons, substantial and new questions of patentability with respect to each of claims 1-20 of the '979 patent have been raised. Reexamination of claims 1-20 is accordingly requested.

Respectfully submitted,

Dated: August 27, 2010

/W. Karl Renner/
W. Karl Renner
Reg. No. 41,265

CERTIFICATE OF SERVICE

I hereby certify that on August 27, 2010, I caused a true and correct copy of the foregoing REQUEST FOR INTER PARTES REEXAMINATION to be served via First Class U.S. Mail on the following:

XEROX CORPORATION
P.O. BOX 1600
STAMFORD, CONNECTICUT 06904

Date: August 27, 2010

/W. Karl Renner/
W. Karl Renner
Reg. No. 41,265