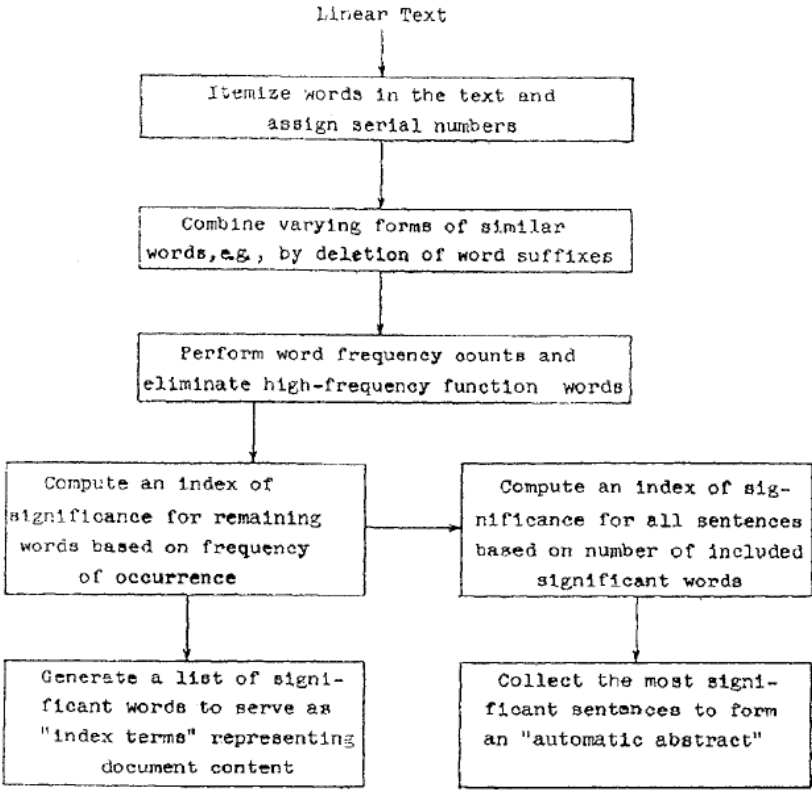


EXHIBIT L

INVALIDITY CLAIM CHART FOR U.S. PATENT NO. 5,544,352
BASED ON SALTON, G., "ASSOCIATIVE DOCUMENT RETRIEVAL TECHNIQUES USING BIBLIOGRAPHIC
INFORMATION," PP. 440-57 (1963). ("SALTON, 1963")

Claim Text from '352 Patent	Salton, 1963
<p>26. A non-semantic method for numerically representing objects in a computer database and for computerized searching of the numerically represented objects in the database, wherein direct and indirect relationships exist between objects in the database, comprising:</p>	<p><i>See, e.g.</i>, Salton, 1963, at Abstract, pp. 443, 446</p> <p>The standard associative retrieval techniques are first briefly reviewed. A computer experiment is then described which tends to confirm they hypothesis that documents exhibiting similar citation sets also deal with similar subject matter. (Salton, 1963, Abstract)</p> <p>The criteria of association used in most automatic programs do not normally require a determination of syntactic or semantic properties. Rather, they are based on simple co-occurrence of words in the same texts or sentences, or on co-occurrence with individual or joint frequencies greater than some given threshold value. (Salton, 1963, p. 443)</p> <p>Because of these and other variations, citation and reference lists have not generally been used as an indication of document content. Rather, such lists are used to detect trends in the literature as a whole, and to serve as adjuncts to certain kinds of literature searches [7, 8]. (Salton, 1963, p. 446)</p>
<p>[26a] Marking objects in the database so that each marked object may be individually identified by a computerized search;</p>	<p><i>See, e.g.</i>, Salton, 1963, at pp. 441, 447</p> <p>Figure 1</p>

Claim Text from '352 Patent	Salton, 1963
	 <pre> graph TD A[Linear Text] --> B[Itemize words in the text and assign serial numbers] B --> C[Combine varying forms of similar words, e.g., by deletion of word suffixes] C --> D[Perform word frequency counts and eliminate high-frequency function words] D --> E[Compute an index of significance for remaining words based on frequency of occurrence] D --> F[Compute an index of significance for all sentences based on number of included significant words] E --> G[Generate a list of significant words to serve as "index terms" representing document content] F --> H[Collect the most significant sentences to form an "automatic abstract"] </pre> <p data-bbox="831 1068 1726 1117">Fig. 1. Typical automatic indexing and abstracting system based on word frequency counts.</p> <p data-bbox="819 1133 1075 1166">(Salton, 1963, p. 441)</p> <p data-bbox="819 1182 1898 1312">Consider a collection of m documents each of which is characterized by the property of being cited by one or more of the other documents in the same collection. Each document can then be represented by an m-dimensional logical vector X_i, where $X_{ji} = 1$ if and only if document i is cited by document j, and $X_{ji} = 0$ otherwise. (Salton, 1963, p. 447)</p>
[26b] creating a first numerical representation for	See, e.g., Salton, 1963, at pp. 446 n.1, 447, 450

Claim Text from '352 Patent

each identified object in the database based upon the object's direct relationship with other objects in the database;

Salton, 1963

A citation index consists of a set of bibliographic references (the set of cited documents), each followed by a list of all those documents (the citing documents) which include the given cited document as a reference. A reference index, on the other hand, lists all cited documents under each citing document. (Salton, 1963, p. 446 n.1)

Consider a collection of m documents each of which is characterized by the property of being cited by one of more of the other documents in the same collection. Each document can then be represented by an m -dimensional logical vector X_i , where $X_{ij} = 1$ if and only if document i is cited by document j , and $X_{ij} = 0$ otherwise. If these m vectors arranged in rows one below the other a square logical incidence matrix is formed similar to the matrix exhibited in Figure 4.

$$\begin{array}{c|ccc}
 \begin{array}{c} \text{Cited} \\ \text{documents} \end{array} & \begin{array}{c} \text{Citing documents} \\ D_1 \quad D_2 \quad \dots \quad D_m \end{array} \\
 \hline
 D_1 & \begin{pmatrix} X_{11} & X_{21} & \dots & X_{m1} \end{pmatrix} \\
 D_2 & \begin{pmatrix} X_{12} & X_{22} & \dots & X_{m2} \end{pmatrix} \\
 \vdots & \begin{pmatrix} \vdots & \vdots & \dots & \vdots \end{pmatrix} \\
 D_m & \begin{pmatrix} X_{1m} & X_{2m} & \dots & X_{mm} \end{pmatrix}
 \end{array} = X$$

($X_j^i = 1 \leftrightarrow$ document D_i is cited by document D_j)

FIG. 4. Matrix X exhibiting direct citations

(Salton, 1963, p. 447)

Figure 5

Claim Text from '352 Patent	Salton, 1963
	<pre> graph TD Start[Consider with each document the set of applicable index terms and the set of applicable citations] --> C[Construct a term-document incidence matrix C listing documents against included terms] Start --> X[Construct a citation incidence matrix X listing each cited document against all citing documents] C --> S[Construct a document-document similarity matrix S based on overlapping index terms] X --> R[Construct a document-document similarity matrix R based on overlapping citations] S --> X_Correl[Compute a cross-correlation vector x and overall cross-correlation coefficient x to measure similarities between document rows R and S, and between the complete matrices, respectively] R --> X_Correl X_Correl --> C_Star[Construct term-term similarity matrix C* and use it to generate new term-document matrices C1, C2, ..., and so on] X_Correl --> X_Powers[Construct squared, cubed, ... incidence matrices X1, X2, ... exhibiting citation links of length two, three, ..., and so on] C_Star -.-> C X_Powers -.-> X </pre> <p>FIG. 5. Comparison of citation similarities with index term similarities</p> <p>(Salton, 1963, p. 450)</p>
[26c] storing the first numerical representations for use in computerized searching;	<p>See, e.g., Salton, 1963, at pp. 446 n.1, 447, 450</p> <p>A citation index consists of a set of bibliographic references (the set of cited documents), each followed by a list of all those documents (the citing documents) which include the given cited document as a reference. A reference index, on the other hand, lists all cited documents under each citing document. (Salton, 1963, p. 446 n.1)</p>

Claim Text from '352 Patent

Salton, 1963

Consider a collection of m documents each of which is characterized by the property of being cited by one of more of the other documents in the same collection. Each document can then be represented by an m -dimensional logical vector X_i , where $X_{ij} = 1$ if and only if document i is cited by document j , and $X_{ij} = 0$ otherwise. If these m vectors arranged in rows one below the other a square logical incidence matrix is formed similar to the matrix exhibited in Figure 4.

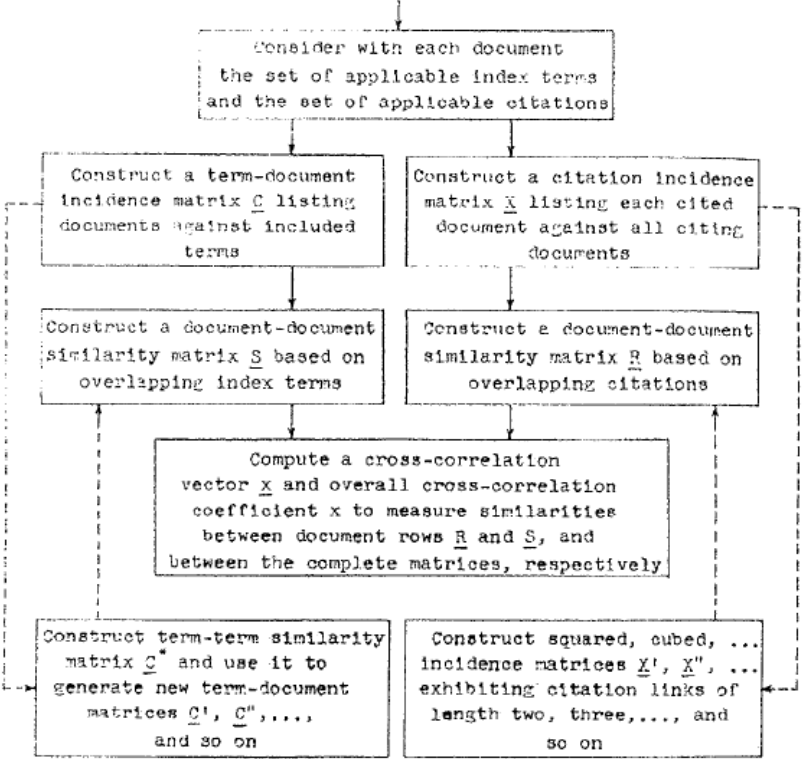
<i>Cited documents</i>		<i>Citing documents</i>				
		D_1	D_2	\dots	D_m	
D_1		X_{11}^1	X_{21}^1	\dots	X_{m1}^1	
D_2		X_{12}^2	X_{22}^2	\dots	X_{m2}^2	
\vdots		\vdots	\vdots	\dots	\vdots	
D_m		X_{1m}^m	X_{2m}^m	\dots	X_{mm}^m	= X

($X_{ij}^i = 1 \leftrightarrow$ document D_i is cited by document D_j)

FIG. 4. Matrix X exhibiting direct citations

(Salton, 1963, p. 447)

Figure 5

Claim Text from '352 Patent	Salton, 1963
	 <pre> graph TD Start[Consider with each document the set of applicable index terms and the set of applicable citations] --> C[Construct a term-document incidence matrix C listing documents against included terms] Start --> X[Construct a citation incidence matrix X listing each cited document against all citing documents] C --> S[Construct a document-document similarity matrix S based on overlapping index terms] X --> R[Construct a document-document similarity matrix R based on overlapping citations] S --> CC[Compute a cross-correlation vector x and overall cross-correlation coefficient alpha to measure similarities between document rows R and S, and between the complete matrices, respectively] R --> CC CC --> C_prime[Construct term-term similarity matrix C' and use it to generate new term-document matrices C', C'', ..., and so on] CC --> X_prime[Construct squared, cubed, ... incidence matrices X', X'', ... exhibiting citation links of length two, three, ..., and so on] C_prime -.-> C X_prime -.-> X </pre> <p data-bbox="911 1040 1625 1062">FIG. 5. Comparison of citation similarities with index term similarities</p> <p data-bbox="816 1094 1075 1122">(Salton, 1963, p. 450)</p>
<p data-bbox="186 1187 804 1279">[26d] analyzing the first numerical representations for indirect relationships existing between or among objects in the database;</p>	<p data-bbox="816 1187 1268 1214"><i>See, e.g.,</i> Salton, 1963, at pp. 448, 450</p> <p data-bbox="816 1273 1913 1370">Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X', X'', etc., exhibiting respectively the existence of paths of length two, three, and so on. Specifically,</p>

Claim Text from '352 Patent	Salton, 1963
	$[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$ <p>Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. $(X')_{ij}$ is then equal to 1 if and only if at least one path of length two exists between documents D_i and D_j; otherwise, $(X')_{ij}$ is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)</p> <p>A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the ith and jth rows (columns) of X.</p> <p>The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448)</p> <p>Figure 5</p>

Claim Text from '352 Patent	Salton, 1963
	<pre> graph TD Start[Consider with each document the set of applicable index terms and the set of applicable citations] --> C[Construct a term-document incidence matrix C listing documents against included terms] Start --> X[Construct a citation incidence matrix X listing each cited document against all citing documents] C --> S[Construct a document-document similarity matrix S based on overlapping index terms] X --> R[Construct a document-document similarity matrix R based on overlapping citations] S --> CC[Compute a cross-correlation vector x and overall cross-correlation coefficient alpha to measure similarities between document rows R and S, and between the complete matrices, respectively] R --> CC CC --> C_prime[Construct term-term similarity matrix C' and use it to generate new term-document matrices C', C'', ..., and so on] CC --> X_prime[Construct squared, cubed, ... incidence matrices X', X'', ... exhibiting citation links of length two, three, ..., and so on] C_prime -.-> C X_prime -.-> X </pre> <p>FIG. 5. Comparison of citation similarities with index term similarities</p> <p>(Salton, 1963, p. 450)</p>
<p>[26e] generating a second numerical representation of each object based on the analysis of the first numerical representation;</p>	<p>See, e.g., Salton, 1963, at pp. 448, 450, 451-52</p> <p>Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X', X'', etc., exhibiting respectively the existence of paths of length two, three, and so on. Specifically,</p>

Claim Text from '352 Patent	Salton, 1963
	$[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$ <p>Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. $(X')_{ij}$ is then equal to 1 if and only if at least one path of length two exists between documents D_i and D_j; otherwise, $(X')_{ij}$ is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)</p> <p>A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the ith and jth rows (columns) of X.</p> <p>The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448)</p> <p>The CITED and CITING similarity matrices of dimension 62 by 62 were obtained from the original citation matrix by row and column comparisons, respectively. The TDCMP similarity matrix, also of dimension 62 by 62, was similarly obtained by column comparisons from the original term-document matrix. Additional citation similarity matrices, designated CTD2, CTD3, CTD4, and CNG2, CNG3, CNG4 were obtained from the squared, cubed, and fourth power logical citation matrices, as previously explained. (Salton, 1963, p. 450)</p>

Figure 5

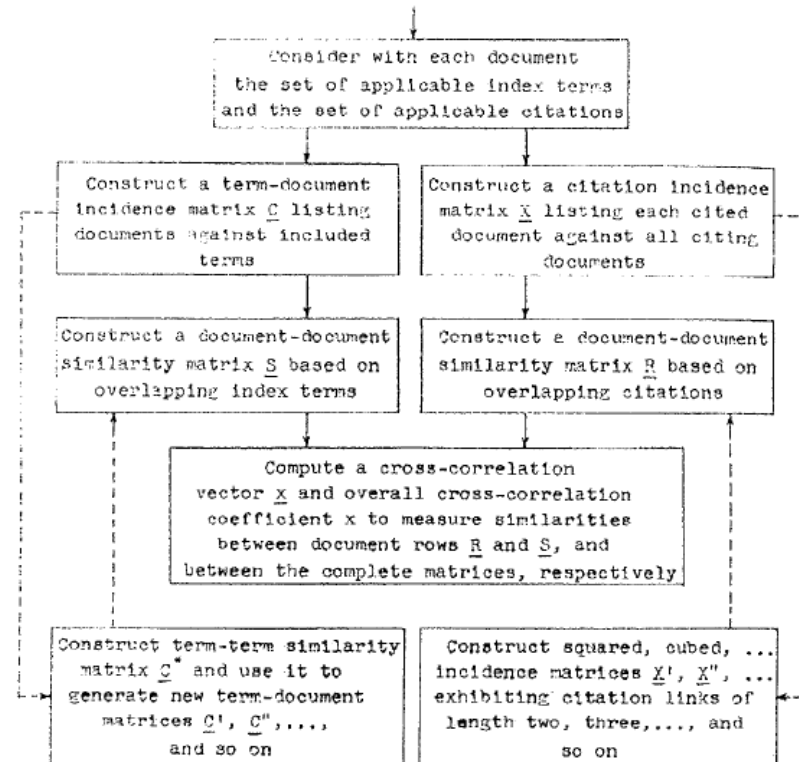


FIG. 5. Comparison of citation similarities with index term similarities

(Salton, 1963, p. 450)

The value of the overall similarity coefficient first rises as the length of the citation links increases, and then drops again as the length of the links becomes still greater [6]. This is due to the fact that as the length of the links increases, the total number of links of any length increases also; an increased number of links results in a larger number of ones in the original logical citation matrix, and thus in a higher probability of overlapping ones and a larger overall similarity coefficient. At the same time, as the length of the links increases, two factors also tend to decrease the magnitude of the overall similarity coefficient. First, the

Claim Text from '352 Patent	Salton, 1963
	<p>number of documents which exhibit citation links of length n but which do not exhibit links of length greater than n increases as n becomes larger. Thus more and more documents will exhibit individual similarity coefficients of zero value, thus tending to decrease the value of the overall coefficient. Second, as the length of the links increases and the citations thus become increasingly less accurate indications of document content, the magnitude of the cross-correlation coefficients obtained from the citation matrix and the term-document matrix would be expected to decrease, even for those documents for which a large number of citation links can still be found. (Salton, 1963, pp. 451-52)</p>
<p>[26f] storing the second numerical representation for use in computerized searching; and</p>	<p>See, e.g., Salton, 1963, at pp. 448, 450</p> <p>Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X', X'', etc., exhibiting respectively the existence of paths of length two, three, and so on. Specifically,</p> $[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$ <p>Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. (X')_{ij} is then equal to 1 if and only if at least one path of length two exists between documents D_i and D_j; otherwise, (X')_{ij} is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)</p> <p>A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the ith and jth rows (columns) of X.</p> <p>The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between</p>

Claim Text from '352 Patent

Salton, 1963

documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448)

Figure 5

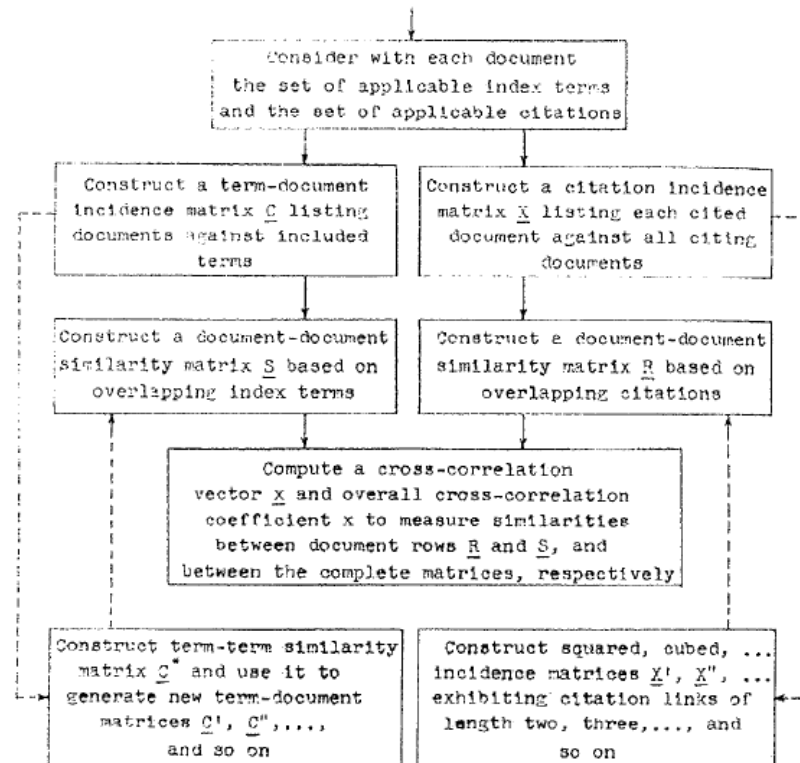


FIG. 5. Comparison of citation similarities with index term similarities

Claim Text from '352 Patent

Salton, 1963

(Salton, 1963, p. 450)

[26g] searching the objects in the database using a computer and the stored second numerical representations, wherein the search identifies one or more of the objects in the database.

See, e.g., Salton, 1963, at pp. 443, 444, 445

Figure 2

$$\begin{array}{c|cccc}
 \text{Terms} & D_1 & D_2 & \dots & D_m \\
 \hline
 W_1 & C_1^1 & C_2^1 & \dots & C_m^1 \\
 W_2 & C_1^2 & C_2^2 & \dots & C_m^2 \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 W_n & C_1^n & C_2^n & \dots & C_m^n
 \end{array} = \mathbf{C}$$

(a) Typical term-document incidence matrix \mathbf{C} ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times)

$$\begin{array}{c|cccc}
 \text{Terms} & W_1 & W_2 & \dots & W_n \\
 \hline
 W_1 & R_1^1 & R_2^1 & \dots & R_n^1 \\
 W_2 & R_1^2 & R_2^2 & \dots & R_n^2 \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 W_n & R_1^n & R_2^n & \dots & R_n^n
 \end{array} = \mathbf{R}$$

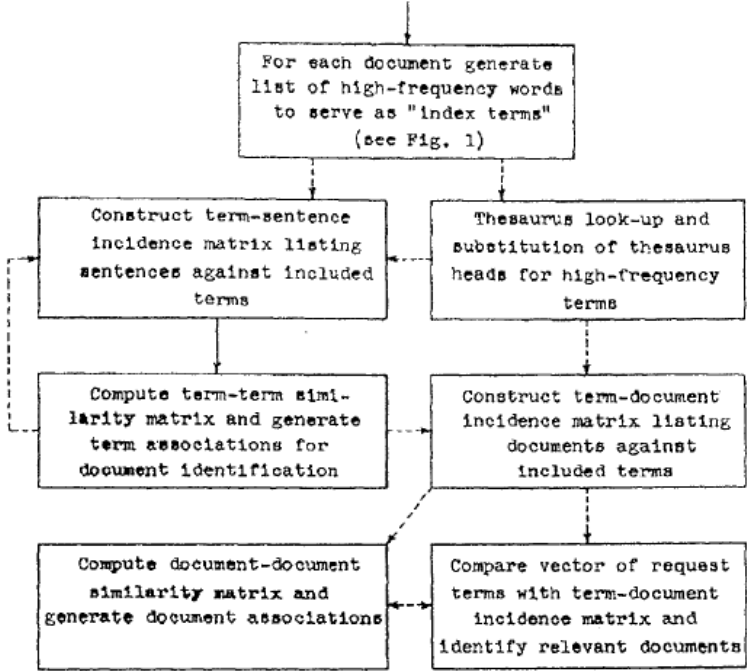
(b) Typical term-term similarity matrix \mathbf{R}

$$\left(R_j^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \sum_{k=1}^m (C_k^j)^2 \right)}} \right)$$

FIG. 2. Matrices used for the generation of term associations

(Salton, 1963, p. 443)

Consider now a typical system for document retrieval using term and document associations as shown in Figure 3. A list of high-frequency terms is first generated for each document by word frequency counting procedures. Normalization may or may not be effected by thesaurus lookup. A term-term similarity matrix is then constructed by using co-occurrence of terms within sentences, rather than within documents, as a criterion. It should be noted that as new term associations are defined, the original incidence matrix can be revised by

Claim Text from '352 Patent	Salton, 1963
	<p data-bbox="816 248 1913 412">inclusion in some of the matrix columns of new, associated terms which are not originally contained in the respective sentences or documents. The revised incidence matrix then gives rise to a new term-term similarity matrix, incorporating second-order associations, and so on. This feedback process is represented by an upward-pointing arrow in Figure 3. (Salton, 1963, p. 444)</p> <p data-bbox="816 427 919 456">Figure 3</p>  <pre data-bbox="919 472 1661 1138"> graph TD Start[For each document generate list of high-frequency words to serve as "index terms" (see Fig. 1)] A[Construct term-sentence incidence matrix listing sentences against included terms] B[Thesaurus look-up and substitution of thesaurus heads for high-frequency terms] C[Compute term-term similarity matrix and generate term associations for document identification] D[Construct term-document incidence matrix listing documents against included terms] E[Compute document-document similarity matrix and generate document associations] F[Compare vector of request terms with term-document incidence matrix and identify relevant documents] Start --> A Start --> B A --> C B --> D C --> E D --> F E --> F F -.-> A F -.-> B F -.-> C F -.-> D F -.-> E </pre> <p data-bbox="877 1154 1713 1203">Fig. 3. Typical automatic document retrieval system using term and document associations → optional paths → compulsory paths</p> <p data-bbox="816 1230 1073 1260">(Salton, 1963, p. 445)</p>
27. The non-semantic method of claim 26, wherein the objects in the database include words, and semantic indexing techniques are used in	<i>See, e.g., Salton, 1963, at Abstract, pp. 442, 446-47, 456-57</i>

Claim Text from '352 Patent	Salton, 1963
<p>combination with the non-semantic method, the method further comprising the step of creating and storing a Boolean word index for the words of the objects in the database.</p>	<p>Automatic documentation systems which use the words contained in the individual documents as a principal source of document identifications may not perform satisfactorily under all circumstances. Methods have therefore been devised within the last few years for computing association measures between words and between documents, and for using such associated words, or information contained in associated documents, to supplement and refine the original document identifications. It is suggested in this study that bibliographic citations may provide a simple means for obtaining associated documents to be incorporated in an automatic documentation system.</p> <p>...</p> <p>Finally, a fully automatic document retrieval system is proposed which uses bibliographic information in addition to other standard criteria for identification of document content, and for the detection of relevant information. (Salton, 1963, Abstract)</p> <p>For this reason, several workers [2, 3, 4, 5] have been interested in automatic procedures designed to supplement the original terms extracted from the documents with new terms related to the old ones in various ways. Indexing techniques which make use of such "associated" terms have come to be known as "associative indexing" and corresponding retrieval operations are known as "associative retrieval."</p> <p>The present report suggests an extension of the usual associative retrieval techniques by taking into account bibliographic citations and other information peculiar to the author of a given document. It is suggested, specifically, that the set of identifying words extracted from the documents be supplemented by new words obtained in part from the bibliographic information provided with the documents; these new expanded sets of index terms may then give a more accurate representation of document content than the original ones and may thus provide a more effective retrieval mechanism. (Salton, 1963, p. 442)</p> <p>If it could be shown that citations were usable as content indicators, then the associative techniques described in Section 2 could be further refined by adding to the term-document matrix illustrated in Figure 2(a) further document columns representing cited documents, citing documents, or documents written by the same author. These new documents would then provide new associated terms which might be equally as important as the term associations derived from other documents in the same collection. (Salton, 1963, pp. 446-47)</p> <p>Figure 9</p>

Claim Text from '352 Patent

Salton, 1963

$$\begin{array}{c}
 \begin{array}{l}
 \text{Original terms} \\
 \left\{ \begin{array}{l} W_1 \\ W_2 \\ \vdots \\ W_n \end{array} \right. \\
 \\
 \begin{array}{l}
 \text{New terms} \\
 \text{provided by} \\
 \text{related} \\
 \text{documents} \\
 \left\{ \begin{array}{l} W_{n+1} \\ \vdots \\ W_r \end{array} \right.
 \end{array}
 \end{array}
 \begin{array}{c}
 \left(\begin{array}{c}
 \begin{array}{c}
 \text{Original} \\
 \text{documents} \\
 D_1 \quad D_2 \quad \dots \quad D_m
 \end{array} \\
 \begin{array}{c}
 \text{Related} \\
 \text{documents} \\
 \text{through} \\
 \text{citations} \\
 D_{m+1} \dots D_p \quad D_{p+1} \dots D_q
 \end{array}
 \end{array}
 \right)
 \begin{array}{c}
 \begin{array}{c}
 \text{Related} \\
 \text{documents} \\
 \text{through} \\
 \text{authorship}
 \end{array}
 \end{array}
 \end{array}
 = C$$

FIG. 9. Basic term-document incidence matrix usable for extended associative retrieval

(Salton, 1963, p. 456)

The following tentative conclusions can be drawn from the foregoing experiment: the similarity coefficients obtained by comparing overlapping citations for a sample document collection with overlapping, manually generated index terms are much larger than those obtained by assuming a random assignment of citations and terms to the documents; relatively large similarity coefficients are generated for nearly all documents which exhibit at least a minimum number of citations. If the foregoing results were confirmed by experiments with other document collections, citations could provide a large number of relevant index terms not originally available with a given document collection, and thereby create a much more flexible retrieval process. Presently available programs for associative retrieval could be used unchanged in an extended system. (Salton, 1963, pp. 456-57)

28. The non-semantic method of claim 26 wherein the first and second numerical representations are vectors that are arranged in first and second matrices;

See, e.g., Salton, 1963, at pp. 443-44, 445, 447, 448, 449

Figure 2

Terms	Documents					
	D_1	D_2	...		D_m	
W_1	C_1^1	C_2^1	...	C_j^1	...	C_m^1
W_2	C_1^2	C_2^2	...	C_j^2	...	C_m^2
\vdots						
W_n	C_1^n	C_2^n	...	C_j^n	...	C_m^n

$$= C$$

(a) Typical term-document incidence matrix C ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times)

Terms	Terms			
	W_1	W_2	...	W_n
W_1	R_1^1	R_2^1	...	R_n^1
W_2	R_1^2	R_2^2	...	R_n^2
\vdots				
W_n	R_1^n	R_2^n	...	R_n^n

$$= R$$

(b) Typical term-term similarity matrix R

$$\left(R_j^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \sum_{k=1}^m (C_k^j)^2 \right)}} \right)$$

FIG. 2. Matrices used for the generation of term associations

(Salton, 1963, p. 443)

Many different types of similarity coefficients have been suggested in the literature; a simple coefficient of similarity between rows of a numeric matrix, and one which may be as meaningful as any of the others, is the cosine of the angle between the corresponding m -dimensional vectors. The similarity coefficients can be displayed in an $n \times n$ symmetric term-similarity matrix R , where the coefficient of similarity R_{ji} between term W_i and term W_j is

$$R_j^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \sum_{k=1}^m (C_k^j)^2 \right)}}$$

Document similarities are therefore obtained by comparing pairs of columns (instead of rows) of the term-document matrix C, and a document-document similarity matrix is constructed and used in the same way as the previously described term-term matrix R. (Salton, 1963, pp. 443-44)

Figure 3

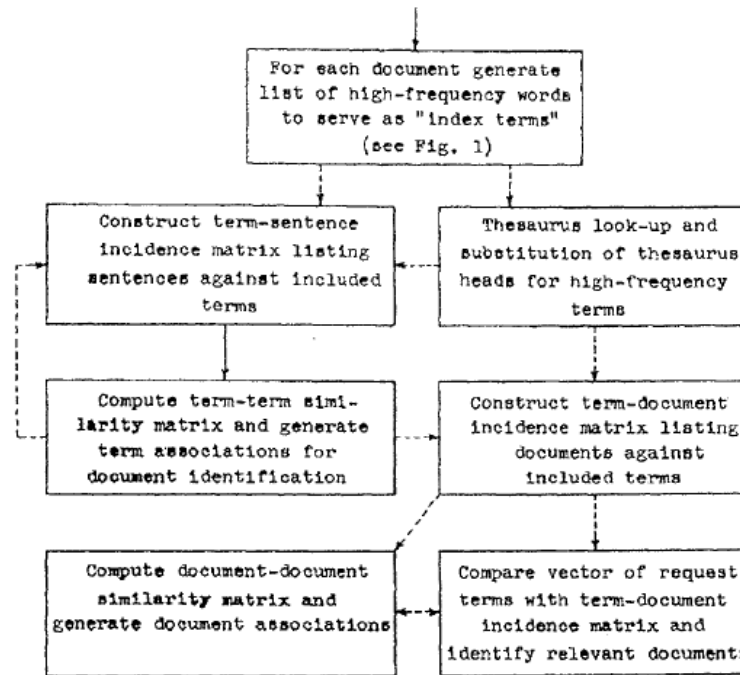


FIG. 3. Typical automatic document retrieval system using term and document associations
 → optional paths → compulsory paths

(Salton, 1963, p. 445)

Consider a collection of m documents each of which is characterized by the property of being cited by one or more of the other documents in the same collection. Each document can then be represented by an m -dimensional logical vector X_i , where $X_{ij} = 1$ if and only if document i is cited by document j , and $X_{ij} = 0$ otherwise. If these m vectors arranged in rows one below the other a square logical incidence matrix is formed similar to the matrix

exhibited in Figure 4.

$$\begin{array}{c} \text{Cited} \\ \text{documents} \end{array} \left| \begin{array}{cccc} \text{Citing documents} \\ D_1 & D_2 & \dots & D_m \end{array} \right. \\ \hline \begin{array}{c} D_1 \\ D_2 \\ \vdots \\ D_m \end{array} \begin{pmatrix} X_{11}^1 & X_{21}^1 & \dots & X_{m1}^1 \\ X_{11}^2 & X_{21}^2 & \dots & X_{m1}^2 \\ \vdots & \vdots & \dots & \vdots \\ X_{11}^m & X_{21}^m & \dots & X_{m1}^m \end{pmatrix} = X$$

($X_{ij}^k = 1 \leftrightarrow$ document D_i is cited by document D_j)

FIG. 4. Matrix X exhibiting direct citations

(Salton, 1963, p. 447)

Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X' , X'' , etc., exhibiting respectively the existence of paths of length two, three, and so on.

Specifically,

$$[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$$

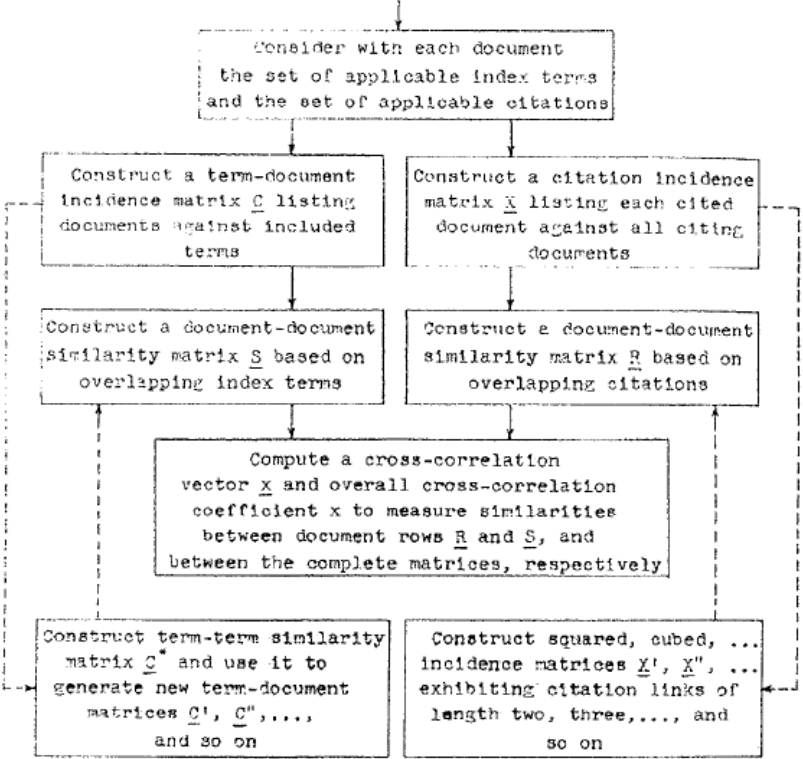
$$[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$$

Boolean multiplication is used, since the new connection matrices X' , X'' , etc., are again defined as logical matrices. $(X')_{ij}$ is then equal to 1 if and only if at least one path of length two exists between documents D_i and D_j ; otherwise, $(X')_{ij}$ is equal to 0. It may be noted that X' , unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)

Since the term-document matrix C is not in general a square matrix, matrix multiplication cannot be used to obtain second order effects, similar to the citation links of length two or more. Instead, it is first necessary to compare the index terms by performing a row comparison of the rows of C. This produces a new n symmetric term matrix C^* which displays similarity between index terms. This matrix can be used to eliminate from the set of index terms those terms which exhibit a large number of joint occurrences with other terms. A reduced set of index terms can then be formed and a new term-document matrix C?

Claim Text from '352 Patent	Salton, 1963
	constructed, from which a new correlation matrix S is formed. (Salton, 1963, p. 449)
[28a] the direct relationships are express references from a one object to another object in the database;	<p><i>See, e.g.</i>, Salton, 1963, at Abstract, pp. 443, 446, 447, 450</p> <p>The standard associative retrieval techniques are first briefly reviewed. A computer experiment is then described which tends to confirm they hypothesis that documents exhibiting similar citation sets also deal with similar subject matter. (Salton, 1963, Abstract)</p> <p>The criteria of association used in most automatic programs do not normally require a determination of syntactic or semantic properties. Rather, they are based on simple co-occurrence of words in the same texts or sentences, or on co-occurrence with individual or joint frequencies greater than some given threshold value. (Salton, 1963, p. 443)</p> <p>Because of these and other variations, citation and reference lists have not generally been used as an indication of document content. Rather, such lists are used to detect trends in the literature as a whole, and to serve as adjuncts to certain kinds of literature searches [7, 8]. (Salton, 1963, p. 446)</p> <p>A citation index consists of a set of bibliographic references (the set of cited documents), each followed by a list of all those documents (the citing documents) which include the given cited document as a reference. A reference index, on the other hand, lists all cited documents under each citing document. (Salton, 1963, p. 446 n.1)</p> <p>Consider a collection of m documents each of which is characterized by the property of being cited by one of more of the other documents in the same collection. Each document can then be represented by an m-dimensional logical vector X_i, where $X_{ij} = 1$ if and only if document i is cited by document j, and $X_{ij} = 0$ otherwise. If these m vectors arranged in rows one below the other a square logical incidence matrix is formed similar to the matrix exhibited in Figure 4.</p>

Claim Text from '352 Patent	Salton, 1963															
	<div style="text-align: center;"> <table style="margin: auto;"> <tr> <td style="border-right: 1px solid black; padding: 5px;"><i>Cited documents</i></td> <td style="padding: 5px;"><i>Citing documents</i></td> <td></td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;"></td> <td style="padding: 5px;">D_1 D_2 \dots D_m</td> <td></td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;">D_1</td> <td style="padding: 5px;">X_1^1 X_2^1 \dots X_m^1</td> <td rowspan="4" style="vertical-align: middle; padding: 0 10px;">= X</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;">D_2</td> <td style="padding: 5px;">X_1^2 X_2^2 \dots X_m^2</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;">\vdots</td> <td style="padding: 5px;">\vdots</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;">D_m</td> <td style="padding: 5px;">X_1^m X_2^m \dots X_m^m</td> </tr> </table> <p style="text-align: center;">($X_i^j = 1 \leftrightarrow$ document D_i is cited by document D_j)</p> <p style="text-align: center;">FIG. 4. Matrix X exhibiting direct citations</p> <p>(Salton, 1963, p. 447)</p> <p>Figure 5</p> </div>	<i>Cited documents</i>	<i>Citing documents</i>			D_1 D_2 \dots D_m		D_1	X_1^1 X_2^1 \dots X_m^1	= X	D_2	X_1^2 X_2^2 \dots X_m^2	\vdots	\vdots	D_m	X_1^m X_2^m \dots X_m^m
<i>Cited documents</i>	<i>Citing documents</i>															
	D_1 D_2 \dots D_m															
D_1	X_1^1 X_2^1 \dots X_m^1	= X														
D_2	X_1^2 X_2^2 \dots X_m^2															
\vdots	\vdots															
D_m	X_1^m X_2^m \dots X_m^m															

Claim Text from '352 Patent	Salton, 1963
	 <p>The flowchart illustrates the process of comparing citation similarities with index term similarities. It begins with a step: 'Consider with each document the set of applicable index terms and the set of applicable citations'. This leads to two parallel paths: <ul style="list-style-type: none"> Left path: 'Construct a term-document incidence matrix \underline{C} listing documents against included terms' followed by 'Construct a document-document similarity matrix \underline{S} based on overlapping index terms'. Right path: 'Construct a citation incidence matrix \underline{X} listing each cited document against all citing documents' followed by 'Construct a document-document similarity matrix \underline{R} based on overlapping citations'. A central step, 'Compute a cross-correlation vector \underline{x} and overall cross-correlation coefficient x to measure similarities between document rows \underline{R} and \underline{S}, and between the complete matrices, respectively', receives input from both similarity matrices. Below this, two final steps are shown: <ul style="list-style-type: none"> Left: 'Construct term-term similarity matrix \underline{C}^* and use it to generate new term-document matrices $\underline{C}^1, \underline{C}^2, \dots$ and so on'. A dashed arrow points from this step back to the term-document incidence matrix \underline{C}. Right: 'Construct squared, cubed, ... incidence matrices $\underline{X}^1, \underline{X}^2, \dots$ exhibiting citation links of length two, three, ... and so on'. A dashed arrow points from this step back to the citation incidence matrix \underline{X}. </p> <p>FIG. 5. Comparison of citation similarities with index term similarities</p> <p>(Salton, 1963, p. 450)</p>
<p>[28b] the objects in the database are assigned chronological data;</p>	<p>See, e.g., Salton, 1963, at p. 446</p> <p>A second important criterion is the availability of the cited document. Thus, reports included in certain books or in important journals are likely to be cited more often than those not generally available to the public. By the same token, unclassified papers are cited more freely than classified ones. The date of publication is a related factor which also affects the</p>

Claim Text from '352 Patent	Salton, 1963
	<p>probability of being cited. Very recent documents which have not had a chance to circulate, and very old ones which no longer circulate are, in general, cited more rarely than current articles which have been distributed within the recent past. (Salton, 1963, p. 446)</p>
<p>[28c] and wherein the step of searching comprises the steps of matrix searching of the second matrices;</p>	<p>See, e.g., Salton, 1963, at pp. 443-45, 448, 450, 451-52</p> <p>Figure 2</p> $ \begin{array}{c cccc} \text{Terms} & D_1 & D_2 & \dots & D_m \\ \hline W_1 & C_1^1 & C_2^1 & \dots & C_m^1 \\ W_2 & C_1^2 & C_2^2 & \dots & C_m^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W_n & C_1^n & C_2^n & \dots & C_m^n \end{array} = \mathbf{C} $ <p>(a) Typical term-document incidence matrix \mathbf{C} ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times)</p> $ \begin{array}{c cccc} \text{Terms} & W_1 & W_2 & \dots & W_n \\ \hline W_1 & R_1^1 & R_2^1 & \dots & R_n^1 \\ W_2 & R_1^2 & R_2^2 & \dots & R_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W_n & R_1^n & R_2^n & \dots & R_n^n \end{array} = \mathbf{R} $ <p>(b) Typical term-term similarity matrix \mathbf{R}</p> $ \left(R_j^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \right) \left(\sum_{k=1}^m (C_k^j)^2 \right)}} \right) $ <p>FIG. 2. Matrices used for the generation of term associations</p> <p>(Salton, 1963, p. 443)</p> <p>Consider now a typical system for document retrieval using term and document associations as shown in Figure 3. A list of high-frequency terms is first generated for each document by word frequency counting procedures. Normalization may or may not be effected by thesaurus lookup. A term-term similarity matrix is then constructed by using co-occurrence</p>

Claim Text from '352 Patent

Salton, 1963

of terms within sentences, rather than within documents, as a criterion. It should be noted that as new term associations are defined, the original incidence matrix can be revised by inclusion in some of the matrix columns of new, associated terms which are not originally contained in the respective sentences or documents. The revised incidence matrix then gives rise to a new term-term similarity matrix, incorporating second-order associations, and so on. This feedback process is represented by an upward-pointing arrow in Figure 3. (Salton, 1963, p. 444)

Figure 3

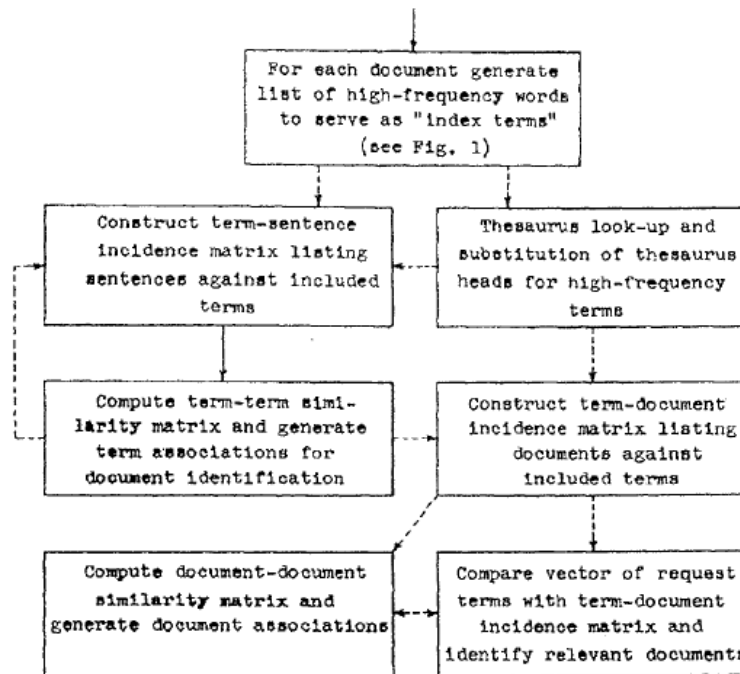


FIG. 3. Typical automatic document retrieval system using term and document associations
 → optional paths → compulsory paths

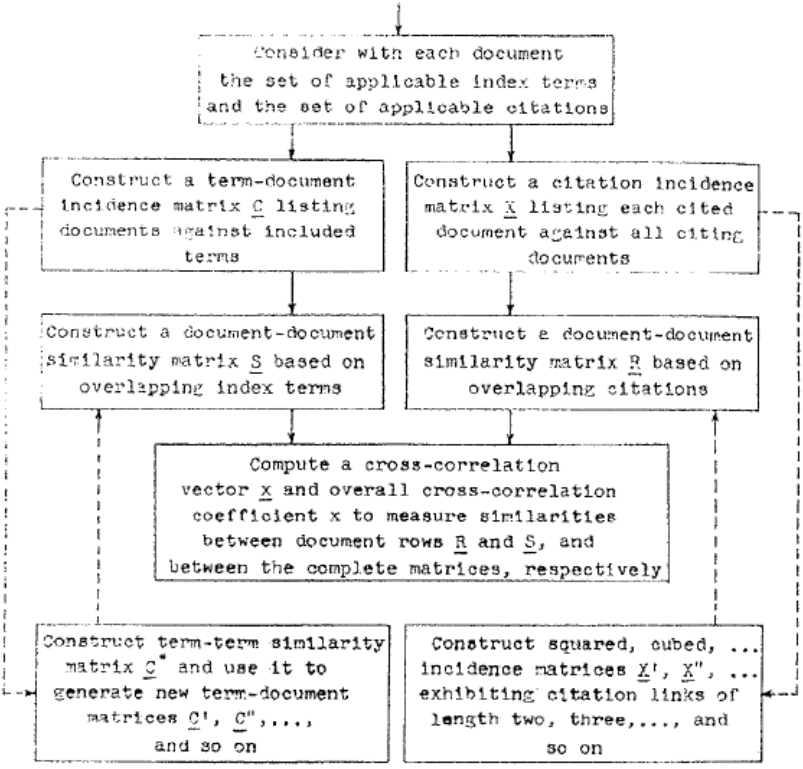
(Salton, 1963, p. 445)

Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X' , X'' , etc., exhibiting respectively the existence of paths of length two, three, and so on.

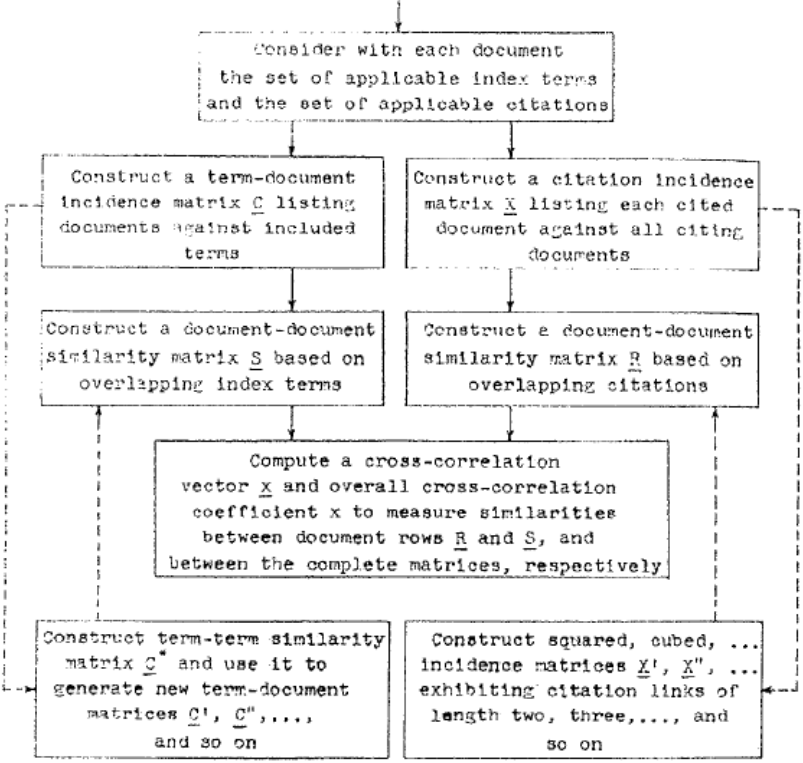
Claim Text from '352 Patent	Salton, 1963
	<p>Specifically,</p> $[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$ <p>Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. $(X')_{ij}$ is then equal to 1 if and only if at least one path of length two exists between documents D_i and D_j; otherwise, $(X')_{ij}$ is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)</p> <p>The CITED and CITING similarity matrices of dimension 62 by 62 were obtained from the original citation matrix by row and column comparisons, respectively. The TDCMP similarity matrix, also of dimension 62 by 62, was similarly obtained by column comparisons from the original term-document matrix. Additional citation similarity matrices, designated CTD2, CTD3, CTD4, and CNG2, CNG3, CNG4 were obtained from the squared, cubed, and fourth power logical citation matrices, as previously explained. (Salton, 1963, p. 450)</p> <p>The value of the overall similarity coefficient first rises as the length of the citation links increases, and then drops again as the length of the links becomes still greater [6]. This is due to the fact that as the length of the links increases, the total number of links of any length increases also; an increased number of links results in a larger number of ones in the original logical citation matrix, and thus in a higher probability of overlapping ones and a larger overall similarity coefficient. At the same time, as the length of the links increases, two factors also tend to decrease the magnitude of the overall similarity coefficient. First, the number of documents which exhibit citation links of length n but which do not exhibit links of length greater than n increases as n becomes larger. Thus more and more documents will exhibit individual similarity coefficients of zero value, thus tending to decrease the value of the overall coefficient. Second, as the length of the links increases and the citations thus become increasingly less accurate indications of document content, the magnitude of the cross-correlation coefficients obtained from the citation matrix and the term-document matrix would be expected to decrease, even for those documents for which a large number of citation links can still be found. (Salton, 1963, pp. 451-52)</p>

Claim Text from '352 Patent	Salton, 1963
[28d] and examining the chronological data.	Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.
29. The non-semantic method of claim 26 wherein the step of analyzing the first numerical representation further comprises: examining the first numerical representation for patterns which indicate the indirect relationships.	<p><i>See, e.g.,</i> Salton, 1963, at pp. 447-48, 450</p> <p>To test the significance of bibliographic citations, a comparison was made between citation similarities and index term similarities for an indexed document collection. Specifically, a measure of similarity was computed between each pair of documents in the collection, based on the number of overlapping index terms a similar measure was then computed for the same pairs of documents, based on the number of overlapping citations; finally, the similarity measures obtained from index terms and citations respectively were compared by calculating a similarity index between citation similarities and index term similarities. An overall measure was also computed for the complete document collection by taking into account the similarity measures between all document pairs. (Salton, 1963, p. 447)</p> <p>Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X', X'', etc., exhibiting respectively the existence of paths of length two, three, and so on. Specifically,</p> $[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$ <p>Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. (X')_{ij} is then equal to 1 if and only if at least one path of length two exists between documents Di and Dj; otherwise, (X')_{ij} is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)</p>

Claim Text from '352 Patent	Salton, 1963
	<p>A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the ith and jth rows (columns) of X.</p> <p>The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448)</p> <p>Figure 5</p>

Claim Text from '352 Patent	Salton, 1963
	 <p>The flowchart illustrates the process of comparing citation similarities with index term similarities. It begins with a central step: "Consider with each document the set of applicable index terms and the set of applicable citations". This leads to two parallel paths. The left path involves constructing a term-document incidence matrix C and a document-document similarity matrix S based on overlapping index terms. The right path involves constructing a citation incidence matrix X and a document-document similarity matrix R based on overlapping citations. A central step, "Compute a cross-correlation vector x and overall cross-correlation coefficient x to measure similarities between document rows R and S, and between the complete matrices, respectively", receives input from both S and R. Below this, two final steps are shown: "Construct term-term similarity matrix C^* and use it to generate new term-document matrices C^1, C^2, \dots and so on" (receiving input from C and C^*), and "Construct squared, cubed, ... incidence matrices X^1, X^2, \dots exhibiting citation links of length two, three, ..., and so on" (receiving input from X and X^1, X^2, \dots). Dashed lines indicate feedback loops from the final steps back to the initial construction steps.</p> <p>FIG. 5. Comparison of citation similarities with index term similarities</p> <p>(Salton, 1963, p. 450)</p>
<p>30. The non-semantic method of claim 29, given that object A occurs before object B and object c occurs before object A, and wherein the step of creating a first numerical representation comprises examining for the direct relationship B cites A and wherein the step of examining for patterns further comprises the step of examining for the following</p>	<p>See, e.g., Salton, 1963, at pp. 447-48, 450</p> <p>To test the significance of bibliographic citations, a comparison was made between citation similarities and index term similarities for an indexed document collection. Specifically, a measure of similarity was computed between each pair of documents in the collection, based on the number of overlapping index terms a similar measure was then computed for the same</p>

Claim Text from '352 Patent	Salton, 1963
<p>pattern: A cites c, and B cites c.</p>	<p>pairs of documents, based on the number of overlapping citations; finally, the similarity measures obtained from index terms and citations respectively were compared by calculating a similarity index between citation similarities and index term similarities. An overall measure was also computed for the complete document collection by taking into account the similarity measures between all document pairs. (Salton, 1963, p. 447)</p> <p>Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X', X'', etc., exhibiting respectively the existence of paths of length two, three, and so on. Specifically,</p> $[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$ <p>Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. (X')_{ij} is then equal to 1 if and only if at least one path of length two exists between documents D_i and D_j; otherwise, (X')_{ij} is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)</p> <p>A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the ith and jth rows (columns) of X.</p> <p>The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of</p>

Claim Text from '352 Patent	Salton, 1963
	<p>length three, and so on. (Salton, 1963, p. 448)</p> <p>Figure 5</p>  <pre> graph TD Start[Consider with each document the set of applicable index terms and the set of applicable citations] --> C[Construct a term-document incidence matrix C listing documents against included terms] Start --> X[Construct a citation incidence matrix X listing each cited document against all citing documents] C --> S[Construct a document-document similarity matrix S based on overlapping index terms] X --> R[Construct a document-document similarity matrix R based on overlapping citations] S --> CC[Compute a cross-correlation vector x and overall cross-correlation coefficient x to measure similarities between document rows R and S, and between the complete matrices, respectively] R --> CC CC --> C_prime[Construct term-term similarity matrix C' and use it to generate new term-document matrices C', C'', ..., and so on] CC --> X_prime[Construct squared, cubed, ... incidence matrices X', X'', ... exhibiting citation links of length two, three, ..., and so on] C_prime -.-> C X_prime -.-> X </pre> <p>FIG. 5. Comparison of citation similarities with index term similarities</p> <p>(Salton, 1963, p. 450)</p>
<p>31. The non-semantical method of claim 29, wherein a, b, c, A, d, e, f, B, g, h, and i are objects in the database and given that; a, b, and c occur before A;</p>	<p>See, e.g., Salton, 1963, at pp. 447-48, 450</p> <p>To test the significance of bibliographic citations, a comparison was made between citation similarities and index term similarities for an indexed document collection. Specifically, a</p>

Claim Text from '352 Patent	Salton, 1963
<p>A occurs before d, e, and f, which occur before B; and</p> <p>B occurs before g, h, and i;</p> <p>and wherein the step of examining for patterns further comprises the step of examining for one or more of the following patterns:</p> <p>(i) g cites A, and g cites B;</p> <p>(ii) B cites f, and f cites A;</p> <p>(iii) B cites f, f cites e, and e cites A;</p> <p>(iv) B cites f, f cites e, e cites d, and d cites A;</p> <p>(v) g cites A, h cites B, g cites a, and h cites a;</p> <p>(vi) i cites B, i cites f (or g), and f (or g) cites A;</p> <p>(vii) i cites g, i cites A, and g cites B;</p> <p>(viii) i cites g (or d), i cites h, g (or d) cites A, and h cites B;</p> <p>(ix) i cites a, i cites B, and A cites a;</p> <p>(x) i cites A, i cites e, B cites e;</p> <p>(xi) g cites A, g cites a, A cites a, h cites B, and h cites a;</p> <p>(xii) A cites a, B cites d, i cites a, and i cites d;</p> <p>(xiii) i cites B, i cites d, A cites a, and d cites a;</p> <p>(xiv) A cites b, B cites d (or c), and d (or c) cites b;</p> <p>(xv) A cites b, B cites d, b cites a, and d cites a;</p> <p>(xvi) A cites a, B cites b, d (or c) cites a, and d (or c) cites b.</p>	<p>measure of similarity was computed between each pair of documents in the collection, based on the number of overlapping index terms a similar measure was then computed for the same pairs of documents, based on the number of overlapping citations; finally, the similarity measures obtained from index terms and citations respectively were compared by calculating a similarity index between citation similarities and index term similarities. An overall measure was also computed for the complete document collection by taking into account the similarity measures between all document pairs. (Salton, 1963, p. 447)</p> <p>Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X', X'', etc., exhibiting respectively the existence of paths of length two, three, and so on. Specifically,</p> $[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$ <p>Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. (X')_{ij} is then equal to 1 if and only if at least one path of length two exists between documents Di and Dj; otherwise, (X')_{ij} is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)</p> <p>A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where Rij is the value of the similarity coefficient between the ith and jth rows (columns) of X.</p> <p>The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct link exists between A and B. Instead, A and B are then linked by a path of length two, since</p>

Claim Text from '352 Patent

Salton, 1963

an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448)

Figure 5

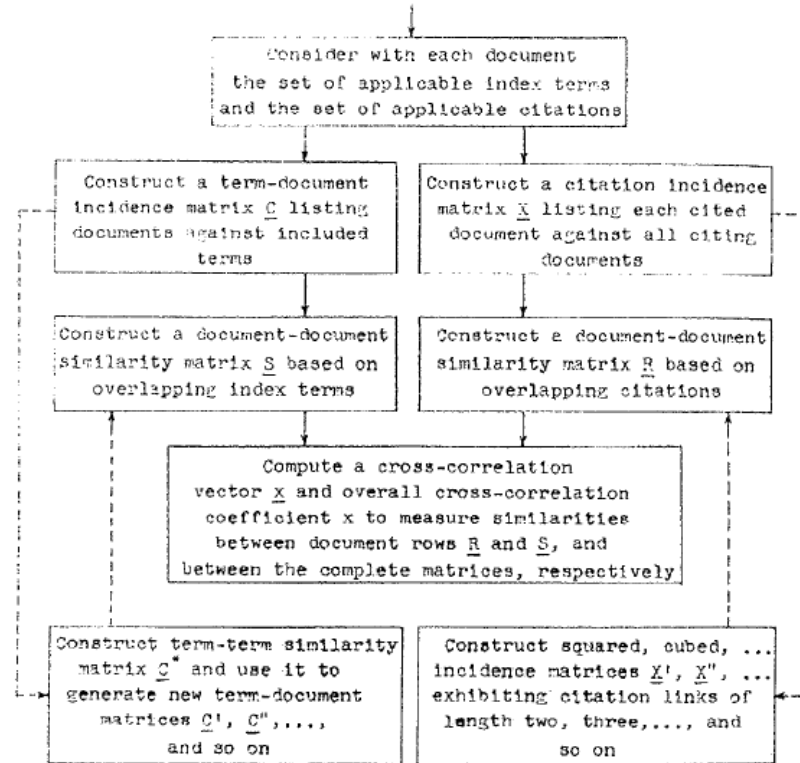


FIG. 5. Comparison of citation similarities with index term similarities

(Salton, 1963, p. 450)

32. The non-semantic method of claim 26, wherein the step of analyzing further comprises the

See, e.g., Salton, 1963, at pp. 444, 448, 450, 451-52

Claim Text from '352 Patent	Salton, 1963
<p>step of weighing, wherein some indirect relationships are weighed more heavily than other indirect relationships.</p>	<p>To retrieve documents in answer to search requests, the programs already available can be used by adding to the term-document matrix C a new column Cm+1, representing the request terms. Specifically, element Ckm+1 is set equal to w if term Wk is used in the search request with weight w; if word Wk is not used in the given search request Ckm+1 is set equal to 0. If no weights are specified by the requestor the values of the elements of column Cm+1 are restricted to 0 and 1. (Salton, 1963, p. 444)</p> <p>Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X', X'', etc., exhibiting respectively the existence of paths of length two, three, and so on. Specifically,</p> $[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$ <p>Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. (X')ij is then equal to 1 if and only if at least one path of length two exists between documents Di and Dj; otherwise, (X')ij is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)</p> <p>The CITED and CITING similarity matrices of dimension 62 by 62 were obtained from the original citation matrix by row and column comparisons, respectively. The TDCMP similarity matrix, also of dimension 62 by 62, was similarly obtained by column comparisons from the original term-document matrix. Additional citation similarity matrices, designated CTD2, CTD3, CTD4, and CNG2, CNG3, CNG4 were obtained from the squared, cubed, and fourth power logical citation matrices, as previously explained. (Salton, 1963, p. 450)</p> <p>The value of the overall similarity coefficient first rises as the length of the citation links increases, and then drops again as the length of the links becomes still greater [6]. This is due to the fact that as the length of the links increases, the total number of links of any length increases also; an increased number of links results in a larger number of ones in the original logical citation matrix, and thus in a higher probability of overlapping ones and a larger overall similarity coefficient. At the same time, as the length of the links increases, two factors also tend to decrease the magnitude of the overall similarity coefficient. First, the</p>

Claim Text from '352 Patent	Salton, 1963
	<p>number of documents which exhibit citation links of length n but which do not exhibit links of length greater than n increases as n becomes larger. Thus more and more documents will exhibit individual similarity coefficients of zero value, thus tending to decrease the value of the overall coefficient. Second, as the length of the links increases and the citations thus become increasingly less accurate indications of document content, the magnitude of the cross-correlation coefficients obtained from the citation matrix and the term-document matrix would be expected to decrease, even for those documents for which a large number of citation links can still be found. (Salton, 1963, pp. 451-52)</p>
<p>33. The non-semantic method of claim 26, wherein the step of analyzing the first numerical representations for indirect relationships further comprises:</p>	<p><i>See, e.g.</i>, Salton, 1963, at pp. 448, 450</p> <p>Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X', X'', etc., exhibiting respectively the existence of paths of length two, three, and so on. Specifically,</p> $[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$ <p>Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. (X')_{ij} is then equal to 1 if and only if at least one path of length two exists between documents D_i and D_j; otherwise, (X')_{ij} is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)</p> <p>A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the ith and jth rows (columns) of X.</p> <p>The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between</p>

Claim Text from '352 Patent

Salton, 1963

documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448)

Figure 5

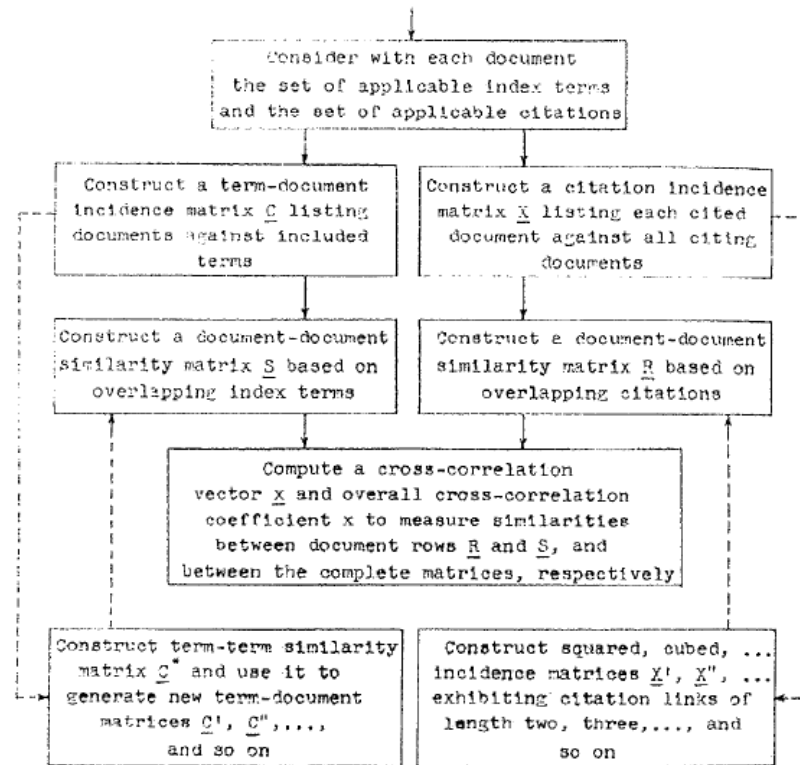


FIG. 5. Comparison of citation similarities with index term similarities

Claim Text from '352 Patent

Salton, 1963

(Salton, 1963, p. 450)

[33a] creating an interim vector representing each object; and wherein the step of generating a second numerical representation uses coefficients of similarity and further comprises:

See, e.g., Salton, 1963, at pp. 443-45, 447-50

Figure 2

Terms	Documents					
	D_1	D_2	...		D_m	
W_1	C_1^1	C_2^1	...	C_j^1	...	C_m^1
W_2	C_1^2	C_2^2	...	C_j^2	...	C_m^2
\vdots						
W_n	C_1^n	C_2^n	...	C_j^n	...	C_m^n

$$= C$$

(a) Typical term-document incidence matrix C ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times)

Terms	Terms			
	W_1	W_2	...	W_n
W_1	R_1^1	R_2^1	...	R_n^1
W_2	R_1^2	R_2^2	...	R_n^2
\vdots				
W_n	R_1^n	R_2^n	...	R_n^n

$$= R$$

(b) Typical term-term similarity matrix R

$$\left(R_j^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \sum_{k=1}^m (C_k^j)^2 \right)}} \right)$$

FIG. 2. Matrices used for the generation of term associations

(Salton, 1963, p. 443)

Many different types of similarity coefficients have been suggested in the literature; a simple coefficient of similarity between rows of a numeric matrix, and one which may be as meaningful as any of the others, is the cosine of the angle between the corresponding m-dimensional vectors. The similarity coefficients can be displayed in an $n \times n$ symmetric term-similarity matrix R , where the coefficient of similarity R_{ji} between term W_i and term W_j is

Claim Text from '352 Patent	Salton, 1963
	$R_i^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \sum_{k=1}^m (C_k^j)^2\right)}}$ <p data-bbox="810 467 1848 597">Document similarities are therefore obtained by comparing pairs of columns (instead of rows) of the term-document matrix C, and a document-document similarity matrix is constructed and used in the same way as the previously described term-term matrix R. (Salton, 1963, pp. 443-44)</p> <p data-bbox="810 613 1919 844">A term-term similarity matrix is then constructed by using co-occurrence of terms within sentences rather than within documents as a criterion. It should be noted that as new term associations are defined, the original incidence matrix can be revised by inclusion in some of the matrix columns of new, associated terms which are not originally contained in the respective sentences or documents. The revised incidence matrix then gives rise to a new term-term similarity matrix, incorporating second-order associations, and so on. (Salton, 1963, p. 444)</p> <p data-bbox="810 857 919 889">Figure 3</p>

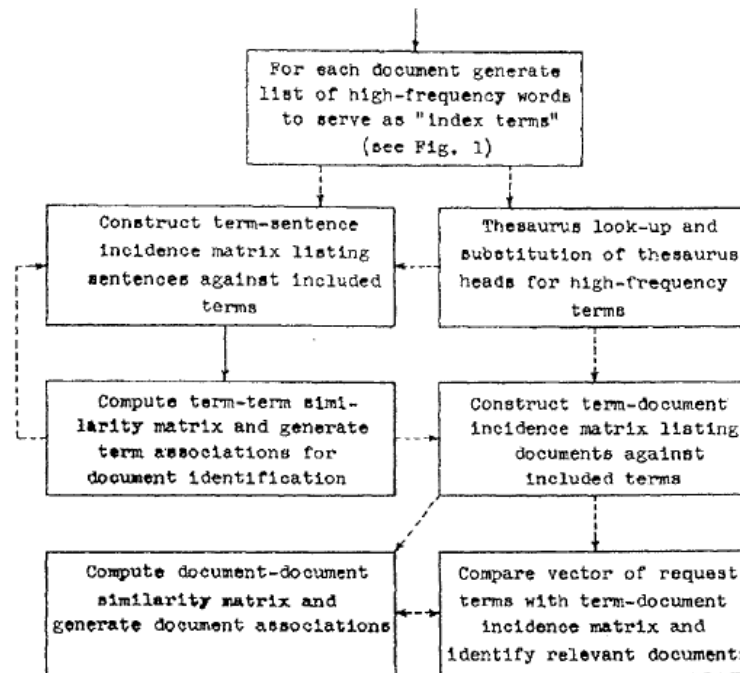


Fig. 3. Typical automatic document retrieval system using term and document associations
 → optional paths → compulsory paths

(Salton, 1963, p. 445)

Consider a collection of m documents each of which is characterized by the property of being cited by one or more of the other documents in the same collection. Each document can then be represented by an m -dimensional logical vector X_i , where $X_{ij} = 1$ if and only if document i is cited by document j , and $X_{ij} = 0$ otherwise. If these m vectors arranged in rows one below the other a square logical incidence matrix is formed similar to the matrix exhibited in Figure 4.

<i>Cited documents</i>	<i>Citing documents</i>	
	D_1 D_2 \dots D_m	
D_1	X_1^1 X_2^1 \dots X_m^1	= X
D_2	X_1^2 X_2^2 \dots X_m^2	
\vdots	\vdots	
D_m	X_1^m X_2^m \dots X_m^m	

$(X_j^i = 1 \leftrightarrow \text{document } D_i \text{ is cited by document } D_j)$

FIG. 4. Matrix X exhibiting direct citations

(Salton, 1963, p. 447)

Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X', X'', etc., exhibiting respectively the existence of paths of length two, three, and so on. Specifically,

$$[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$$

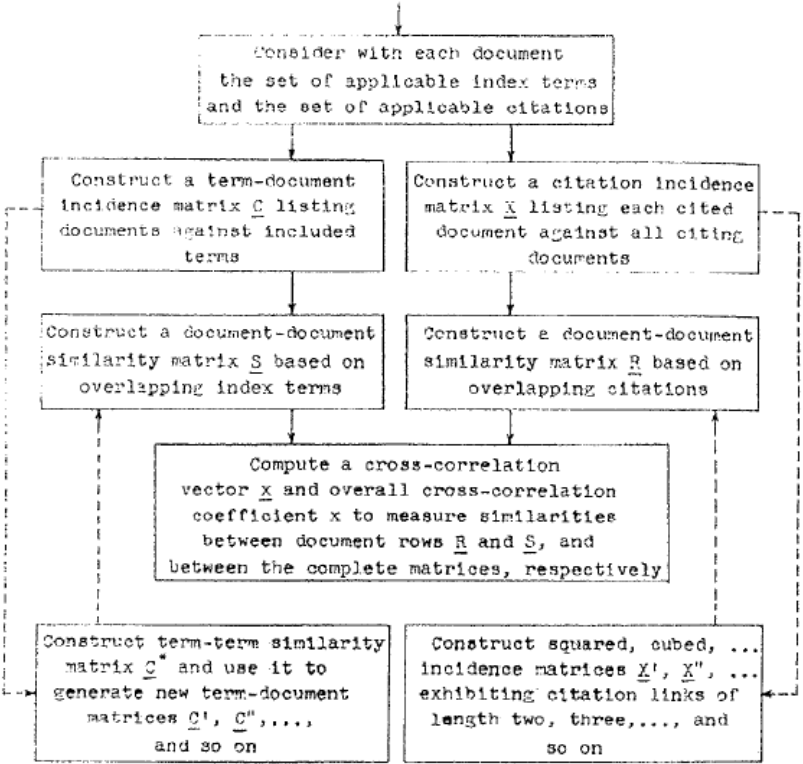
$$[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$$

Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. $(X')_{ij}$ is then equal to 1 if and only if at least one path of length two exists between documents D_i and D_j ; otherwise, $(X')_{ij}$ is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)

A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the i th and j th rows (columns) of X.

The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between

Claim Text from '352 Patent	Salton, 1963
	<p>documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448)</p> <p>Since the term-document matrix C is not in general a square matrix, matrix multiplication cannot be used to obtain second order effects, similar to the citation links of length two or more. Instead, it is first necessary to compare the index terms by performing a row comparison of the rows of C. This produces a new n symmetric term matrix C^* which displays similarity between index terms. This matrix can be used to eliminate from the set of index terms those terms which exhibit a large number of joint occurrences with other terms. A reduced set of index terms can then be formed and a new term-document matrix C' constructed, from which a new correlation matrix S' is formed. (Salton, 1963, p. 449)</p> <p>Figure 5</p>

Claim Text from '352 Patent	Salton, 1963
	 <p>The flowchart illustrates the process of comparing citation similarities with index term similarities. It begins with a box: "Consider with each document the set of applicable index terms and the set of applicable citations". This leads to two parallel paths. The left path: "Construct a term-document incidence matrix \underline{C} listing documents against included terms" leads to "Construct a document-document similarity matrix \underline{S} based on overlapping index terms". The right path: "Construct a citation incidence matrix \underline{X} listing each cited document against all citing documents" leads to "Construct a document-document similarity matrix \underline{R} based on overlapping citations". Both paths converge at a central box: "Compute a cross-correlation vector \underline{x} and overall cross-correlation coefficient x to measure similarities between document rows \underline{R} and \underline{S}, and between the complete matrices, respectively". From this central box, two dashed arrows point to two final boxes. The left box: "Construct term-term similarity matrix \underline{C}^* and use it to generate new term-document matrices \underline{C}', \underline{C}'', ..., and so on". The right box: "Construct squared, cubed, ... incidence matrices \underline{X}', \underline{X}'', ... exhibiting citation links of length two, three, ..., and so on".</p> <p>FIG. 5. Comparison of citation similarities with index term similarities</p> <p>(Salton, 1963, p. 450)</p>
<p>[33b] calculating Euclidean distances between interim vector representations of each object;</p>	<p>See, e.g., Salton, 1963, at pp. 443-44, 447, 448</p> <p>Many different types of similarity coefficients have been suggested in the literature; a simple coefficient of similarity between rows of a numeric matrix, and one which may be as meaningful as any of the others, is the cosine of the angle between the corresponding m-dimensional vectors. The similarity coefficients can be displayed in an $n \times n$ symmetric</p>

Claim Text from '352 Patent	Salton, 1963
	<p>term-similarity matrix R, where the coefficient of similarity R_{ji} between term W_i and term W_j is</p> $R_i^j = R_j^i = \frac{\sum_{k=1}^m c_k^i c_k^j}{\sqrt{\left(\sum_{k=1}^m (c_k^i)^2\right) \left(\sum_{k=1}^m (c_k^j)^2\right)}}$ <p>Document similarities are therefore obtained by comparing pairs of columns (instead of rows) of the term-document matrix C, and a document-document similarity matrix is constructed and used in the same way as the previously described term-term matrix R. (Salton, 1963, pp. 443-44)</p> <p>To test the significance of bibliographic citations, a comparison was made between citation similarities and index term similarities for an indexed document collection. Specifically, a measure of similarity was computed between each pair of documents in the collection, based on the number of overlapping index terms a similar measure was then computed for the same pairs of documents, based on the number of overlapping citations; finally, the similarity measures obtained from index terms and citations respectively were compared by calculating a similarity index between citation similarities and index term similarities. An overall measure was also computed for the complete document collection by taking into account the similarity measures between all document pairs. (Salton, 1963, p. 447)</p> <p>A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the ith and jth rows (columns) of X. (Salton, 1963, p. 448)</p> <p>Further, disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.</p>

Claim Text from '352 Patent	Salton, 1963
<p>[33c] creating proximity vectors representing the objects using the calculated Euclidean distances; and</p>	<p>See, e.g., Salton, 1963, at pp. 443-44, 447, 448</p> <p>Many different types of similarity coefficients have been suggested in the literature; a simple coefficient of similarity between rows of a numeric matrix, and one which may be as meaningful as any of the others, is the cosine of the angle between the corresponding m-dimensional vectors. The similarity coefficients can be displayed in an n x n symmetric term-similarity matrix R, where the coefficient of similarity R_{ij} between term W_i and term W_j is</p> $R_i^j = R_j^i = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \sum_{k=1}^m (C_k^j)^2\right)}}$ <p>...</p> <p>Document similarities are therefore obtained by comparing pairs of columns (instead of rows) of the term-document matrix C, and a document-document similarity matrix is constructed and used in the same way as the previously described term-term matrix R. (Salton, 1963, pp. 443-44)</p> <p>To test the significance of bibliographic citations, a comparison was made between citation similarities and index term similarities for an indexed document collection. Specifically, a measure of similarity was computed between each pair of documents in the collection, based on the number of overlapping index terms a similar measure was then computed for the same pairs of documents, based on the number of overlapping citations; finally, the similarity measures obtained from index terms and citations respectively were compared by calculating a similarity index between citation similarities and index term similarities. An overall measure was also computed for the complete document collection by taking into account the similarity measures between all document pairs. (Salton, 1963, p. 447)</p> <p>A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that</p>

Claim Text from '352 Patent	Salton, 1963																																																																						
	<p>shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the ith and jth rows (columns) of X. (Salton, 1963, p. 448)</p> <p>Further, disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.</p>																																																																						
<p>[33d] using the proximity vectors and using coefficients of similarity to calculate the second numerical representations.</p>	<p>See, e.g., Salton, 1963, at pp. 443, 444, 445, 448, 449, 450</p> <p>Figure 2</p> <div style="text-align: center;"> <table border="1" style="margin: auto;"> <thead> <tr> <th style="border-right: 1px solid black;">Terms</th> <th colspan="5">Documents</th> </tr> <tr> <th style="border-right: 1px solid black;"></th> <th>D_1</th> <th>D_2</th> <th>...</th> <th></th> <th>D_m</th> </tr> </thead> <tbody> <tr> <td style="border-right: 1px solid black;">W_1</td> <td>C_1^1</td> <td>C_2^1</td> <td>...</td> <td>C_j^1</td> <td>...</td> <td>C_m^1</td> </tr> <tr> <td style="border-right: 1px solid black;">W_2</td> <td>C_1^2</td> <td>C_2^2</td> <td>...</td> <td>C_j^2</td> <td>...</td> <td>C_m^2</td> </tr> <tr> <td style="border-right: 1px solid black;">\vdots</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td style="border-right: 1px solid black;">W_n</td> <td>C_1^n</td> <td>C_2^n</td> <td>...</td> <td>C_j^n</td> <td>...</td> <td>C_m^n</td> </tr> </tbody> </table> $= C$ <p>(a) Typical term-document incidence matrix C ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times)</p> <table border="1" style="margin: auto;"> <thead> <tr> <th style="border-right: 1px solid black;">Terms</th> <th colspan="4">Terms</th> </tr> <tr> <th style="border-right: 1px solid black;"></th> <th>W_1</th> <th>W_2</th> <th>...</th> <th>W_n</th> </tr> </thead> <tbody> <tr> <td style="border-right: 1px solid black;">W_1</td> <td>R_1^1</td> <td>R_2^1</td> <td>...</td> <td>R_n^1</td> </tr> <tr> <td style="border-right: 1px solid black;">W_2</td> <td>R_1^2</td> <td>R_2^2</td> <td>...</td> <td>R_n^2</td> </tr> <tr> <td style="border-right: 1px solid black;">\vdots</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td style="border-right: 1px solid black;">W_n</td> <td>R_1^n</td> <td>R_2^n</td> <td>...</td> <td>R_n^n</td> </tr> </tbody> </table> $= R$ <p>(b) Typical term-term similarity matrix R</p> $\left(R_i^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \right) \left(\sum_{k=1}^m (C_k^j)^2 \right)}} \right)$ </div> <p>FIG. 2. Matrices used for the generation of term associations</p>	Terms	Documents						D_1	D_2	...		D_m	W_1	C_1^1	C_2^1	...	C_j^1	...	C_m^1	W_2	C_1^2	C_2^2	...	C_j^2	...	C_m^2	\vdots							W_n	C_1^n	C_2^n	...	C_j^n	...	C_m^n	Terms	Terms					W_1	W_2	...	W_n	W_1	R_1^1	R_2^1	...	R_n^1	W_2	R_1^2	R_2^2	...	R_n^2	\vdots					W_n	R_1^n	R_2^n	...	R_n^n
Terms	Documents																																																																						
	D_1	D_2	...		D_m																																																																		
W_1	C_1^1	C_2^1	...	C_j^1	...	C_m^1																																																																	
W_2	C_1^2	C_2^2	...	C_j^2	...	C_m^2																																																																	
\vdots																																																																							
W_n	C_1^n	C_2^n	...	C_j^n	...	C_m^n																																																																	
Terms	Terms																																																																						
	W_1	W_2	...	W_n																																																																			
W_1	R_1^1	R_2^1	...	R_n^1																																																																			
W_2	R_1^2	R_2^2	...	R_n^2																																																																			
\vdots																																																																							
W_n	R_1^n	R_2^n	...	R_n^n																																																																			

(Salton, 1963, p. 443)

A term-term similarity matrix is then constructed by using co-occurrence of terms within sentences rather than within documents as a criterion. It should be noted that as new term associations are defined, the original incidence matrix can be revised by inclusion in some of the matrix columns of new, associated terms which are not originally contained in the respective sentences or documents. The revised incidence matrix then gives rise to a new term-term similarity matrix, incorporating second-order associations, and so on. (Salton, 1963, p. 444)

Figure 3

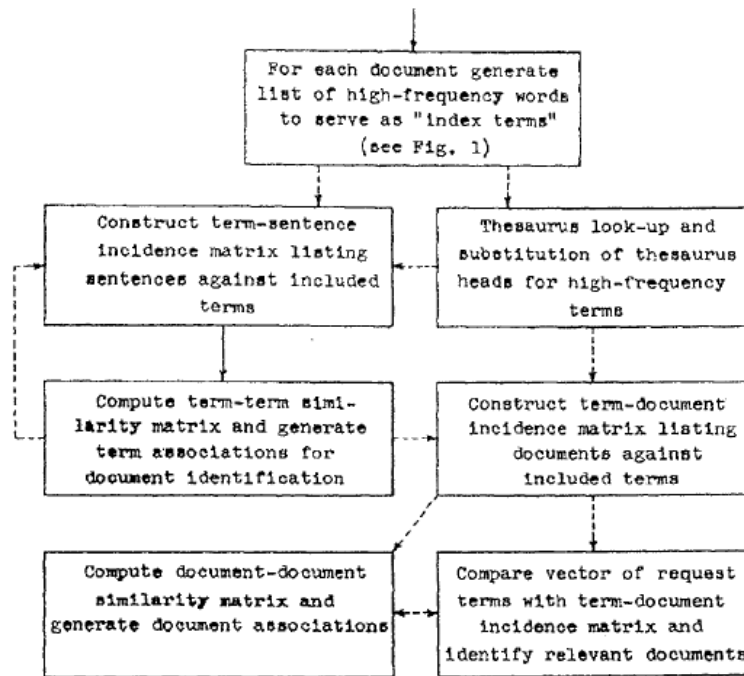
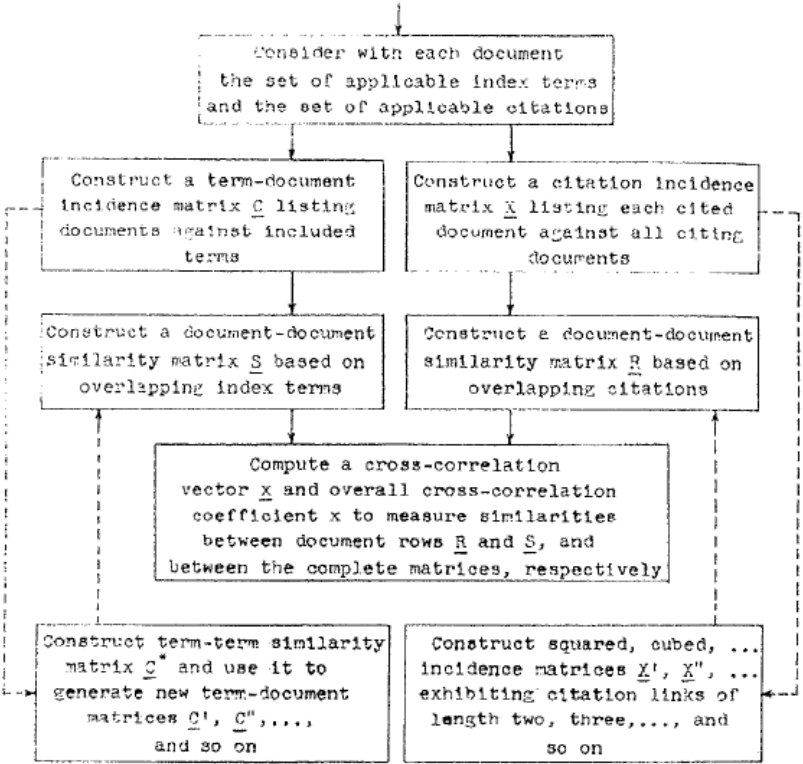


FIG. 3. Typical automatic document retrieval system using term and document associations
 → optional paths → compulsory paths

(Salton, 1963, p. 445)

A measure of similarity between row (column) vectors can be obtained by calculating the

Claim Text from '352 Patent	Salton, 1963
	<p>cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the ith and jth rows (columns) of X.</p> <p>The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448)</p> <p>Since the term-document matrix C is not in general a square matrix, matrix multiplication cannot be used to obtain second order effects, similar to the citation links of length two or more. Instead, it is first necessary to compare the index terms by performing a row comparison of the rows of C. This produces a new n symmetric term matrix C^* which displays similarity between index terms. This matrix can be used to eliminate from the set of index terms those terms which exhibit a large number of joint occurrences with other terms. A reduced set of index terms can then be formed and a new term-document matrix C' constructed, from which a new correlation matrix S' is formed. (Salton, 1963, p. 449)</p> <p>Figure 5</p>

Claim Text from '352 Patent	Salton, 1963
	 <p>The flowchart illustrates the process of comparing citation similarities with index term similarities. It begins with a step: "Consider with each document the set of applicable index terms and the set of applicable citations". This leads to two parallel paths: <ul style="list-style-type: none"> Left path: "Construct a term-document incidence matrix \underline{C} listing documents against included terms" followed by "Construct a document-document similarity matrix \underline{S} based on overlapping index terms". Right path: "Construct a citation incidence matrix \underline{X} listing each cited document against all citing documents" followed by "Construct a document-document similarity matrix \underline{R} based on overlapping citations". A central step: "Compute a cross-correlation vector \underline{x} and overall cross-correlation coefficient x to measure similarities between document rows \underline{R} and \underline{S}, and between the complete matrices, respectively" receives input from both similarity matrices. Below this, two more steps are shown: <ul style="list-style-type: none"> Left: "Construct term-term similarity matrix \underline{C}^* and use it to generate new term-document matrices $\underline{C}^1, \underline{C}^2, \dots$ and so on". Right: "Construct squared, cubed, ... incidence matrices $\underline{X}^1, \underline{X}^2, \dots$ exhibiting citation links of length two, three, ... and so on". Dashed feedback arrows connect the bottom steps back to the top of their respective paths.</p> <p>FIG. 5. Comparison of citation similarities with index term similarities</p> <p>(Salton, 1963, p. 450)</p>
<p>34. The non-semantic method of claim 26, wherein objects in the database may be divided into subsets and wherein the marking step includes the step of marking subsets of objects in the database and wherein relationships exist between or among subsets of objects in the database.</p>	<p>See, e.g., Salton, 1963, at pp. 441, 444, 447</p> <p>Figure 1</p>

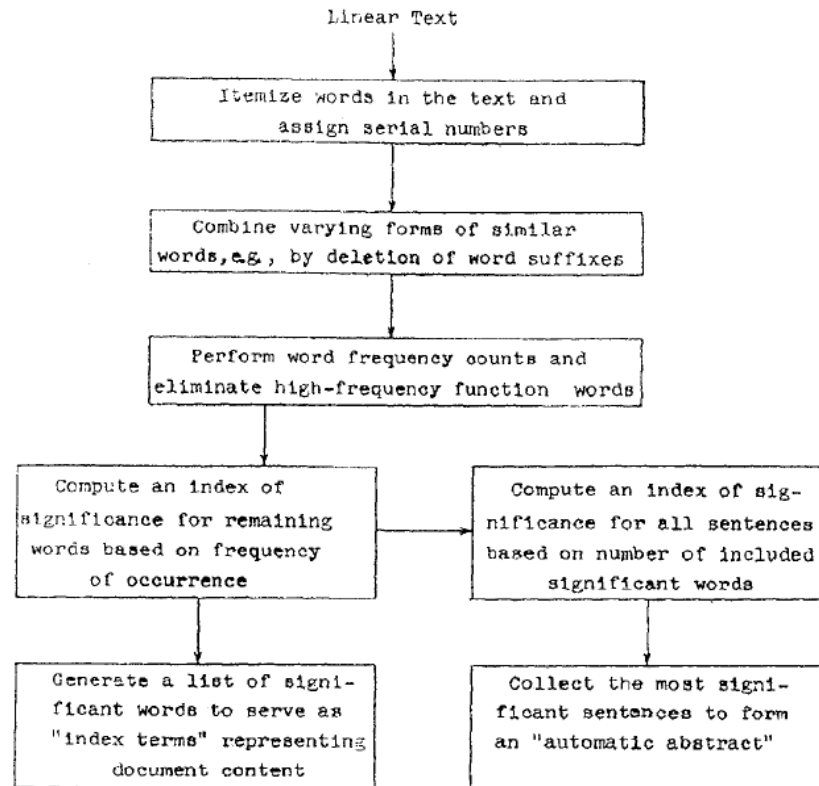


FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts.

(Salton, 1963, p. 441)

To generate document associations instead of term associations the same procedures can be used, since the strength of association between documents may be conveniently assumed to be a function of the number and frequencies of the shared terms in their respective term lists. Document similarities are therefore obtained by comparing pairs of columns (instead of rows) of the term-document matrix C , and a document-document similarity matrix is constructed and used in the same way as the previously described term-term matrix R .

(Salton, 1963, p. 444)

Claim Text from '352 Patent	Salton, 1963
	<p>Consider a collection of m documents each of which is characterized by the property of being cited by one or more of the other documents in the same collection. Each document can then be represented by an m-dimensional logical vector X_i, where $X_{ji} = 1$ if and only if document i is cited by document j, and $X_{ji} = 0$ otherwise. (Salton, 1963, p. 447)</p>
<p>35. The non-semantic method of claim 34 wherein the objects are textual objects with paragraphs and the subsets are the paragraphs of the textual objects, the method further comprising the steps of:</p>	<p><i>See, e.g.,</i> Salton, 1963, at pp. 441, 444</p> <p>Figure 1</p>

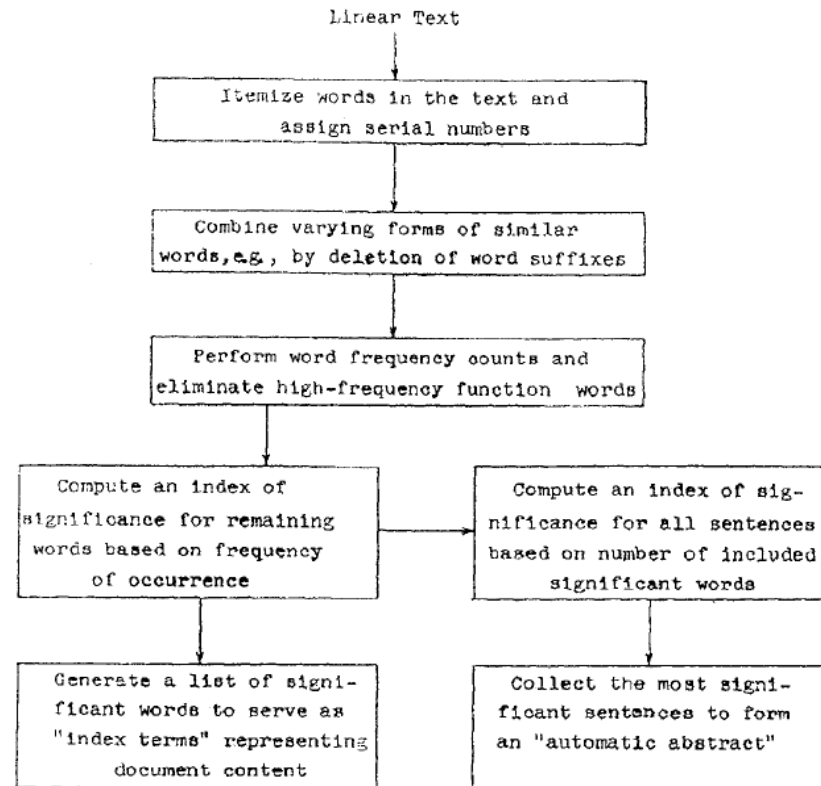


FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts.

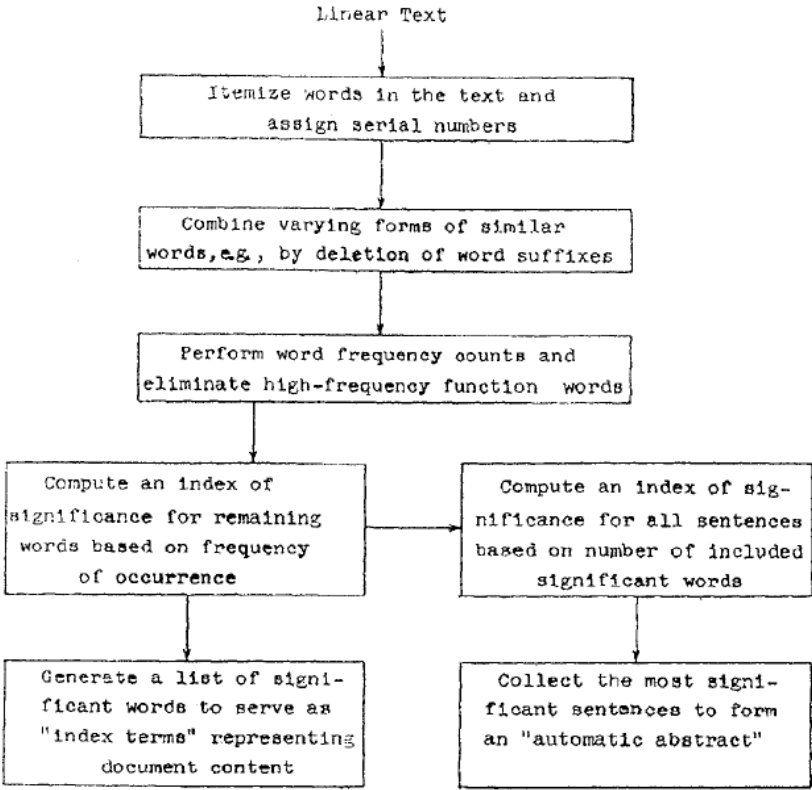
(Salton, 1963, p. 441)

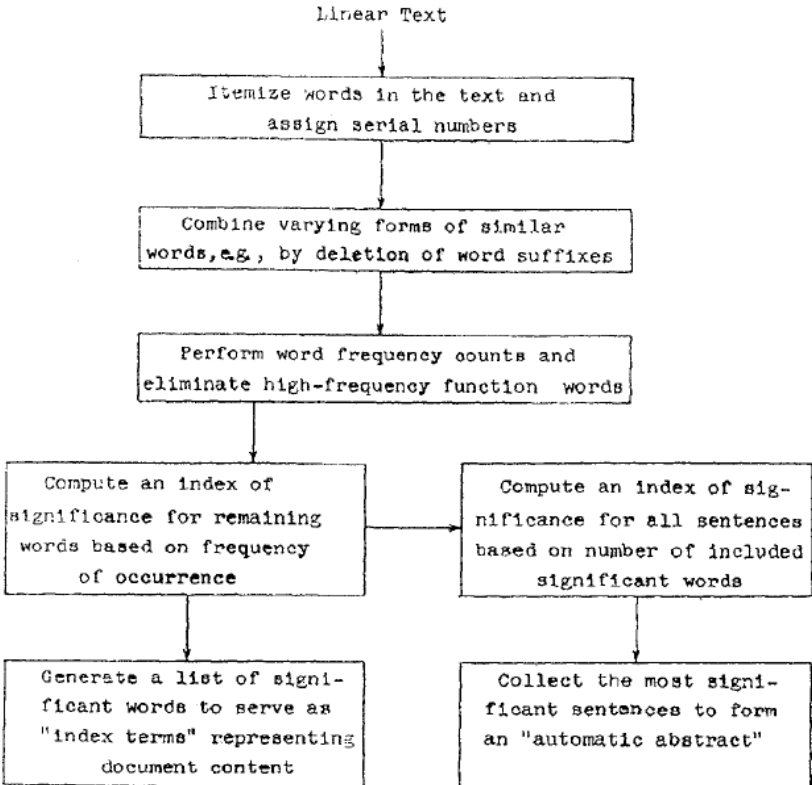
To generate document associations instead of term associations the same procedures can be used, since the strength of association between documents may be conveniently assumed to be a function of the number and frequencies of the shared terms in their respective term lists. Document similarities are therefore obtained by comparing pairs of columns (instead of rows) of the term-document matrix C , and a document-document similarity matrix is constructed and used in the same way as the previously described term-term matrix R .

(Salton, 1963, p. 444)

Claim Text from '352 Patent	Salton, 1963
<p>[35a] creating a subset numerical representation for each subset based upon the relationships between or among subsets;</p>	<p><i>See, e.g.</i>, Salton, 1963, at p. 444</p> <p>To generate document associations instead of term associations the same procedures can be used, since the strength of association between documents may be conveniently assumed to be a function of the number and frequencies of the shared terms in their respective term lists. Document similarities are therefore obtained by comparing pairs of columns (instead of rows) of the term-document matrix <i>C</i>, and a document-document similarity matrix is constructed and used in the same way as the previously described term-term matrix <i>R</i>. (Salton, 1963, p. 444)</p>
<p>[35b] analyzing the subset numerical representations;</p>	<p><i>See, e.g.</i>, Salton, 1963, at p. 441</p> <p>Figure 1</p>

Claim Text from '352 Patent	Salton, 1963
	<pre> graph TD A[Linear Text] --> B[Itemize words in the text and assign serial numbers] B --> C[Combine varying forms of similar words, e.g., by deletion of word suffixes] C --> D[Perform word frequency counts and eliminate high-frequency function words] D --> E[Compute an index of significance for remaining words based on frequency of occurrence] D --> F[Compute an index of significance for all sentences based on number of included significant words] E --> G[Generate a list of significant words to serve as "index terms" representing document content] F --> H[Collect the most significant sentences to form an "automatic abstract"] </pre> <p>FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts.</p> <p>(Salton, 1963, p. 441)</p>
[35c] clustering the subsets into sections based upon the subset analysis; and	<p>See, e.g., Salton, 1963, at p. 441</p> <p>Figure 1</p>

Claim Text from '352 Patent	Salton, 1963
	 <pre> graph TD A[Linear Text] --> B[Itemize words in the text and assign serial numbers] B --> C[Combine varying forms of similar words, e.g., by deletion of word suffixes] C --> D[Perform word frequency counts and eliminate high-frequency function words] D --> E[Compute an index of significance for remaining words based on frequency of occurrence] D --> F[Compute an index of significance for all sentences based on number of included significant words] E --> G[Generate a list of significant words to serve as "index terms" representing document content] F --> H[Collect the most significant sentences to form an "automatic abstract"] </pre> <p data-bbox="829 1063 1722 1112">FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts.</p> <p data-bbox="819 1128 1081 1161">(Salton, 1963, p. 441)</p>
<p data-bbox="184 1226 798 1323">[35d] generating a section numerical representation for each section, wherein the section numerical representations are available for searching.</p>	<p data-bbox="819 1226 1197 1258"><i>See, e.g.,</i> Salton, 1963, at p. 441</p> <p data-bbox="819 1307 924 1339">Figure 1</p>

Claim Text from '352 Patent	Salton, 1963
	 <pre> graph TD A[Linear Text] --> B[Itemize words in the text and assign serial numbers] B --> C[Combine varying forms of similar words, e.g., by deletion of word suffixes] C --> D[Perform word frequency counts and eliminate high-frequency function words] D --> E[Compute an index of significance for remaining words based on frequency of occurrence] D --> F[Compute an index of significance for all sentences based on number of included significant words] E --> G[Generate a list of significant words to serve as "index terms" representing document content] F --> H[Collect the most significant sentences to form an "automatic abstract"] </pre> <p data-bbox="829 1063 1722 1112">FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts.</p> <p data-bbox="819 1128 1081 1161">(Salton, 1963, p. 441)</p>
<p data-bbox="184 1226 798 1388">36. The non-semantic method of claim 26, wherein the step of searching the objects comprises the steps of: selecting an object; using the second numerical representation to search for objects similar to the selected object.</p>	<p data-bbox="819 1226 1333 1258"><i>See, e.g.,</i> Salton, 1963, at pp. 443, 444, 445</p> <p data-bbox="819 1307 924 1339">Figure 2</p>

Terms	Documents			
	D_1	D_2	...	D_m
W_1	C_1^1	C_2^1	...	C_m^1
W_2	C_1^2	C_2^2	...	C_m^2
\vdots	\vdots	\vdots	\vdots	\vdots
W_n	C_1^n	C_2^n	...	C_m^n

$$= C$$

(a) Typical term-document incidence matrix C ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times)

Terms	Terms			
	W_1	W_2	...	W_n
W_1	R_1^1	R_2^1	...	R_n^1
W_2	R_1^2	R_2^2	...	R_n^2
\vdots	\vdots	\vdots	\vdots	\vdots
W_n	R_1^n	R_2^n	...	R_n^n

$$= R$$

(b) Typical term-term similarity matrix R

$$\left(R_j^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \sum_{k=1}^m (C_k^j)^2 \right)}} \right)$$

FIG. 2. Matrices used for the generation of term associations

(Salton, 1963, p. 443)

Consider now a typical system for document retrieval using term and document associations as shown in Figure 3. A list of high-frequency terms is first generated for each document by word frequency counting procedures. Normalization may or may not be effected by thesaurus lookup. A term-term similarity matrix is then constructed by using co-occurrence of terms within sentences, rather than within documents, as a criterion. It should be noted that as new term associations are defined, the original incidence matrix can be revised by inclusion in some of the matrix columns of new, associated terms which are not originally contained in the respective sentences or documents. The revised incidence matrix then gives rise to a new term-term similarity matrix, incorporating second-order associations, and so on. This feedback process is represented by an upward-pointing arrow in Figure 3. (Salton, 1963, p. 444)

Figure 3

Claim Text from '352 Patent	Salton, 1963
	<pre> graph TD Start[For each document generate list of high-frequency words to serve as "index terms" (see Fig. 1)] A[Construct term-sentence incidence matrix listing sentences against included terms] B[Thesaurus look-up and substitution of thesaurus heads for high-frequency terms] C[Compute term-term similarity matrix and generate term associations for document identification] D[Construct term-document incidence matrix listing documents against included terms] E[Compare vector of request terms with term-document incidence matrix and identify relevant documents] F[Compute document-document similarity matrix and generate document associations] Start --> A Start --> B A --> C B --> D C --> D D --> E E --> F </pre> <p>Fig. 3. Typical automatic document retrieval system using term and document associations → optional paths → compulsory paths</p> <p>(Salton, 1963, p. 445)</p>
<p>37. The non-semantic method of claim 26, wherein the step of searching includes the step of graphically displaying one or more of the identified objects.</p>	<p>Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.</p>
<p>38. The non-semantic method of claim 26, wherein the step of searching includes the step of</p>	<p>Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R.</p>

Claim Text from '352 Patent	Salton, 1963
identifying a paradigm object.	3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.
39. The non-semantic method of claim 26, wherein the step of searching the objects comprises the steps of: selecting a pool of objects;	<p><i>See, e.g.,</i> Salton, 1963, at p. 447</p> <p>Consider a collection of m documents each of which is characterized by the property of being cited by one or more of the other documents in the same collection. Each document can then be represented by an m-dimensional logical vector X_i, where $X_{ji} = 1$ if and only if document i is cited by document j, and $X_{ji} = 0$ otherwise. (Salton, 1963, p. 447)</p>
[39a] pool-similarity searching to identify a similar pool of textual objects, similar in relation to the objects in marked pool; and	<p><i>See, e.g.,</i> Salton, 1963, at p. 441</p> <p>Figure 1</p>

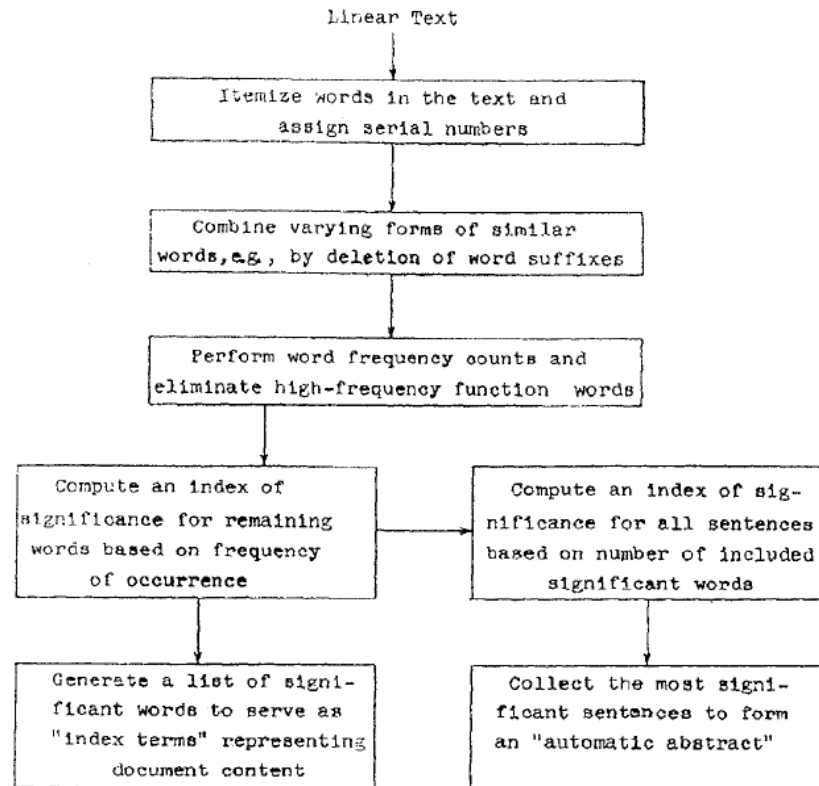


FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts.

(Salton, 1963, p. 441)

Further, disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.

Claim Text from '352 Patent	Salton, 1963
<p>[39b] pool-importance searching to identify an important pool of textual objects, important in relation to the objects in the selected pool.</p>	<p><i>See, e.g.</i>, Salton, 1963, at p. 444</p> <p>An estimate of document relevance is then obtained by computing for each document the similarity coefficient between the request column C_{m+1} and the respective document column. The documents can be arranged in decreasing order of similarity coefficients, and all documents with a sufficiently large coefficient can be judged to be relevant to the given request. (Salton, 1963, p. 444)</p> <p>Further, disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.</p>
<p>40. The non-semantic method of claim 26, the step of searching comprising the steps of: identifying a paradigm pool of objects; and</p>	<p>Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.</p>
<p>[40a] searching for relationships between the objects and the paradigm pool of objects;</p>	<p>Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.</p>
<p>[40b] wherein the searched for relationship is pool importance or pool similarity.</p>	<p>Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are</p>

Claim Text from '352 Patent	Salton, 1963
	incorporated by reference into this chart.
<p>41. A method for the non-semantic indexing of objects stored in a computer database, the method for use in searching the database for the objects, comprising the steps of: extracting, comprising the steps of:</p>	<p><i>See, e.g.</i>, Salton, 1963, at Abstract, pp. 440-43, 446, 450</p> <p>The standard associative retrieval techniques are first briefly reviewed. A computer experiment is then described which tends to confirm the hypothesis that documents exhibiting similar citation sets also deal with similar subject matter. (Salton, 1963, Abstract)</p> <p>It has been suggested [1] that an acceptable system can be generated by extracting from the text and from the information requests those linguistic units which are believed to be representative of document content, and by defining a standard of comparison between words extracted from documents and words used in the requests for documents. To determine which words are particularly significant as an indication of document content a variety of criteria may be used, including the position of the words in the texts, the word types, the vocabulary size, and most importantly the frequency of occurrence of the individual words. The most significant words are then used as "index terms" to characterize the documents, and the most significant sentences, that is, those containing a large number of significant words, are used as abstracts for the documents.</p> <p>A typical automatic indexing and abstracting system based on word frequency counts is shown in Figure 1. (Salton, 1963, pp. 440-41)</p> <p>Figure 1</p>

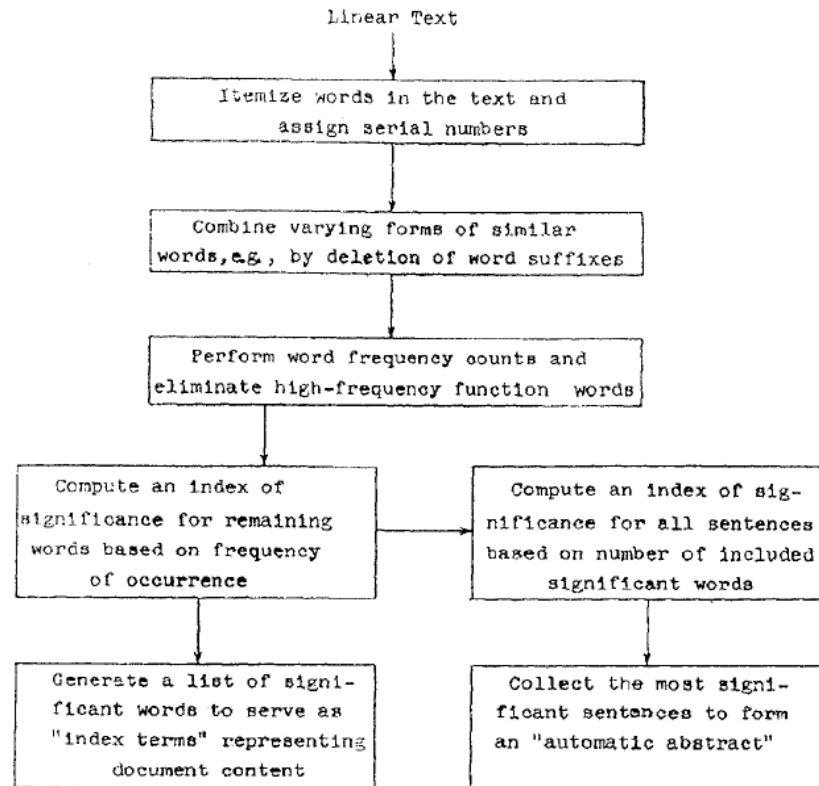


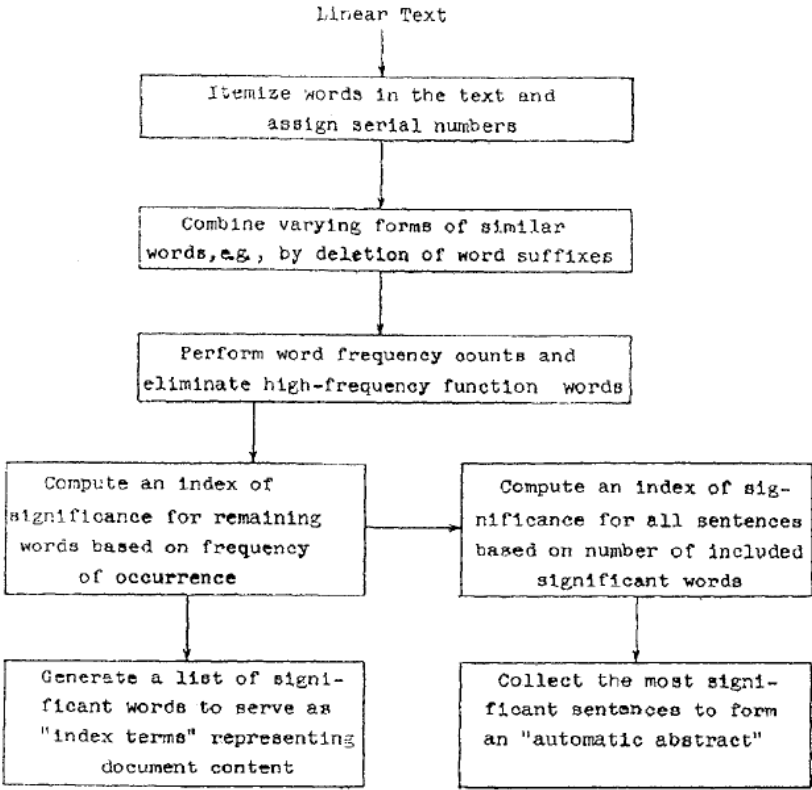
FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts.

(Salton, 1963, p. 441)

Most associative retrieval systems are based on the statistical word frequency counting procedures previously illustrated in Figure 1. Thus, given a document collection, it is possible to extract a set of n distinct high-frequency words W_1, W_2, \dots, W_n , such that each document within the collection is initially identified by some subset of the set of n given words. (Salton, 1963, p. 442)

The criteria of association used in most automatic programs do not normally require a determination of syntactic or semantic properties. Rather, they are based on simple co-

Claim Text from '352 Patent	Salton, 1963
	<p>occurrence of words in the same texts or sentences, or on co-occurrence with individual or joint frequencies greater than some given threshold value. (Salton, 1963, p. 443)</p> <p>Because of these and other variations, citation and reference lists have not generally been used as an indication of document content. Rather, such lists are used to detect trends in the literature as a whole, and to serve as adjuncts to certain kinds of literature searches [7, 8]. (Salton, 1963, p. 446)</p> <p>The complete procedure is summarized in the flow-chart of Figure 5. For the actual experiment, a collection of sixty-two documents dealing with linguistics and machine translation was chosen. A set of fifty-six index terms was used for manual indexing of the documents. The two basic inputs used for the computer experiments were thus logical matrices of dimension 62 by 62 and 62 by 56, listing, respectively, cited versus citing documents, and documents versus terms. (Salton, 1963, p. 450)</p>
[41a] labeling objects with a first numerical representation; and	<p><i>See, e.g.</i>, Salton, 1963, at pp. 441, 447</p> <p>Figure 1</p>

Claim Text from '352 Patent	Salton, 1963
	 <pre> graph TD A[Linear Text] --> B[Itemize words in the text and assign serial numbers] B --> C[Combine varying forms of similar words, e.g., by deletion of word suffixes] C --> D[Perform word frequency counts and eliminate high-frequency function words] D --> E[Compute an index of significance for remaining words based on frequency of occurrence] D --> F[Compute an index of significance for all sentences based on number of included significant words] E --> G[Generate a list of significant words to serve as "index terms" representing document content] F --> H[Collect the most significant sentences to form an "automatic abstract"] </pre> <p data-bbox="829 1063 1722 1112">FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts.</p> <p data-bbox="816 1128 1075 1161">(Salton, 1963, p. 441)</p> <p data-bbox="816 1177 1900 1307">Consider a collection of m documents each of which is characterized by the property of being cited by one or more of the other documents in the same collection. Each document can then be represented by an m-dimensional logical vector X_i, where $X_{ji} = 1$ if and only if document i is cited by document j, and $X_{ji} = 0$ otherwise. (Salton, 1963, p. 447)</p>
[41b] generating a second numerical representation	See, e.g., Salton, 1963, at pp. 446 n.1, 447, 450

Claim Text from '352 Patent

Salton, 1963

for each object based on each object's references to other objects;

A citation index consists of a set of bibliographic references (the set of cited documents), each followed by a list of all those documents (the citing documents) which include the given cited document as a reference. A reference index, on the other hand, lists all cited documents under each citing document. (Salton, 1963, p. 446 n.1)

Consider a collection of m documents each of which is characterized by the property of being cited by one of more of the other documents in the same collection. Each document can then be represented by an m -dimensional logical vector X_i , where $X_{ij} = 1$ if and only if document i is cited by document j , and $X_{ij} = 0$ otherwise. If these m vectors arranged in rows one below the other a square logical incidence matrix is formed similar to the matrix exhibited in Figure 4.

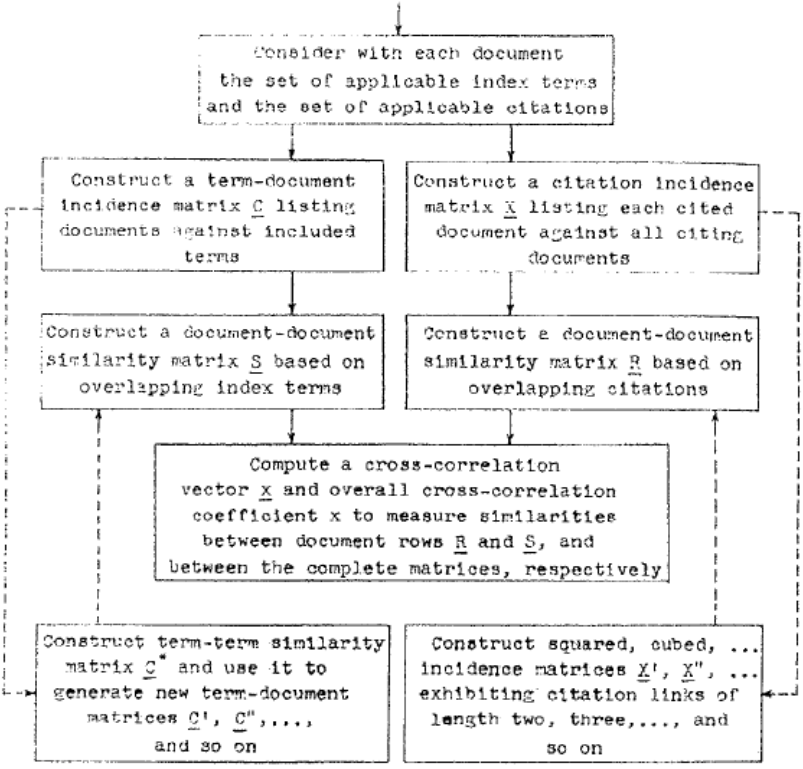
$$\begin{array}{c|ccc}
 \begin{array}{c} \text{Cited} \\ \text{documents} \end{array} & \begin{array}{c} \text{Citing documents} \\ D_1 \quad D_2 \quad \dots \quad D_m \end{array} \\
 \hline
 D_1 & \begin{pmatrix} X_{11} & X_{21} & \dots & X_{m1} \end{pmatrix} \\
 D_2 & \begin{pmatrix} X_{12} & X_{22} & \dots & X_{m2} \end{pmatrix} \\
 \vdots & \begin{pmatrix} \vdots & \vdots & \dots & \vdots \end{pmatrix} \\
 D_m & \begin{pmatrix} X_{1m} & X_{2m} & \dots & X_{mm} \end{pmatrix}
 \end{array} = X$$

($X_j^i = 1 \leftrightarrow$ document D_i is cited by document D_j)

FIG. 4. Matrix X exhibiting direct citations

(Salton, 1963, p. 447)

Figure 5

Claim Text from '352 Patent	Salton, 1963
	 <p>The flowchart illustrates the process of comparing citation similarities with index term similarities. It begins with a step: 'Consider with each document the set of applicable index terms and the set of applicable citations'. This leads to two parallel paths: <ul style="list-style-type: none"> Left path: 'Construct a term-document incidence matrix \underline{C} listing documents against included terms' followed by 'Construct a document-document similarity matrix \underline{S} based on overlapping index terms'. Right path: 'Construct a citation incidence matrix \underline{X} listing each cited document against all citing documents' followed by 'Construct a document-document similarity matrix \underline{R} based on overlapping citations'. A central step, 'Compute a cross-correlation vector \underline{x} and overall cross-correlation coefficient x to measure similarities between document rows \underline{R} and \underline{S}, and between the complete matrices, respectively', receives input from both similarity matrices. Below this, two more steps are shown: <ul style="list-style-type: none"> Left: 'Construct term-term similarity matrix \underline{C}^* and use it to generate new term-document matrices $\underline{C}^1, \underline{C}^2, \dots$ and so on'. A dashed arrow points from this step back to the \underline{C} matrix construction step. Right: 'Construct squared, cubed, ... incidence matrices $\underline{X}^1, \underline{X}^2, \dots$ exhibiting citation links of length two, three, ... and so on'. A dashed arrow points from this step back to the \underline{X} matrix construction step. </p> <p>FIG. 5. Comparison of citation similarities with index term similarities</p> <p>(Salton, 1963, p. 450)</p>
<p>[41c] patterning, comprising the step of creating a third numerical representation for each object using the second numerical representations, wherein the third numerical representation for each object is determined from an examination of the second numerical representations for occurrences of patterns that define indirect relations between or</p>	<p>See, e.g., Salton, 1963, at pp. 448, 450, 451-52</p> <p>Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X^1, X^2, \dots, exhibiting respectively the existence of paths of length two, three, and so on. Specifically,</p>

Claim Text from '352 Patent	Salton, 1963
among objects;	$[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$ <p>Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. $(X')_{ij}$ is then equal to 1 if and only if at least one path of length two exists between documents D_i and D_j; otherwise, $(X')_{ij}$ is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)</p> <p>A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the ith and jth rows (columns) of X.</p> <p>The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448)</p> <p>The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct</p>

Claim Text from '352 Patent	Salton, 1963
	<p>link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448)</p> <p>The CITED and CITING similarity matrices of dimension 62 by 62 were obtained from the original citation matrix by row and column comparisons, respectively. The TDCMP similarity matrix, also of dimension 62 by 62, was similarly obtained by column comparisons from the original term-document matrix. Additional citation similarity matrices, designated CTD2, CTD3, CTD4, and CNG2, CNG3, CNG4 were obtained from the squared, cubed, and fourth power logical citation matrices, as previously explained. (Salton, 1963, p. 450)</p> <p>Figure 5</p>

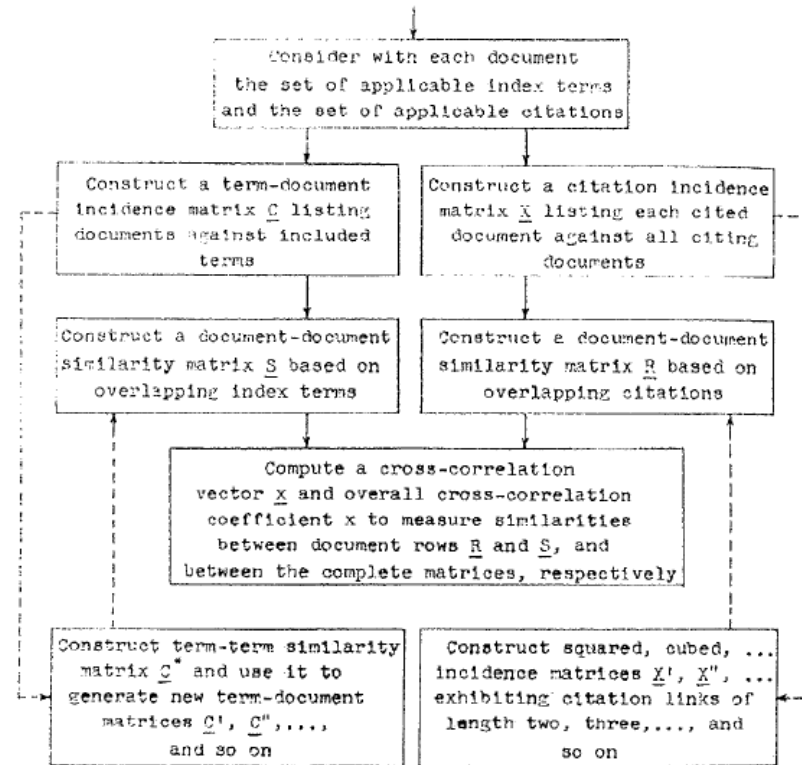


FIG. 5. Comparison of citation similarities with index term similarities

(Salton, 1963, p. 450)

The value of the overall similarity coefficient first rises as the length of the citation links increases, and then drops again as the length of the links becomes still greater [6]. This is due to the fact that as the length of the links increases, the total number of links of any length increases also; an increased number of links results in a larger number of ones in the original logical citation matrix, and thus in a higher probability of overlapping ones and a larger overall similarity coefficient. At the same time, as the length of the links increases, two factors also tend to decrease the magnitude of the overall similarity coefficient. First, the number of documents which exhibit citation links of length n but which do not exhibit links

Claim Text from '352 Patent	Salton, 1963
	<p>of length greater than n increases as n becomes larger. Thus more and more documents will exhibit individual similarity coefficients of zero value, thus tending to decrease the value of the overall coefficient. Second, as the length of the links increases and the citations thus become increasingly less accurate indications of document content, the magnitude of the cross-correlation coefficients obtained from the citation matrix and the term-document matrix would be expected to decrease, even for those documents for which a large number of citation links can still be found. (Salton, 1963, pp. 451-52)</p>
<p>[41d] weaving, comprising the steps of: calculating a fourth numerical representation for each object based on the euclidean distances between the third numerical representations; and</p>	<p>See, e.g., Salton, 1963, at pp. 443-45, 447-48</p> <p>Figure 2</p> $ \begin{array}{c cccc} \text{Terms} & & \text{Documents} & & \\ & D_1 & D_2 & \dots & D_m \\ \hline W_1 & C_1^1 & C_2^1 & \dots & C_m^1 \\ W_2 & C_1^2 & C_2^2 & \dots & C_m^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W_n & C_1^n & C_2^n & \dots & C_m^n \end{array} = C $ <p>(a) Typical term-document incidence matrix C ($C_i^j = n \leftrightarrow$ document D_j contains term W_i exactly n times)</p> $ \begin{array}{c cccc} \text{Terms} & & \text{Terms} & & \\ & W_1 & W_2 & \dots & W_n \\ \hline W_1 & R_1^1 & R_2^1 & \dots & R_n^1 \\ W_2 & R_1^2 & R_2^2 & \dots & R_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W_n & R_1^n & R_2^n & \dots & R_n^n \end{array} = R $ <p>(b) Typical term-term similarity matrix R</p> $ \left(R_i^j = R_j^i = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \right) \left(\sum_{k=1}^m (C_k^j)^2 \right)}} \right) $ <p>FIG. 2. Matrices used for the generation of term associations</p> <p>(Salton, 1963, p. 443)</p>

Claim Text from '352 Patent	Salton, 1963
	<p data-bbox="816 251 1906 451">Many different types of similarity coefficients have been suggested in the literature; a simple coefficient of similarity between rows of a numeric matrix, and one which may be as meaningful as any of the others, is the cosine of the angle between the corresponding m-dimensional vectors. The similarity coefficients can be displayed in an n x n symmetric term-similarity matrix R, where the coefficient of similarity R_{ij} between term W_i and term W_j is</p> $R_i^i = R_i^j = \frac{\sum_{k=1}^m c_k^i c_k^j}{\sqrt{\left(\sum_{k=1}^m (c_k^i)^2 \sum_{k=1}^m (c_k^j)^2\right)}}$ <p data-bbox="816 683 848 703">...</p> <p data-bbox="816 716 1845 846">Document similarities are therefore obtained by comparing pairs of columns (instead of rows) of the term-document matrix C, and a document-document similarity matrix is constructed and used in the same way as the previously described term-term matrix R. (Salton, 1963, pp. 443-44)</p> <p data-bbox="816 862 1906 1227">Consider now a typical system for document retrieval using term and document associations as shown in Figure 3. A list of high-frequency terms is first generated for each document by word frequency counting procedures. Normalization may or may not be effected by thesaurus lookup. A term-term similarity matrix is then constructed by using co-occurrence of terms within sentences, rather than within documents, as a criterion. It should be noted that as new term associations are defined, the original incidence matrix can be revised by inclusion in some of the matrix columns of new, associated terms which are not originally contained in the respective sentences or documents. The revised incidence matrix then gives rise to a new term-term similarity matrix, incorporating second-order associations, and so on. This feedback process is represented by an upward-pointing arrow in Figure 3. (Salton, 1963, p. 444)</p> <p data-bbox="816 1240 919 1268">Figure 3</p>

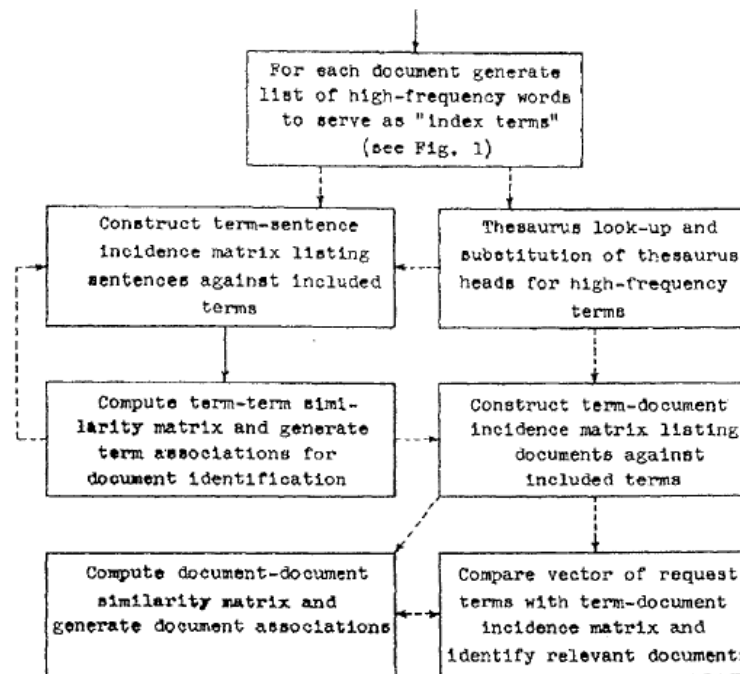


Fig. 3. Typical automatic document retrieval system using term and document associations
 → optional paths → compulsory paths

(Salton, 1963, p. 445)

To test the significance of bibliographic citations, a comparison was made between citation similarities and index term similarities for an indexed document collection. Specifically, a measure of similarity was computed between each pair of documents in the collection, based on the number of overlapping index terms a similar measure was then computed for the same pairs of documents, based on the number of overlapping citations; finally, the similarity measures obtained from index terms and citations respectively were compared by calculating a similarity index between citation similarities and index term similarities. An overall measure was also computed for the complete document collection by taking into account the similarity measures between all document pairs. (Salton, 1963, p. 447)

A measure of similarity between row (column) vectors can be obtained by calculating the

Claim Text from '352 Patent

Salton, 1963

cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the i th and j th rows (columns) of X. (Salton, 1963, p. 448)

[41e] determining a fifth numerical representation for each object by processing the fourth numerical representations through similarity processing; and

See, e.g., Salton, 1963, at pp. 443-44, 447-50

Figure 2

$$\begin{array}{c|cccc}
 \text{Terms} & D_1 & D_2 & \dots & D_m \\
 \hline
 W_1 & C_1^1 & C_2^1 & \dots & C_m^1 \\
 W_2 & C_1^2 & C_2^2 & \dots & C_m^2 \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 W_n & C_1^n & C_2^n & \dots & C_m^n
 \end{array} = C$$

(a) Typical term-document incidence matrix C ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times)

$$\begin{array}{c|cccc}
 \text{Terms} & W_1 & W_2 & \dots & W_n \\
 \hline
 W_1 & R_1^1 & R_2^1 & \dots & R_n^1 \\
 W_2 & R_1^2 & R_2^2 & \dots & R_n^2 \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 W_n & R_1^n & R_2^n & \dots & R_n^n
 \end{array} = R$$

(b) Typical term-term similarity matrix R

$$\left(R_j^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \right) \left(\sum_{k=1}^m (C_k^j)^2 \right)}} \right)$$

FIG. 2. Matrices used for the generation of term associations

(Salton, 1963, p. 443)

Many different types of similarity coefficients have been suggested in the literature; a simple coefficient of similarity between rows of a numeric matrix, and one which may be as meaningful as any of the others, is the cosine of the angle between the corresponding m-

Claim Text from '352 Patent	Salton, 1963
	<p>dimensional vectors. The similarity coefficients can be displayed in an n x n symmetric term-similarity matrix R, where the coefficient of similarity R_{ji} between term W_i and term W_j is</p> $R_i^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \sum_{k=1}^m (C_k^j)^2\right)}}.$ <p>...</p> <p>Document similarities are therefore obtained by comparing pairs of columns (instead of rows) of the term-document matrix C, and a document-document similarity matrix is constructed and used in the same way as the previously described term-term matrix R. (Salton, 1963, pp. 443-44)</p> <p>A term-term similarity matrix is then constructed by using co-occurrence of terms within sentences rather than within documents as a criterion. It should be noted that as new term associations are defined, the original incidence matrix can be revised by inclusion in some of the matrix columns of new, associated terms which are not originally contained in the respective sentences or documents. The revised incidence matrix then gives rise to a new term-term similarity matrix, incorporating second-order associations, and so on. (Salton, 1963, p. 444)</p> <p>Figure 3</p>

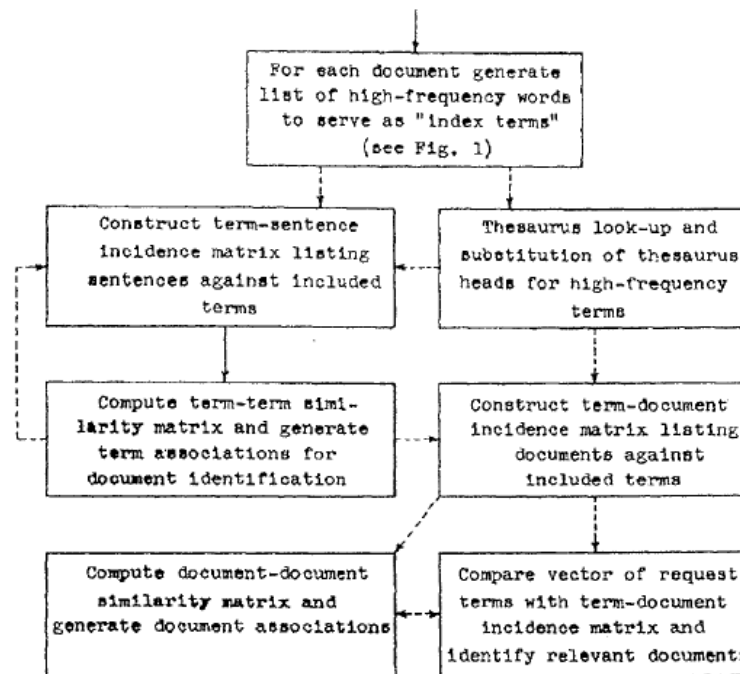


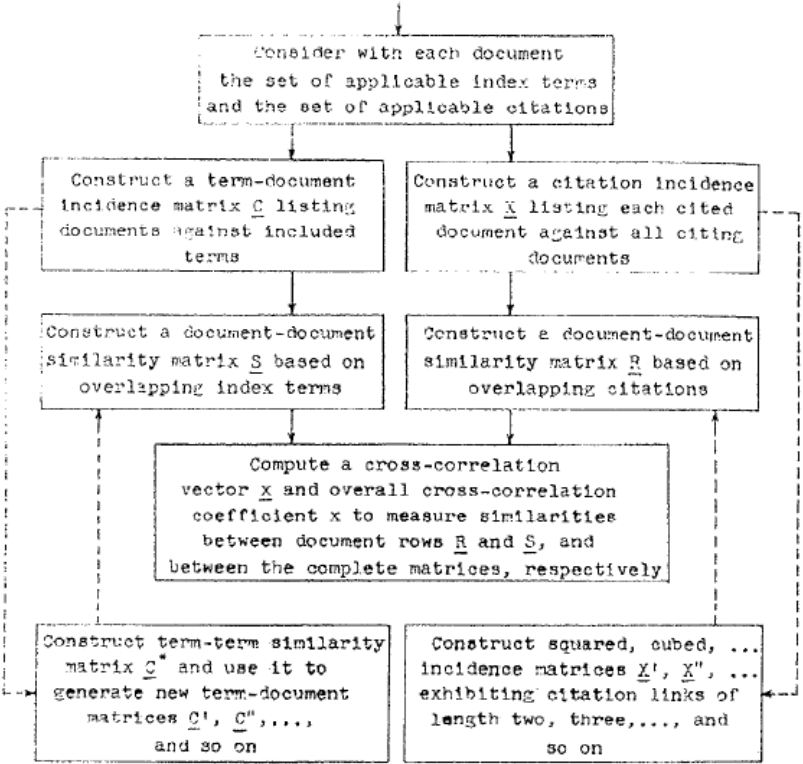
Fig. 3. Typical automatic document retrieval system using term and document associations
 → optional paths → compulsory paths

(Salton, 1963, p. 445)

To test the significance of bibliographic citations, a comparison was made between citation similarities and index term similarities for an indexed document collection. Specifically, a measure of similarity was computed between each pair of documents in the collection, based on the number of overlapping index terms a similar measure was then computed for the same pairs of documents, based on the number of overlapping citations; finally, the similarity measures obtained from index terms and citations respectively were compared by calculating a similarity index between citation similarities and index term similarities. An overall measure was also computed for the complete document collection by taking into account the similarity measures between all document pairs. (Salton, 1963, p. 447)

A measure of similarity between row (column) vectors can be obtained by calculating the

Claim Text from '352 Patent	Salton, 1963
	<p>cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the ith and jth rows (columns) of X. (Salton, 1963, p. 448)</p> <p>A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the ith and jth rows (columns) of X.</p> <p>The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448)</p> <p>Since the term-document matrix C is not in general a square matrix, matrix multiplication cannot be used to obtain second order effects, similar to the citation links of length two or more. Instead, it is first necessary to compare the index terms by performing a row comparison of the rows of C. This produces a new n symmetric term matrix C^* which displays similarity between index terms. This matrix can be used to eliminate from the set of index terms those terms which exhibit a large number of joint occurrences with other terms. A reduced set of index terms can then be formed and a new term-document matrix C' constructed, from which a new correlation matrix S' is formed. (Salton, 1963, p. 449)</p> <p>Figure 5</p>

Claim Text from '352 Patent	Salton, 1963
	 <p>The flowchart illustrates the process of comparing citation similarities with index term similarities. It begins with a central step: 'Consider with each document the set of applicable index terms and the set of applicable citations'. This leads to two parallel paths. The left path involves constructing a term-document incidence matrix \underline{C} and a document-document similarity matrix \underline{S} based on overlapping index terms. The right path involves constructing a citation incidence matrix \underline{X} and a document-document similarity matrix \underline{R} based on overlapping citations. A central step, 'Compute a cross-correlation vector \underline{x} and overall cross-correlation coefficient x to measure similarities between document rows \underline{R} and \underline{S}, and between the complete matrices, respectively', receives input from both \underline{S} and \underline{R}. Below this, two final steps are shown: 'Construct term-term similarity matrix \underline{C}^* and use it to generate new term-document matrices $\underline{C}^1, \underline{C}^2, \dots$ and so on' (receiving input from \underline{C}^* and \underline{C}), and 'Construct squared, cubed, ... incidence matrices $\underline{X}^1, \underline{X}^2, \dots$ exhibiting citation links of length two, three, ..., and so on' (receiving input from \underline{X} and \underline{X}). Dashed lines indicate feedback loops from the final steps back to the initial construction steps.</p> <p>FIG. 5. Comparison of citation similarities with index term similarities</p> <p>(Salton, 1963, p. 450)</p>
<p>[41f] storing the fifth numerical representations in the computer database as the index for use in searching for objects in the database.</p>	<p>See, e.g., Salton, 1963, at pp. 440-41, 442, 450</p> <p>It has been suggested [1] that an acceptable system can be generated by extracting from the text and from the information requests those linguistic units which are believed to be representative of document content, and by defining a standard of comparison between words extracted from documents and words used in the requests for documents. To</p>

Claim Text from '352 Patent	Salton, 1963
	<p>determine which words are particularly significant as an indication of document content a variety of criteria may be used, including the position of the words in the texts, the word types, the vocabulary size, and most importantly the frequency of occurrence of the individual words. The most significant words are then used as “index terms” to characterize the documents, and the most significant sentences, that is, those containing a large number of significant words, are used as abstracts for the documents.</p> <p>A typical automatic indexing and abstracting system based on word frequency counts is shown in Figure 1. (Salton, 1963, pp. 440-41)</p> <p>Figure 1</p>

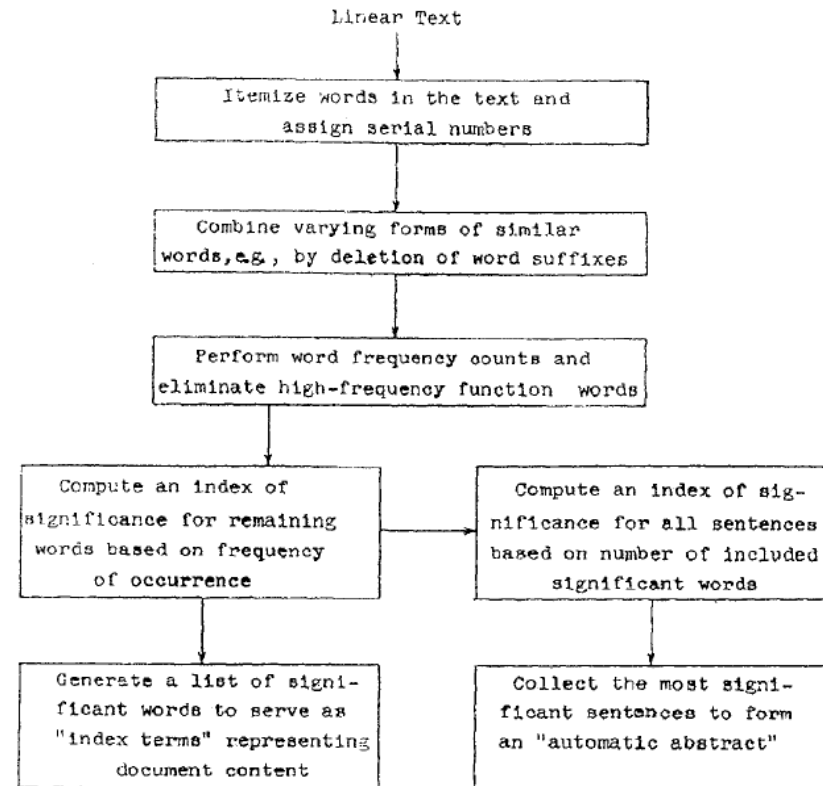


FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts.

(Salton, 1963, p. 441)

Most associative retrieval systems are based on the statistical word frequency counting procedures previously illustrated in Figure 1. Thus, given a document collection, it is possible to extract a set of n distinct high-frequency words W_1, W_2, \dots, W_n , such that each document within the collection is initially identified by some subset of the set of n given words. (Salton, 1963, p. 442)

Claim Text from '352 Patent	Salton, 1963
	<p>The complete procedure is summarized in the flow-chart of Figure 5. For the actual experiment, a collection of sixty-two documents dealing with linguistics and machine translation was chosen. A set of fifty-six index terms was used for manual indexing of the documents. The two basic inputs used for the computer experiments were thus logical matrices of dimension 62 by 62 and 62 by 56, listing, respectively, cited versus citing documents, and documents versus terms. (Salton, 1963, p. 450)</p>
<p>42. The method of claim 41 wherein the first through fifth numerical representations are vector representations and further comprises the step of clustering objects having similar characteristics.</p>	<p><i>See, e.g.</i>, Salton, 1963, at pp. 441, 443-45, 447, 449-52</p> <p>Figure 1</p>

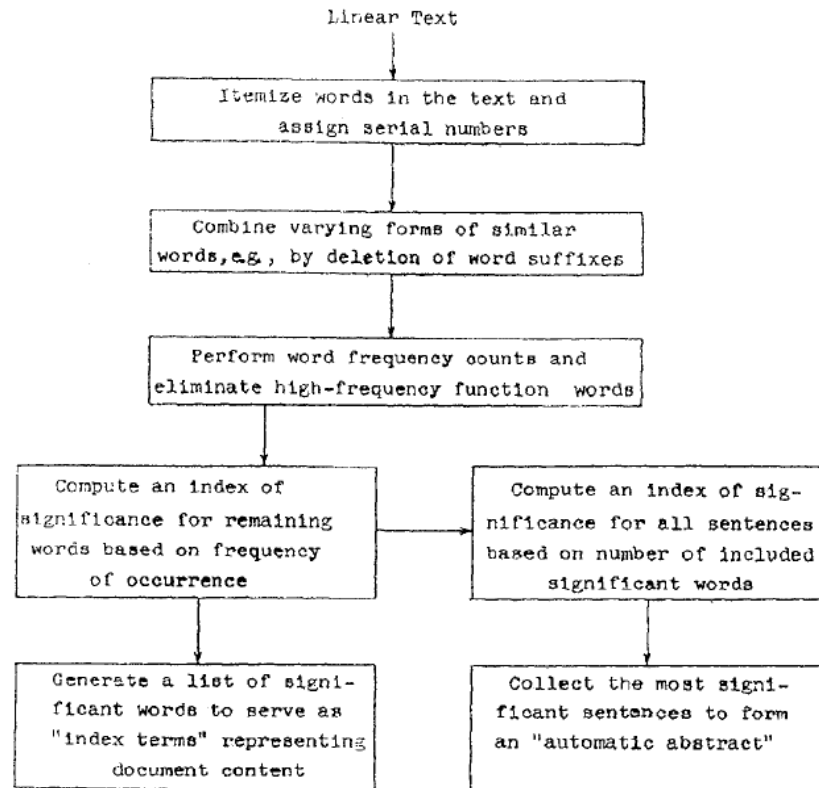


FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts.

(Salton, 1963, p. 441)

Figure 2

Terms	Documents			
	D_1	D_2	...	D_m
W_1	C_1^1	C_2^1	...	C_m^1
W_2	C_1^2	C_2^2	...	C_m^2
\vdots	\vdots	\vdots	\vdots	\vdots
W_n	C_1^n	C_2^n	...	C_m^n

$$= C$$

(a) Typical term-document incidence matrix C ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times)

Terms	Terms			
	W_1	W_2	...	W_n
W_1	R_1^1	R_2^1	...	R_n^1
W_2	R_1^2	R_2^2	...	R_n^2
\vdots	\vdots	\vdots	\vdots	\vdots
W_n	R_1^n	R_2^n	...	R_n^n

$$= R$$

(b) Typical term-term similarity matrix R

$$\left(R_j^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \sum_{k=1}^m (C_k^j)^2 \right)}} \right)$$

FIG. 2. Matrices used for the generation of term associations

(Salton, 1963, p. 443)

Many different types of similarity coefficients have been suggested in the literature; a simple coefficient of similarity between rows of a numeric matrix, and one which may be as meaningful as any of the others, is the cosine of the angle between the corresponding m-dimensional vectors. The similarity coefficients can be displayed in an $n \times n$ symmetric term-similarity matrix R , where the coefficient of similarity R_{ji} between term W_i and term W_j is

$$R_j^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \sum_{k=1}^m (C_k^j)^2 \right)}}$$

...

Document similarities are therefore obtained by comparing pairs of columns (instead of rows) of the term-document matrix C, and a document-document similarity matrix is constructed and used in the same way as the previously described term-term matrix R. (Salton, 1963, pp. 443-44)

Figure 3

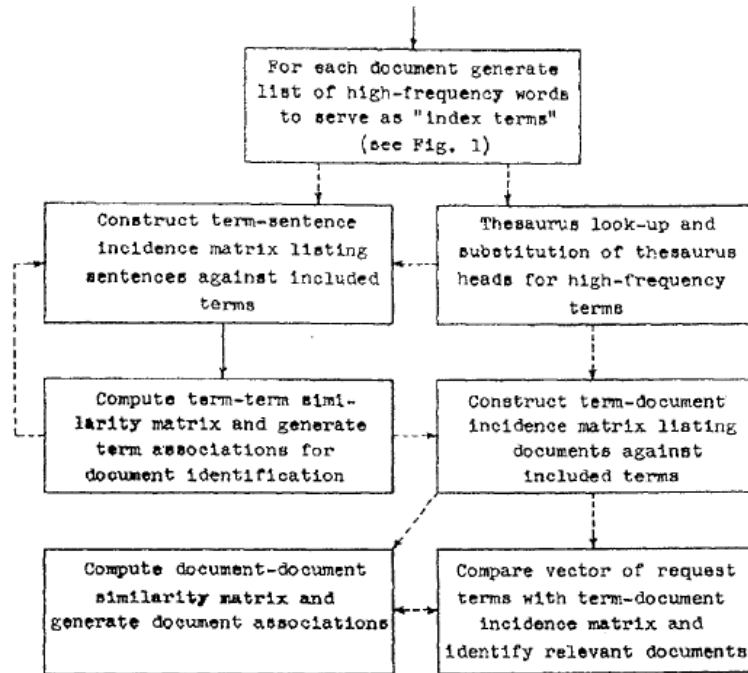


FIG. 3. Typical automatic document retrieval system using term and document associations
 → optional paths → compulsory paths

(Salton, 1963, p. 445)

Consider a collection of m documents each of which is characterized by the property of being cited by one or more of the other documents in the same collection. Each document can then be represented by an m -dimensional logical vector X_i , where $X_{ij} = 1$ if and only if document i is cited by document j , and $X_{ij} = 0$ otherwise. If these m vectors arranged in

rows one below the other a square logical incidence matrix is formed similar to the matrix exhibited in Figure 4.

<i>Cited documents</i>		<i>Citing documents</i>				
		D_1	D_2	\dots	D_m	
D_1		X_{11}^1	X_{21}^1	\dots	X_{m1}^1	
D_2		X_{12}^2	X_{22}^2	\dots	X_{m2}^2	
\vdots		\vdots	\vdots	\vdots	\vdots	
D_m		X_{1m}^m	X_{2m}^m	\dots	X_{mm}^m	

$$= X$$

($X_{ij}^i = 1 \leftrightarrow$ document D_i is cited by document D_j)

FIG. 4. Matrix X exhibiting direct citations

(Salton, 1963, p. 447)

Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X', X'', etc., exhibiting respectively the existence of paths of length two, three, and so on. Specifically,

$$[X']_{ij}^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$$

$$[X'']_{ij}^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_{jk}^k), \text{ and so on.}$$

Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. (X')_{ij} is then equal to 1 if and only if at least one path of length two exists between documents Di and Dj; otherwise, (X')_{ij} is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)

Since the term-document matrix C is not in general a square matrix, matrix multiplication cannot be used to obtain second order effects, similar to the citation links of length two or more. Instead, it is first necessary to compare the index terms by performing a row comparison of the rows of C. This produces a new n symmetric term matrix C* which displays similarity between index terms. This matrix can be used to eliminate from the set of index terms those terms which exhibit a large number of joint occurrences with other terms.

Claim Text from '352 Patent	Salton, 1963
	<p>A reduced set of index terms can then be formed and a new term-document matrix C? constructed, from which a new correlation matrix S? is formed. (Salton, 1963, p. 449)</p> <p>The CITED and CITING similarity matrices of dimension 62 by 62 were obtained from the original citation matrix by row and column comparisons, respectively. The TDCMP similarity matrix, also of dimension 62 by 62, was similarly obtained by column comparisons from the original term-document matrix. Additional citation similarity matrices, designated CTD2, CTD3, CTD4, and CNG2, CNG3, CNG4 were obtained from the squared, cubed, and fourth power logical citation matrices, as previously explained. (Salton, 1963, p. 450)</p> <p>The value of the overall similarity coefficient first rises as the length of the citation links increases, and then drops again as the length of the links becomes still greater [6]. This is due to the fact that as the length of the links increases, the total number of links of any length increases also; an increased number of links results in a larger number of ones in the original logical citation matrix, and thus in a higher probability of overlapping ones and a larger overall similarity coefficient. At the same time, as the length of the links increases, two factors also tend to decrease the magnitude of the overall similarity coefficient. First, the number of documents which exhibit citation links of length n but which do not exhibit links of length greater than n increases as n becomes larger. Thus more and more documents will exhibit individual similarity coefficients of zero value, thus tending to decrease the value of the overall coefficient. Second, as the length of the links increases and the citations thus become increasingly less accurate indications of document content, the magnitude of the cross-correlation coefficients obtained from the citation matrix and the term-document matrix would be expected to decrease, even for those documents for which a large number of citation links can still be found. (Salton, 1963, pp. 451-52)</p>
<p>44. The method of claim 41 wherein the step of creating the third numerical representations further comprises the steps of:</p>	<p><i>See, e.g.,</i> Salton, 1963, at pp. 448, 450-52</p> <p>The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A</p>

Claim Text from '352 Patent	Salton, 1963
	<p>does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448)</p> <p>Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X', X'', etc., exhibiting respectively the existence of paths of length two, three, and so on. Specifically,</p> $[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$ <p>Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. (X')_{ij} is then equal to 1 if and only if at least one path of length two exists between documents D_i and D_j; otherwise, (X')_{ij} is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)</p> <p>The CITED and CITING similarity matrices of dimension 62 by 62 were obtained from the original citation matrix by row and column comparisons, respectively. The TDCMP similarity matrix, also of dimension 62 by 62, was similarly obtained by column comparisons from the original term-document matrix. Additional citation similarity matrices, designated CTD2, CTD3, CTD4, and CNG2, CNG3, CNG4 were obtained from the squared, cubed, and fourth power logical citation matrices, as previously explained. (Salton, 1963, p. 450)</p> <p>The value of the overall similarity coefficient first rises as the length of the citation links increases, and then drops again as the length of the links becomes still greater [6]. This is due to the fact that as the length of the links increases, the total number of links of any length increases also; an increased number of links results in a larger number of ones in the original logical citation matrix, and thus in a higher probability of overlapping ones and a larger overall similarity coefficient. At the same time, as the length of the links increases, two factors also tend to decrease the magnitude of the overall similarity coefficient. First, the number of documents which exhibit citation links of length n but which do not exhibit links</p>

Claim Text from '352 Patent	Salton, 1963
	<p>of length greater than n increases as n becomes larger. Thus more and more documents will exhibit individual similarity coefficients of zero value, thus tending to decrease the value of the overall coefficient. Second, as the length of the links increases and the citations thus become increasingly less accurate indications of document content, the magnitude of the cross-correlation coefficients obtained from the citation matrix and the term-document matrix would be expected to decrease, even for those documents for which a large number of citation links can still be found. (Salton, 1963, pp. 451-52)</p>
<p>[44a] analyzing the second numerical representation against a plurality of empirically defined patterns, wherein certain patterns are more important than others; and</p>	<p><i>See, e.g.</i>, Salton, 1963, at pp. 444, 448</p> <p>To retrieve documents in answer to search requests, the programs already available can be used by adding to the term-document matrix C a new column C_{m+1}, representing the request terms. Specifically, element C_{km+1} is set equal to w if term W_k is used in the search request with weight w; if word W_k is not used in the given search request C_{km+1} is set equal to 0. If no weights are specified by the requestor the values of the elements of column C_{m+1} are restricted to 0 and 1. (Salton, 1963, p. 444)</p> <p>The coefficients of R now represent a measure of similarity between documents, based on the number of overlapping direct citations. This concept may be extended by using as a basis for the calculation of similarity coefficients not the existence of direct links between documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, if A does not cite B, but A cites (or is cited by) C which in turn cites (or is cited by) B, no direct link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448)</p>
<p>[44b] weighing the analyzed second numerical representations according to the importance of the patterns.</p>	<p><i>See, e.g.</i>, Salton, 1963, at p. 444</p> <p>To retrieve documents in answer to search requests, the programs already available can be used by adding to the term-document matrix C a new column C_{m+1}, representing the request</p>

Claim Text from '352 Patent	Salton, 1963
	<p>terms. Specifically, element C_{k+1} is set equal to w if term W_k is used in the search request with weight w; if word W_k is not used in the given search request C_{k+1} is set equal to 0. If no weights are specified by the requestor the values of the elements of column C_{k+1} are restricted to 0 and 1. (Salton, 1963, p. 444)</p>
<p>45. A method for searching indexed objects, wherein the index is stored, comprising the steps of:</p>	<p><i>See, e.g.,</i> Salton, 1963, at pp. 440-41, 442, 450</p> <p>It has been suggested [1] that an acceptable system can be generated by extracting from the text and from the information requests those linguistic units which are believed to be representative of document content, and by defining a standard of comparison between words extracted from documents and words used in the requests for documents. To determine which words are particularly significant as an indication of document content a variety of criteria may be used, including the position of the words in the texts, the word types, the vocabulary size, and most importantly the frequency of occurrence of the individual words. The most significant words are then used as "index terms" to characterize the documents, and the most significant sentences, that is, those containing a large number of significant words, are used as abstracts for the documents.</p> <p>A typical automatic indexing and abstracting system based on word frequency counts is shown in Figure 1. (Salton, 1963, pp. 440-41)</p> <p>Figure 1</p>

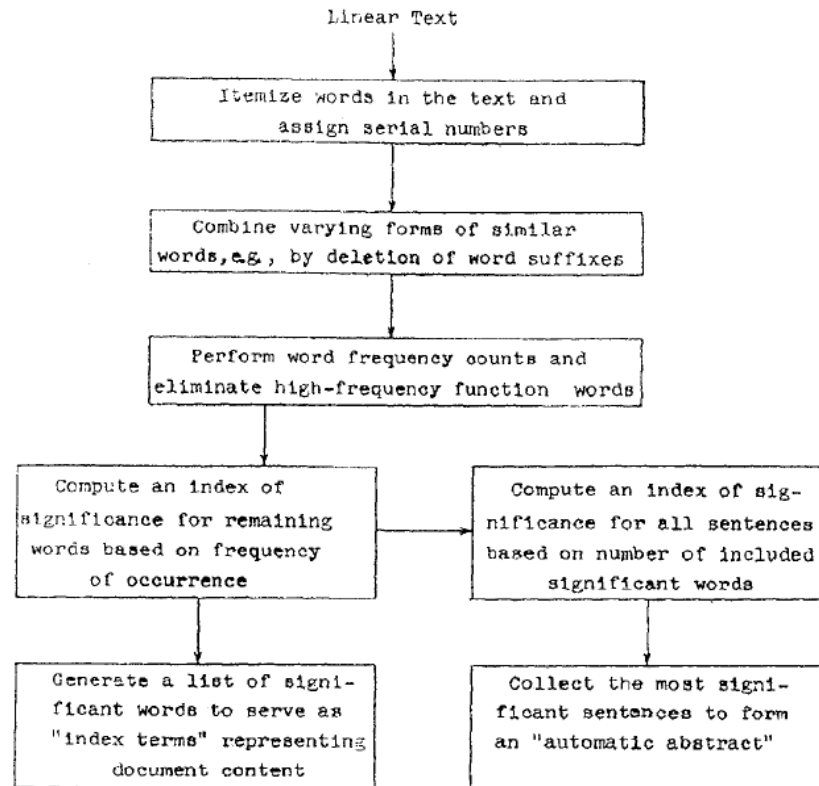


FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts.

(Salton, 1963, p. 441)

Most associative retrieval systems are based on the statistical word frequency counting procedures previously illustrated in Figure 1. Thus, given a document collection, it is possible to extract a set of n distinct high-frequency words W_1, W_2, \dots, W_n , such that each document within the collection is initially identified by some subset of the set of n given words. (Salton, 1963, p. 442)

Claim Text from '352 Patent	Salton, 1963
	<p>The complete procedure is summarized in the flow-chart of Figure 5. For the actual experiment, a collection of sixty-two documents dealing with linguistics and machine translation was chosen. A set of fifty-six index terms was used for manual indexing of the documents. The two basic inputs used for the computer experiments were thus logical matrices of dimension 62 by 62 and 62 by 56, listing, respectively, cited versus citing documents, and documents versus terms. (Salton, 1963, p. 450)</p>
<p>[45a] entering search commands;</p>	<p><i>See, e.g.</i>, Salton, 1963, at p. 442</p> <p>In practical retrieval systems, it becomes useful to provide for some additional flexibility. For example, given a search request expressed in terms of words in the natural language, it may be convenient to alter somewhat the original request, either by making it more specific and thus presumably reducing the size of the document set which fulfils the request or, alternatively, by making it more general. In the same way, given a set of terms identifying a specified document, it may be useful to alter somewhat the original set by deletion of old terms or addition of new ones in such a way that documents dealing with similar subject matter are identified by similar sets of index terms. (Salton, 1963, p. 442)</p>
<p>[45b] processing the search commands with a processor;</p>	<p><i>See, e.g.</i>, Salton, 1963, at p. 442</p> <p>In practical retrieval systems, it becomes useful to provide for some additional flexibility. For example, given a search request expressed in terms of words in the natural language, it may be convenient to alter somewhat the original request, either by making it more specific and thus presumably reducing the size of the document set which fulfils the request or, alternatively, by making it more general. In the same way, given a set of terms identifying a specified document, it may be useful to alter somewhat the original set by deletion of old terms or addition of new ones in such a way that documents dealing with similar subject matter are identified by similar sets of index terms. (Salton, 1963, p. 442)</p>
<p>[45c] retrieving the stored index using the processor;</p>	<p><i>See, e.g.</i>, Salton, 1963, at pp. 442-46, 449</p>

An analogous problem arises in connection with the document sets which are obtained in answer to certain search requests. It is often useful to alter these document sets by addition of further documents which may also have some relevance or, alternatively, by deletion of documents which are not directly relevant. Both questions can be treated by determining a measure of association between words or index terms on the one hand and between documents on the other, and by using the association measure for the alteration of the corresponding index term and document subsets. (Salton, 1963, p. 442)

Figure 2

Terms	Documents					
	D_1	D_2	...	D_j	...	D_m
W_1	C_1^1	C_2^1	...	C_j^1	...	C_m^1
W_2	C_1^2	C_2^2	...	C_j^2	...	C_m^2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
W_n	C_1^n	C_2^n	...	C_j^n	...	C_m^n

$$= \mathbf{C}$$

(a) Typical term-document incidence matrix \mathbf{C} ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times)

Terms	Terms			
	W_1	W_2	...	W_n
W_1	R_1^1	R_2^1	...	R_n^1
W_2	R_1^2	R_2^2	...	R_n^2
\vdots	\vdots	\vdots	\vdots	\vdots
W_n	R_1^n	R_2^n	...	R_n^n

$$= \mathbf{R}$$

(b) Typical term-term similarity matrix \mathbf{R}

$$\left(R_j^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \sum_{k=1}^m (C_k^j)^2 \right)}} \right)$$

FIG. 2. Matrices used for the generation of term associations

(Salton, 1963, p. 443)

Consider now a typical system for document retrieval using term and document associations as shown in Figure 3. A list of high-frequency terms is first generated for each document by word frequency counting procedures. Normalization may or may not be effected by thesaurus lookup. A term-term similarity matrix is then constructed by using co-occurrence

Claim Text from '352 Patent

Salton, 1963

of terms within sentences, rather than within documents, as a criterion. It should be noted that as new term associations are defined, the original incidence matrix can be revised by inclusion in some of the matrix columns of new, associated terms which are not originally contained in the respective sentences or documents. The revised incidence matrix then gives rise to a new term-term similarity matrix, incorporating second-order associations, and so on. This feedback process is represented by an upward-pointing arrow in Figure 3. (Salton, 1963, p. 444)

Figure 3

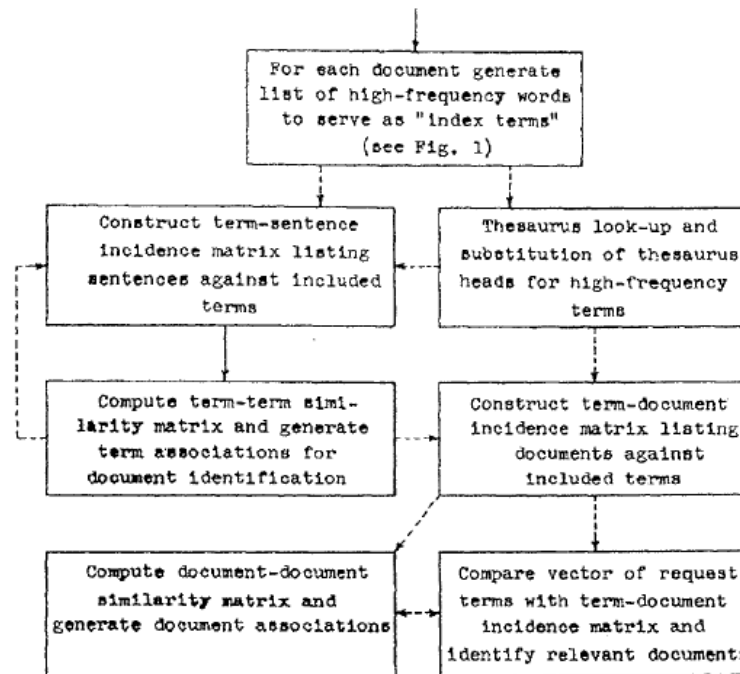


FIG. 3. Typical automatic document retrieval system using term and document associations
 → optional paths → compulsory paths

(Salton, 1963, p. 445)

Because of these and other variations, citation and reference lists have not generally been used as an indication of document content. Rather, such lists are used to detect trends in the

Claim Text from '352 Patent	Salton, 1963
	<p>literature as a whole, and to serve as adjuncts to certain kinds of literature searches [7, 8]. (Salton, 1963, p. 446)</p> <p>Since the term-document matrix C is not in general a square matrix, matrix multiplication cannot be used to obtain second order effects, similar to the citation links of length two or more. Instead, it is first necessary to compare the index terms by performing a row comparison of the rows of C. This produces a new n symmetric term matrix C^* which displays similarity between index terms. This matrix can be used to eliminate from the set of index terms those terms which exhibit a large number of joint occurrences with other terms. A reduced set of index terms can then be formed and a new term-document matrix C' constructed, from which a new correlation matrix S' is formed. (Salton, 1963, p. 449)</p>
<p>[45d] Analyzing the index to identify a pool of objects, comprising the steps of:</p>	<p><i>See, e.g.</i>, Salton, 1963, at pp. 442-46, 449</p> <p>An analogous problem arises in connection with the document sets which are obtained in answer to certain search requests. It is often useful to alter these document sets by addition of further documents which may also have some relevance or, alternatively, by deletion of documents which are not directly relevant. Both questions can be treated by determining a measure of association between words or index terms on the one hand and between documents on the other, and by using the association measure for the alteration of the corresponding index term and document subsets. (Salton, 1963, p. 442)</p> <p>Figure 2</p>

Terms	Documents					
	D_1	D_2	...		D_m	
W_1	C_1^1	C_2^1	...	C_j^1	...	C_m^1
W_2	C_1^2	C_2^2	...	C_j^2	...	C_m^2
\vdots						
W_n	C_1^n	C_2^n	...	C_j^n	...	C_m^n

(a) Typical term-document incidence matrix C ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times)

Terms	Terms			
	W_1	W_2	...	W_n
W_1	R_1^1	R_2^1	...	R_n^1
W_2	R_1^2	R_2^2	...	R_n^2
\vdots				
W_n	R_1^n	R_2^n	...	R_n^n

(b) Typical term-term similarity matrix R

$$\left(R_j^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \sum_{k=1}^m (C_k^j)^2 \right)}} \right)$$

FIG. 2. Matrices used for the generation of term associations

(Salton, 1963, p. 443)

Consider now a typical system for document retrieval using term and document associations as shown in Figure 3. A list of high-frequency terms is first generated for each document by word frequency counting procedures. Normalization may or may not be effected by thesaurus lookup. A term-term similarity matrix is then constructed by using co-occurrence of terms within sentences, rather than within documents, as a criterion. It should be noted that as new term associations are defined, the original incidence matrix can be revised by inclusion in some of the matrix columns of new, associated terms which are not originally contained in the respective sentences or documents. The revised incidence matrix then gives rise to a new term-term similarity matrix, incorporating second-order associations, and so on. This feedback process is represented by an upward-pointing arrow in Figure 3. (Salton, 1963, p. 444)

Figure 3

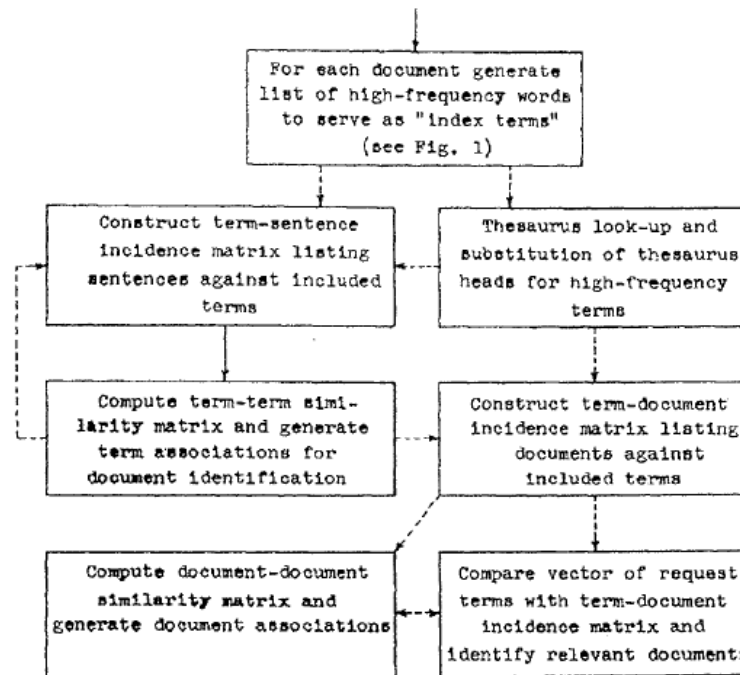


Fig. 3. Typical automatic document retrieval system using term and document associations
 → optional paths → compulsory paths

(Salton, 1963, p. 445)

Because of these and other variations, citation and reference lists have not generally been used as an indication of document content. Rather, such lists are used to detect trends in the literature as a whole, and to serve as adjuncts to certain kinds of literature searches [7, 8].

(Salton, 1963, p. 446)

Since the term-document matrix C is not in general a square matrix, matrix multiplication cannot be used to obtain second order effects, similar to the citation links of length two or more. Instead, it is first necessary to compare the index terms by performing a row comparison of the rows of C . This produces a new n symmetric term matrix C^* which displays similarity between index terms. This matrix can be used to eliminate from the set of index terms those terms which exhibit a large number of joint occurrences with other terms.

Claim Text from '352 Patent	Salton, 1963
	A reduced set of index terms can then be formed and a new term-document matrix C? constructed, from which a new correlation matrix S? is formed. (Salton, 1963, p. 449)
[45e] interpreting the processed searched commands as a selection of an object;	<p><i>See, e.g.,</i> Salton, 1963, at pp. 442, 447</p> <p>In practical retrieval systems, it becomes useful to provide for some additional flexibility. For example, given a search request expressed in terms of words in the natural language, it may be convenient to alter somewhat the original request, either by making it more specific and thus presumably reducing the size of the document set which fulfils the request or, alternatively, by making it more general. In the same way, given a set of terms identifying a specified document, it may be useful to alter somewhat the original set by deletion of old terms or addition of new ones in such a way that documents dealing with similar subject matter are identified by similar sets of index terms. (Salton, 1963, p. 442)</p> <p>Consider a collection of m documents each of which is characterized by the property of being cited by one or more of the other documents in the same collection. Each document can then be represented by an m-dimensional logical vector X_i, where $X_{ji} = 1$ if and only if document i is cited by document j, and $X_{ji} = 0$ otherwise. (Salton, 1963, p. 447)</p>
[45f] identifying a group of objects that have a relationship to the selected object, wherein the step of identifying comprises the steps of:	<p><i>See, e.g.,</i> Salton, 1963, at pp. 443-48, 450</p> <p>Many different types of similarity coefficients have been suggested in the literature; a simple coefficient of similarity between rows of a numeric matrix, and one which may be as meaningful as any of the others, is the cosine of the angle between the corresponding m-dimensional vectors. The similarity coefficients can be displayed in an $n \times n$ symmetric term-similarity matrix R, where the coefficient of similarity R_{ji} between term W_i and term W_j is</p>

Claim Text from '352 Patent	Salton, 1963
	$R_i^i = R_i^j = \frac{\sum_{k=1}^m C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^m (C_k^i)^2 \sum_{k=1}^m (C_k^j)^2\right)}}$ <p>...</p> <p>Document similarities are therefore obtained by comparing pairs of columns (instead of rows) of the term-document matrix C, and a document-document similarity matrix is constructed and used in the same way as the previously described term-term matrix R. (Salton, 1963, pp. 443-44)</p> <p>In particular, it may be conjectured that information associated with the author of a given document, for example data contained in related publications of the same author, may furnish usable content indicators. The same considerations may also apply to information obtained from publications cited by a given author in his list of references, or from those citing the given document. (Salton, 1963, p. 445)</p> <p>A citation index consists of a set of bibliographic references (the set of cited documents), each followed by a list of all those documents (the citing documents) which include the given cited document as a reference. A reference index, on the other hand, lists all cited documents under each citing document. (Salton, 1963, p. 446 n.1)</p> <p>To test the significance of bibliographic citations, a comparison was made between citation similarities and index term similarities for an indexed document collection. Specifically, a measure of similarity was computed between each pair of documents in the collection, based on the number of overlapping index terms a similar measure was then computed for the same pairs of documents, based on the number of overlapping citations; finally, the similarity measures obtained from index terms and citations respectively were compared by calculating a similarity index between citation similarities and index term similarities. An overall measure was also computed for the complete document collection by taking into account the similarity measures between all document pairs. (Salton, 1963, p. 447)</p> <p>A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar to that shown in Figure 2(b), where R_{ij} is the value of the similarity coefficient between the ith and</p>

Claim Text from '352 Patent	Salton, 1963
	<p>jth rows (columns) of X. (Salton, 1963, p. 448)</p> <p>The CITED and CITING similarity matrices of dimension 62 by 62 were obtained from the original citation matrix by row and column comparisons, respectively. The TDCMP similarity matrix, also of dimension 62 by 62, was similarly obtained by column comparisons from the original term-document matrix. Additional citation similarity matrices, designated CTD2, CTD3, CTD4, and CNG2, CNG3, CNG4 were obtained from the squared, cubed, and fourth power logical citation matrices, as previously explained. (Salton, 1963, p. 450)</p>
<p>[45g] Identifying objects that are referred to by the selected object; and</p>	<p><i>See, e.g.</i>, Salton, 1963, at pp. 445, 446 n.1, 450</p> <p>In particular, it may be conjectured that information associated with the author of a given document, for example data contained in related publications of the same author, may furnish usable content indicators. The same considerations may also apply to information obtained from publications cited by a given author in his list of references, or from those citing the given document. (Salton, 1963, p. 445)</p> <p>A citation index consists of a set of bibliographic references (the set of cited documents), each followed by a list of all those documents (the citing documents) which include the given cited document as a reference. A reference index, on the other hand, lists all cited documents under each citing document. (Salton, 1963, p. 446 n.1)</p> <p>The CITED and CITING similarity matrices of dimension 62 by 62 were obtained from the original citation matrix by row and column comparisons, respectively. The TDCMP similarity matrix, also of dimension 62 by 62, was similarly obtained by column comparisons from the original term-document matrix. Additional citation similarity matrices, designated CTD2, CTD3, CTD4, and CNG2, CNG3, CNG4 were obtained from the squared, cubed, and fourth power logical citation matrices, as previously explained. (Salton, 1963, p. 450)</p>
<p>[45h] Identifying objects that refer to the selected object</p>	<p><i>See, e.g.</i>, Salton, 1963, at pp. 445, 446 n.1, 450</p> <p>In particular, it may be conjectured that information associated with the author of a given</p>

Claim Text from '352 Patent	Salton, 1963
	<p>document, for example data contained in related publications of the same author, may furnish usable content indicators. The same considerations may also apply to information obtained from publications cited by a given author in his list of references, or from those citing the given document. (Salton, 1963, p. 445)</p> <p>A citation index consists of a set of bibliographic references (the set of cited documents), each followed by a list of all those documents (the citing documents) which include the given cited document as a reference. A reference index, on the other hand, lists all cited documents under each citing document. (Salton, 1963, p. 446 n.1)</p> <p>The CITED and CITING similarity matrices of dimension 62 by 62 were obtained from the original citation matrix by row and column comparisons, respectively. The TDCMP similarity matrix, also of dimension 62 by 62, was similarly obtained by column comparisons from the original term-document matrix. Additional citation similarity matrices, designated CTD2, CTD3, CTD4, and CNG2, CNG3, CNG4 were obtained from the squared, cubed, and fourth power logical citation matrices, as previously explained. (Salton, 1963, p. 450)</p>
<p>[45i] quantifying the relationship of the selected object to each object in the group of objects; and</p>	<p><i>See, e.g.</i>, Salton, 1963, at pp. 444, 448, 450-52</p> <p>An estimate of document relevance is then obtained by computing for each document the similarity coefficient between the request column C_{m+1} and the respective document column. The documents can be arranged in decreasing order of similarity coefficients, and all documents with a sufficiently large coefficient can be judged to be relevant to the given request. (Salton, 1963, p. 444)</p> <p>To retrieve documents in answer to search requests, the programs already available can be used by adding to the term-document matrix C a new column C_{m+1}, representing the request terms. Specifically, element C_{km+1} is set equal to w if term W_k is used in the search request with weight w; if word W_k is not used in the given search request C_{km+1} is set equal to 0. If no weights are specified by the requestor the values of the elements of column C_{m+1} are restricted to 0 and 1. (Salton, 1963, p. 444)</p> <p>Given a square citation matrix X it is possible by matrix multiplication to obtain matrices X', X'', etc., exhibiting respectively the existence of paths of length two, three, and so on. Specifically,</p>

Claim Text from '352 Patent	Salton, 1963
	$[X']_j^i = \bigvee_{k=1}^m (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^m (X_k^i \wedge (X')_j^k), \text{ and so on.}$ <p>Boolean multiplication is used, since the new connection matrices X', X'', etc., are again defined as logical matrices. $(X')_{ij}$ is then equal to 1 if and only if at least one path of length two exists between documents D_i and D_j; otherwise, $(X')_{ij}$ is equal to 0. It may be noted that X', unlike X, can have nonzero diagonal elements, corresponding to the case where two documents mutually cite each other. (Salton, 1963, p. 448)</p> <p>The CITED and CITING similarity matrices of dimension 62 by 62 were obtained from the original citation matrix by row and column comparisons, respectively. The TDCMP similarity matrix, also of dimension 62 by 62, was similarly obtained by column comparisons from the original term-document matrix. Additional citation similarity matrices, designated CTD2, CTD3, CTD4, and CNG2, CNG3, CNG4 were obtained from the squared, cubed, and fourth power logical citation matrices, as previously explained. (Salton, 1963, p. 450)</p> <p>The value of the overall similarity coefficient first rises as the length of the citation links increases, and then drops again as the length of the links becomes still greater [6]. This is due to the fact that as the length of the links increases, the total number of links of any length increases also; an increased number of links results in a larger number of ones in the original logical citation matrix, and thus in a higher probability of overlapping ones and a larger overall similarity coefficient. At the same time, as the length of the links increases, two factors also tend to decrease the magnitude of the overall similarity coefficient. First, the number of documents which exhibit citation links of length n but which do not exhibit links of length greater than n increases as n becomes larger. Thus more and more documents will exhibit individual similarity coefficients of zero value, thus tending to decrease the value of the overall coefficient. Second, as the length of the links increases and the citations thus become increasingly less accurate indications of document content, the magnitude of the cross-correlation coefficients obtained from the citation matrix and the term-document matrix would be expected to decrease, even for those documents for which a large number of citation links can still be found. (Salton, 1963, pp. 451-52)</p>

Claim Text from '352 Patent	Salton, 1963
[45j] ranking the objects in the group of objects in accordance to the quantified relationship to the selected object; and	<p><i>See, e.g.</i>, Salton, 1963, at p. 444</p> <p>An estimate of document relevance is then obtained by computing for each document the similarity coefficient between the request column C_{m+1} and the respective document column. The documents can be arranged in decreasing order of similarity coefficients, and all documents with a sufficiently large coefficient can be judged to be relevant to the given request. (Salton, 1963, p. 444)</p>
[45k] presenting one or more objects from the group of objects in ranked order.	<p><i>See, e.g.</i>, Salton, 1963, at p. 444</p> <p>An estimate of document relevance is then obtained by computing for each document the similarity coefficient between the request column C_{m+1} and the respective document column. The documents can be arranged in decreasing order of similarity coefficients, and all documents with a sufficiently large coefficient can be judged to be relevant to the given request. (Salton, 1963, p. 444)</p>

Defendants reserve the right to revise this contention chart concerning the invalidity of the asserted claims, as appropriate, for example depending upon the Court's construction of the asserted claims, any findings as to the priority date of the asserted claims, and/or positions that Plaintiff or its expert witness(es) may take concerning claim interpretation, construction, infringement, and/or invalidity issues.

Plaintiff's Infringement Contentions are based on an apparent construction of the claim terms. Defendants disagree with these apparent constructions. Nothing stated herein shall be treated as an admission or suggestion that Defendants agree with Plaintiff regarding either the scope of any of the asserted claims or the claim constructions advanced by Plaintiff in its Infringement Contentions or anywhere else, or that any of Defendants' accused technology meets any limitations of the claims. Nothing stated herein shall be construed as an admission or a waiver of any particular construction of any claim term. Defendants also reserve all their rights to challenge any of the claim terms herein under 35 U.S.C. § 112, including by arguing that they are indefinite, not supported by the written description and/or not enabled. Accordingly, nothing stated herein shall be construed as a waiver of any argument available under 35 U.S.C. § 112.

INVALIDITY CLAIM CHART FOR U.S. PATENT NO. 5,832,494
BASED ON EDWARD ALAN FOX, “EXTENDING THE BOOLEAN AND VECTOR SPACE MODELS OF INFORMATION RETRIEVAL WITH P-NORM QUERIES AND MULTIPLE CONCEPT TYPES” (“FOX THESIS, 1983”)

Claim Text for '494 Patent	Fox Thesis, 1983
1. A method of analyzing a database with indirect relationships, using links and nodes, comprising the steps of:	<i>See infra; see also, e.g.</i> , Chapters 1, & 6-9.
Selecting a node for analysis;	<i>Id.</i> at Chapter 1 (e.g., pp. 16-18, 19 (“The use of multiple concept types to generalize the vector representation of documents provides a second method for performance improvement. By including more information in the document representation and by judiciously utilizing that information through the relevance feedback cycle, improved retrieval can result.”), Chapter 5, Chapter 6, e.g., pp. 157-158, 159: “Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed . . . The bibliographic measures described here have been useful in both retrieval and clustering applications.”), 160-172, 173: “The use of <i>bc</i> , <i>cc</i> , and <i>ln</i> submatrices seems justified as an initial approach to better incorporating bibliographic data in the vector space model. Experiments in later chapters will contrast the utility of these measures and see how they can best be combined to aid retrieval system performance. The first requisite for such utilization, however, is an effective means to include the appropriate subvectors when computing similarities.”, 174-182), Chapter 7, Chapter 8, p. 205-206; <i>see also, e.g.</i> , Chapters 1, 7-9.
Generating candidate cluster links for the selected node, wherein the step of generating comprises an analysis of one or more indirect relationships in the database;	<i>Id.</i> at Chapter 6 (e.g., pp. 159-164, pp. 167-168: “B and C are bibliographically coupled if some document, say E, is referred to by both B and C. Here, a computer can count how many articles provide a coupling connection in a similar fashion to E - in Figure 6.2 there are no more - and define the degree of bibliographic coupling. thus, for arbitrary documents i and j,

Claim Text for '494 Patent	Fox Thesis, 1983																																																
	<p data-bbox="953 272 1073 297">$bc_{ij} = D'$</p> <p data-bbox="921 326 972 345">where</p> <p data-bbox="953 375 1220 399">$D_k \in D' \Leftrightarrow D_i \rightarrow D_k \text{ and } D_j \rightarrow D_k$</p> <p data-bbox="921 428 1419 448">and D' is restricted to the document set of definition, e.g., O.</p> <p data-bbox="863 448 1843 578">In the example of Figure 6.3, $bc_{B,C} = 1$ and $bc_{D,E} = 2$ since one document, E, is referred to by both B and C, while two documents, F and G, are each referred to by both D and E. Thus, B->E, C->E and D->F, E->F, D->G, E->G.”</p> <p data-bbox="926 618 1371 703">Table 6.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)</p> <table border="1" data-bbox="926 740 1430 1016"> <thead> <tr> <th>Citing Doc.</th> <th></th> <th>Cited Doc.</th> <th>Citing Doc.</th> <th></th> <th>Cited Doc.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>→</td> <td>D</td> <td>E</td> <td>→</td> <td>G</td> </tr> <tr> <td>B</td> <td>→</td> <td>E</td> <td>E</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>E</td> <td>G</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>G</td> <td>H</td> <td>→</td> <td>E</td> </tr> <tr> <td>D</td> <td>→</td> <td>F</td> <td>H</td> <td>→</td> <td>G</td> </tr> <tr> <td>D</td> <td>→</td> <td>G</td> <td>I</td> <td>→</td> <td>G</td> </tr> <tr> <td>E</td> <td>→</td> <td>F</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Citing Doc.		Cited Doc.	Citing Doc.		Cited Doc.	A	→	D	E	→	G	B	→	E	E	→	J	C	→	E	G	→	J	C	→	G	H	→	E	D	→	F	H	→	G	D	→	G	I	→	G	E	→	F			
Citing Doc.		Cited Doc.	Citing Doc.		Cited Doc.																																												
A	→	D	E	→	G																																												
B	→	E	E	→	J																																												
C	→	E	G	→	J																																												
C	→	G	H	→	E																																												
D	→	F	H	→	G																																												
D	→	G	I	→	G																																												
E	→	F																																															

Figure 6.3: b_c Submatrix

	A	B	C	D	E	F	G
A	1						
B		1	1				
C		1	2	1	1		
D			1	2	2		
E			1	2	3		
F						0	
G							1

Note: $bc_{E,G} \neq 1$ since J is not $\in O$.

,” p. 168: “F and G are co-cited [Small 1973] if some document, say D, refers to both of them in its bibliography. One can count the total number of articles that each refer to both F and G. For arbitrary documents i and j, the co-citation strength is then given by

$$cc_{ij} = |D''|$$

where

$$D'' \subseteq C,$$

the source set of documents considered, and

$$D_k \in D'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j.$$

Note that cc_{ii} is simply the number of articles that cite document i, that is, its citation count. That value can be used for normalizing other cc values or to gauge the importance of the given article. In the example, then, one observes that $cc_{E,G} = 2$ $cc_{F,G} = 2$ $cc_{F,J} = 1$,”

Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.
A	→	D	E	→	G
B	→	E	E	→	J
C	→	E	G	→	J
C	→	G	H	→	E
D	→	F	H	→	G
D	→	G	I	→	G
E	→	F			

Figure 6.4: cc Submatrix

	A	B	C	D	E	F	G
A	0						
B		0					
C			0				
D				1			
E					3		2
F						2	2
G					2	2	5

Note: this includes the fact that H cites E,G when $cc_{E,G}$ is computed.
The reason is that H is in the source set C for co-citations.

,” p. 170:

Claim Text for '494 Patent	Fox Thesis, 1983
	<p>For each document it is straightforward using the definitions of the last section to determine values of the linkage, bibliographic coupling, and co-citation measures between that document and any other document. Rather than using a dictionary to provide concept numbers, the document numbers themselves can be used so that</p> $M_{bc} = M_{cc} = M_{ln} = N \quad (6-21)$ <p>and submatrices BC, CC, and LN will each be of size $N \times N$. Note that according to the definitions of the various measures all diagonal entries will be non-zero but in general the submatrices will be sparsely populated.</p> <p>To obtain some intuition as to the meaning of these submatrices, consider the subvectors $b\vec{c}_i$, $c\vec{c}_i$, and $l\vec{n}_i$ for the i^{th} document. Diagonal entries are</p> $\begin{aligned} b\vec{c}_i &= \text{no. of references in bibliography of } i \\ c\vec{c}_i &= \text{no. of articles that refer to } i \\ l\vec{n}_i &= 1 \end{aligned} \quad (6-22)$ <p>where another way to understand $c\vec{c}_i$ is to view it as the incoming citation count.</p> <p>Off diagonal entries show how the i^{th} document relates to other documents. Thus, the j^{th} column of each submatrix shows how documents relate to the j^{th} document - one in effect treats a document as a "bibliographic concept". Off diagonal values have the following significance:</p> $\begin{aligned} b\vec{c}_{ij} &= \text{no. of articles referred to by both } i, j \\ c\vec{c}_{ij} &= \text{no. of articles that each refer to both } i, j \\ l\vec{n}_{ij} &= 1 \text{ if the } i^{th} \text{ doc. refers to the } j^{th}, \text{ or vice versa} \end{aligned} \quad (6-23)$ <p>," pp. 171-182, 205-206, 240 (Figure 8.2, Sample computations of inner products); <i>see also, e.g.</i>, Chapter 1, Chapters 6-9.</p>
Deriving actual cluster links from the candidate cluster links;	<p><i>Id.</i> at 192: "Later, Bichteler and Eaton [1980] demonstrated that for retrieval purposes using a similarity formula combining bibliographic coupling and co-citations was better than if bibliographic coupling alone was included. And, though on a small scale, they did do a certain amount of grouping of documents based on the resulting combined similarity values.," p. 192: "The algorithm produces a hierarchical clustering where all N documents in a collection end up as leaves of a multilevel tree.," pp. 199-201 ("Clustering Process"), 205-206; <i>see also, e.g.</i>, Chapters 1, & 6-9, Charts for the preceding limitation (including the quotations and descriptions set forth therein, which are incorporated by reference herein).</p>
identifying one or more nodes for display; and	<p><i>Id.</i> at Chapter 1 (e.g., pp. 16-18, 19 ("The use of multiple concept types to generalize the vector representation of documents provides a second method for performance</p>

Claim Text for '494 Patent	Fox Thesis, 1983
	improvement. By including more information in the document representation and by judiciously utilizing that information through the relevance feedback cycle, improved retrieval can result.”), Chapter 5, Chapter 6, e.g., pp. 157-158, 159: “Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed . . . The bibliographic measures described here have been useful in both retrieval and clustering applications.”), 160-172, 173: “The use of <i>bc</i> , <i>cc</i> , and <i>ln</i> submatrices seems justified as an initial approach to better incorporating bibliographic data in the vector space model. Experiments in later chapters will contrast the utility of these measures and see how they can best be combined to aid retrieval system performance. The first requisite for such utilization, however, is an effective means to include the appropriate subvectors when computing similarities.”, 174-182), Chapter 7, Chapter 8; <i>see also, e.g.</i> , Chapters 1, 7-9, Charts for the preceding limitation (including the quotations and descriptions set forth therein, which are incorporated by reference herein).
displaying the identity of one or more nodes using the actual cluster links.	<i>Id.</i> at 6: “In addition to being able to locate documents of interest, the user may be able to retrieve and/or examine paragraphs, passages, sentences, or single word occurrences (in context).” p. 219: “Note that exactly 30 documents are shown to the user,” p. 326: “First, it should be noted that at Syracuse an entire search was carried out, where various sets were retrieved and eventually the results of one of the sets was selected for printing.” <i>See</i> chapters 5 and 8; <i>see also, e.g.</i> , Chapters 1, & 6-9.
2. The method of claim 1 wherein each link is given a length, the step of generating the candidate cluster links comprises the steps of:	<i>See infra; see also, e.g.</i> , Chapters 1, & 6-9, Chart for Claim 1, <i>supra</i> (including the quotations and descriptions set forth therein, which are incorporated by reference herein).
Choosing a number as the maximum number of link lengths that will be examined; and	Chapters 1, & 6-9, Chart for Claim 1, <i>supra</i> (including the quotations and descriptions set forth therein, which are incorporated by reference herein).
examining only those links which are less than the maximum number of link lengths.	Chapters 1, & 6-9, Chart for Claim 1, <i>supra</i> (including the quotations and descriptions set forth therein, which are incorporated by reference herein).
3. The method of claim 1 wherein the step of deriving actual cluster links comprises the step of: selecting the top rated candidate cluster links, wherein the top rated candidate cluster links are those which are most closely linked to the node	<i>Id.</i> at Chapter 1 (e.g., pp. 16-18, 19 (“The use of multiple concept types to generalize the vector representation of documents provides a second method for performance improvement. By including more information in the document representation and by judiciously utilizing that information through the relevance feedback cycle, improved retrieval can result.”), Chapter 5, Chapter 6, e.g., pp. 157-158, 159: “Of

Claim Text for '494 Patent	Fox Thesis, 1983
under analysis.	more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed . . . The bibliographic measures described here have been useful in both retrieval and clustering applications.”), 160-172, 173: “The use of <i>bc</i> , <i>cc</i> , and <i>ln</i> submatrices seems justified as an initial approach to better incorporating bibliographic data in the vector space model. Experiments in later chapters will contrast the utility of these measures and see how they can best be combined to aid retrieval system performance. The first requisite for such utilization, however, is an effective means to include the appropriate subvectors when computing similarities.”, 174-182), Chapter 7, Chapter 8; <i>see also, e.g.</i> , Chapters 1, 7-9, Chart for Claim 1, <i>supra</i> (including the quotations and descriptions set forth therein, which are incorporated by reference herein).
5. The method of claim 1 wherein the step of generating the candidate cluster links comprises the step of: eliminating candidate cluster links, wherein the number of candidate cluster links is limited and the closest candidate cluster links are chosen over the remaining links.	<i>Id.</i> at Chapter 1 (e.g., pp. 16-18, 19 (“The use of multiple concept types to generalize the vector representation of documents provides a second method for performance improvement. By including more information in the document representation and by judiciously utilizing that information through the relevance feedback cycle, improved retrieval can result.”), Chapter 5, Chapter 6, e.g., pp. 157-158, 159: “Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed . . . The bibliographic measures described here have been useful in both retrieval and clustering applications.”), 160-172, 173: “The use of <i>bc</i> , <i>cc</i> , and <i>ln</i> submatrices seems justified as an initial approach to better incorporating bibliographic data in the vector space model. Experiments in later chapters will contrast the utility of these measures and see how they can best be combined to aid retrieval system performance. The first requisite for such utilization, however, is an effective means to include the appropriate subvectors when computing similarities.”, 174-182), Chapter 7, Chapter 8; <i>see also, e.g.</i> , Chapters 1, 7-9, Chart for Claim 1, <i>supra</i> (including the quotations and descriptions set forth therein, which are incorporated by reference herein).
7. The method of claim 1, wherein one or more nodes provide external connections to objects external to the database, the method further comprising the steps of:	Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.

Claim Text for '494 Patent	Fox Thesis, 1983
Activating the desired node; and	Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.
Accessing the external object linked to the node.	Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.
8. The method of claim 7, wherein the external object is an independent application which can be executed in background, the method further comprising the step of:	Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.
executing the independent application.	Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.
9. The method of claim 8, wherein one or more nodes provide links to more than one independent application which can be executed as an extension, the method further comprising the steps of:	Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.
displaying a list of independent applications linked to the node, wherein the step of accessing accesses an independent application.	Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.
10. The method of claim 8, wherein the connection provides the independent application access to the	Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art

Claim Text for '494 Patent	Fox Thesis, 1983
information stored within the database.	at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.
11. The method of claim 7, wherein the external connection is to another computer, wherein information is located that can be accessed, the step of accessing further comprising the step of:	Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.
accessing the information located within the computer.	Disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination with the state of the art at the time of the alleged invention, as evidenced by substantial other references identified in Defendants' P. R. 3-3 statement and accompanying charts. Rather than repeat those disclosures here, they are incorporated by reference into this chart.
12. A method for determining the proximity of an object in a stored database to another object in the stored database using indirect relationships, links, and a display, comprising:	<i>See infra; see also, e.g.</i> , Chapters 1, & 6-9.
Selecting an object to determine the proximity of other objects to the selected object;	<i>Id.</i> at Chapter 1 (e.g., pp. 16-18, 19 (“The use of multiple concept types to generalize the vector representation of documents provides a second method for performance improvement. By including more information in the document representation and by judiciously utilizing that information through the relevance feedback cycle, improved retrieval can result.”), Chapter 5, Chapter 6, e.g., pp. 157-158, 159: “Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed . . . The bibliographic measures described here have been useful in both retrieval and clustering applications.”), 160-172, 173: “The use of <i>bc</i> , <i>cc</i> , and <i>ln</i> submatrices seems justified as an initial approach to better incorporating bibliographic data in the vector space model. Experiments in later chapters will contrast the utility of these measures and see how they can best be combined to aid retrieval system performance. The first requisite for such utilization, however, is an effective means to include the appropriate subvectors when computing similarities.”, 174-182), p. 205-206, Chapter 7, Chapter 8; <i>see also, e.g.</i> , Chapters 1, 7-9.

Claim Text for '494 Patent	Fox Thesis, 1983																																																
<p>generating a candidate cluster link set for the selected object, wherein the generating step includes an analysis of one or more indirect relationships in the database;</p>	<p><i>Id.</i> at Chapter 6 (e.g., pp. 159-164, pp. 167-168: “B and C are bibliographically coupled if some document, say E, is referred to by both B and C. Here, a computer can count how many articles provide a coupling connection in a similar fashion to E - in Figure 6.2 there are no more - and define the degree of bibliographic coupling. thus, for arbitrary documents i and j,</p> $bc_{ij} = D' $ <p>where</p> $D_k \in D' \Leftrightarrow D_i \rightarrow D_k \text{ and } D_j \rightarrow D_k$ <p>and D' is restricted to the document set of definition, e.g., O.</p> <p>In the example of Figure 6.3, $bc_{B,C} = 1$ and $bc_{D,E} = 2$ since one document, E, is referred to by both B and C, while two documents, F and G, are each referred to by both D and E. Thus, B->E, C->E and D->F, E->F, D->G, E->G.”,</p> <p>Table 6.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)</p> <table border="1" data-bbox="919 917 1430 1193"> <thead> <tr> <th>Citing Doc.</th> <th></th> <th>Cited Doc.</th> <th>Citing Doc.</th> <th></th> <th>Cited Doc.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>→</td> <td>D</td> <td>E</td> <td>→</td> <td>G</td> </tr> <tr> <td>B</td> <td>→</td> <td>E</td> <td>E</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>E</td> <td>G</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>G</td> <td>H</td> <td>→</td> <td>E</td> </tr> <tr> <td>D</td> <td>→</td> <td>F</td> <td>H</td> <td>→</td> <td>G</td> </tr> <tr> <td>D</td> <td>→</td> <td>G</td> <td>I</td> <td>→</td> <td>G</td> </tr> <tr> <td>E</td> <td>→</td> <td>F</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Citing Doc.		Cited Doc.	Citing Doc.		Cited Doc.	A	→	D	E	→	G	B	→	E	E	→	J	C	→	E	G	→	J	C	→	G	H	→	E	D	→	F	H	→	G	D	→	G	I	→	G	E	→	F			
Citing Doc.		Cited Doc.	Citing Doc.		Cited Doc.																																												
A	→	D	E	→	G																																												
B	→	E	E	→	J																																												
C	→	E	G	→	J																																												
C	→	G	H	→	E																																												
D	→	F	H	→	G																																												
D	→	G	I	→	G																																												
E	→	F																																															

Figure 6.3: b_c Submatrix

	A	B	C	D	E	F	G
A	1						
B		1	1				
C		1	2	1	1		
D			1	2	2		
E			1	2	3		
F						0	
G							1

Note: $bc_{E,G} \neq 1$ since J is not $\in O$.

,” p. 168: “F and G are co-cited [Small 1973] if some document, say D, refers to both of them in its bibliography. One can count the total number of articles that each refer to both F and G. For arbitrary documents i and j, the co-citation strength is then given by

$$cc_{ij} = |D''|$$

where

$$D'' \subseteq C,$$

the source set of documents considered, and

$$D_k \in D'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j.$$

Note that cc_{ii} is simply the number of articles that cite document i, that is, its citation count. That value can be used for normalizing other cc values or to gauge the importance of the given article. In the example, then, one observes that $cc_{E,G} = 2$ $cc_{F,G} = 2$ $cc_{F,J} = 1$,”

Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.
A	→	D	E	→	G
B	→	E	E	→	J
C	→	E	G	→	J
C	→	G	H	→	E
D	→	F	H	→	G
D	→	G	I	→	G
E	→	F			

Figure 6.4: cc Submatrix

	A	B	C	D	E	F	G
A	0						
B		0					
C			0				
D				1			
E					3		2
F						2	2
G					2	2	5

Note: this includes the fact that H cites E,G when $cc_{E,G}$ is computed.
The reason is that H is in the source set C for co-citations.

,” p. 170:

Claim Text for '494 Patent	Fox Thesis, 1983
	<p>For each document it is straightforward using the definitions of the last section to determine values of the linkage, bibliographic coupling, and co-citation measures between that document and any other document. Rather than using a dictionary to provide concept numbers, the document numbers themselves can be used so that</p> $M_{bc} = M_{cc} = M_{ln} = N \quad (6-21)$ <p>and submatrices BC, CC, and LN will each be of size $N \times N$. Note that according to the definitions of the various measures all diagonal entries will be non-zero but in general the submatrices will be sparsely populated.</p> <p>To obtain some intuition as to the meaning of these submatrices, consider the subvectors $b\vec{c}_i$, $c\vec{c}_i$, and $l\vec{n}_i$ for the i^{th} document. Diagonal entries are</p> $\begin{aligned} bc_{ii} &= \text{no. of references in bibliography of } i \\ cc_{ii} &= \text{no. of articles that refer to } i \\ ln_{ii} &= 1 \end{aligned} \quad (6-22)$ <p>where another way to understand cc_{ii} is to view it as the incoming citation count.</p> <p>Off diagonal entries show how the i^{th} document relates to other documents. Thus, the j^{th} column of each submatrix shows how documents relate to the j^{th} document - one in effect treats a document as a "bibliographic concept". Off diagonal values have the following significance:</p> $\begin{aligned} bc_{ij} &= \text{no. of articles referred to by both } i, j \\ cc_{ij} &= \text{no. of articles that each refer to both } i, j \\ ln_{ij} &= 1 \text{ if the } i^{th} \text{ doc. refers to the } j^{th}, \text{ or vice versa} \end{aligned} \quad (6-23)$ <p>,” pp. 171-182, 205-206, p. 240 (Figure 8.2, Sample computations of inner products); <i>see also</i>, e.g., Chapter 1, Chapters 6-9.</p>
<p>Deriving an actual cluster link set for the selected object using the generated candidate cluster link set; and</p>	<p><i>Id.</i> at Chapter 1 (e.g., pp. 16-18, 19 (“The use of multiple concept types to generalize the vector representation of documents provides a second method for performance improvement. By including more information in the document representation and by judiciously utilizing that information through the relevance feedback cycle, improved retrieval can result.”), Chapter 5, Chapter 6, e.g., pp. 157-158, 159: “Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed . . . The bibliographic measures described here have been useful in both retrieval and clustering applications.”), 160-172, 173: “The use of bc, cc, and ln submatrices seems justified as an initial approach to better incorporating bibliographic data in the vector space model. Experiments in later chapters will contrast the utility of these measures and see how they can best be</p>

Claim Text for '494 Patent	Fox Thesis, 1983
	combined to aid retrieval system performance. The first requisite for such utilization, however, is an effective means to include the appropriate subvectors when computing similarities.”, 174-182), p. 205-206, Chapter 7, Chapter 8; <i>see also, e.g.</i> , Chapters 1, 7-9, Charts for the preceding limitation (including the quotations and descriptions set forth therein, which are incorporated by reference herein).
Displaying one or more of the objects in the database, referred to in the actual cluster link set, on a display.	<i>Id.</i> at 6: “In addition to being able to locate documents of interest, the user may be able to retrieve and/or examine paragraphs, passages, sentences, or single word occurrences (in context).,” p. 219: “Note that exactly 30 documents are shown to the user,” p. 326: “First, it should be noted that at Syracuse an entire search was carried out, where various sets were retrieved and eventually the results of one of the sets was selected for printing.” <i>See</i> chapters 5 and 8; <i>see also, e.g.</i> , Chapters 1, & 6-9.
13. The method of 12 wherein a set of direct links exists for the database, and wherein the step of generating a candidate cluster link set comprises: recursively analyzing portions of the set of direct links for indirect links.	<p><i>Id.</i> at Chapter 6 (<i>e.g., pp. 159-164, pp. 167-168</i>): “B and C are bibliographically coupled if some document, say E, is referred to by both B and C. Here, a computer can count how many articles provide a coupling connection in a similar fashion to E - in Figure 6.2 there are no more - and define the degree of bibliographic coupling. thus, for arbitrary documents i and j,</p> $bc_{ij} = D' $ <p>where</p> $D_k \in D' \Leftrightarrow D_i \rightarrow D_k \text{ and } D_j \rightarrow D_k$ <p>and D' is restricted to the document set of definition, <i>e.g., O</i>.</p> <p>In the example of Figure 6.3, $bc_{B,C} = 1$ and $bc_{D,E} = 2$ since one document, E, is referred to by both B and C, while two documents, F and G, are each referred to by both D and E. Thus, B->E, C->E and D->F, E->F, D->G, E->G.”,</p>

Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.
A	→	D	E	→	G
B	→	E	E	→	J
C	→	E	G	→	J
C	→	G	H	→	E
D	→	F	H	→	G
D	→	G	I	→	G
E	→	F			

Figure 6.3: bc Submatrix

	A	B	C	D	E	F	G
A	1						
B		1	1				
C		1	2	1	1		
D			1	2	2		
E			1	2	3		
F						0	
G							1

Note: $bc_{E,G} \neq 1$ since J is not $\in O$.

[Small 1973] if some document, say D, refers to both of them in its bibliography. One can count the total number of articles that each refer to both F and G. For arbitrary documents i and j, the co-citation strength is then given by
 ,” p. 168: “F and G are co-cited

Claim Text for '494 Patent	Fox Thesis, 1983																																																
	<p data-bbox="976 267 1144 300">$cc_{ij} = D''$</p> <p data-bbox="924 341 997 365">where</p> <p data-bbox="966 406 1081 438">$D'' \subseteq C,$</p> <p data-bbox="924 479 1407 511">the source set of documents considered, and</p> <p data-bbox="966 552 1365 584">$D_k \in D'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j.$</p> <p data-bbox="861 600 1858 706">Note that cc_{ii} is simply the number of articles that cite document i, that is, its citation count. That value can be used for normalizing other cc values or to gauge the importance of the given article. In the example, then, one observes that</p> <p data-bbox="861 714 1186 747">$cc_{E,G} = 2 \quad cc_{F,G} = 2 \quad cc_{E,J} = 1,$</p> <p data-bbox="924 787 1375 876">Table 8.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)</p> <table border="1" data-bbox="924 909 1438 1185"> <thead> <tr> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>→</td> <td>D</td> <td>E</td> <td>→</td> <td>G</td> </tr> <tr> <td>B</td> <td>→</td> <td>E</td> <td>E</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>E</td> <td>G</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>G</td> <td>H</td> <td>→</td> <td>E</td> </tr> <tr> <td>D</td> <td>→</td> <td>F</td> <td>H</td> <td>→</td> <td>G</td> </tr> <tr> <td>D</td> <td>→</td> <td>G</td> <td>I</td> <td>→</td> <td>G</td> </tr> <tr> <td>E</td> <td>→</td> <td>F</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.	A	→	D	E	→	G	B	→	E	E	→	J	C	→	E	G	→	J	C	→	G	H	→	E	D	→	F	H	→	G	D	→	G	I	→	G	E	→	F			
Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.																																												
A	→	D	E	→	G																																												
B	→	E	E	→	J																																												
C	→	E	G	→	J																																												
C	→	G	H	→	E																																												
D	→	F	H	→	G																																												
D	→	G	I	→	G																																												
E	→	F																																															

Figure 6.4: cc Submatrix

	A	B	C	D	E	F	G
A	0						
B		0					
C			0				
D				1			
E					3		2
F						2	2
G					2	2	5

Note: this includes the fact that H cites E,G when $cc_{E,G}$ is computed. The reason is that H is in the source set C for co-citations.

,” p. 170:

For each document it is straightforward using the definitions of the last section to determine values of the linkage, bibliographic coupling, and co-citation measures between that document and any other document. Rather than using a dictionary to provide concept numbers, the document numbers themselves can be used so that

$$M_{bc} = M_{cc} = M_{ln} = N \tag{6-21}$$

and submatrices BC , CC , and LN will each be of size $N \times N$. Note that according to the definitions of the various measures all diagonal entries will be non-zero but in general the submatrices will be sparsely populated.

To obtain some intuition as to the meaning of these submatrices, consider the subvectors \vec{bc}_i , \vec{cc}_i , and \vec{ln}_i for the i^{th} document. Diagonal entries are

$$\begin{aligned} bc_{ii} &= \text{no. of references in bibliography of } i \\ cc_{ii} &= \text{no. of articles that refer to } i \\ ln_{ii} &= 1 \end{aligned} \tag{6-22}$$

where another way to understand cc_{ii} is to view it as the incoming citation count.

Off diagonal entries show how the i^{th} document relates to other documents. Thus, the j^{th} column of each submatrix shows how documents relate to the j^{th} document – one in effect treats a document as a “bibliographic concept”. Off diagonal values have the following significance:

$$\begin{aligned} bc_{ij} &= \text{no. of articles referred to by both } i, j \\ cc_{ij} &= \text{no. of articles that each refer to both } i, j \\ ln_{ij} &= 1 \text{ if the } i^{th} \text{ doc. refers to the } j^{th}, \text{ or vice versa} \end{aligned} \tag{6-23}$$

, pp. 171-182, p. 193, 205-206,

Id. at Chapter 7 (e.g., p. 192: “The algorithm produces a hierarchical clustering where all N documents in a collection end up as leaves of a multilevel tree.” pp.

Claim Text for '494 Patent	Fox Thesis, 1983																																																
	199-201 (“Clustering Process”); <i>see also, e.g.</i> , Chapters 1, 6, 8-9.																																																
14. A method for representing the relationship between nodes using stored direct links, paths, and candidate cluster links, comprising the steps of:	<i>See infra; see also, e.g.</i> , Chapters 1, & 6-9.																																																
initializing a set of candidate cluster links;	<p><i>Id.</i> at Chapter 6 (<i>e.g.</i>, pp. 159-164, pp. 167-168: “B and C are bibliographically coupled if some document, say E, is referred to by both B and C. Here, a computer can count how many articles provide a coupling connection in a similar fashion to E - in Figure 6.2 there are no more - and define the degree of bibliographic coupling. thus, for arbitrary documents i and j,</p> $bc_{ij} = D' $ <p>where</p> $D_k \in D' \Leftrightarrow D_i \rightarrow D_k \text{ and } D_j \rightarrow D_k$ <p>and D' is restricted to the document set of definition, <i>e.g.</i>, O.</p> <p>In the example of Figure 6.3, $bc_{B,C} = 1$ and $bc_{D,E} = 2$ since one document, E, is referred to by both B and C, while two documents, F and G, are each referred to by both D and E. Thus, B->E, C->E and D->F, E->F, D->G, E->G.”</p> <p>Table 6.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)</p> <table border="1" data-bbox="919 1068 1432 1351"> <thead> <tr> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>→</td> <td>D</td> <td>E</td> <td>→</td> <td>G</td> </tr> <tr> <td>B</td> <td>→</td> <td>E</td> <td>E</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>E</td> <td>G</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>G</td> <td>H</td> <td>→</td> <td>E</td> </tr> <tr> <td>D</td> <td>→</td> <td>F</td> <td>H</td> <td>→</td> <td>G</td> </tr> <tr> <td>D</td> <td>→</td> <td>G</td> <td>I</td> <td>→</td> <td>G</td> </tr> <tr> <td>E</td> <td>→</td> <td>F</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.	A	→	D	E	→	G	B	→	E	E	→	J	C	→	E	G	→	J	C	→	G	H	→	E	D	→	F	H	→	G	D	→	G	I	→	G	E	→	F			
Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.																																												
A	→	D	E	→	G																																												
B	→	E	E	→	J																																												
C	→	E	G	→	J																																												
C	→	G	H	→	E																																												
D	→	F	H	→	G																																												
D	→	G	I	→	G																																												
E	→	F																																															

Figure 6.3: b_c Submatrix

	A	B	C	D	E	F	G
A	1						
B		1	1				
C		1	2	1	1		
D			1	2	2		
E			1	2	3		
F						0	
G							1

Note: $bc_{E,G} \neq 1$ since J is not $\in O$.

,” p. 168: “F and G are co-cited [Small 1973] if some document, say D, refers to both of them in its bibliography. One can count the total number of articles that each refer to both F and G. For arbitrary documents i and j, the co-citation strength is then given by

$$cc_{ij} = |D''|$$

where

$$D'' \subseteq C,$$

the source set of documents considered, and

$$D_k \in D'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j.$$

Note that cc_{ii} is simply the number of articles that cite document i, that is, its citation count. That value can be used for normalizing other cc values or to gauge the importance of the given article. In the example, then, one observes that $cc_{E,G} = 2$ $cc_{F,G} = 2$ $cc_{F,J} = 1$,”

Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.
A	→	D	E	→	G
B	→	E	E	→	J
C	→	E	G	→	J
C	→	G	H	→	E
D	→	F	H	→	G
D	→	G	I	→	G
E	→	F			

Figure 6.4: cc Submatrix

	A	B	C	D	E	F	G
A	0						
B		0					
C			0				
D				1			
E					3		2
F						2	2
G					2	2	5

Note: this includes the fact that H cites E,G when $cc_{E,G}$ is computed.
The reason is that H is in the source set C for co-citations.

,” p. 170:

Claim Text for '494 Patent	Fox Thesis, 1983
	<p>For each document it is straightforward using the definitions of the last section to determine values of the linkage, bibliographic coupling, and co-citation measures between that document and any other document. Rather than using a dictionary to provide concept numbers, the document numbers themselves can be used so that</p> $M_{bc} = M_{cc} = M_{ln} = N \quad (6-21)$ <p>and submatrices BC, CC, and LN will each be of size $N \times N$. Note that according to the definitions of the various measures all diagonal entries will be non-zero but in general the submatrices will be sparsely populated.</p> <p>To obtain some intuition as to the meaning of these submatrices, consider the subvectors $b\vec{c}_i$, $c\vec{c}_i$, and $l\vec{n}_i$ for the i^{th} document. Diagonal entries are</p> $\begin{aligned} b\vec{c}_i &= \text{no. of references in bibliography of } i \\ c\vec{c}_i &= \text{no. of articles that refer to } i \\ l\vec{n}_i &= 1 \end{aligned} \quad (6-22)$ <p>where another way to understand $c\vec{c}_i$ is to view it as the incoming citation count.</p> <p>Off diagonal entries show how the i^{th} document relates to other documents. Thus, the j^{th} column of each submatrix shows how documents relate to the j^{th} document - one in effect treats a document as a "bibliographic concept". Off diagonal values have the following significance:</p> $\begin{aligned} b\vec{c}_{ij} &= \text{no. of articles referred to by both } i, j \\ c\vec{c}_{ij} &= \text{no. of articles that each refer to both } i, j \\ l\vec{n}_{ij} &= 1 \text{ if the } i^{\text{th}} \text{ doc. refers to the } j^{\text{th}}, \text{ or vice versa} \end{aligned} \quad (6-23)$ <p>,” pp. 171-182, 205-206, p. 240 (Figure 8.2, Sample computations of inner products); <i>see also, e.g.</i>, Chapter 1, Chapters 6-9.</p>
<p>Selecting the destination node of a path as the selected node to analyze;</p>	<p><i>Id.</i> at 159: “In addition to terms and authors, other types of information are available in many collections. Dates and controlled vocabulary terms may be properly separated from regular terms. Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed,” pp. 166-167: “Based on the reference pattern for a set of documents, one may define various derived measures of the interconnection between those documents. The relevant notation and definitions follow, using the data of Figure 6-3 to illustrate each point:</p>

Claim Text for '494 Patent	Fox Thesis, 1983																																																
	<p>(6-16) $A \rightarrow D$ Direct Reference when A refers to (cites) document D, so that D is referred to (cited by) A. By definition, $D \rightarrow D$ always holds.</p> <p>(6-17) $A \rightarrow^k G$ Indirect Reference when A indirectly refers to (cites) G (e.g., at distance $k=2$), so that G is indirectly referred to (cited by) A.</p> <p>,” p. 169: “A and D are linked if either $A \rightarrow D$ or $D \rightarrow A$ [Salton 1963]. This definition allows the computer to symmetrically view citation connections between documents, regardless of the ordering of the articles based on time of publication. More formally,</p> $ln_{i,j} = \begin{cases} 1 & \text{if } D_i \rightarrow D_j \\ 1 & \text{if } D_j \rightarrow D_i \\ 1 & \text{if } i = j, \text{ by definition} \\ 0 & \text{otherwise.} \end{cases}$ <p>In the example, there are ln_{ij} values of 1 for pairs such as A and D or C and G.</p> <p>,” Figure 6.5:</p> <p>Table 6.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)</p> <table border="1" data-bbox="919 862 1430 1138"> <thead> <tr> <th>Citing Doc.</th> <th></th> <th>Cited Doc.</th> <th>Citing Doc.</th> <th></th> <th>Cited Doc.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>→</td> <td>D</td> <td>E</td> <td>→</td> <td>G</td> </tr> <tr> <td>B</td> <td>→</td> <td>E</td> <td>E</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>E</td> <td>G</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>G</td> <td>H</td> <td>→</td> <td>E</td> </tr> <tr> <td>D</td> <td>→</td> <td>F</td> <td>H</td> <td>→</td> <td>G</td> </tr> <tr> <td>D</td> <td>→</td> <td>G</td> <td>I</td> <td>→</td> <td>G</td> </tr> <tr> <td>E</td> <td>→</td> <td>F</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Citing Doc.		Cited Doc.	Citing Doc.		Cited Doc.	A	→	D	E	→	G	B	→	E	E	→	J	C	→	E	G	→	J	C	→	G	H	→	E	D	→	F	H	→	G	D	→	G	I	→	G	E	→	F			
Citing Doc.		Cited Doc.	Citing Doc.		Cited Doc.																																												
A	→	D	E	→	G																																												
B	→	E	E	→	J																																												
C	→	E	G	→	J																																												
C	→	G	H	→	E																																												
D	→	F	H	→	G																																												
D	→	G	I	→	G																																												
E	→	F																																															

Claim Text for '494 Patent	Fox Thesis, 1983																																																																
	<p data-bbox="926 261 1192 289">Figure 6.5: <i>I_n</i> Submatrix</p> <table border="1" data-bbox="926 321 1341 586"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> </tr> </thead> <tbody> <tr> <th>A</th> <td>1</td> <td></td> <td></td> <td>1</td> <td></td> <td></td> <td></td> </tr> <tr> <th>B</th> <td></td> <td>1</td> <td></td> <td></td> <td>1</td> <td></td> <td></td> </tr> <tr> <th>C</th> <td></td> <td></td> <td>1</td> <td></td> <td>1</td> <td></td> <td>1</td> </tr> <tr> <th>D</th> <td>1</td> <td></td> <td></td> <td>1</td> <td></td> <td>1</td> <td>1</td> </tr> <tr> <th>E</th> <td></td> <td>1</td> <td>1</td> <td></td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <th>F</th> <td></td> <td></td> <td></td> <td>1</td> <td>1</td> <td>1</td> <td></td> </tr> <tr> <th>G</th> <td></td> <td></td> <td>1</td> <td>1</td> <td>1</td> <td></td> <td>1</td> </tr> </tbody> </table> <p data-bbox="863 613 1856 678">, p. 170 “$I_{ij} = 1$ if the i^{th} doc. refers to the j^{th}, or vice versa.”; <i>see also, e.g.</i>, Chapters 1, & 6-9.</p>		A	B	C	D	E	F	G	A	1			1				B		1			1			C			1		1		1	D	1			1		1	1	E		1	1		1	1	1	F				1	1	1		G			1	1	1		1
	A	B	C	D	E	F	G																																																										
A	1			1																																																													
B		1			1																																																												
C			1		1		1																																																										
D	1			1		1	1																																																										
E		1	1		1	1	1																																																										
F				1	1	1																																																											
G			1	1	1		1																																																										
<p data-bbox="233 699 814 760">retrieving the set of direct links from the selected node to any other node in the database;</p>	<p data-bbox="863 699 1866 963"><i>Id.</i> at 159: “In addition to terms and authors, other types of information are available in many collections. Dates and controlled vocabulary terms may be properly separated from regular terms. Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed,” pp. 166-167: “Based on the reference pattern for a set of documents, one may define various derived measures of the interconnection between those documents. The relevant notation and definitions follow, using the data of Figure 6-3 to illustrate each point:</p> <p data-bbox="877 976 1444 1036">(6-16) $A \rightarrow D$ Direct Reference when A refers to (cites) document D, so that D is referred to (cited by) A. By definition, $D \rightarrow D$ always holds.</p> <p data-bbox="877 1060 1444 1120">(6-17) $A \rightarrow^k G$ Indirect Reference when A indirectly refers to (cites) G (e.g., at distance $k=2$), so that G is indirectly referred to (cited by) A.</p> <p data-bbox="863 1097 1866 1227">,” p. 169: “A and D are linked if either $A \rightarrow D$ or $D \rightarrow A$ [Salton 1963]. This definition allows the computer to symmetrically view citation connections between documents, regardless of the ordering of the articles based on time of publication. More formally,</p>																																																																

$$ln_{i,j} = \begin{cases} 1 & \text{if } D_i \rightarrow D_j \\ 1 & \text{if } D_j \rightarrow D_i \\ 1 & \text{if } i = j, \text{ by definition} \\ 0 & \text{otherwise.} \end{cases}$$

In the example, there are ln_{ij} values of 1 for pairs such as A and D or C and G.

,” Figure 6.5:

Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

Citing Doc.		Cited Doc.	Citing Doc.		Cited Doc.
A	→	D	E	→	G
B	→	E	E	→	J
C	→	E	G	→	J
C	→	G	H	→	E
D	→	F	H	→	G
D	→	G	I	→	G
E	→	F			

Figure 6.5: ln Submatrix

	A	B	C	D	E	F	G
A	1			1			
B		1			1		
C			1		1		1
D	1			1		1	1
E		1	1		1	1	1
F				1	1	1	
G			1	1	1		1

, p. 170 “ $ln_{ij} = 1$ if the i^{th} doc. refers to the j^{th} , or vice versa.”; *see also, e.g.*, Chapters 1, & 6-9.

Determining the weight of the path using the retrieved direct links;

Id. at 158: “Incidentally, the various subvectors could be construed using different weighting schemes; an additional column in Table 6.1 could show that, for example,

Claim Text for '494 Patent	Fox Thesis, 1983
	<p>term weights were computed using the scheme $tf*idf$ while author entries were given binary weights,” p. 168: “Now, citing directly as given in (6-16) or indirectly as in (6-17) are binary events -- either they occur or not. On the other hand, the next two definitions can result in an assignment of weights that are based upon integer counts.</p> <p>(6-18) B and C are bibliographically coupled [Kessler 1962] if some document, say E, is referred to by both B and C.</p> <p>Hence a computer can count how many articles provide a coupling connection in a similar fashion to E -- in Figure 6.2 there are no more -- and define the degree of bibliographic coupling,” p. 179: “Weighting methods may vary for different subvectors. Dates should undoubtedly receive binary weights, whereas terms benefit from applying an inverse document frequency (idf) factor. Bibliographic submatrices should also use some type of weighting,” p. 168: “F and G are co-cited [Small 1973] if some document, say D, refers to both of them in its bibliography. One can count the total number of articles that each refer to both F and G. For arbitrary documents i and j, the co-citation strength is then given by</p> $cc_{ij} = D'' $ <p>where</p> $D'' \subseteq C,$ <p>the source set of documents considered, and</p> $D_k \in D'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j.$ <p>Note that cc_{ii} is simply the number of articles that cite document i, that is, its citation count. That value can be used for normalizing other cc values or to gauge the importance of the given article. In the example, then, one observes that $cc_{E,G} = 2 \text{ } cc_{F,G} = 2 \text{ } cc_{F,I} = 1,$”</p>

Claim Text for '494 Patent	Fox Thesis, 1983																																																																																																																
	<p data-bbox="926 277 1369 363">Table 6.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)</p> <table border="1" data-bbox="919 402 1430 678"> <thead> <tr> <th>Citing Doc.</th> <th></th> <th>Cited Doc.</th> <th>Citing Doc.</th> <th></th> <th>Cited Doc.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>→</td> <td>D</td> <td>E</td> <td>→</td> <td>G</td> </tr> <tr> <td>B</td> <td>→</td> <td>E</td> <td>E</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>E</td> <td>G</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>G</td> <td>H</td> <td>→</td> <td>E</td> </tr> <tr> <td>D</td> <td>→</td> <td>F</td> <td>H</td> <td>→</td> <td>G</td> </tr> <tr> <td>D</td> <td>→</td> <td>G</td> <td>I</td> <td>→</td> <td>G</td> </tr> <tr> <td>E</td> <td>→</td> <td>F</td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p data-bbox="900 729 1113 750">Figure 6.4: cc Submatrix</p> <table border="1" data-bbox="905 777 1226 987"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> </tr> </thead> <tbody> <tr> <th>A</th> <td>0</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>B</th> <td></td> <td>0</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>C</th> <td></td> <td></td> <td>0</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>D</th> <td></td> <td></td> <td></td> <td>1</td> <td></td> <td></td> <td></td> </tr> <tr> <th>E</th> <td></td> <td></td> <td></td> <td></td> <td>3</td> <td></td> <td>2</td> </tr> <tr> <th>F</th> <td></td> <td></td> <td></td> <td></td> <td></td> <td>2</td> <td>2</td> </tr> <tr> <th>G</th> <td></td> <td></td> <td></td> <td></td> <td>2</td> <td>2</td> <td>5</td> </tr> </tbody> </table> <p data-bbox="900 992 1476 1036">Note: this includes the fact that H cites E,G when $cc_{E,G}$ is computed. The reason is that H is in the source set C for co-citations.</p> <p data-bbox="863 1081 1251 1110"><i>See also, e.g.,</i> Chapters 1, & 6-9.</p>	Citing Doc.		Cited Doc.	Citing Doc.		Cited Doc.	A	→	D	E	→	G	B	→	E	E	→	J	C	→	E	G	→	J	C	→	G	H	→	E	D	→	F	H	→	G	D	→	G	I	→	G	E	→	F					A	B	C	D	E	F	G	A	0							B		0						C			0					D				1				E					3		2	F						2	2	G					2	2	5
Citing Doc.		Cited Doc.	Citing Doc.		Cited Doc.																																																																																																												
A	→	D	E	→	G																																																																																																												
B	→	E	E	→	J																																																																																																												
C	→	E	G	→	J																																																																																																												
C	→	G	H	→	E																																																																																																												
D	→	F	H	→	G																																																																																																												
D	→	G	I	→	G																																																																																																												
E	→	F																																																																																																															
	A	B	C	D	E	F	G																																																																																																										
A	0																																																																																																																
B		0																																																																																																															
C			0																																																																																																														
D				1																																																																																																													
E					3		2																																																																																																										
F						2	2																																																																																																										
G					2	2	5																																																																																																										
repeating steps b through d for each path; and	<i>See</i> Charts for previous limitations.																																																																																																																
Storing the determined weights as candidate cluster links.	<p data-bbox="863 1174 1860 1416"><i>Id.</i> at 158: “Incidentally, the various subvectors could be construed using different weighting schemes; an additional column in Table 6.1 could show that, for example, term weights were computed using the scheme $tf*idf$ while author entries were given binary weights,” p. 168: “Now, citing directly as given in (6-16) or indirectly as in (6-17) are binary events -- either they occur or not. On the other hand, the next two definitions can result in an assignment of weights that are based upon integer counts. (6-18) B and C are bibliographically coupled [Kessler 1962] if some document, say</p>																																																																																																																

Claim Text for '494 Patent	Fox Thesis, 1983
	<p>E, is referred to by both B and C.</p> <p>Hence a computer can count how many articles provide a coupling connection in a similar fashion to E -- in Figure 6.2 there are no more -- and define the degree of bibliographic coupling," p. 179: "Weighting methods may vary for different subvectors. Dates should undoubtedly receive binary weights, whereas terms benefit from applying an inverse document frequency (idf) factor. Bibliographic submatrices should also use some type of weighting.," p. 168: "F and G are co-cited [Small 1973] if some document, say D, refers to both of them in its bibliography. One can count the total number of articles that each refer to both F and G. For arbitrary documents i and j, the co-citation strength is then given by</p> $cc_{ij} = D'' $ <p>where</p> $D'' \subseteq C,$ <p>the source set of documents considered, and</p> $D_k \in D'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j.$ <p>Note that cc_{ii} is simply the number of articles that cite document i, that is, its citation count. That value can be used for normalizing other cc values or to gauge the importance of the given article. In the example, then, one observes that</p> $cc_{E,G} = 2 \quad cc_{F,G} = 2 \quad cc_{F,J} = 1,$

Claim Text for '494 Patent	Fox Thesis, 1983																																																																																																																
	<p data-bbox="926 277 1371 363">Table 8.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)</p> <table border="1" data-bbox="921 402 1430 675"> <thead> <tr> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>→</td> <td>D</td> <td>E</td> <td>→</td> <td>G</td> </tr> <tr> <td>B</td> <td>→</td> <td>E</td> <td>E</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>E</td> <td>G</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>G</td> <td>H</td> <td>→</td> <td>E</td> </tr> <tr> <td>D</td> <td>→</td> <td>F</td> <td>H</td> <td>→</td> <td>G</td> </tr> <tr> <td>D</td> <td>→</td> <td>G</td> <td>I</td> <td>→</td> <td>G</td> </tr> <tr> <td>E</td> <td>→</td> <td>F</td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p data-bbox="900 729 1115 750">Figure 6.4: cc Submatrix</p> <table border="1" data-bbox="905 777 1226 987"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> </tr> </thead> <tbody> <tr> <th>A</th> <td>0</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>B</th> <td></td> <td>0</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>C</th> <td></td> <td></td> <td>0</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>D</th> <td></td> <td></td> <td></td> <td>1</td> <td></td> <td></td> <td></td> </tr> <tr> <th>E</th> <td></td> <td></td> <td></td> <td></td> <td>3</td> <td></td> <td>2</td> </tr> <tr> <th>F</th> <td></td> <td></td> <td></td> <td></td> <td></td> <td>2</td> <td>2</td> </tr> <tr> <th>G</th> <td></td> <td></td> <td></td> <td></td> <td>2</td> <td>2</td> <td>5</td> </tr> </tbody> </table> <p data-bbox="900 992 1476 1036">Note: this includes the fact that H cites E,G when $cc_{E,G}$ is computed. The reason is that H is in the source set C for co-citations.</p> <p data-bbox="863 1081 1262 1110"><i>See also, e.g.,</i> Chapters 1, & 6-9).</p>	Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.	A	→	D	E	→	G	B	→	E	E	→	J	C	→	E	G	→	J	C	→	G	H	→	E	D	→	F	H	→	G	D	→	G	I	→	G	E	→	F					A	B	C	D	E	F	G	A	0							B		0						C			0					D				1				E					3		2	F						2	2	G					2	2	5
Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.																																																																																																												
A	→	D	E	→	G																																																																																																												
B	→	E	E	→	J																																																																																																												
C	→	E	G	→	J																																																																																																												
C	→	G	H	→	E																																																																																																												
D	→	F	H	→	G																																																																																																												
D	→	G	I	→	G																																																																																																												
E	→	F																																																																																																															
	A	B	C	D	E	F	G																																																																																																										
A	0																																																																																																																
B		0																																																																																																															
C			0																																																																																																														
D				1																																																																																																													
E					3		2																																																																																																										
F						2	2																																																																																																										
G					2	2	5																																																																																																										
15. The method of claim 14 further comprising the step of deriving the actual cluster links wherein the actual cluster links are a subset of the candidate cluster links.	<i>See id.</i> at Chapter 7; <i>see also, e.g.,</i> Chapters 1, & 6-9, Chart for Claim 14, <i>supra</i> (including the quotations and descriptions set forth therein, which are incorporated by reference herein).																																																																																																																
16. The method of claim 15 wherein the step of deriving comprises the step of choosing the top rated candidate cluster links.	<i>Id.</i> at 193: “The algorithm produces a hierarchical clustering where all N documents in a collection end up as leaves of a multilevel tree. Interior nodes are associated with cluster centroids which represent all the documents in the subtree below them. Viewed another way, a given centroid summarizes all the information contained in																																																																																																																

Claim Text for '494 Patent	Fox Thesis, 1983
	<p>the children immediately below regardless of whether those are documents or other centroids.</p> <p>Clustering proceeds by adding documents one by one starting with an initially empty tree. The addition process involves a search for the proper place to insert the new document and a subsequent adjustment of the three to first include the new entry and secondly conform to the various constraints enforced during the build operation.</p> <p>Table 7.1 gives specific parameters required to handle clustering of extended vectors. The first three values indicate choices specifying how the overall similarity between documents can be determined based on available subvectors -- relative weighting method, similarity function used, and whether real valued weights are allowed. The last two parameters relate to special processing when a centroid subvector gets too long and must be shortened to fit available space.</p> <p>Table 7.1: Combined Retrieval Parameters for Each Concept Type</p> <p>similarity coefficient = coefficient used for a given concept type before adding it to arrive at overall similarity, based on formula:</p> $\text{combined similarity} = \sum_{\text{all types } t} \text{coeff}_t \cdot \text{sim}_t$ <p>similarity computation method = specification of function to compute similarity: cos correlation, inner product, normalized inner product (i.e., divided by sum of vector values)</p> <p>weighting method = use binary or real values</p> <p>maximum subvector length = length of this subvector that must not be exceeded; if it is, then low frequency values in the subvector are deleted to shorten it to within bounds</p> <p>subvector deletion frequency: initial value and increment = when subvector must be shortened, all entries below the initial value are deleted, and for subsequent deletions the increment is added to the cutoff previously used</p> <p>”; see also, e.g., Chapters 1, & 6-9.</p>
<p>18. A method of analyzing a database having objects and a first numerical representation of direct relationships in the database, comprising the steps of:</p>	<p><i>Id.</i> at Chapter 6 (e.g., p. 155: “it seems to be practically and conceptually better to more clearly separate the extended vector into two subvectors. Representing the term subvector for the i^{th} subvector as tm_i, and the author subvector as au_i, the i^{th} document is described as</p> $\vec{D}_i^t = (tm_i, au_i). \tag{6-4}$ <p>Expanded, the subvectors have the equivalent form</p> $\vec{D}_i^t = (tm_{i1}, \dots, tm_{iM_m}, au_{i1}, \dots, au_{iM_a}). \tag{6-5}$ <p>”, p. 159: “In addition to terms and authors, other types of information are available in many collections. Dates and</p>

Claim Text for '494 Patent	Fox Thesis, 1983																																																
	<p>controlled vocabulary terms may be properly separated from regular terms. Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed,” pp. 166-67: “Based on the reference pattern for a set of documents, one may define various derived measures of the interconnection between those documents. The relevant notation and definitions follow, using the data of Figure 6-3 to illustrate each point:</p> <p>(6-16) $A \rightarrow D$ Direct Reference when A refers to (cites) document D, so that D is referred to (cited by) A. By definition, $D \rightarrow D$ always holds.</p> <p>(6-17) $A \rightarrow^k G$ Indirect Reference when A indirectly refers to (cites) G (e.g., at distance $k=2$), so that G is indirectly referred to (cited by) A.</p> <p>,” p. 169: “A and D are linked if either $A \rightarrow D$ or $D \rightarrow A$ [Salton 1963]. This definition allows the computer to symmetrically view citation connections between documents, regardless of the ordering of the articles based on time of publication. More formally,</p> $m_{i,j} = \begin{cases} 1 & \text{if } D_i \rightarrow D_j \\ 1 & \text{if } D_j \rightarrow D_i \\ 1 & \text{if } i = j, \text{ by definition} \\ 0 & \text{otherwise.} \end{cases}$ <p>In the example, there are m_{ij} values of 1 for pairs such as A and D or C and G. ” Figure 6.5:</p> <p>Table 6.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)</p> <table border="1" data-bbox="919 1101 1430 1377"> <thead> <tr> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>→</td> <td>D</td> <td>E</td> <td>→</td> <td>G</td> </tr> <tr> <td>B</td> <td>→</td> <td>E</td> <td>E</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>E</td> <td>G</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>G</td> <td>H</td> <td>→</td> <td>E</td> </tr> <tr> <td>D</td> <td>→</td> <td>F</td> <td>H</td> <td>→</td> <td>G</td> </tr> <tr> <td>D</td> <td>→</td> <td>G</td> <td>I</td> <td>→</td> <td>G</td> </tr> <tr> <td>E</td> <td>→</td> <td>F</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.	A	→	D	E	→	G	B	→	E	E	→	J	C	→	E	G	→	J	C	→	G	H	→	E	D	→	F	H	→	G	D	→	G	I	→	G	E	→	F			
Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.																																												
A	→	D	E	→	G																																												
B	→	E	E	→	J																																												
C	→	E	G	→	J																																												
C	→	G	H	→	E																																												
D	→	F	H	→	G																																												
D	→	G	I	→	G																																												
E	→	F																																															

Claim Text for '494 Patent	Fox Thesis, 1983																																																																
	<p data-bbox="926 261 1192 285">Figure 6.5: I_n Submatrix</p> <table border="1" data-bbox="926 321 1341 586"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> </tr> </thead> <tbody> <tr> <th>A</th> <td>1</td> <td></td> <td></td> <td>1</td> <td></td> <td></td> <td></td> </tr> <tr> <th>B</th> <td></td> <td>1</td> <td></td> <td></td> <td>1</td> <td></td> <td></td> </tr> <tr> <th>C</th> <td></td> <td></td> <td>1</td> <td></td> <td>1</td> <td></td> <td>1</td> </tr> <tr> <th>D</th> <td>1</td> <td></td> <td></td> <td>1</td> <td></td> <td>1</td> <td>1</td> </tr> <tr> <th>E</th> <td></td> <td>1</td> <td>1</td> <td></td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <th>F</th> <td></td> <td></td> <td></td> <td>1</td> <td>1</td> <td>1</td> <td></td> </tr> <tr> <th>G</th> <td></td> <td></td> <td>1</td> <td>1</td> <td>1</td> <td></td> <td>1</td> </tr> </tbody> </table> <p data-bbox="863 613 1856 683">, p. 170 “$I_{ij} = 1$ if the i^{th} doc. refers to the j^{th}, or vice versa.”; <i>see also, e.g.</i>, Chapters 1, 7-9, Appendix C.</p>		A	B	C	D	E	F	G	A	1			1				B		1			1			C			1		1		1	D	1			1		1	1	E		1	1		1	1	1	F				1	1	1		G			1	1	1		1
	A	B	C	D	E	F	G																																																										
A	1			1																																																													
B		1			1																																																												
C			1		1		1																																																										
D	1			1		1	1																																																										
E		1	1		1	1	1																																																										
F				1	1	1																																																											
G			1	1	1		1																																																										
<p data-bbox="233 699 837 829">generating a second numerical representation using the first numerical representation, wherein the second numerical representation accounts for indirect relationships in the database;</p>	<p data-bbox="863 699 1856 862"><i>Id.</i> at Chapter 6 (<i>e.g.</i>, pp. 159-164, pp. 167-168: “B and C are bibliographically coupled if some document, say E, is referred to by both B and C. Here, a computer can count how many articles provide a coupling connection in a similar fashion to E - in Figure 6.2 there are no more - and define the degree of bibliographic coupling. thus, for arbitrary documents i and j,</p> $bc_{ij} = D' $ <p data-bbox="919 951 972 967">where</p> $D_k \in D' \Leftrightarrow D_i \rightarrow D_k \text{ and } D_j \rightarrow D_k$ <p data-bbox="919 1057 1423 1073">and D' is restricted to the document set of definition, <i>e.g.</i>, O.</p> <p data-bbox="863 1073 1856 1203">In the example of Figure 6.3, $bc_{B,C} = 1$ and $bc_{D,E} = 2$ since one document, E, is referred to by both B and C, while two documents, F and G, are each referred to by both D and E. Thus, B->E, C->E and D->F, E->F, D->G, E->G.”</p>																																																																

Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.
A	→	D	E	→	G
B	→	E	E	→	J
C	→	E	G	→	J
C	→	G	H	→	E
D	→	F	H	→	G
D	→	G	I	→	G
E	→	F			

Figure 6.3: b_c Submatrix

	A	B	C	D	E	F	G
A	1						
B		1	1				
C		1	2	1	1		
D			1	2	2		
E			1	2	3		
F						0	
G							1

Note: $b_{c_{E,G}} \neq 1$ since J is not $\in O$.

[Small 1973] , p. 168: "F and G are co-cited if some document, say D, refers to both of them in its bibliography. One can count the total number of articles that each refer to both F and G. For arbitrary documents i and j, the co-citation strength is then given by

$$cc_{ij} = |D''|$$

where

$$D'' \subseteq C,$$

the source set of documents considered, and

$$D_k \in D'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j.$$

Note that cc_{ii} is simply the number of articles that cite document i , that is, its citation count. That value can be used for normalizing other cc values or to gauge the importance of the given article. In the example, then, one observes that

$$cc_{E,G} = 2 \quad cc_{F,G} = 2 \quad cc_{F,J} = 1,$$

Figure 6.4: cc Submatrix

	A	B	C	D	E	F	G
A	0						
B		0					
C			0				
D				1			
E					3		2
F						2	2
G					2	2	5

Note: this includes the fact that H cites E,G when $cc_{E,G}$ is computed. The reason is that H is in the source set C for co-citations.

, p. 170:

Claim Text for '494 Patent	Fox Thesis, 1983
	<p>For each document it is straightforward using the definitions of the last section to determine values of the linkage, bibliographic coupling, and co-citation measures between that document and any other document. Rather than using a dictionary to provide concept numbers, the document numbers themselves can be used so that</p> $M_{bc} = M_{cc} = M_{ln} = N \quad (6-21)$ <p>and submatrices BC, CC, and LN will each be of size $N \times N$. Note that according to the definitions of the various measures all diagonal entries will be non-zero but in general the submatrices will be sparsely populated.</p> <p>To obtain some intuition as to the meaning of these submatrices, consider the subvectors \vec{bc}_i, \vec{cc}_i, and \vec{ln}_i for the i^{th} document. Diagonal entries are</p> $\begin{aligned} bc_{ii} &= \text{no. of references in bibliography of } i \\ cc_{ii} &= \text{no. of articles that refer to } i \\ ln_{ii} &= 1 \end{aligned} \quad (6-22)$ <p>where another way to understand cc_{ij} is to view it as the incoming citation count.</p> <p>Off diagonal entries show how the i^{th} document relates to other documents. Thus, the j^{th} column of each submatrix shows how documents relate to the j^{th} document - one in effect treats a document as a "bibliographic concept". Off diagonal values have the following significance:</p> $\begin{aligned} bc_{ij} &= \text{no. of articles referred to by both } i, j \\ cc_{ij} &= \text{no. of articles that each refer to both } i, j \\ ln_{ij} &= 1 \text{ if the } i^{\text{th}} \text{ doc. refers to the } j^{\text{th}}, \text{ or vice versa} \end{aligned} \quad (6-23)$ <p style="text-align: right;">, pp. 171-182, 205-206, p. 240</p> <p>(Figure 8.2, Sample computations of inner products); see also, <i>e.g.</i>, Chapters 1, & 6-9.</p>
storing the second numerical representation;	<p><i>Id.</i> at Chapter 6 (<i>e.g.</i>, pp. 159-164, pp. 167-168: "B and C are bibliographically coupled if some document, say E, is referred to by both B and C. Here, a computer can count how many articles provide a coupling connection in a similar fashion to E - in Figure 6.2 there are no more - and define the degree of bibliographic coupling. thus, for arbitrary documents i and j,</p> $bc_{ij} = D' $ <p>where</p> $D_k \in D' \Leftrightarrow D_i \rightarrow D_k \text{ and } D_j \rightarrow D_k$ <p>and D' is restricted to the document set of definition, <i>e.g.</i>, O.</p> <p style="text-align: right;">In the example of Figure 6.3,</p>

Claim Text for '494 Patent	Fox Thesis, 1983																																																																																																																
	<p data-bbox="865 250 1843 344">$bc_{B,C} = 1$ and $bc_{D,E} = 2$ since one document, E, is referred to by both B and C, while two documents, F and G, are each referred to by both D and E. Thus, B->E, C->E and D->F, E->F, D->G, E->G.”</p> <p data-bbox="926 386 1369 472">Table 6.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)</p> <table border="1" data-bbox="926 508 1430 781"> <thead> <tr> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>→</td> <td>D</td> <td>E</td> <td>→</td> <td>G</td> </tr> <tr> <td>B</td> <td>→</td> <td>E</td> <td>E</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>E</td> <td>G</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>G</td> <td>H</td> <td>→</td> <td>E</td> </tr> <tr> <td>D</td> <td>→</td> <td>F</td> <td>H</td> <td>→</td> <td>G</td> </tr> <tr> <td>D</td> <td>→</td> <td>G</td> <td>I</td> <td>→</td> <td>G</td> </tr> <tr> <td>E</td> <td>→</td> <td>F</td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p data-bbox="942 878 1257 906">Figure 6.3: bc Submatrix</p> <table border="1" data-bbox="947 951 1434 1260"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> </tr> </thead> <tbody> <tr> <th>A</th> <td>1</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>B</th> <td></td> <td>1</td> <td>1</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>C</th> <td></td> <td>1</td> <td>2</td> <td>1</td> <td>1</td> <td></td> <td></td> </tr> <tr> <th>D</th> <td></td> <td></td> <td>1</td> <td>2</td> <td>2</td> <td></td> <td></td> </tr> <tr> <th>E</th> <td></td> <td></td> <td>1</td> <td>2</td> <td>3</td> <td></td> <td></td> </tr> <tr> <th>F</th> <td></td> <td></td> <td></td> <td></td> <td></td> <td>0</td> <td></td> </tr> <tr> <th>G</th> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td>1</td> </tr> </tbody> </table> <p data-bbox="947 1271 1388 1300">Note: $bc_{E,G} \neq 1$ since J is not $\in O$.</p> <p data-bbox="865 1317 1843 1414">[Small 1973] if some document, say D, refers to both of them in its bibliography. One can count the total number of articles that each refer to both F and G. For , p. 168: “F and G are co-cited</p>	Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.	A	→	D	E	→	G	B	→	E	E	→	J	C	→	E	G	→	J	C	→	G	H	→	E	D	→	F	H	→	G	D	→	G	I	→	G	E	→	F					A	B	C	D	E	F	G	A	1							B		1	1					C		1	2	1	1			D			1	2	2			E			1	2	3			F						0		G							1
Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.																																																																																																												
A	→	D	E	→	G																																																																																																												
B	→	E	E	→	J																																																																																																												
C	→	E	G	→	J																																																																																																												
C	→	G	H	→	E																																																																																																												
D	→	F	H	→	G																																																																																																												
D	→	G	I	→	G																																																																																																												
E	→	F																																																																																																															
	A	B	C	D	E	F	G																																																																																																										
A	1																																																																																																																
B		1	1																																																																																																														
C		1	2	1	1																																																																																																												
D			1	2	2																																																																																																												
E			1	2	3																																																																																																												
F						0																																																																																																											
G							1																																																																																																										

Claim Text for '494 Patent

Fox Thesis, 1983

arbitrary documents i and j , the co-citation strength is then given by

$$cc_{ij} = |D''|$$

where

$$D'' \subseteq C,$$

the source set of documents considered, and

$$D_k \in D'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j.$$

Note that cc_{ii} is simply the number of articles that cite document i , that is, its citation count. That value can be used for normalizing other cc values or to gauge the importance of the given article. In the example, then, one observes that

$$cc_{E,G} = 2 \quad cc_{F,G} = 2 \quad cc_{F,I} = 1,$$

Figure 6.4: cc Submatrix

	A	B	C	D	E	F	G
A	0						
B		0					
C			0				
D				1			
E					3		2
F						2	2
G					2	2	5

Note: this includes the fact that H cites E,G when $cc_{E,G}$ is computed. The reason is that H is in the source set C for co-citations.

, p. 170:

Claim Text for '494 Patent	Fox Thesis, 1983
	<p>For each document it is straightforward using the definitions of the last section to determine values of the linkage, bibliographic coupling, and co-citation measures between that document and any other document. Rather than using a dictionary to provide concept numbers, the document numbers themselves can be used so that</p> $M_{bc} = M_{cc} = M_{ln} = N \quad (6-21)$ <p>and submatrices BC, CC, and LN will each be of size $N \times N$. Note that according to the definitions of the various measures all diagonal entries will be non-zero but in general the submatrices will be sparsely populated.</p> <p>To obtain some intuition as to the meaning of these submatrices, consider the subvectors \vec{bc}_i, \vec{cc}_i, and \vec{ln}_i for the i^{th} document. Diagonal entries are</p> $\begin{aligned} bc_{ii} &= \text{no. of references in bibliography of } i \\ cc_{ii} &= \text{no. of articles that refer to } i \\ ln_{ii} &= 1 \end{aligned} \quad (6-22)$ <p>where another way to understand cc_{ij} is to view it as the incoming citation count.</p> <p>Off diagonal entries show how the i^{th} document relates to other documents. Thus, the j^{th} column of each submatrix shows how documents relate to the j^{th} document - one in effect treats a document as a "bibliographic concept". Off diagonal values have the following significance:</p> $\begin{aligned} bc_{ij} &= \text{no. of articles referred to by both } i, j \\ cc_{ij} &= \text{no. of articles that each refer to both } i, j \\ ln_{ij} &= 1 \text{ if the } i^{th} \text{ doc. refers to the } j^{th}, \text{ or vice versa} \end{aligned} \quad (6-23)$ <p style="text-align: right;">, pp. 171-182, 205-206, p. 240</p> <p>(Figure 8.2, Sample computations of inner products); see also, <i>e.g.</i>, Chapters 1, & 6-9.</p>

Claim Text for '494 Patent	Fox Thesis, 1983
<p>identifying at least one object in the database, wherein the stored numerical representation is used to identify objects; and</p>	<p><i>Id.</i> at Chapter 1 (e.g., pp. 16-18, 19 (“The use of multiple concept types to generalize the vector representation of documents provides a second method for performance improvement. By including more information in the document representation and by judiciously utilizing that information through the relevance feedback cycle, improved retrieval can result.”), Chapter 5, Chapter 6, e.g., pp. 157-158, 159: “Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed . . . The bibliographic measures described here have been useful in both retrieval and clustering applications.”), 160-172, 173: “The use of <i>bc</i>, <i>cc</i>, and <i>ln</i> submatrices seems justified as an initial approach to better incorporating bibliographic data in the vector space model. Experiments in later chapters will contrast the utility of these measures and see how they can best be combined to aid retrieval system performance. The first requisite for such utilization, however, is an effective means to include the appropriate subvectors when computing similarities.”, 174-182), Chapter 7, Chapter 8; <i>see also</i>, e.g., Chapters 1, 7-9.</p>
<p>displaying one or more identified objects from the database.</p>	<p><i>Id.</i> at 6: “In addition to being able to locate documents of interest, the user may be able to retrieve and/or examine paragraphs, passages, sentences, or single word occurrences (in context).” p. 219: “Note that exactly 30 documents are shown to the user,” p. 326: “First, it should be noted that at Syracuse an entire search was carried out, where various sets were retrieved and eventually the results of one of the sets was selected for printing.” <i>See</i> chapters 5 and 8; <i>see also</i>, e.g., Chapters 1, & 6-9.</p>
<p>19. The method of claim 18 wherein the step of generating a second numerical representation comprises: selecting an object in the database for analysis;</p>	<p><i>Id.</i> at Chapter 1 (e.g., pp. 16-18, 19 (“The use of multiple concept types to generalize the vector representation of documents provides a second method for performance improvement. By including more information in the document representation and by judiciously utilizing that information through the relevance feedback cycle, improved retrieval can result.”), Chapter 5, Chapter 6, e.g., pp. 157-158, 159: “Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed . . . The bibliographic measures described here have been useful in both retrieval and clustering applications.”), 160-172, 173: “The use of <i>bc</i>, <i>cc</i>, and <i>ln</i> submatrices seems justified as an initial approach to better incorporating bibliographic data in the vector space model. Experiments in later chapters will contrast the utility of these measures and see how they can best be</p>

Claim Text for '494 Patent	Fox Thesis, 1983																																																
	<p>combined to aid retrieval system performance. The first requisite for such utilization, however, is an effective means to include the appropriate subvectors when computing similarities.”, 174-182), Chapter 7, Chapter 8; <i>see also, e.g.</i>, Chapters 1, 7-9.</p>																																																
<p>analyzing the direct relationships expressed by the first numerical representation for indirect relationships involving the selected object; and creating a second numerical representation of the direct and indirect relationships involving the selected object.</p>	<p><i>Id.</i> at Chapter 6 (<i>e.g.</i>, pp. 159-164, pp. 167-168: “B and C are bibliographically coupled if some document, say E, is referred to by both B and C. Here, a computer can count how many articles provide a coupling connection in a similar fashion to E - in Figure 6.2 there are no more - and define the degree of bibliographic coupling. thus, for arbitrary documents i and j,</p> $bc_{ij} = D' $ <p>where</p> $D_k \in D' \Leftrightarrow D_i \rightarrow D_k \text{ and } D_j \rightarrow D_k$ <p>and D' is restricted to the document set of definition, <i>e.g.</i>, O.</p> <p>In the example of Figure 6.3, $bc_{B,C} = 1$ and $bc_{D,E} = 2$ since one document, E, is referred to by both B and C, while two documents, F and G, are each referred to by both D and E. Thus, B->E, C->E and D->F, E->F, D->G, E->G.”,</p> <p>Table 6.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)</p> <table border="1" data-bbox="919 1057 1430 1333"> <thead> <tr> <th>Citing Doc.</th> <th></th> <th>Cited Doc.</th> <th>Citing Doc.</th> <th></th> <th>Cited Doc.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>→</td> <td>D</td> <td>E</td> <td>→</td> <td>G</td> </tr> <tr> <td>B</td> <td>→</td> <td>E</td> <td>E</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>E</td> <td>G</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>G</td> <td>H</td> <td>→</td> <td>E</td> </tr> <tr> <td>D</td> <td>→</td> <td>F</td> <td>H</td> <td>→</td> <td>G</td> </tr> <tr> <td>D</td> <td>→</td> <td>G</td> <td>I</td> <td>→</td> <td>G</td> </tr> <tr> <td>E</td> <td>→</td> <td>F</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Citing Doc.		Cited Doc.	Citing Doc.		Cited Doc.	A	→	D	E	→	G	B	→	E	E	→	J	C	→	E	G	→	J	C	→	G	H	→	E	D	→	F	H	→	G	D	→	G	I	→	G	E	→	F			
Citing Doc.		Cited Doc.	Citing Doc.		Cited Doc.																																												
A	→	D	E	→	G																																												
B	→	E	E	→	J																																												
C	→	E	G	→	J																																												
C	→	G	H	→	E																																												
D	→	F	H	→	G																																												
D	→	G	I	→	G																																												
E	→	F																																															

Figure 6.3: b_c Submatrix

	A	B	C	D	E	F	G
A	1						
B		1	1				
C		1	2	1	1		
D			1	2	2		
E			1	2	3		
F						0	
G							1

Note: $bc_{E,G} \neq 1$ since J is not $\in O$.

,” p. 168: “F and G are co-cited [Small 1973] if some document, say D, refers to both of them in its bibliography. One can count the total number of articles that each refer to both F and G. For arbitrary documents i and j, the co-citation strength is then given by

$$cc_{ij} = |D''|$$

where

$$D'' \subseteq C,$$

the source set of documents considered, and

$$D_k \in D'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j.$$

Note that cc_{ii} is simply the number of articles that cite document i, that is, its citation count. That value can be used for normalizing other cc values or to gauge the importance of the given article. In the example, then, one observes that $cc_{E,G} = 2$ $cc_{F,G} = 2$ $cc_{F,J} = 1$,”

Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.
A	→	D	E	→	G
B	→	E	E	→	J
C	→	E	G	→	J
C	→	G	H	→	E
D	→	F	H	→	G
D	→	G	I	→	G
E	→	F			

Figure 6.4: cc Submatrix

	A	B	C	D	E	F	G
A	0						
B		0					
C			0				
D				1			
E					3		2
F						2	2
G					2	2	5

Note: this includes the fact that H cites E,G when $cc_{E,G}$ is computed.
The reason is that H is in the source set C for co-citations.

,” p. 170:

Claim Text for '494 Patent	Fox Thesis, 1983
	<p>For each document it is straightforward using the definitions of the last section to determine values of the linkage, bibliographic coupling, and co-citation measures between that document and any other document. Rather than using a dictionary to provide concept numbers, the document numbers themselves can be used so that</p> $M_{bc} = M_{cc} = M_{ln} = N \quad (6-21)$ <p>and submatrices BC, CC, and LN will each be of size $N \times N$. Note that according to the definitions of the various measures all diagonal entries will be non-zero but in general the submatrices will be sparsely populated.</p> <p>To obtain some intuition as to the meaning of these submatrices, consider the subvectors bc_i, cc_i, and ln_i for the i^{th} document. Diagonal entries are</p> $\begin{aligned} bc_{ii} &= \text{no. of references in bibliography of } i \\ cc_{ii} &= \text{no. of articles that refer to } i \\ ln_{ii} &= 1 \end{aligned} \quad (6-22)$ <p>where another way to understand cc_{ii} is to view it as the incoming citation count.</p> <p>Off diagonal entries show how the i^{th} document relates to other documents. Thus, the j^{th} column of each submatrix shows how documents relate to the j^{th} document - one in effect treats a document as a "bibliographic concept". Off diagonal values have the following significance:</p> $\begin{aligned} bc_{ij} &= \text{no. of articles referred to by both } i, j \\ cc_{ij} &= \text{no. of articles that each refer to both } i, j \\ ln_{ij} &= 1 \text{ if the } i^{th} \text{ doc. refers to the } j^{th}, \text{ or vice versa} \end{aligned} \quad (6-23)$ <p>,” pp. 171-182, 205-206, p. 240 (Figure 8.2, Sample computations of inner products); <i>see also, e.g.</i>, Chapter 1, Chapters 6-9.</p>
<p>20. The method of 18 wherein the step of identifying at least one object in the database comprises: searching for objects in a database using the stored numerical representation, wherein direct and/or indirect relationships are searched.</p>	<p><i>Id.</i> at Chapter 1 (e.g., pp. 16-18, 19 (“The use of multiple concept types to generalize the vector representation of documents provides a second method for performance improvement. By including more information in the document representation and by judiciously utilizing that information through the relevance feedback cycle, improved retrieval can result.”), Chapter 5, Chapter 6, e.g., pp. 157-158, 159: “Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed . . . The bibliographic measures described here have been useful in both retrieval and clustering applications.”), 160-172, 173: “The use of bc, cc, and ln submatrices seems justified as an initial approach to better incorporating bibliographic data in the vector space model. Experiments in later chapters will contrast the utility of these measures and see how they can best be</p>

Claim Text for '494 Patent	Fox Thesis, 1983
	combined to aid retrieval system performance. The first requisite for such utilization, however, is an effective means to include the appropriate subvectors when computing similarities.”, 174-182), Chapter 7, Chapter 8; <i>see also, e.g.</i> , Chapters 1, 7-9.
21. The method of claim 18 wherein the displaying step comprises: generating a graphical display for representing an object in the database.	<i>Id.</i> at 6: “In addition to being able to locate documents of interest, the user may be able to retrieve and/or examine paragraphs, passages, sentences, or single word occurrences (in context).” p. 219: “Note that exactly 30 documents are shown to the user,” p. 326: “First, it should be noted that at Syracuse an entire search was carried out, where various sets were retrieved and eventually the results of one of the sets was selected for printing.” <i>See</i> chapters 5 and 8; <i>see also, e.g.</i> , Chapters 1, & 6-9.
23. A method of representing data in a computer database with relationships, comprising the steps of:	<i>See infra; see also, e.g.</i> , Chapters 1, & 6-9.
assigning nodes node identifications;	<i>Id.</i> at 153: “Consider a collection, <i>C</i> , containing <i>N</i> documents, that is processed by automatic indexing routines which first eliminate stop words and reduce remaining words to their respective stems,” p. 196:

Claim Text for '494 Patent

Fox Thesis, 1983

Table 7.2: Doc. Number, Year, Vol., No., Title, Author for Last 55 Articles

Did	Yr	Vo	No	Title (first part)	Author (first)
3150	79	07	01	Beyond Programming Languages	Winograd, T.
3151	79	07	02	An Optimal Real-Time Algorithm for Plana	Preparata, F.P.
3152	79	07	03	Storage Reorganization Techniques for Ma	Fischer, P.C.
3153	79	07	04	The Control of Response Times in Multi-C	Hine, J.H.
3154	79	07	05	Algorithm = Logic + Control	Kowalski, R.
3155	79	08	01	The Paradigms of Programming	Floyd, R.W.
3156	79	08	02	Computing Connected Components on Parall	Hirschberg, D.S.
3157	79	08	03	Proving Termination with Multiset Orderi	Dershowitz, N.
3158	79	08	04	Secure Personal Computing in an Insecure	Denning, D.E.
3159	79	08	05	Further Remark on Stably Updating Mean a	Nelson, L.S.
3160	79	09	01	Rejuvenating Experimental Computer Scien	Feldman, J.A.
3161	79	09	02	An ACM Executive Committee Position on t	McCracken, D.D.
3162	79	09	03	On Improving the Worst Case Running Time	Galil, Z.
3163	79	09	04	An Optimal Insertion Algorithm for One-S	Raiha, K.J.
3164	79	09	05	Progressive Acyclic Digraphs-A Tool for	Hansen, W.J.
3165	79	09	06	Approximation of Polygonal Maps by Cellu	Nagy, G.
3166	79	09	07	Computing Standard Deviations- Accuracy	Chan, T.F.
3167	79	09	08	Updating Mean and Variance Estimates- An	West, D.H.D.
3168	79	10	01	Comment on "An Optimal Evaluation of Boo	Laird, P.D.
3169	79	10	02	Note on "An Optimal Evaluation of Boolea	Gudes, E.
3170	79	10	03	On the Proof of Correctness of a Calenda	Lampert, L.
3171	79	10	04	Line Numbers Made Cheap	Klint, P.
3172	79	10	05	An Algorithm for Planning Collision-Free	Lozano-Perez, T.
3173	79	11	01	A Psychology of Learning BASIC	Mayer, R.E.
3174	79	11	02	Password Security- A Case History	Morris, R.
3175	79	11	03	Breaking Substitution Ciphers Using a Re	Peleg, S.
3176	79	11	04	Storing a Sparse Table	Tarjan, R.E.
3177	79	11	05	How to Share a Secret	Shamir, A.
3178	79	12	01	Introduction to the EFT Symposium	Kling, R.
3179	79	12	02	Overview of the EFT Symposium	Kraemer, K.L.
3180	79	12	03	Costs of the Current U.S. Payments Syste	Lipis, A.H.
3181	79	12	04	Public Protection and Education with EFT	Long, R.H.
3182	79	12	05	Vulnerabilities of EFTs to Intentionally	Parker, D.B.
3183	79	12	06	Policy, Values, and EFT Research- Anatom	Kraemer, K.L.

p. 198:

Claim Text for '494 Patent	Fox Thesis, 1983																																																																		
	<p data-bbox="894 269 1430 293">Table 7.3: \overline{CR} Subvector Information for Last 55 Articles</p> <table border="1" data-bbox="894 375 1157 1222"> <thead> <tr> <th data-bbox="894 375 961 423"><u>Doc. Id.</u></th> <th data-bbox="974 375 1157 423"><u>List of \overline{C} Category Concept Numbers</u></th> </tr> </thead> <tbody> <tr><td>3150</td><td>105 113 115 127</td></tr> <tr><td>3151</td><td>132 157 164</td></tr> <tr><td>3152</td><td>123 145 157</td></tr> <tr><td>3153</td><td>121 196</td></tr> <tr><td>3154</td><td>85 113 119 153 156</td></tr> <tr><td>3156</td><td>157 164 181</td></tr> <tr><td>3157</td><td>156 174</td></tr> <tr><td>3158</td><td>18 179</td></tr> <tr><td>3159</td><td>150 171</td></tr> <tr><td>3162</td><td>94 127 157</td></tr> <tr><td>3163</td><td>93 94 123 157 163</td></tr> <tr><td>3164</td><td>122 123 164</td></tr> <tr><td>3165</td><td>38 123 198</td></tr> <tr><td>3166</td><td>142 150 171</td></tr> <tr><td>3167</td><td>150 171</td></tr> <tr><td>3168</td><td>74 93 94</td></tr> <tr><td>3169</td><td>70 90 94</td></tr> <tr><td>3170</td><td>156</td></tr> <tr><td>3171</td><td>109 110 113 129</td></tr> <tr><td>3172</td><td>39 85 87 196</td></tr> <tr><td>3173</td><td>7 59 115</td></tr> <tr><td>3174</td><td>24 124</td></tr> <tr><td>3175</td><td>65 84</td></tr> <tr><td>3176</td><td>94 109 123 157</td></tr> <tr><td>3177</td><td>165 173</td></tr> <tr><td>3179</td><td>17 72 73 98</td></tr> <tr><td>3180</td><td>72</td></tr> <tr><td>3181</td><td>18</td></tr> <tr><td>3182</td><td>17 21 72 98</td></tr> <tr><td>3183</td><td>17 21 72 73 98</td></tr> <tr><td>3186</td><td>115 155 156</td></tr> <tr><td>3191</td><td>165</td></tr> </tbody> </table> <p data-bbox="863 1263 1812 1328"><i>See also, e.g., id.</i> at 27, 195, 203, 207, 211-13, 225-26, 229, 230, Tables 7.2, 7.9, Section 6.5.1.5 “Indexing,” Chapters 1, & 6-9.</p>	<u>Doc. Id.</u>	<u>List of \overline{C} Category Concept Numbers</u>	3150	105 113 115 127	3151	132 157 164	3152	123 145 157	3153	121 196	3154	85 113 119 153 156	3156	157 164 181	3157	156 174	3158	18 179	3159	150 171	3162	94 127 157	3163	93 94 123 157 163	3164	122 123 164	3165	38 123 198	3166	142 150 171	3167	150 171	3168	74 93 94	3169	70 90 94	3170	156	3171	109 110 113 129	3172	39 85 87 196	3173	7 59 115	3174	24 124	3175	65 84	3176	94 109 123 157	3177	165 173	3179	17 72 73 98	3180	72	3181	18	3182	17 21 72 98	3183	17 21 72 73 98	3186	115 155 156	3191	165
<u>Doc. Id.</u>	<u>List of \overline{C} Category Concept Numbers</u>																																																																		
3150	105 113 115 127																																																																		
3151	132 157 164																																																																		
3152	123 145 157																																																																		
3153	121 196																																																																		
3154	85 113 119 153 156																																																																		
3156	157 164 181																																																																		
3157	156 174																																																																		
3158	18 179																																																																		
3159	150 171																																																																		
3162	94 127 157																																																																		
3163	93 94 123 157 163																																																																		
3164	122 123 164																																																																		
3165	38 123 198																																																																		
3166	142 150 171																																																																		
3167	150 171																																																																		
3168	74 93 94																																																																		
3169	70 90 94																																																																		
3170	156																																																																		
3171	109 110 113 129																																																																		
3172	39 85 87 196																																																																		
3173	7 59 115																																																																		
3174	24 124																																																																		
3175	65 84																																																																		
3176	94 109 123 157																																																																		
3177	165 173																																																																		
3179	17 72 73 98																																																																		
3180	72																																																																		
3181	18																																																																		
3182	17 21 72 98																																																																		
3183	17 21 72 73 98																																																																		
3186	115 155 156																																																																		
3191	165																																																																		
generating links, wherein each link represents a relationship between two nodes and is identified by	<i>Id.</i> at 159: “In addition to terms and authors, other types of information are available in many collections. Dates and controlled vocabulary terms may be properly																																																																		

Claim Text for '494 Patent	Fox Thesis, 1983																																																
<p>the two nodes in which the relationship exists;</p>	<p>separated from regular terms. Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed,” pp. 166-167: “Based on the reference pattern for a set of documents, one may define various derived measures of the interconnection between those documents. The relevant notation and definitions follow, using the data of Figure 6-3 to illustrate each point:</p> <p>(6-16) $A \rightarrow D$ Direct Reference when A refers to (cites) document D, so that D is referred to (cited by) A. By definition, $D \rightarrow D$ always holds.</p> <p>(6-17) $A \rightarrow^k G$ Indirect Reference when A indirectly refers to (cites) G (e.g., at distance $k=2$), so that G is indirectly referred to (cited by) A.</p> <p>,” p. 169: “A and D are linked if either $A \rightarrow D$ or $D \rightarrow A$ [Salton 1963]. This definition allows the computer to symmetrically view citation connections between documents, regardless of the ordering of the articles based on time of publication. More formally,</p> $ln_{i,j} = \begin{cases} 1 & \text{if } D_i \rightarrow D_j \\ 1 & \text{if } D_j \rightarrow D_i \\ 1 & \text{if } i = j, \text{ by definition} \\ 0 & \text{otherwise.} \end{cases}$ <p>In the example, there are ln_{ij} values of 1 for pairs such as A and D or C and G.</p> <p>,” Figure 6.5:</p> <p>Table 6.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)</p> <table border="1" data-bbox="919 1068 1432 1344"> <thead> <tr> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>→</td> <td>D</td> <td>E</td> <td>→</td> <td>G</td> </tr> <tr> <td>B</td> <td>→</td> <td>E</td> <td>E</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>E</td> <td>G</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>G</td> <td>H</td> <td>→</td> <td>E</td> </tr> <tr> <td>D</td> <td>→</td> <td>F</td> <td>H</td> <td>→</td> <td>G</td> </tr> <tr> <td>D</td> <td>→</td> <td>G</td> <td>I</td> <td>→</td> <td>G</td> </tr> <tr> <td>E</td> <td>→</td> <td>F</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.	A	→	D	E	→	G	B	→	E	E	→	J	C	→	E	G	→	J	C	→	G	H	→	E	D	→	F	H	→	G	D	→	G	I	→	G	E	→	F			
Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.																																												
A	→	D	E	→	G																																												
B	→	E	E	→	J																																												
C	→	E	G	→	J																																												
C	→	G	H	→	E																																												
D	→	F	H	→	G																																												
D	→	G	I	→	G																																												
E	→	F																																															

Claim Text for '494 Patent	Fox Thesis, 1983																																																																
	<p data-bbox="926 261 1192 289">Figure 6.5: <i>ln</i> Submatrix</p> <table border="1" data-bbox="926 321 1341 586"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> </tr> </thead> <tbody> <tr> <th>A</th> <td>1</td> <td></td> <td></td> <td>1</td> <td></td> <td></td> <td></td> </tr> <tr> <th>B</th> <td></td> <td>1</td> <td></td> <td></td> <td>1</td> <td></td> <td></td> </tr> <tr> <th>C</th> <td></td> <td></td> <td>1</td> <td></td> <td>1</td> <td></td> <td>1</td> </tr> <tr> <th>D</th> <td>1</td> <td></td> <td></td> <td>1</td> <td></td> <td>1</td> <td>1</td> </tr> <tr> <th>E</th> <td></td> <td>1</td> <td>1</td> <td></td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <th>F</th> <td></td> <td></td> <td></td> <td>1</td> <td>1</td> <td>1</td> <td></td> </tr> <tr> <th>G</th> <td></td> <td></td> <td>1</td> <td>1</td> <td>1</td> <td></td> <td>1</td> </tr> </tbody> </table> <p data-bbox="863 613 1856 683">, p. 170 “$ln_{ij} = 1$ if the i^{th} doc. refers to the j^{th}, or vice versa.”; <i>see also, e.g.</i>, Chapters 1, & 6-9 & Appendix C.</p>		A	B	C	D	E	F	G	A	1			1				B		1			1			C			1		1		1	D	1			1		1	1	E		1	1		1	1	1	F				1	1	1		G			1	1	1		1
	A	B	C	D	E	F	G																																																										
A	1			1																																																													
B		1			1																																																												
C			1		1		1																																																										
D	1			1		1	1																																																										
E		1	1		1	1	1																																																										
F				1	1	1																																																											
G			1	1	1		1																																																										
<p data-bbox="233 699 842 829">allocating a weight to each link, wherein the weight signifies the strength of the relationship represented by the link relative to the strength of other relationships represented by other links; and</p>	<p data-bbox="863 699 1856 976"><i>Id.</i> at 158: “Incidentally, the various subvectors could be construed using different weighting schemes; an additional column in Table 6.1 could show that, for example, term weights were computed using the scheme $tf*idf$ while author entries were given binary weights,” p. 168: “Now, citing directly as given in (6-16) or indirectly as in (6-17) are binary events -- either they occur or not. On the other hand, the next two definitions can result in an assignment of weights that are based upon integer counts. (6-18) B and C are bibliographically coupled [Kessler 1962] if some document, say E, is referred to by both B and C.</p> <p data-bbox="863 987 1856 1284">Hence a computer can count how many articles provide a coupling connection in a similar fashion to E -- in Figure 6.2 there are no more -- and define the degree of bibliographic coupling,” p. 179: “Weighting methods may vary for different subvectors. Dates should undoubtedly receive binary weights, whereas terms benefit from applying an inverse document frequency (idf) factor. Bibliographic submatrices should also use some type of weighting,” p. 168: “F and G are co-cited [Small 1973] if some document, say D, refers to both of them in its bibliography. One can count the total number of articles that each refer to both F and G. For arbitrary documents i and j, the co-citation strength is then given by</p>																																																																

Claim Text for '494 Patent	Fox Thesis, 1983																																																
	<p data-bbox="976 267 1144 300">$cc_{ij} = D''$</p> <p data-bbox="934 341 997 373">where</p> <p data-bbox="976 406 1081 438">$D'' \subseteq C,$</p> <p data-bbox="934 479 1407 511">the source set of documents considered, and</p> <p data-bbox="976 544 1365 576">$D_k \in D'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j.$</p> <p data-bbox="861 600 1858 706">Note that cc_{ii} is simply the number of articles that cite document i, that is, its citation count. That value can be used for normalizing other cc values or to gauge the importance of the given article. In the example, then, one observes that</p> <p data-bbox="861 714 1186 747">$cc_{E,G} = 2 \quad cc_{F,G} = 2 \quad cc_{E,J} = 1,$</p> <p data-bbox="924 787 1375 876">Table 8.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)</p> <table border="1" data-bbox="924 909 1438 1193"> <thead> <tr> <th>Citing Doc.</th> <th></th> <th>Cited Doc.</th> <th>Citing Doc.</th> <th></th> <th>Cited Doc.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>→</td> <td>D</td> <td>E</td> <td>→</td> <td>G</td> </tr> <tr> <td>B</td> <td>→</td> <td>E</td> <td>E</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>E</td> <td>G</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>G</td> <td>H</td> <td>→</td> <td>E</td> </tr> <tr> <td>D</td> <td>→</td> <td>F</td> <td>H</td> <td>→</td> <td>G</td> </tr> <tr> <td>D</td> <td>→</td> <td>G</td> <td>I</td> <td>→</td> <td>G</td> </tr> <tr> <td>E</td> <td>→</td> <td>F</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Citing Doc.		Cited Doc.	Citing Doc.		Cited Doc.	A	→	D	E	→	G	B	→	E	E	→	J	C	→	E	G	→	J	C	→	G	H	→	E	D	→	F	H	→	G	D	→	G	I	→	G	E	→	F			
Citing Doc.		Cited Doc.	Citing Doc.		Cited Doc.																																												
A	→	D	E	→	G																																												
B	→	E	E	→	J																																												
C	→	E	G	→	J																																												
C	→	G	H	→	E																																												
D	→	F	H	→	G																																												
D	→	G	I	→	G																																												
E	→	F																																															

Claim Text for '494 Patent	Fox Thesis, 1983																																																																
	<p data-bbox="905 266 1115 285">Figure 6.4: cc Submatrix</p> <table border="1" data-bbox="905 315 1228 521"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> </tr> </thead> <tbody> <tr> <th>A</th> <td>0</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>B</th> <td></td> <td>0</td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>C</th> <td></td> <td></td> <td>0</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <th>D</th> <td></td> <td></td> <td></td> <td>1</td> <td></td> <td></td> <td></td> </tr> <tr> <th>E</th> <td></td> <td></td> <td></td> <td></td> <td>3</td> <td></td> <td>2</td> </tr> <tr> <th>F</th> <td></td> <td></td> <td></td> <td></td> <td></td> <td>2</td> <td>2</td> </tr> <tr> <th>G</th> <td></td> <td></td> <td></td> <td></td> <td></td> <td>2</td> <td>5</td> </tr> </tbody> </table> <p data-bbox="905 529 1478 570">Note: this includes the fact that H cites E,G when $cc_{E,G}$ is computed. The reason is that H is in the source set C for co-citations.</p> <p data-bbox="865 618 1251 646"><i>See also, e.g.,</i> Chapters 1, & 6-9.</p>		A	B	C	D	E	F	G	A	0							B		0						C			0					D				1				E					3		2	F						2	2	G						2	5
	A	B	C	D	E	F	G																																																										
A	0																																																																
B		0																																																															
C			0																																																														
D				1																																																													
E					3		2																																																										
F						2	2																																																										
G						2	5																																																										
displaying a node identification.	<p data-bbox="865 662 1866 894"><i>Id.</i> at 6: “In addition to being able to locate documents of interest, the user may be able to retrieve and/or examine paragraphs, passages, sentences, or single word occurrences (in context).” p. 219: “Note that exactly 30 documents are shown to the user,” p. 326: “First, it should be noted that at Syracuse an entire search was carried out, where various sets were retrieved and eventually the results of one of the sets was selected for printing”; <i>see also</i> chapters 5 and 8; <i>see also, e.g.,</i> Chapters 1, & 6-9.</p>																																																																
24. The method of claim 23, wherein the data in the database is objects, wherein the nodes represent objects and each object is assigned a node identification, and wherein the relationships that exist comprise direct relationships between objects, further comprising the step of: searching generated links, wherein nodes are located by searching the generated links.	<p data-bbox="865 911 1866 1179"><i>See</i> Sections 7.1.3 (“Clustering with Bibliographic Data”), 7.2 (“Algorithms”) p. 192: “Having previously attempted clustering with bibliographic coupling data, Schiminovich [1971] developed what was termed a ‘pattern discovery algorithm’ to directly utilize links between documents. Afterwards, Bichteler and Parsons [1974] modified that method for document retrieval Later, Bichteler and Eaton [1980] demonstrated that for retrieval purposes using a similarity formula combining bibliographic coupling and co-citations was better than if bibliographic coupling alone was included,” Section 7.2.2 (“Searching”); <i>see also, e.g.,</i> Chapters 1, & 6-9.</p>																																																																
25. The method of claim 23 further comprising the step of: generating link sub-types, comprising the steps of:	<p data-bbox="865 1203 1866 1396"><i>Id.</i> at 214; <i>see also, e.g.,</i> Chapters 1, & 6-9, Chart for Claim 1, <i>supra</i> (including the quotations and descriptions set forth therein, which are incorporated by reference herein); <i>see also</i>, p. 182 (“the CACM collection used seven different concept types, including ones based on textual terms (tm), ones of factual information (au, bi), ones derived from bibliographic references (bc, cc, and ln), and one based on indexer interpretation (cr).”).</p>																																																																

Claim Text for '494 Patent	Fox Thesis, 1983
identifying each link sub-type with a name; and	<i>Id.</i> at 214; <i>see also, e.g.</i> , Chapters 1, & 6-9, Chart for Claim 1, <i>supra</i> (including the quotations and descriptions set forth therein, which are incorporated by reference herein); <i>see also</i> , p. 182 (“the CACM collection used seven different concept types, including ones based on textual terms (tm), ones of factual information (au, bi), ones derived from bibliographic references (bc, cc, and ln), and one based on indexer interpretation (cr).”).
Providing a comment to one or more link subtypes.	<i>Id.</i> at 214; <i>see also, e.g.</i> , Chapters 1, & 6-9, Chart for Claim 1, <i>supra</i> (including the quotations and descriptions set forth therein, which are incorporated by reference herein); <i>see also</i> , p. 182 (“the CACM collection used seven different concept types, including ones based on textual terms (tm), ones of factual information (au, bi), ones derived from bibliographic references (bc, cc, and ln), and one based on indexer interpretation (cr).”).
31. The method of claim 23 wherein attributes are assigned to nodes.	<i>Id.</i> at 19: “The use of multiple concept types to generalize the vector representation of documents provides a second method for performance improvement” p. 153: “Consequently, Sections 6.1 through 6.4 describe a new extended model, demonstrating in a step-by-step fashion how additional types of concepts can be added to the usual terms only vector,” pp. 154-158 (“Additional Information -- Authors”); <i>see also, e.g.</i> , Chapters 1, & 6-9.
32. The method of claim 31 further comprising the step of: generating node sub-types wherein the node sub-types are assigned information.	<i>Id.</i> at 19: “The use of multiple concept types to generalize the vector representation of documents provides a second method for performance improvement” p. 153: “Consequently, Sections 6.1 through 6.4 describe a new extended model, demonstrating in a step-by-step fashion how additional types of concepts can be added to the usual terms only vector,” pp. 154-158 (“Additional Information -- Authors”); <i>see also, e.g.</i> , Chapters 1, & 6-9.
33. A method of representing data in a computer database and for computerized searching of the data, wherein relationships exist in the database, comprising:	<i>See infra</i> ; <i>see also, e.g.</i> , Chapters 1, & 6-9.
assigning links to represent relationships in the database;	<i>Id.</i> at Chapter 6 (<i>e.g.</i> , p. 155: “it seems to be practically and conceptually better to more clearly separate the extended vector into two subvectors. Representing the term subvector for the i^{th} subvector as tm_i , and the author subvector as au_i , the i^{th} document is described as

Claim Text for '494 Patent	Fox Thesis, 1983
	<p> $\vec{D}_i' = (tm_i, \vec{a}u_i). \tag{6-4}$ </p> <p>Expanded, the subvectors have the equivalent form</p> <p> $\vec{D}_i' = (tm_{i1}, \dots, tm_{iM_m}, au_{i1}, \dots, au_{iM_u}). \tag{6-5}$ </p> <p>” p. 159: “In addition to terms and authors, other types of information are available in many collections. Dates and controlled vocabulary terms may be properly separated from regular terms. Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed,” pp. 166-67: “Based on the reference pattern for a set of documents, one may define various derived measures of the interconnection between those documents. The relevant notation and definitions follow, using the data of Figure 6-3 to illustrate each point:</p> <p>(6-16) $A \rightarrow D$ Direct Reference when A refers to (cites) document D, so that D is referred to (cited by) A. By definition, $D \rightarrow D$ always holds.</p> <p>(6-17) $A \rightarrow^k G$ Indirect Reference when A indirectly refers to (cites) G (e.g., at distance $k=2$), so that G is indirectly referred to (cited by) A.</p> <p>” p. 169: “A and D are linked if either $A \rightarrow D$ or $D \rightarrow A$ [Salton 1963]. This definition allows the computer to symmetrically view citation connections between documents, regardless of the ordering of the articles based on time of publication. More formally,</p> $ln_{i,j} = \begin{cases} 1 & \text{if } D_i \rightarrow D_j \\ 1 & \text{if } D_j \rightarrow D_i \\ 1 & \text{if } i = j, \text{ by definition} \\ 0 & \text{otherwise.} \end{cases}$ <p>In the example, there are ln_{ij} values of 1 for pairs such as A and D or C and G.</p> <p>” Figure 6.5:</p>

Claim Text for '494 Patent

Fox Thesis, 1983

Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.
A	→	D	E	→	G
B	→	E	E	→	J
C	→	E	G	→	J
C	→	G	H	→	E
D	→	F	H	→	G
D	→	G	I	→	G
E	→	F			

Figure 6.5: ln Submatrix

	A	B	C	D	E	F	G
A	1			1			
B		1			1		
C			1		1		1
D	1			1		1	1
E		1	1		1	1	1
F				1	1	1	
G			1	1	1		1

, p. 170 " $ln_{ij} = 1$ if the i^{th} doc. refers to the j^{th} , or vice versa."); *see also, e.g.*, Chapters 1, & 6-9 & Appendix C.

generating node identifications based upon the assigned links, wherein node identifications are generated so that each link represents a relationship between two identified nodes;

Id. at Chapter 6 (*e.g.*, p. 155: "it seems to be practically and conceptually better to more clearly separate the extended vector into two subvectors. Representing the term subvector for the i^{th} subvector as tm_i , and the author subvector as au_i , the i^{th} document is described as

Claim Text for '494 Patent	Fox Thesis, 1983
	<p> $\vec{D}_i' = (tm_i, \vec{a}u_i). \tag{6-4}$ </p> <p>Expanded, the subvectors have the equivalent form</p> <p> $\vec{D}_i' = (tm_{i1}, \dots, tm_{iM_m}, au_{i1}, \dots, au_{iM_u}). \tag{6-5}$ </p> <p> ,” p. 159: “In addition to terms and authors, other types of information are available in many collections. Dates and controlled vocabulary terms may be properly separated from regular terms. Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed,” pp. 166-67: “Based on the reference pattern for a set of documents, one may define various derived measures of the interconnection between those documents. The relevant notation and definitions follow, using the data of Figure 6-3 to illustrate each point: </p> <p> (6-16) $A \rightarrow D$ Direct Reference when A refers to (cites) document D, so that D is referred to (cited by) A. By definition, $D \rightarrow D$ always holds. </p> <p> (6-17) $A \rightarrow^k G$ Indirect Reference when A indirectly refers to (cites) G (e.g., at distance $k=2$), so that G is indirectly referred to (cited by) A. </p> <p> ,” p. 169: “A and D are linked if either $A \rightarrow D$ or $D \rightarrow A$ [Salton 1963]. This definition allows the computer to symmetrically view citation connections between documents, regardless of the ordering of the articles based on time of publication. More formally, </p> $ln_{i,j} = \begin{cases} 1 & \text{if } D_i \rightarrow D_j \\ 1 & \text{if } D_j \rightarrow D_i \\ 1 & \text{if } i = j, \text{ by definition} \\ 0 & \text{otherwise.} \end{cases}$ <p> In the example, there are ln_{ij} values of 1 for pairs such as A and D or C and G. </p> <p> ,” Figure 6.5: </p>

Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.
A	→	D	E	→	G
B	→	E	E	→	J
C	→	E	G	→	J
C	→	G	H	→	E
D	→	F	H	→	G
D	→	G	I	→	G
E	→	F			

Figure 6.5: ln Submatrix

	A	B	C	D	E	F	G
A	1			1			
B		1			1		
C			1		1		1
D	1			1		1	1
E		1	1		1	1	1
F				1	1	1	
G			1	1	1		1

, p. 170 " $ln_{ij} = 1$ if the i^{th} doc. refers to the j^{th} , or vice versa."); *see also, e.g.*, Chapters 1, & 6-9 & Appendix C.

storing the links and node identifications, wherein the links and nodes may be retrieved;

Id. at Chapter 6 (*e.g.*, p. 155: "it seems to be practically and conceptually better to more clearly separate the extended vector into two subvectors. Representing the term subvector for the i^{th} subvector as tm_i , and the author subvector as au_i , the i^{th} document is described as

Claim Text for '494 Patent	Fox Thesis, 1983
	<p> $\vec{D}_i' = (tm_i, \vec{a}u_i). \tag{6-4}$ </p> <p>Expanded, the subvectors have the equivalent form</p> <p> $\vec{D}_i' = (tm_{i1}, \dots, tm_{iM_m}, au_{i1}, \dots, au_{iM_u}). \tag{6-5}$ </p> <p>,” p. 159: “In addition to terms and authors, other types of information are available in many collections. Dates and controlled vocabulary terms may be properly separated from regular terms. Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed,” pp. 166-67: “Based on the reference pattern for a set of documents, one may define various derived measures of the interconnection between those documents. The relevant notation and definitions follow, using the data of Figure 6-3 to illustrate each point:</p> <p>(6-16) $A \rightarrow D$ Direct Reference when A refers to (cites) document D, so that D is referred to (cited by) A. By definition, $D \rightarrow D$ always holds.</p> <p>(6-17) $A \rightarrow^k G$ Indirect Reference when A indirectly refers to (cites) G (e.g., at distance $k=2$), so that G is indirectly referred to (cited by) A. ” p. 169: “A and D are linked if either $A \rightarrow D$ or $D \rightarrow A$ [Salton 1963]. This definition allows the computer to symmetrically view citation connections between documents, regardless of the ordering of the articles based on time of publication. More formally,</p> $ln_{i,j} = \begin{cases} 1 & \text{if } D_i \rightarrow D_j \\ 1 & \text{if } D_j \rightarrow D_i \\ 1 & \text{if } i = j, \text{ by definition} \\ 0 & \text{otherwise.} \end{cases}$ <p>In the example, there are ln_{ij} values of 1 for pairs such as A and D or C and G. ” Figure 6.5:</p>

Claim Text for '494 Patent	Fox Thesis, 1983																																																																																																																
	<p data-bbox="926 277 1371 363">Table 6.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)</p> <table border="1" data-bbox="926 402 1430 675"> <thead> <tr> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> <th>Citing Doc.</th> <th>→</th> <th>Cited Doc.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>→</td> <td>D</td> <td>E</td> <td>→</td> <td>G</td> </tr> <tr> <td>B</td> <td>→</td> <td>E</td> <td>E</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>E</td> <td>G</td> <td>→</td> <td>J</td> </tr> <tr> <td>C</td> <td>→</td> <td>G</td> <td>H</td> <td>→</td> <td>E</td> </tr> <tr> <td>D</td> <td>→</td> <td>F</td> <td>H</td> <td>→</td> <td>G</td> </tr> <tr> <td>D</td> <td>→</td> <td>G</td> <td>I</td> <td>→</td> <td>G</td> </tr> <tr> <td>E</td> <td>→</td> <td>F</td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p data-bbox="926 721 1192 745">Figure 6.5: <i>ln</i> Submatrix</p> <table border="1" data-bbox="926 781 1341 1045"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C</th> <th>D</th> <th>E</th> <th>F</th> <th>G</th> </tr> </thead> <tbody> <tr> <th>A</th> <td>1</td> <td></td> <td></td> <td>1</td> <td></td> <td></td> <td></td> </tr> <tr> <th>B</th> <td></td> <td>1</td> <td></td> <td></td> <td>1</td> <td></td> <td></td> </tr> <tr> <th>C</th> <td></td> <td></td> <td>1</td> <td></td> <td>1</td> <td></td> <td>1</td> </tr> <tr> <th>D</th> <td>1</td> <td></td> <td></td> <td>1</td> <td></td> <td>1</td> <td>1</td> </tr> <tr> <th>E</th> <td></td> <td>1</td> <td>1</td> <td></td> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <th>F</th> <td></td> <td></td> <td></td> <td>1</td> <td>1</td> <td>1</td> <td></td> </tr> <tr> <th>G</th> <td></td> <td></td> <td>1</td> <td>1</td> <td>1</td> <td></td> <td>1</td> </tr> </tbody> </table> <p data-bbox="863 1073 1864 1138">, p. 170 “$ln_{ij} = 1$ if the i^{th} doc. refers to the j^{th}, or vice versa.”); <i>see also, e.g.</i>, Chapters 1, & 6-9 & Appendix C.</p>	Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.	A	→	D	E	→	G	B	→	E	E	→	J	C	→	E	G	→	J	C	→	G	H	→	E	D	→	F	H	→	G	D	→	G	I	→	G	E	→	F					A	B	C	D	E	F	G	A	1			1				B		1			1			C			1		1		1	D	1			1		1	1	E		1	1		1	1	1	F				1	1	1		G			1	1	1		1
Citing Doc.	→	Cited Doc.	Citing Doc.	→	Cited Doc.																																																																																																												
A	→	D	E	→	G																																																																																																												
B	→	E	E	→	J																																																																																																												
C	→	E	G	→	J																																																																																																												
C	→	G	H	→	E																																																																																																												
D	→	F	H	→	G																																																																																																												
D	→	G	I	→	G																																																																																																												
E	→	F																																																																																																															
	A	B	C	D	E	F	G																																																																																																										
A	1			1																																																																																																													
B		1			1																																																																																																												
C			1		1		1																																																																																																										
D	1			1		1	1																																																																																																										
E		1	1		1	1	1																																																																																																										
F				1	1	1																																																																																																											
G			1	1	1		1																																																																																																										
<p data-bbox="233 1157 825 1219">searching for node identifications using the stored links; and</p>	<p data-bbox="863 1157 1854 1422"><i>Id.</i> at Chapter 6, <i>e.g.</i>, pp. 157-158, p. 159: “Of more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed . . . The bibliographic measures described here have been useful in both retrieval and clustering applications.”), pp. 160-172, 173: “The use of <i>bc</i>, <i>cc</i>, and <i>ln</i> submatrices seems justified as an initial approach to better incorporating bibliographic data in the vector space model. Experiments in later chapters will contrast the utility of these measures and see how they can best be combined to aid</p>																																																																																																																

Claim Text for '494 Patent	Fox Thesis, 1983
	retrieval system performance. The first requisite for such utilization, however, is an effective means to include the appropriate subvectors when computing similarities,” pp. 174-182, Chapter 7, Chapter 8; <i>see also, e.g.</i> , Chapters 1,6 & 9.
displaying node identifications, wherein the displayed node identifications are located in the searching step.	<i>Id.</i> at 6: “In addition to being able to locate documents of interest, the user may be able to retrieve and/or examine paragraphs, passages, sentences, or single word occurrences (in context).,” p. 219: “Note that exactly 30 documents are shown to the user,” p. 326: “First, it should be noted that at Syracuse an entire search was carried out, where various sets were retrieved and eventually the results of one of the sets was selected for printing”; <i>see also</i> Chapters 5 and 8; <i>see also, e.g.</i> , Chapters 1, & 6-9.

Defendants reserve the right to revise this contention chart concerning the invalidity of the asserted claims, as appropriate, for example depending upon the Court’s construction of the asserted claims, any findings as to the priority date of the asserted claims, and/or positions that Plaintiff or its expert witness(es) may take concerning claim interpretation, construction, infringement, and/or invalidity issues.

Plaintiff’s Infringement Contentions are based on an apparent construction of the claim terms. Defendants disagree with these apparent constructions. Nothing stated herein shall be treated as an admission or suggestion that Defendants agree with Plaintiff regarding either the scope of any of the asserted claims or the claim constructions advanced by Plaintiff in its Infringement Contentions or anywhere else, or that any of Defendants’ accused technology meets any limitations of the claims. Nothing stated herein shall be construed as an admission or a waiver of any particular construction of any claim term. Defendants also reserve all their rights to challenge any of the claim terms herein under 35 U.S.C. § 112, including by arguing that they are indefinite, not supported by the written description and/or not enabled. Accordingly, nothing stated herein shall be construed as a waiver of any argument available under 35 U.S.C. § 112.