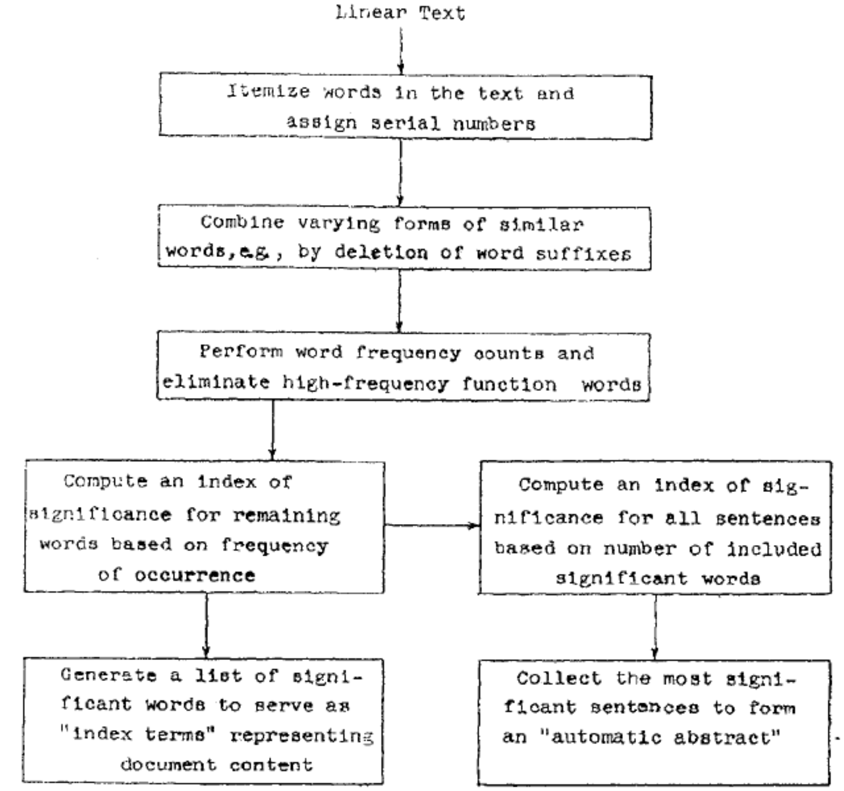
EXHIBIT L

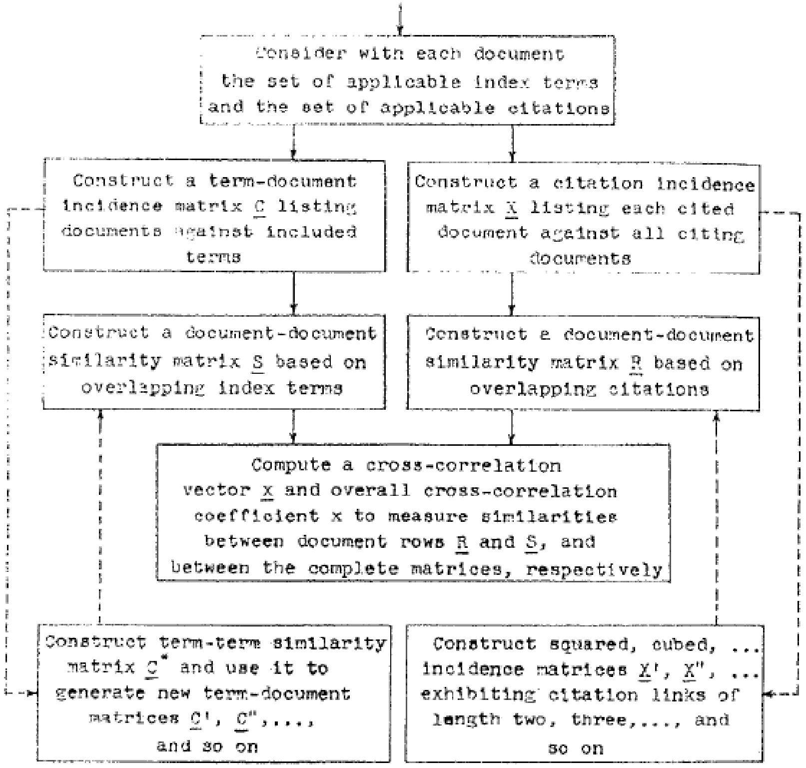
INVALIDITY CLAIM CHART FOR U.S. PATENT NO. 5,544,352

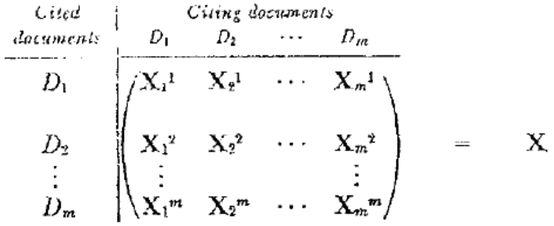
BASED ON SALTON, G., "ASSOCIATIVE DOCUMENT RETRIEVAL TECHNIQUES USING BIBLIOGRAPHIC INFORMATION," PP. 440-57 (1963). ("SALTON, 1963")

Claim Text from '352 Patent	Salton, 1963
26. A non-semantical 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:	*See, e.g.,* Salton, 1963, at Abstract, pp. 443, 446 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) 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) 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)
[26a] Marking objects in the database so that each marked object may be individually identified by a computerized search;	*See, e.g.,* Salton, 1963, at pp. 441, 447 Figure 1

Claim Text from '352 Patent	Salton, 1963
	 FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts. (Salton, 1963, p. 441) 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 Xi , where Xji = 1 if and only if document i is cited by document j, and Xji = 1 otherwise. (Salton, 1963, p. 447)
[26b] creating a first numerical representation for	*See, e.g.,* Salton, 1963, at pp. 446 n.1, 447, 450

Claim Text from ‘352 Patent	Salton, 1963
each identified object in the database based upon the object’s direct relationship with other objects in the database;	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 Xi, where Xij = 1 if and only if document i is cited by document j, and Xij = 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. $$\begin{array}{c\|c} \text{Cited documents} & \begin{array}{cccc} \multicolumn{4}{c}{\text{Citing documents}} \\ D_1 & D_2 & \cdots & D_m \end{array} \\ \hline \begin{array}{c} D_1 \\ \\ D_2 \\ \vdots \\ D_m \end{array} & \begin{pmatrix} \mathbf{X}_1{}^1 & \mathbf{X}_2{}^1 & \cdots & \mathbf{X}_m{}^1 \\ \\ \mathbf{X}_1{}^2 & \mathbf{X}_2{}^2 & \cdots & \mathbf{X}_m{}^2 \\ \vdots & & & \vdots \\ \mathbf{X}_1{}^m & \mathbf{X}_2{}^m & \cdots & \mathbf{X}_m{}^m \end{pmatrix} \end{array} = \mathbf{X}$$ $(\mathbf{X}_j{}^i = 1 \leftrightarrow$ 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
	 FIG. 5. Comparison of citation similarities with index term similarities (Salton, 1963, p. 450)
[26c] storing the first numerical representations for use in computerized searching;	*See, e.g.*, Salton, 1963, at pp. 446 n.1, 447, 450 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)

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 Xi, where Xij = 1 if and only if document i is cited by document j, and Xij = 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.  $\begin{array}{c:cccc} \text{Cited documents} & \multicolumn{4}{c}{\text{Citing documents}} \\ & D_1 & D_2 & \cdots & D_m \\ \hline D_1 & X_1{}^1 & X_2{}^1 & \cdots & X_m{}^1 \\ D_2 & X_1{}^2 & X_2{}^2 & \cdots & X_m{}^2 \\ \vdots & \vdots & & & \vdots \\ D_m & X_1{}^m & X_2{}^m & \cdots & X_m{}^m \end{array} \quad = \quad X$ $(X_j{}^i = 1 \leftrightarrow$ document D_i is cited by document $D_j)$ FIG. 4. Matrix X exhibiting direct citations (Salton, 1963, p. 447) Figure 5

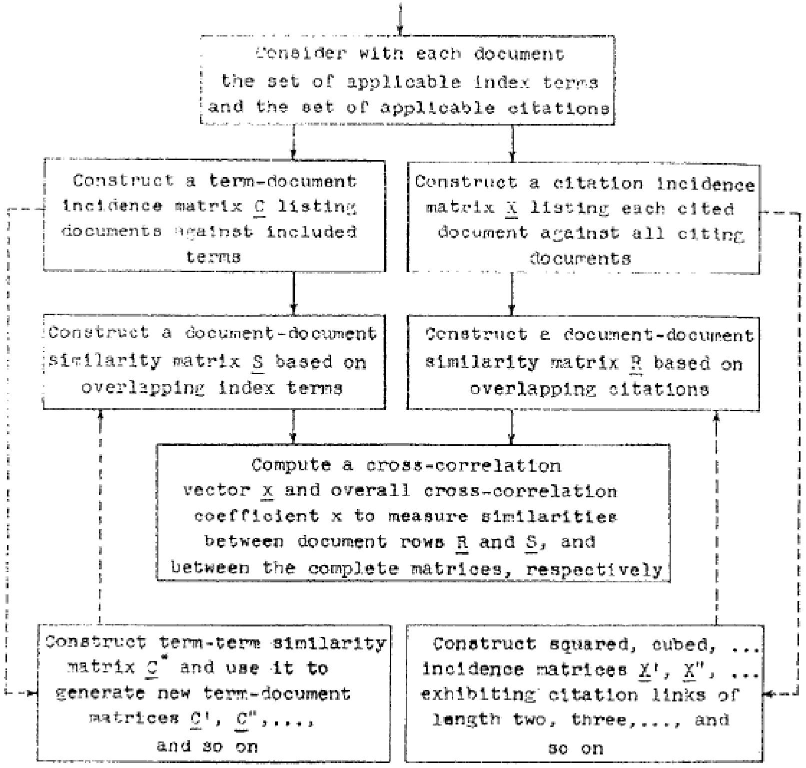
<table>
<tr><th>Claim Text from ‘352 Patent</th><th>Salton, 1963</th></tr>
<tr><td></td><td>


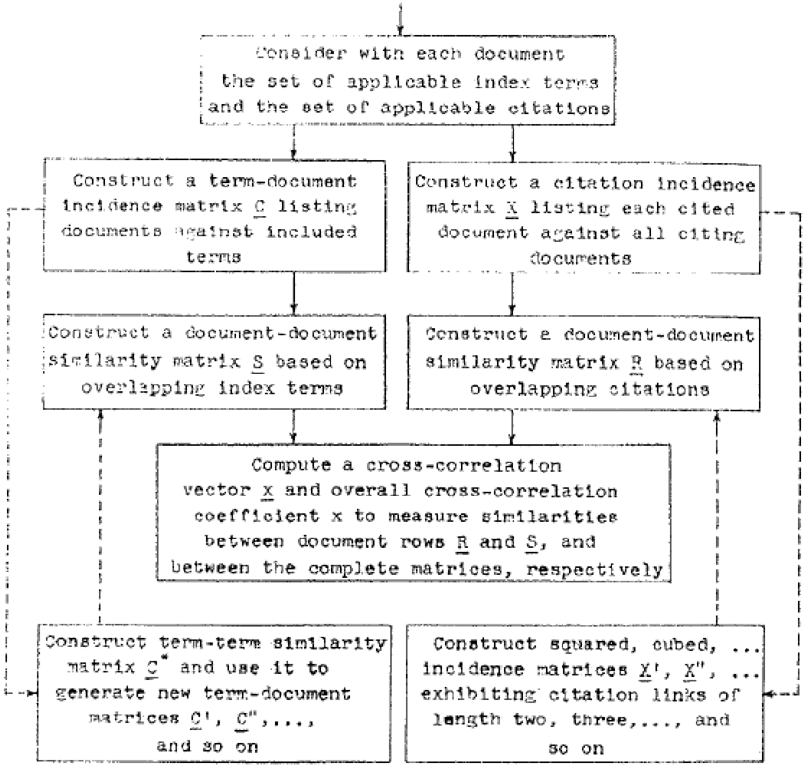
FIG. 5. Comparison of citation similarities with index term similarities

(Salton, 1963, p. 450)
</td></tr>
<tr><td>[26d] analyzing the first numerical representations for indirect relationships existing between or among objects in the database;</td><td>

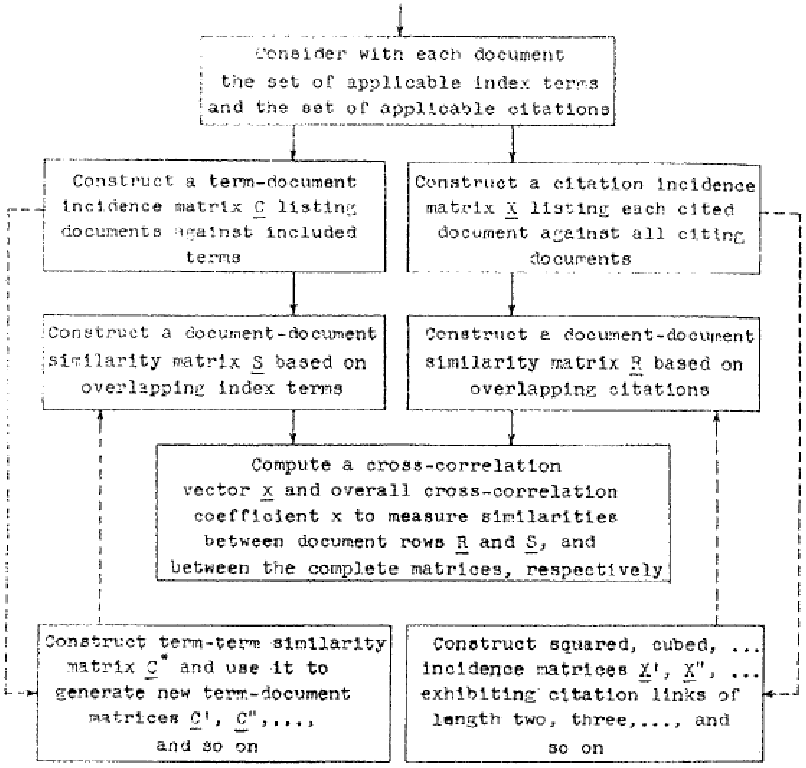
See, e.g., Salton, 1963, at pp. 448, 450

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,
</td></tr>
</table>

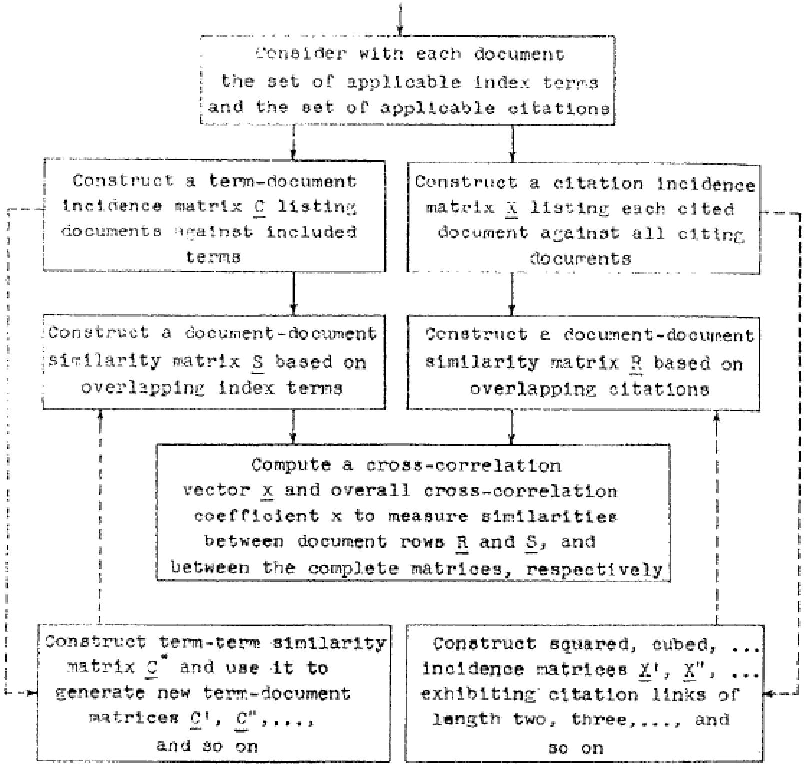
Claim Text from '352 Patent	Salton, 1963
	$[\mathbf{X}']_j^i = \bigvee_{k=1}^{m} (\mathbf{X}_k^i \wedge \mathbf{X}_j^k)$, $[\mathbf{X}'']_j^i = \bigvee_{k=1}^{m} (\mathbf{X}_k^i \wedge (\mathbf{X}')_j^k)$, and so on. Boolean multiplication is used, since the new connection matrices X', X'', etc., are 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) 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. 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) Figure 5

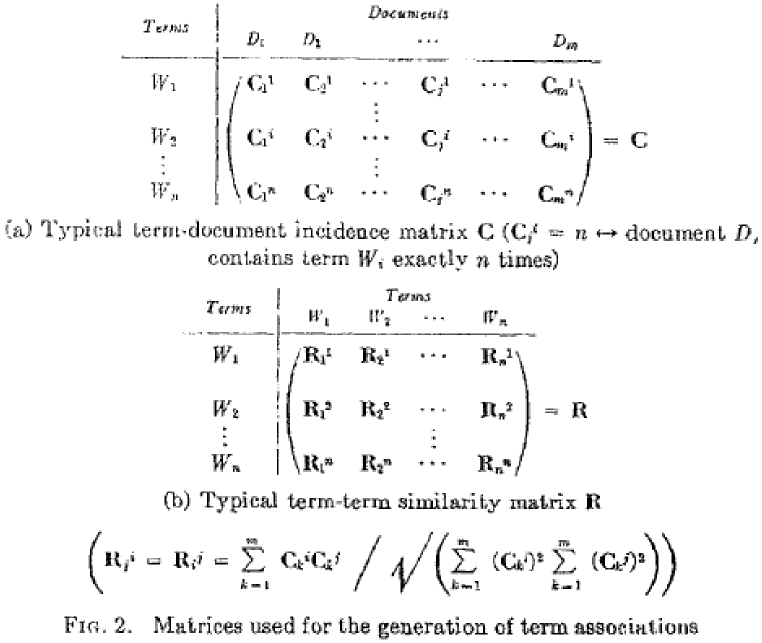
Claim Text from '352 Patent	Salton, 1963
	 FIG. 5. Comparison of citation similarities with index term similarities (Salton, 1963, p. 450)
[26e] generating a second numerical representation of each object based on the analysis of the first numerical representation;	*See, e.g.*, Salton, 1963, at pp. 448, 450, 451-52 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,

Claim Text from ‘352 Patent	**Salton, 1963**
	$[\mathbf{X}']_j^{\,i} = \bigvee_{k=1}^{m} (\mathbf{X}_k^{\,i} \wedge \mathbf{X}_j^{\,k}),$ $[\mathbf{X}'']_j^{\,i} = \bigvee_{k=1}^{m} (\mathbf{X}_k^{\,i} \wedge (\mathbf{X}')_j^{\,k}),$ 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) 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. 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) 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)

Claim Text from ‘352 Patent	Salton, 1963
	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
	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)
[26f] storing the second numerical representation for use in computerized searching; and	*See, e.g.,* Salton, 1963, at pp. 448, 450 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, $[\mathbf{X}']_j{}^i = \bigvee_{k=1}^{m} (\mathbf{X}_k{}^i \wedge \mathbf{X}_j{}^k),$ $[\mathbf{X}'']_j{}^i = \bigvee_{k=1}^{m} (\mathbf{X}_k{}^i \wedge (\mathbf{X}')_j{}^k),$ 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) 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. 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
	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  $\begin{array}{c\|cccccc} \text{Terms} & & & \text{Documents} & & & \\ & D_1 & D_2 & & \cdots & & D_m \\ \hline W_1 & C_1^1 & C_2^1 & \cdots & C_j^1 & \cdots & C_m^1 \\ W_2 & C_1^i & C_2^i & \cdots & C_j^i & \cdots & C_m^i \\ \vdots & & & \vdots & & & \\ W_n & C_1^n & C_2^n & \cdots & C_j^n & \cdots & C_m^n \end{array} = \mathbf{C}$ (a) Typical term-document incidence matrix **C** ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times) $\begin{array}{c\|cccc} \text{Terms} & & \text{Terms} & & \\ & W_1 & W_2 & \cdots & W_n \\ \hline W_1 & R_1^1 & R_2^1 & \cdots & R_n^1 \\ W_2 & R_1^2 & R_2^2 & \cdots & R_n^2 \\ \vdots & & & \vdots & \\ W_n & R_1^n & R_2^n & \cdots & R_n^n \end{array} = \mathbf{R}$ (b) Typical term-term similarity matrix **R** $\left(R_j^i = R_i^j = \sum_{k=1}^{m} C_k^i C_k^j \Big/ \sqrt{\left(\sum_{k=1}^{m} (C_k^i)^2 \sum_{k=1}^{m} (C_k^j)^2\right)}\right)$ 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

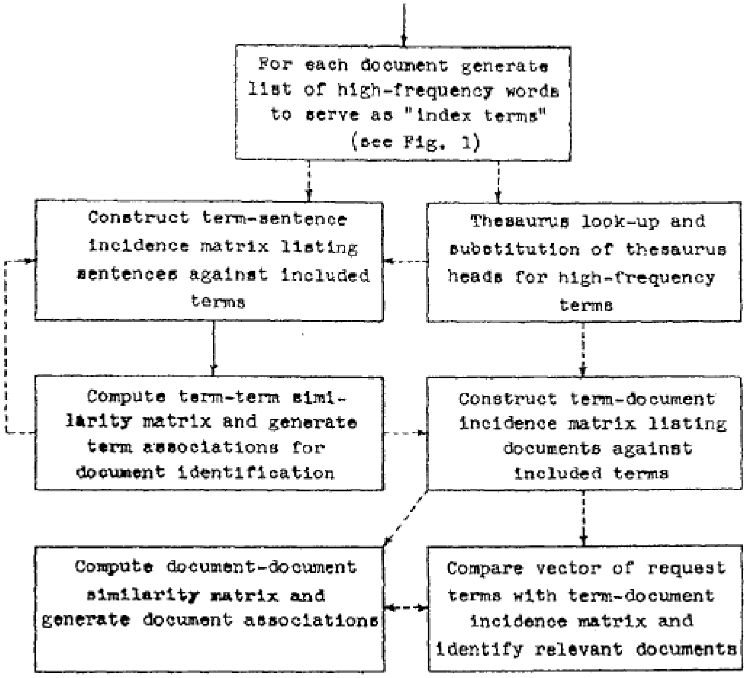
<table>
<tr><th>Claim Text from ‘352 Patent</th><th>Salton, 1963</th></tr>
<tr><td></td><td>inclusion in some of the matrix columns of new, associated terms which are not originally contained in the respective sentences or documents. The revised incidence matrix then gives rise to a new term-term similarity 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)</td></tr>
<tr><td>27. The non-semantical method of claim 26, wherein the objects in the database include words, and semantic indexing techniques are used in</td><td>*See, e.g.,* Salton, 1963, at Abstract, pp. 442, 446-47, 456-57</td></tr>
</table>

Claim Text from '352 Patent	Salton, 1963
combination with the non-semantical method, the method further comprising the step of creating and storing a Boolean word index for the words of the objects in the database.	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. … 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) 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." 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) 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) Figure 9

<table>
<tr><th>Claim Text from ‘352 Patent</th><th>Salton, 1963</th></tr>
<tr><td></td><td>

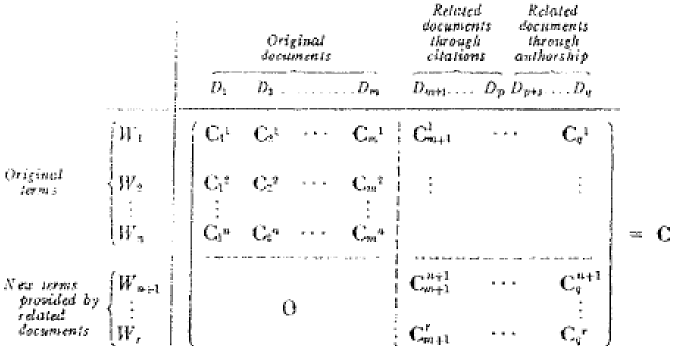


FIG. 9. Basic term-document incidence matrix usable for extended associative retrieval

(Salton, 1963, p. 456)

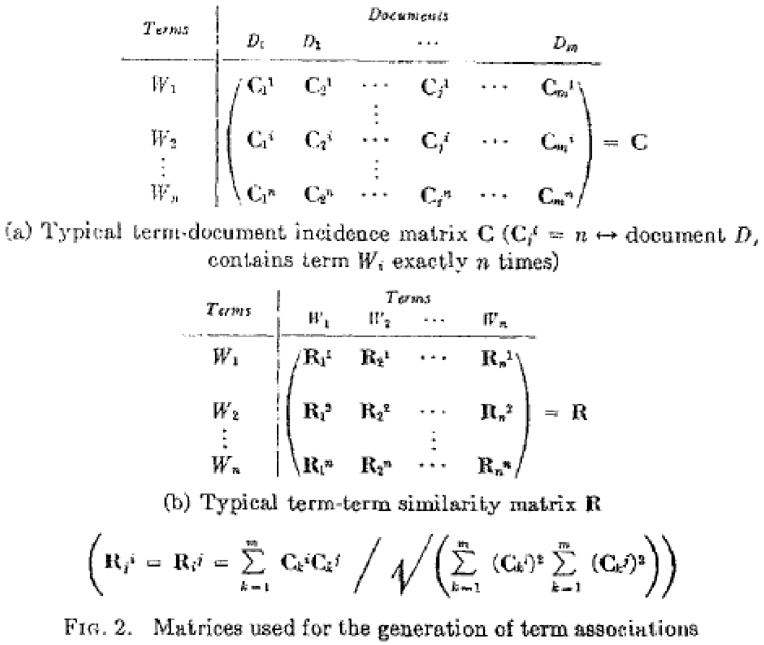
The following tentative conclusions can be drawn from the foregoing experiment: the similarity coefficients obtained by comparing overlapping citations for a sample document collection with overlapping, manually generated index terms are much larger than those obtained by assuming a random assignment 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)

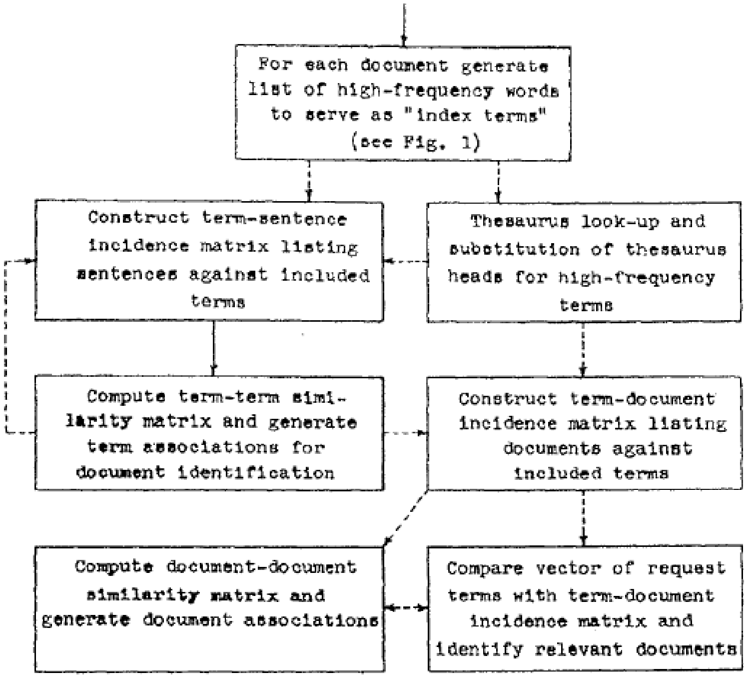
</td></tr>
<tr><td>28. The non-semantical method of claim 26 wherein the first and second numerical representations are vectors that are arranged in first and second matrices;</td><td>

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

Figure 2

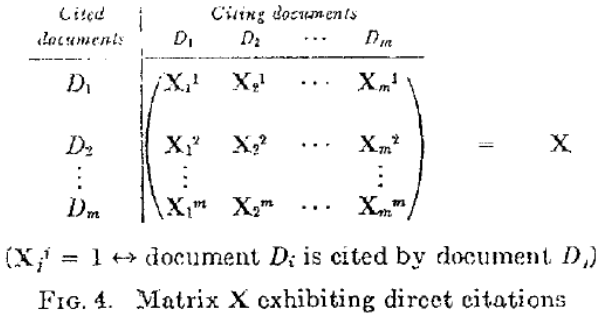
</td></tr>
</table>

Claim Text from ‘352 Patent	Salton, 1963
	 $\begin{array}{c\|ccccccc} \text{Terms} & & & \text{Documents} & & & \\ & D_1 & D_2 & & \cdots & & D_m \\ \hline W_1 & C_1^1 & C_2^1 & \cdots & C_j^1 & \cdots & C_m^1 \\ \vdots & & & \vdots & & & \\ W_i & C_1^i & C_2^i & \cdots & C_j^i & \cdots & C_m^i \\ \vdots & & & \vdots & & & \\ W_n & C_1^n & C_2^n & \cdots & C_j^n & \cdots & C_m^n \end{array} = \mathbf{C}$ (a) Typical term-document incidence matrix C ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times) $\begin{array}{c\|cccc} \text{Terms} & & \text{Terms} & & \\ & W_1 & W_2 & \cdots & W_n \\ \hline W_1 & R_1^1 & R_2^1 & \cdots & R_n^1 \\ W_2 & R_1^2 & R_2^2 & \cdots & R_n^2 \\ \vdots & & & \vdots & \\ W_n & R_1^n & R_2^n & \cdots & R_n^n \end{array} = \mathbf{R}$ (b) Typical term-term similarity matrix R $\left(R_j^i = R_i^j = \sum_{k=1}^{m} C_k^i C_k^j \Big/ \sqrt{\left(\sum_{k=1}^{m} (C_k^i)^2 \sum_{k=1}^{m} (C_k^j)^2\right)}\right)$ 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 x n symmetric term-similarity matrix R, where the coefficient of similarity Rji between term Wi and term Wj is $R_j^i = R_i^j = \dfrac{\sum_{k=1}^{m} C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^{m} (C_k^i)^2 \sum_{k=1}^{m} (C_k^j)^2\right)}}.$ …

Claim Text from '352 Patent	Salton, 1963
	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 of more of the other documents in the same collection. Each document can then be represented by an m-dimensional logical vector Xi, where Xij = 1 if and only if document i is cited by document j, and Xij = 0 otherwise. If these m vectors arranged in rows one below the other a square logical incidence matrix is formed similar to the matrix

Claim Text from ‘352 Patent	Salton, 1963
	exhibited in Figure 4. $\begin{array}{ccccc} \text{Cited} & \text{Citing documents} & & & \\ \text{documents} & D_1 & D_2 & \cdots & D_m \\ \hline D_1 & X_1^1 & X_2^1 & \cdots & X_m^1 \\ D_2 & X_1^2 & X_2^2 & \cdots & X_m^2 \\ \vdots & \vdots & & & \vdots \\ D_m & X_1^m & X_2^m & \cdots & X_m^m \end{array} = X$ $(X_j^i = 1 \leftrightarrow$ document D_i is cited by document $D_j)$ FIG. 4. Matrix X exhibiting direct citations (Salton, 1963, p. 447) 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, $[X']_j^i = \bigvee_{k=1}^{m} (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^{m} (X_k^i \wedge (X')_j^k),$ 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. 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;	*See, e.g.,* Salton, 1963, at Abstract, pp. 443, 446, 447, 450 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) 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) 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) 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 Xi, where Xij = 1 if and only if document i is cited by document j, and Xij = 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.

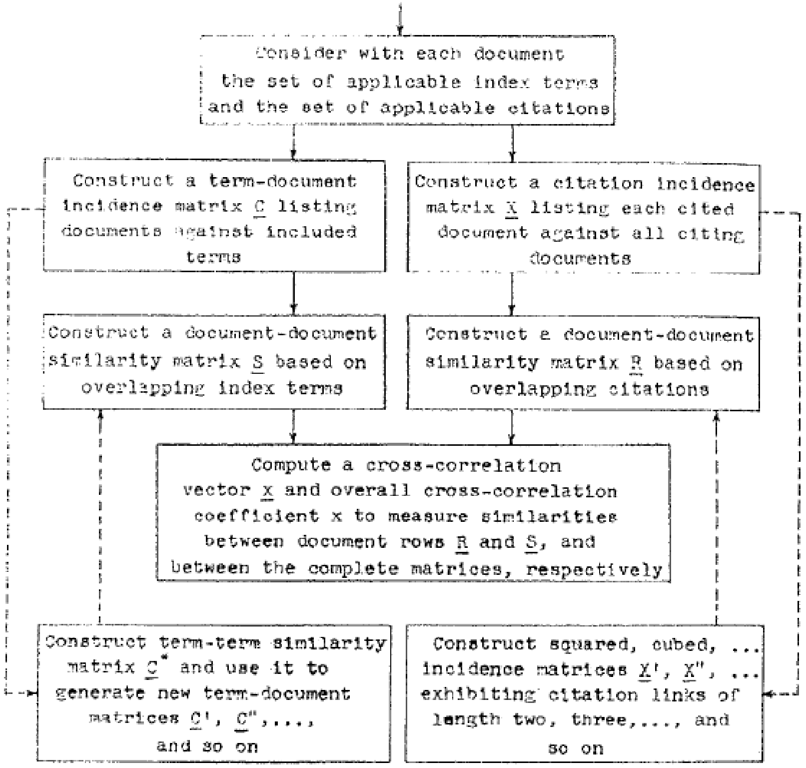
<table>
<tr><th>Claim Text from ‘352 Patent</th><th>Salton, 1963</th></tr>
<tr><td></td><td>

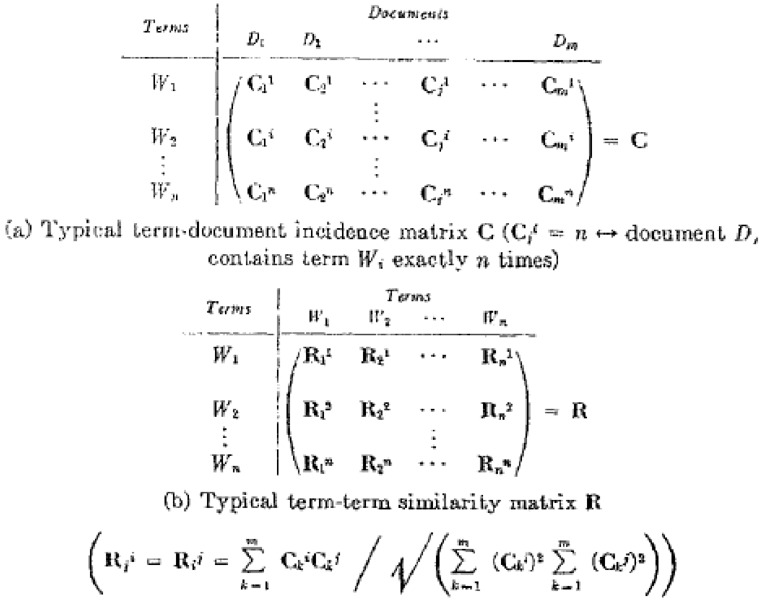
$$\begin{array}{c|cccc} \text{Cited documents} & \multicolumn{4}{c}{\text{Citing documents}} \\ & D_1 & D_2 & \cdots & D_m \\ \hline D_1 & X_1^1 & X_2^1 & \cdots & X_m^1 \\ D_2 & X_1^2 & X_2^2 & \cdots & X_m^2 \\ \vdots & \vdots & & & \vdots \\ D_m & X_1^m & X_2^m & \cdots & X_m^m \end{array} = X$$

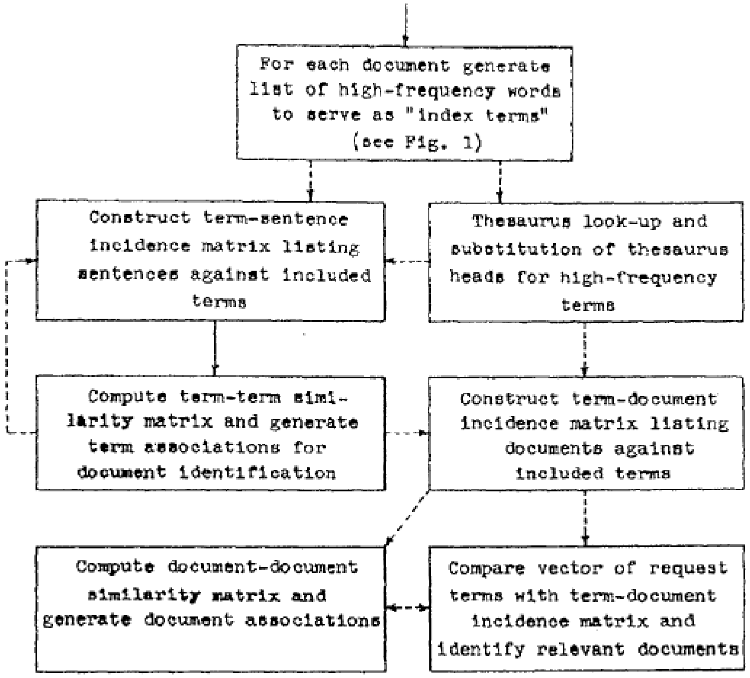
$(X_j^i = 1 \leftrightarrow$ document D_i is cited by document $D_j)$

FIG. 4. Matrix X exhibiting direct citations

(Salton, 1963, p. 447)
Figure 5</td></tr>
</table>

Claim Text from ‘352 Patent	Salton, 1963
	 FIG. 5. Comparison of citation similarities with index term similarities (Salton, 1963, p. 450)
[28b] the objects in the database are assigned chronological data;	*See, e.g.,* Salton, 1963, at p. 446 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

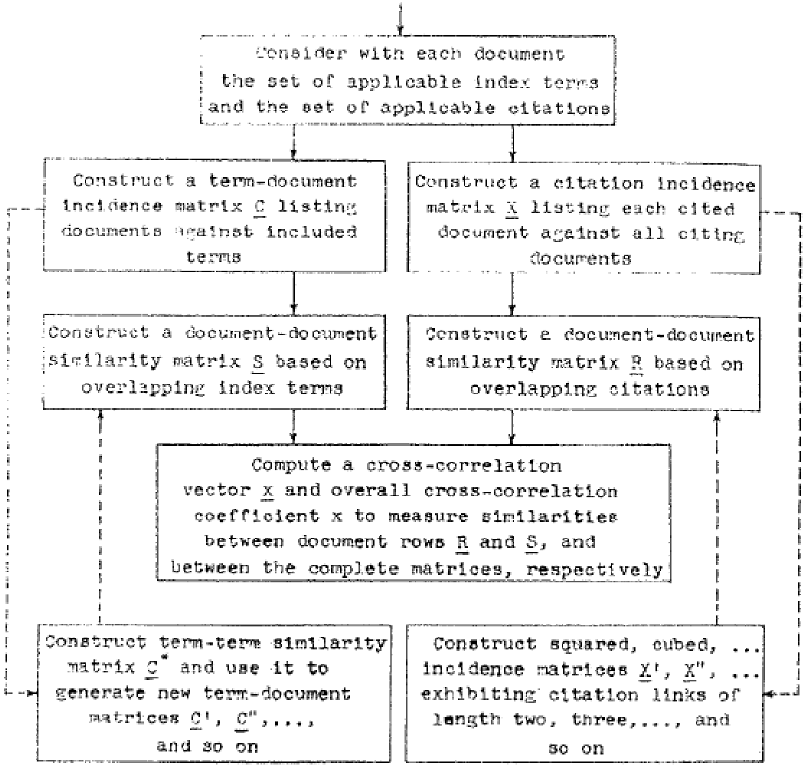
Claim Text from '352 Patent	Salton, 1963
	probability of being cited. Very recent documents which have not had a chance to circulate, and very old ones which no longer circulate are, in general, cited more rarely than current articles which have been distributed within the recent past. (Salton, 1963, p. 446)
[28c] and wherein the step of searching comprises the steps of matrix searching of the second matrices;	*See, e.g.,* Salton, 1963, at pp. 443-45, 448, 450, 451-52 Figure 2  $\begin{array}{ccccccc} \text{Terms} & & & \text{Documents} & & & \\ & D_1 & D_2 & & \cdots & & D_m \\ \hline W_1 & \mathbf{C}_1^1 & \mathbf{C}_2^1 & \cdots & \mathbf{C}_j^1 & \cdots & \mathbf{C}_m^1 \\ W_2 & \mathbf{C}_1^i & \mathbf{C}_2^i & \cdots & \mathbf{C}_j^i & \cdots & \mathbf{C}_m^i \\ \vdots & & & \vdots & & & \\ W_n & \mathbf{C}_1^n & \mathbf{C}_2^n & \cdots & \mathbf{C}_j^n & \cdots & \mathbf{C}_m^n \end{array} = \mathbf{C}$ (a) Typical term-document incidence matrix **C** ($\mathbf{C}_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times) $\begin{array}{ccccc} \text{Terms} & & \text{Terms} & & \\ & W_1 & W_2 & \cdots & W_n \\ \hline W_1 & \mathbf{R}_1^1 & \mathbf{R}_2^1 & \cdots & \mathbf{R}_n^1 \\ W_2 & \mathbf{R}_1^2 & \mathbf{R}_2^2 & \cdots & \mathbf{R}_n^2 \\ \vdots & & & \vdots & \\ W_n & \mathbf{R}_1^n & \mathbf{R}_2^n & \cdots & \mathbf{R}_n^n \end{array} = \mathbf{R}$ (b) Typical term-term similarity matrix **R** $\left(\mathbf{R}_j^i = \mathbf{R}_i^j = \sum_{k=1}^{m} \mathbf{C}_k^i \mathbf{C}_k^j \Big/ \sqrt{\left(\sum_{k=1}^{m} (\mathbf{C}_k^i)^2 \sum_{k=1}^{m} (\mathbf{C}_k^j)^2\right)}\right)$ 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) 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.

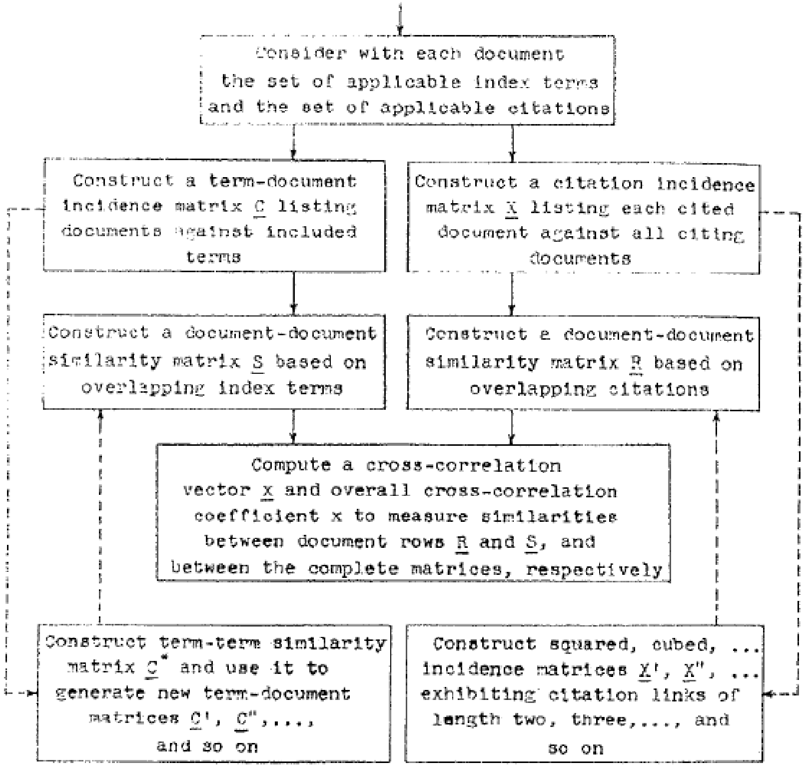
Claim Text from ‘352 Patent	Salton, 1963
	Specifically, $[\mathbf{X}']_j^{\,i} = \bigvee_{k=1}^{m} (\mathbf{X}_k^{\,i} \wedge \mathbf{X}_j^{\,k}),$ $[\mathbf{X}'']_j^{\,i} = \bigvee_{k=1}^{m} (\mathbf{X}_k^{\,i} \wedge (\mathbf{X}')_j^{\,k}),$ and so on. Boolean multiplication is used, since the new 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) 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) 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)

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-semantical 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.	*See*, *e.g.*, Salton, 1963, at pp. 447-48, 450 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) 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, $[\mathbf{X}']_j^{\,i} = \bigvee_{k=1}^{m} (\mathbf{X}_k^{\,i} \wedge \mathbf{X}_j^{\,k}),$ $[\mathbf{X}'']_j^{\,i} = \bigvee_{k=1}^{m} (\mathbf{X}_k^{\,i} \wedge (\mathbf{X}')_j^{\,k}),$ 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)

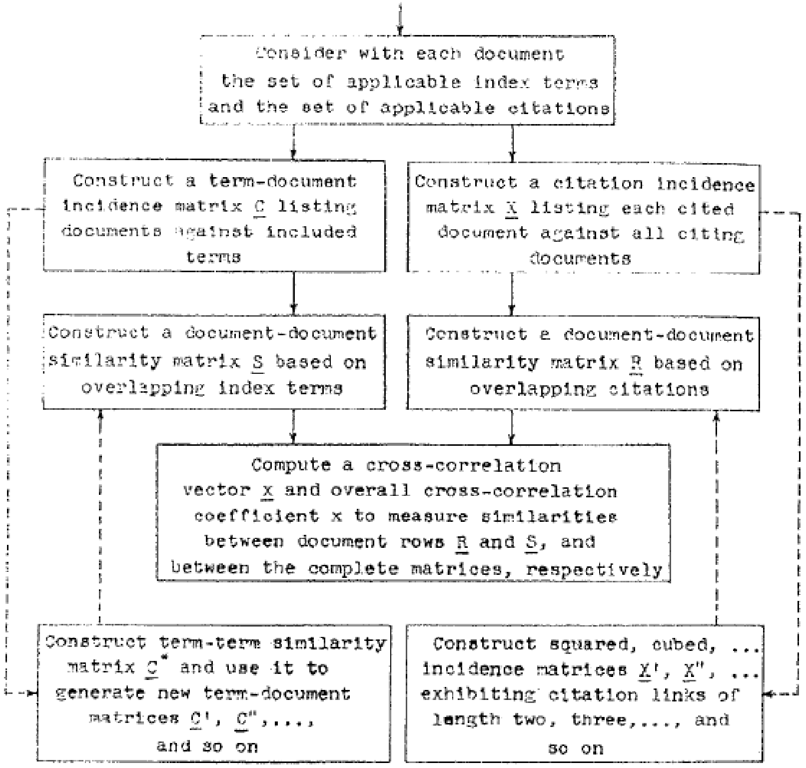
Claim Text from ‘352 Patent	Salton, 1963
	A measure of similarity between row (column) vectors can be obtained by calculating the cosine factor, previously exhibited in Section 2, for each pair of rows (columns). The result of such a computation can again be represented by a similarity matrix R, similar 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. 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) Figure 5

Claim Text from '352 Patent	Salton, 1963
	 FIG. 5. Comparison of citation similarities with index term similarities (Salton, 1963, p. 450)
30. The non-semantical 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	*See, e.g.,* Salton, 1963, at pp. 447-48, 450 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

Claim Text from ‘352 Patent	Salton, 1963
pattern: A cites c, and B cites c.	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) 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, $[\mathbf{X}']_j^i = \bigvee_{k=1}^{m} (\mathbf{X}_k^i \wedge \mathbf{X}_j^k),$ $[\mathbf{X}'']_j^i = \bigvee_{k=1}^{m} (\mathbf{X}_k^i \wedge (\mathbf{X}')_j^k)$, 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) 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. 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

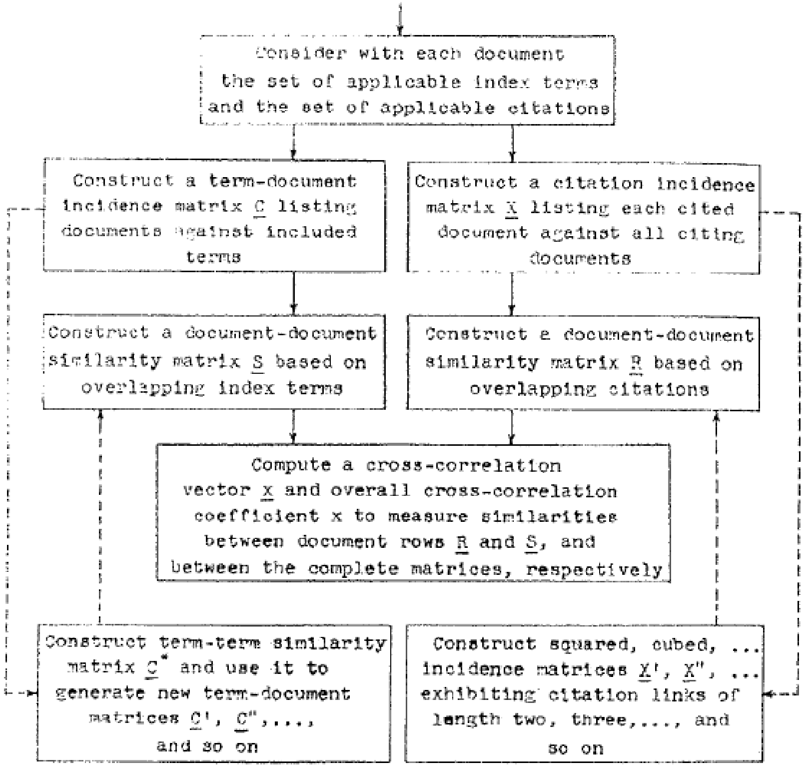
Claim Text from ‘352 Patent	Salton, 1963
	length three, and so on. (Salton, 1963, p. 448) Figure 5  FIG. 5. Comparison of citation similarities with index term similarities (Salton, 1963, p. 450)
31. The non-semantical method of claim 29, wherein a, b, c, A, d, e, f, B, g, h, and i are objects in the database and given that; a, b, and c occur before A;	*See, e.g.*, Salton, 1963, at pp. 447-48, 450 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

Claim Text from ‘352 Patent	Salton, 1963
A occurs before d, e, and f, which occur before B; and B occurs before g, h, and i; and wherein the step of examining for patterns further comprises the step of examining for one or more of the following patterns: (i) g cites A, and g cites B; (ii) B cites f, and f cites A; (iii) B cites f, f cites e, and e cites A; (iv) B cites f, f cites e, e cites d, and d cites A; (v) g cites A, h cites B, g cites a, and h cites a; (vi) i cites B, i cites f (or g), and f (or g) cites A; (vii) i cites g, i cites A, and g cites B; (viii) i cites g (or d), i cites h, g (or d) cites A, and h cites B; (ix) i cites a, i cites B, and A cites a; (x) i cites A, i cites e, B cites e; (xi) g cites A, g cites a, A cites a, h cites B, and h cites a; (xii) A cites a, B cites d, i cites a, and i cites d; (xiii) i cites B, i cites d, A cites a, and d cites a; (xiv) A cites b, B cites d (or c), and d (or c) cites b; (xv) A cites b, B cites d, b cites a, and d cites a; (xvi) A cites a, B cites b, d (or c) cites a, and d (or c) cites b.	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) 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, $[\mathbf{X}']_j{}^i = \bigvee_{k=1}^{m} (\mathbf{X}_k{}^i \wedge \mathbf{X}_j{}^k),$ $[\mathbf{X}'']_j{}^i = \bigvee_{k=1}^{m} (\mathbf{X}_k{}^i \wedge (\mathbf{X}')_j{}^k),$ 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) 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. 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

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-semantical method of claim 26, wherein the step of analyzing further comprises the	*See, e.g.*, Salton, 1963, at pp. 444, 448, 450, 451-52

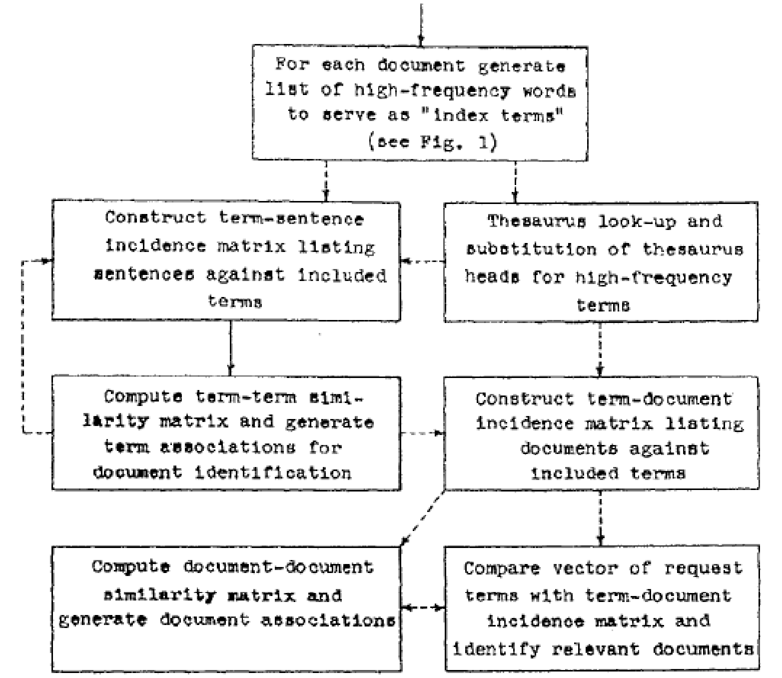
Claim Text from ‘352 Patent	Salton, 1963
step of weighing, wherein some indirect relationships are weighed more heavily than other indirect relationships.	To retrieve documents in answer to search requests, the programs already available can be used by adding to the term-document matrix C a new column 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) 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, $[\mathbf{X}']_j^i = \bigvee_{k=1}^{m} (\mathbf{X}_k^i \wedge \mathbf{X}_j^k),$ $[\mathbf{X}'']_j^i = \bigvee_{k=1}^{m} (\mathbf{X}_k^i \wedge (\mathbf{X}')_j^k),$ 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) 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) 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
	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)
33. The non-semantical method of claim 26, wherein the step of analyzing the first numerical representations for indirect relationships further comprises:	*See, e.g.,* Salton, 1963, at pp. 448, 450 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, $[\mathbf{X}']_j^{\,i} = \bigvee_{k=1}^{m} (\mathbf{X}_k^{\,i} \wedge \mathbf{X}_j^{\,k})$, $[\mathbf{X}'']_j^{\,i} = \bigvee_{k=1}^{m} (\mathbf{X}_k^{\,i} \wedge (\mathbf{X}')_j^{\,k})$, 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) 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. 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
	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

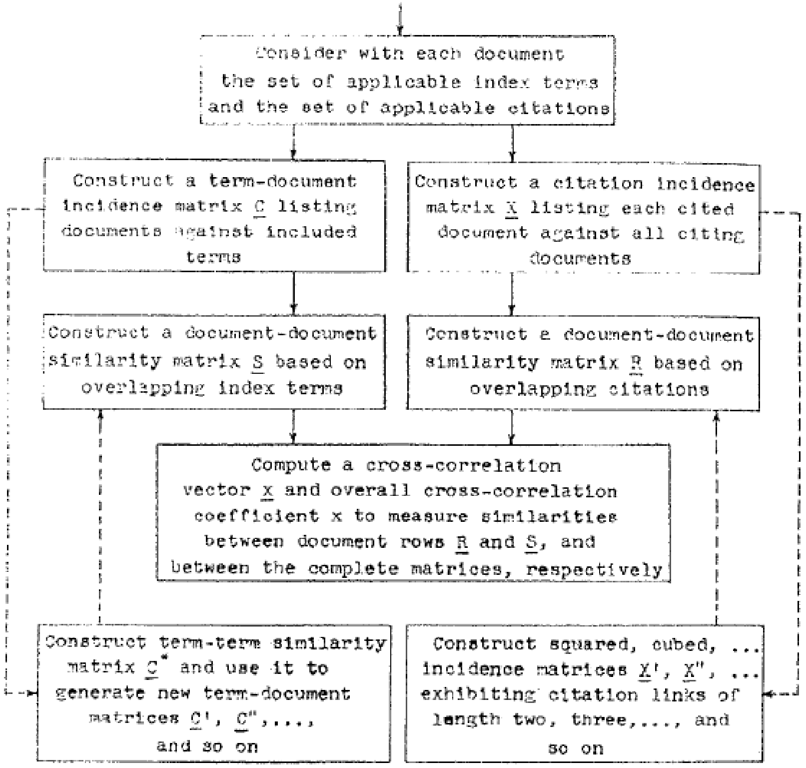
Claim Text from ‘352 Patent	Salton, 1963
	(Salton, 1963, p. 450)
[33a] creating an interim vector representing each object; and wherein the step of generating a second numerical representation uses coefficients of similarity and further comprises:	*See, e.g.,* Salton, 1963, at pp. 443-45, 447-50 Figure 2 $\begin{array}{c\|cccccc} & & & \text{Documents} & & & \\ \text{Terms} & D_1 & D_2 & & \cdots & & D_m \\ \hline W_1 & \mathbf{C}_1^1 & \mathbf{C}_2^1 & \cdots & \mathbf{C}_j^1 & \cdots & \mathbf{C}_m^1 \\ & & & \vdots & & & \\ W_2 & \mathbf{C}_1^i & \mathbf{C}_2^i & \cdots & \mathbf{C}_j^i & \cdots & \mathbf{C}_m^i \\ \vdots & & & \vdots & & & \\ W_n & \mathbf{C}_1^n & \mathbf{C}_2^n & \cdots & \mathbf{C}_j^n & \cdots & \mathbf{C}_m^n \end{array} = \mathbf{C}$ (a) Typical term-document incidence matrix **C** ($\mathbf{C}_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times) $\begin{array}{c\|cccc} & & \text{Terms} & & \\ \text{Terms} & W_1 & W_2 & \cdots & W_n \\ \hline W_1 & \mathbf{R}_1^1 & \mathbf{R}_2^1 & \cdots & \mathbf{R}_n^1 \\ W_2 & \mathbf{R}_1^2 & \mathbf{R}_2^2 & \cdots & \mathbf{R}_n^2 \\ \vdots & & & \vdots & \\ W_n & \mathbf{R}_1^n & \mathbf{R}_2^n & \cdots & \mathbf{R}_n^n \end{array} = \mathbf{R}$ (b) Typical term-term similarity matrix **R** $\left(\mathbf{R}_j^i = \mathbf{R}_i^j = \sum_{k=1}^{m} \mathbf{C}_k^i \mathbf{C}_k^j \Big/ \sqrt{\left(\sum_{k=1}^{m} (\mathbf{C}_k^i)^2 \sum_{k=1}^{m} (\mathbf{C}_k^j)^2\right)}\right)$ 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 x n symmetric term-similarity matrix R, where the coefficient of similarity Rji between term Wi and term Wj is

Claim Text from ‘352 Patent	Salton, 1963
	$R_j^i = R_i^j = \frac{\sum_{k=1}^{m} C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^{m} (C_k^i)^2 \sum_{k=1}^{m} (C_k^j)^2\right)}}.$ … 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) A term-term similarity matrix is then constructed by using co-occurrence if 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

Claim Text from ‘352 Patent	Salton, 1963
	 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 of more of the other documents in the same collection. Each document can then be represented by an m-dimensional logical vector Xi, where Xij = 1 if and only if document i is cited by document j, and Xij = 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.

Claim Text from '352 Patent	Salton, 1963
	$\begin{array}{c|cccc} \text{Cited documents} & \multicolumn{4}{c}{\text{Citing documents}} \\ & D_1 & D_2 & \cdots & D_m \\ \hline D_1 & \mathbf{X}_1^1 & \mathbf{X}_2^1 & \cdots & \mathbf{X}_m^1 \\ D_2 & \mathbf{X}_1^2 & \mathbf{X}_2^2 & \cdots & \mathbf{X}_m^2 \\ \vdots & \vdots & & & \vdots \\ D_m & \mathbf{X}_1^m & \mathbf{X}_2^m & \cdots & \mathbf{X}_m^m \end{array} = \mathbf{X}$ $(\mathbf{X}_j^i = 1 \leftrightarrow$ document D_i is cited by document $D_j)$ Fig. 4. Matrix X exhibiting direct citations (Salton, 1963, p. 447) 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, $[\mathbf{X}']_j^i = \bigvee_{k=1}^{m} (\mathbf{X}_k^i \wedge \mathbf{X}_j^k),$ $[\mathbf{X}'']_j^i = \bigvee_{k=1}^{m} (\mathbf{X}_k^i \wedge (\mathbf{X}')_j^k),$ 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) 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. 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

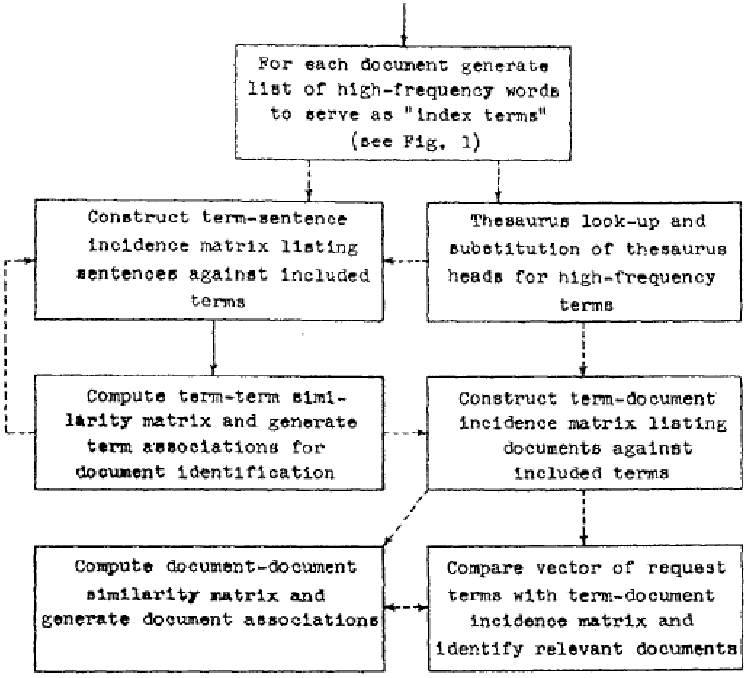
Claim Text from ‘352 Patent	Salton, 1963
	documents (links of length one), but links of length two, three, four, or more. Consider, as an example, a document collection in which document A cites document B, or B cites A. The corresponding documents are then said to be linked directly. On the other hand, 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) 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) Figure 5

Claim Text from '352 Patent	Salton, 1963
	 FIG. 5. Comparison of citation similarities with index term similarities (Salton, 1963, p. 450)
[33b] calculating Euclidean distances between interim vector representations of each object;	*See, e.g.,* Salton, 1963, at pp. 443-44, 447, 448 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

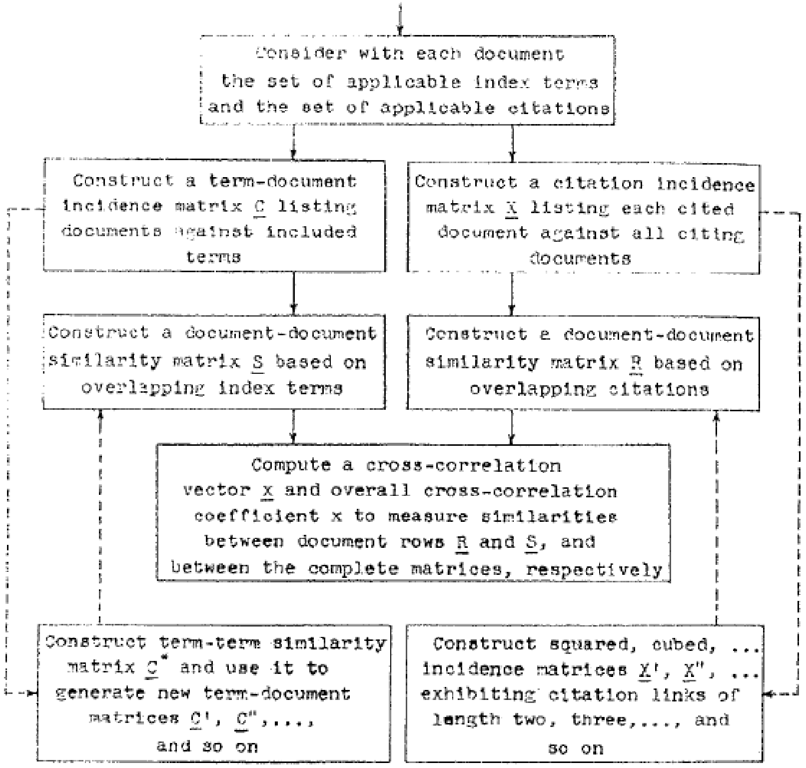
Claim Text from ‘352 Patent	Salton, 1963
	term-similarity matrix R, where the coefficient of similarity Rji between term Wi and term Wj is $R_j^i = R_i^j = \frac{\sum_{k=1}^{m} C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^{m} (C_k^i)^2 \sum_{k=1}^{m} (C_k^j)^2\right)}}$. … Document similarities are therefore 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) 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 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. (Salton, 1963, p. 448) 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.

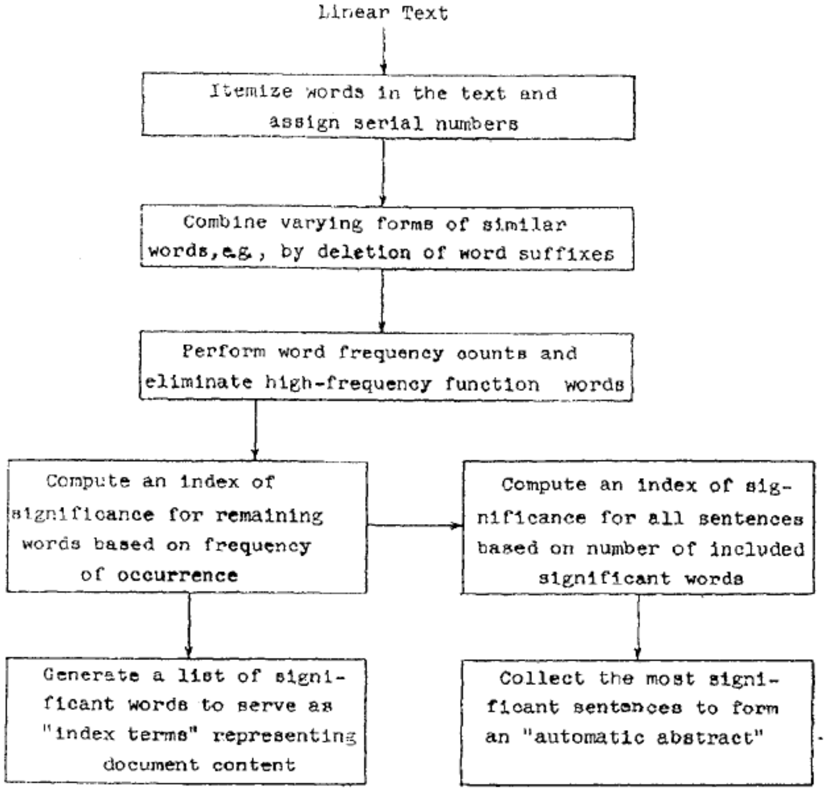
Claim Text from ‘352 Patent	Salton, 1963
[33c] creating proximity vectors representing the objects using the calculated Euclidean distances; and	*See, e.g.,* Salton, 1963, at pp. 443-44, 447, 448 Many different types of similarity coefficients have been suggested in the literature; a simple coefficient 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 Rji between term Wi and term Wj is $\mathrm{R}_j^{\,i} = \mathrm{R}_i^{\,j} = \frac{\sum_{k=1}^{m} \mathrm{C}_k^{\,i}\mathrm{C}_k^{\,j}}{\sqrt{\left(\sum_{k=1}^{m} (\mathrm{C}_k^{\,i})^2 \sum_{k=1}^{m} (\mathrm{C}_k^{\,j})^2\right)}}.$ … 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) 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 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

Claim Text from ‘352 Patent	Salton, 1963
	shown in Figure 2(b), where Rij is the value of the similarity coefficient between the ith and jth rows (columns) of X. (Salton, 1963, p. 448) 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.
[33d] using the proximity vectors and using coefficients of similarity to calculate the second numerical representations.	*See, e.g.,* Salton, 1963, at pp. 443, 444, 445, 448, 449, 450 Figure 2 $\begin{array}{lcccccc} \text{Terms} & & & & \text{Documents} & & \\ & D_1 & D_2 & & \cdots & & D_m \\ W_1 & C_1^1 & C_2^1 & \cdots & C_j^1 & \cdots & C_m^1 \\ & & & \vdots & & & \\ W_2 & C_1^i & C_2^i & \cdots & C_j^i & \cdots & C_m^i \\ \vdots & & & \vdots & & & \\ W_n & C_1^n & C_2^n & \cdots & C_j^n & \cdots & C_m^n \end{array} = \mathbf{C}$ (a) Typical term-document incidence matrix C ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times) $\begin{array}{lcccc} \text{Terms} & & \text{Terms} & & \\ & W_1 & W_2 & \cdots & W_n \\ W_1 & R_1^1 & R_2^1 & \cdots & R_n^1 \\ W_2 & R_1^2 & R_2^2 & \cdots & R_n^2 \\ \vdots & & & \vdots & \\ W_n & R_1^n & R_2^n & \cdots & R_n^n \end{array} = \mathbf{R}$ (b) Typical term-term similarity matrix R $\left(R_j^i = R_i^j = \sum_{k=1}^{m} C_k^i C_k^j \Big/ \sqrt{\left(\sum_{k=1}^{m} (C_k^i)^2 \sum_{k=1}^{m} (C_k^j)^2\right)}\right)$ FIG. 2. Matrices used for the generation of term associations

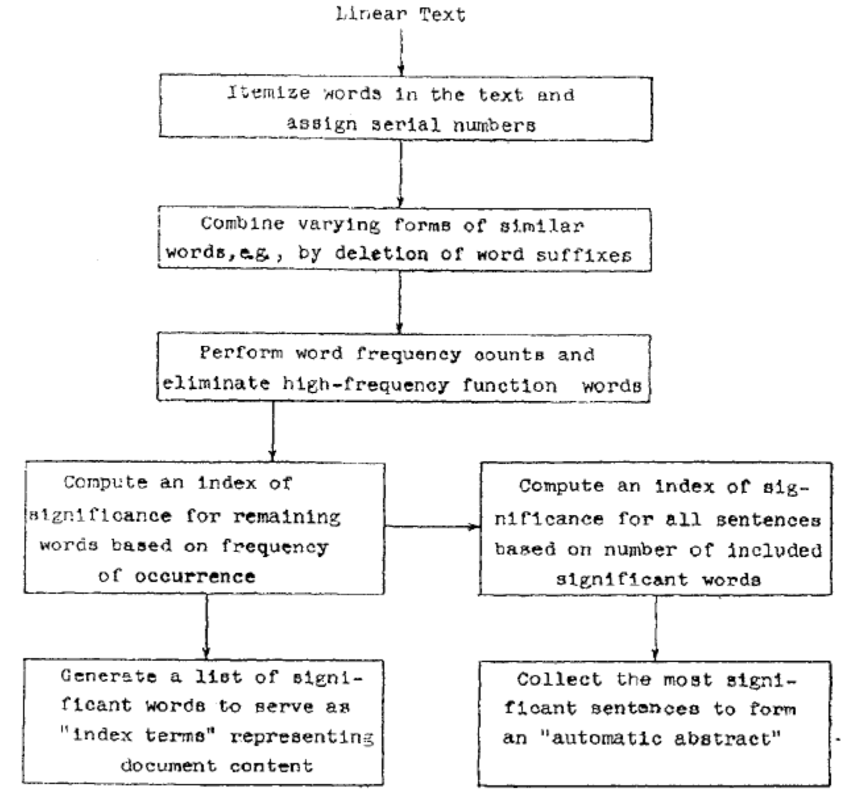
Claim Text from '352 Patent	Salton, 1963
	(Salton, 1963, p. 443) A term-term similarity matrix is then constructed by using co-occurrence if terms within sentences rather than within documents as a criterion. It should be noted that as new term associations are defined, the original incidence matrix 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
	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. 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) 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) Figure 5

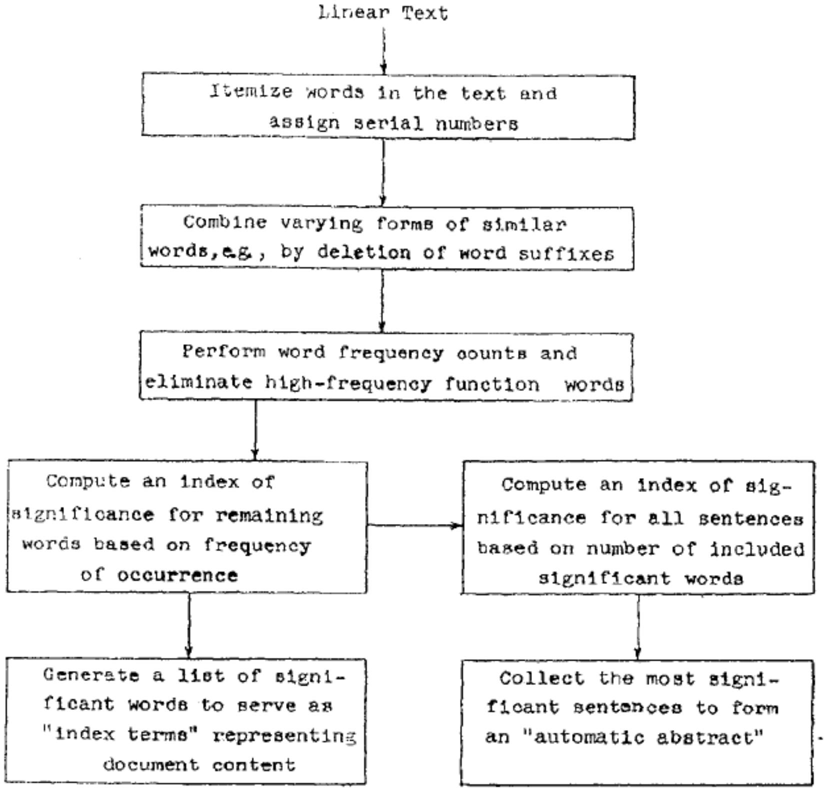
Claim Text from ‘352 Patent	Salton, 1963
	 FIG. 5. Comparison of citation similarities with index term similarities (Salton, 1963, p. 450)
34. The non-semantical method of claim 26, wherein objects in the database may be divided into subsets and wherein the marking step includes the step of marking subsets of objects in the database and wherein relationships exist between or among subsets of objects in the database.	*See, e.g.*, Salton, 1963, at pp. 441, 444, 447 Figure 1

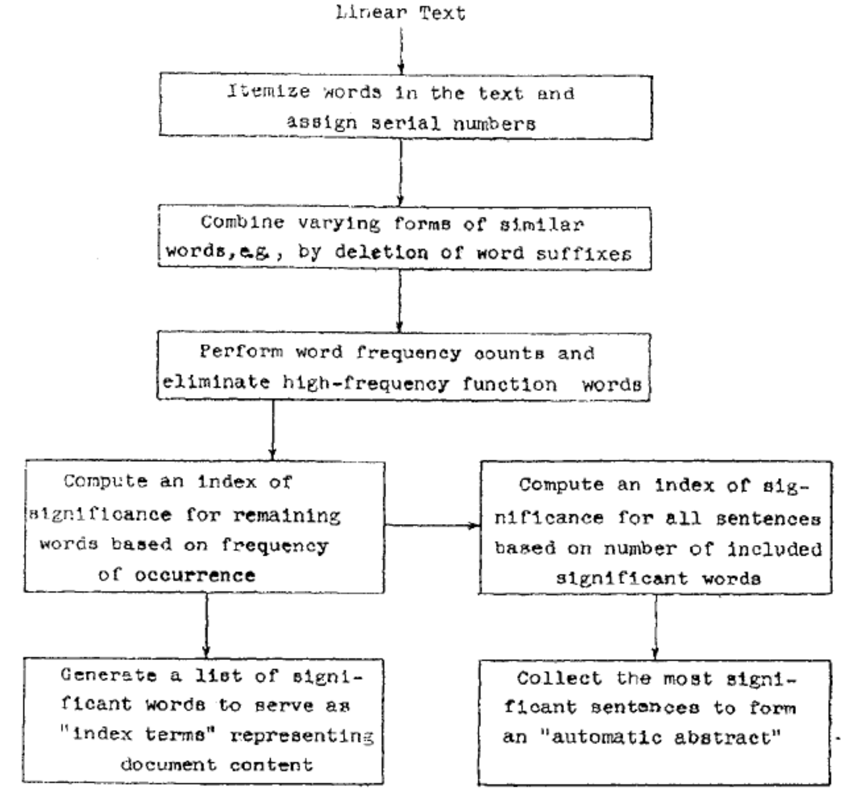
Claim Text from ‘352 Patent	Salton, 1963
	 FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts. (Salton, 1963, p. 441) To generate document associations instead of term associations the same procedures can be used, since the strength of association between documents 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)

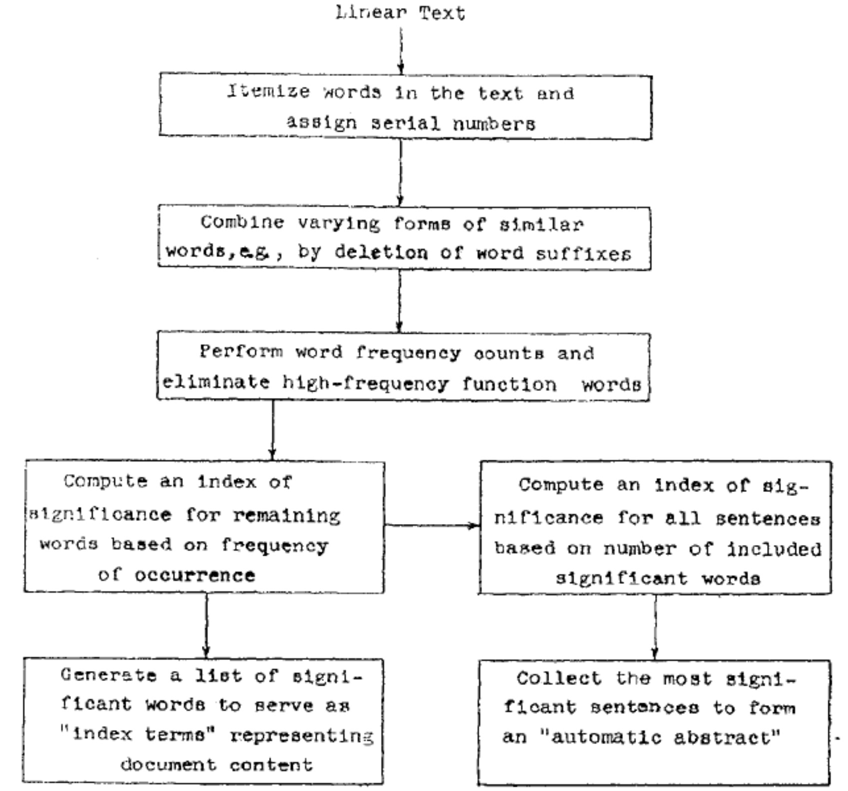
Claim Text from '352 Patent	Salton, 1963
	Consider a collection of m documents each of which is characterized by the property of being cited by one or more of the other documents in the same collection. Each document can then be represented by an m-dimensional logical vector Xi , where Xji = 1 if and only if document i is cited by document j, and Xji = 1 otherwise. (Salton, 1963, p. 447)
35. The non-semantical 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:	*See, e.g.*, Salton, 1963, at pp. 441, 444 Figure 1

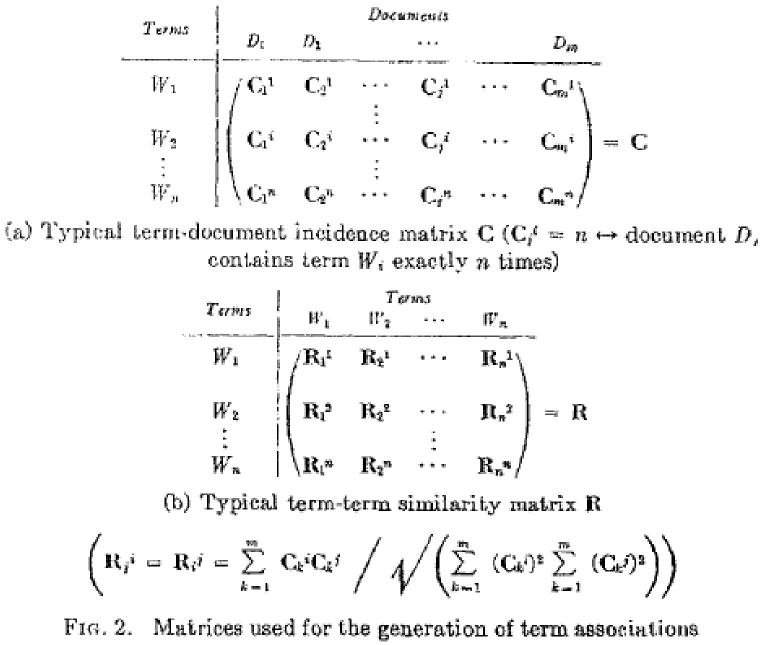
Claim Text from '352 Patent	Salton, 1963
	 FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts. (Salton, 1963, p. 441) To generate document associations instead of term associations the same procedures can be used, since the strength of association between documents 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)

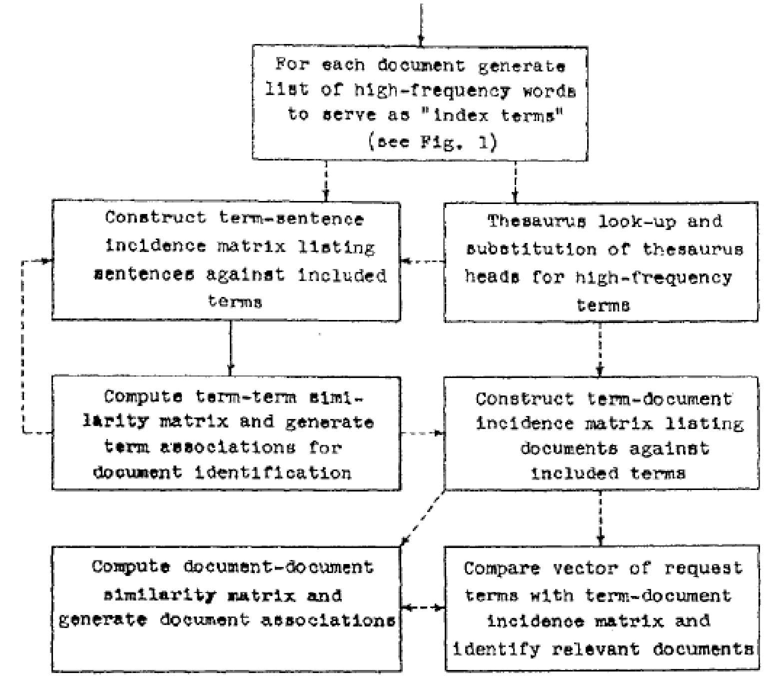
Claim Text from ‘352 Patent	Salton, 1963
[35a] creating a subset numerical representation for each subset based upon the relationships between or among subsets;	*See, e.g.*, Salton, 1963, at p. 444 To generate document associations instead of term associations the same procedures can be used, since the 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)
[35b] analyzing the subset numerical representations;	*See, e.g.*, Salton, 1963, at p. 441 Figure 1

Claim Text from '352 Patent	Salton, 1963
	 FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts. (Salton, 1963, p. 441)
[35c] clustering the subsets into sections based upon the subset analysis; and	*See, e.g.,* Salton, 1963, at p. 441 Figure 1

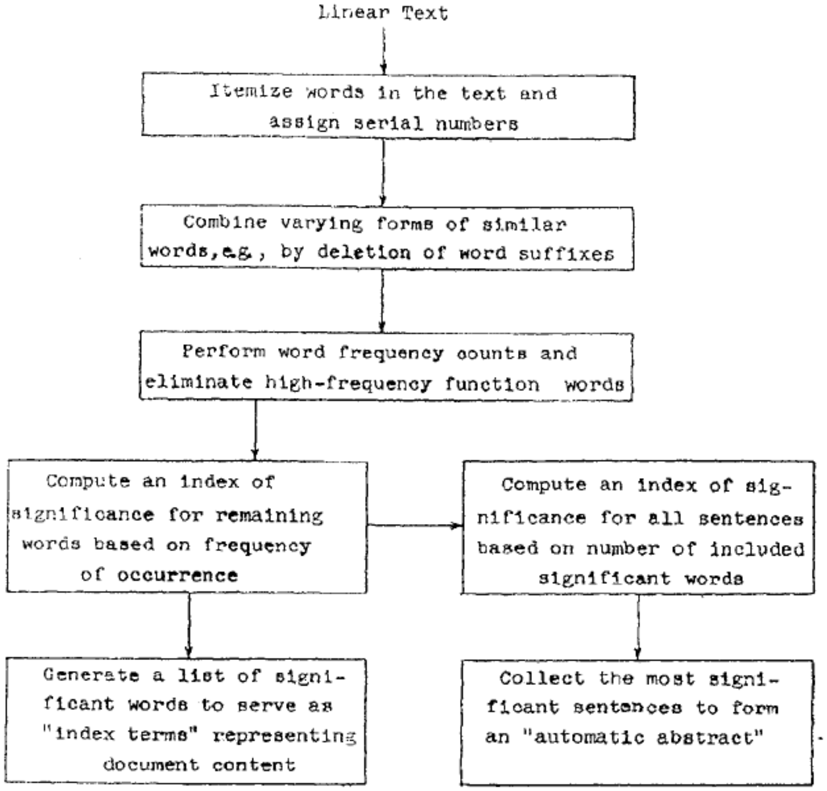
Claim Text from '352 Patent	Salton, 1963
	 FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts. (Salton, 1963, p. 441)
[35d] generating a section numerical representation for each section, wherein the section numerical representations are available for searching.	*See, e.g.,* Salton, 1963, at p. 441 Figure 1

Claim Text from ‘352 Patent	Salton, 1963
	 FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts. (Salton, 1963, p. 441)
36. The non-semantical 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.	*See, e.g.,* Salton, 1963, at pp. 443, 444, 445 Figure 2

Claim Text from ‘352 Patent	Salton, 1963
	 $\begin{array}{cccccccc} \text{Terms} & & & \text{Documents} & & & & \\ & D_1 & D_2 & & \cdots & & D_m & \\ W_1 & C_1^1 & C_2^1 & \cdots & C_j^1 & \cdots & C_m^1 & \\ W_2 & C_1^i & C_2^i & \cdots & C_j^i & \cdots & C_m^i & = \mathbf{C} \\ \vdots & & & \vdots & & & & \\ W_n & C_1^n & C_2^n & \cdots & C_j^n & \cdots & C_m^n & \end{array}$ (a) Typical term-document incidence matrix C ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times) $\begin{array}{cccccc} \text{Terms} & & \text{Terms} & & & \\ & W_1 & W_2 & \cdots & W_n & \\ W_1 & R_1^1 & R_2^1 & \cdots & R_n^1 & \\ W_2 & R_1^2 & R_2^2 & \cdots & R_n^2 & = \mathbf{R} \\ \vdots & & & \vdots & & \\ W_n & R_1^n & R_2^n & \cdots & R_n^n & \end{array}$ (b) Typical term-term similarity matrix R $\left(R_j^i = R_i^j = \sum_{k=1}^{m} C_k^i C_k^j \Big/ \sqrt{\left(\sum_{k=1}^{m} (C_k^i)^2 \sum_{k=1}^{m} (C_k^j)^2\right)}\right)$ 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

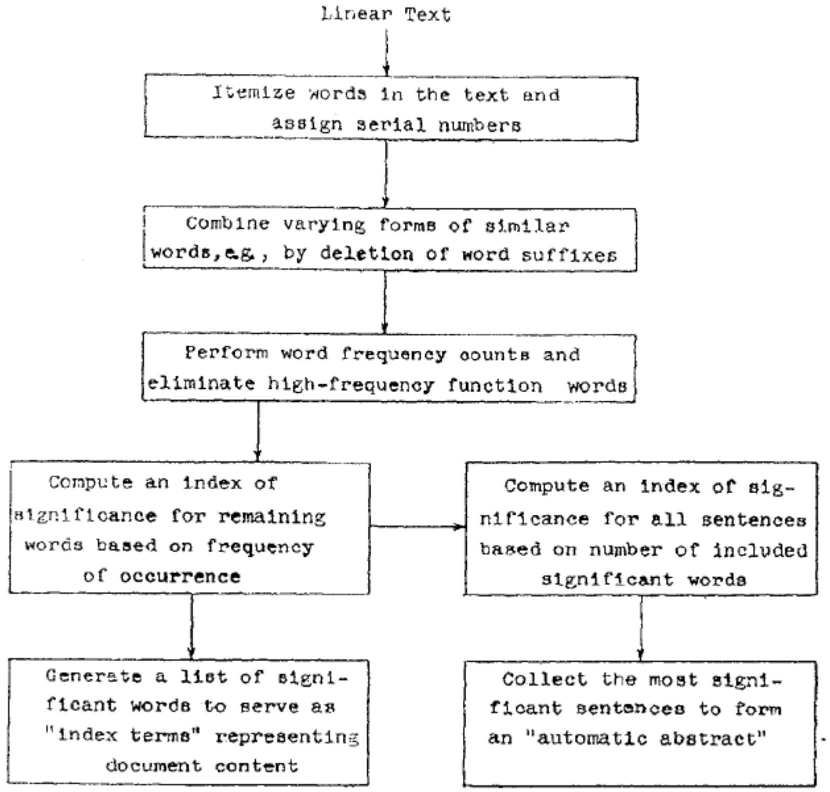
Claim Text from '352 Patent	Salton, 1963
	 FIG. 3. Typical automatic document retrieval system using term and document associations ⇢ optional paths → compulsory paths (Salton, 1963, p. 445)
37. The non-semantical method of claim 26, wherein the step of searching includes the step of graphically displaying one or more of the identified objects.	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.
38. The non-semantical method of claim 26, wherein the step of searching includes 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.

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-semantical method of claim 26, wherein the step of searching the objects comprises the steps of: selecting a pool of objects;	*See, e.g.,* Salton, 1963, at p. 447 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 Xi , where Xji = 1 if and only if document i is cited by document j, and Xji = 1 otherwise. (Salton, 1963, p. 447)
[39a] pool-similarity searching to identify a similar pool of textual objects, similar in relation to the objects in marked pool; and	*See, e.g.,* Salton, 1963, at p. 441 Figure 1

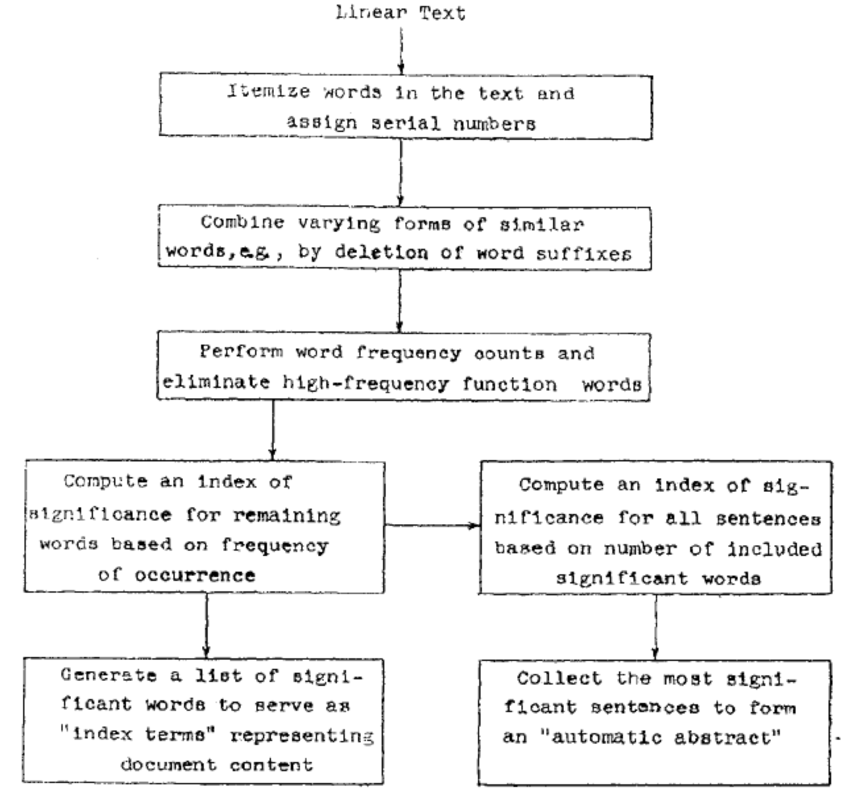
Claim Text from '352 Patent	Salton, 1963
	 FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts. (Salton, 1963, p. 441) Further, disclosed either expressly or inherently in the teachings of the reference and its incorporated disclosures taken as a whole, or in combination 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.

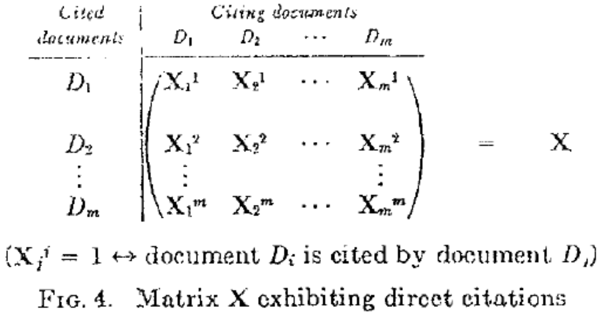
Claim Text from '352 Patent	Salton, 1963
[39b] pool-importance searching to identify an important pool of textual objects, important in relation to the objects in the selected pool.	*See, e.g.,* Salton, 1963, at p. 444 An estimate of document relevance is then obtained by computing for each document the similarity coefficient between the request column Cm+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) 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.
40. The non-semantical method of claim 26, the step of searching comprising the steps of: identifying a paradigm pool of objects; 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.
[40a] searching for relationships between the objects and the paradigm pool of objects;	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.
[40b] wherein the searched for relationship is pool importance or pool similarity.	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

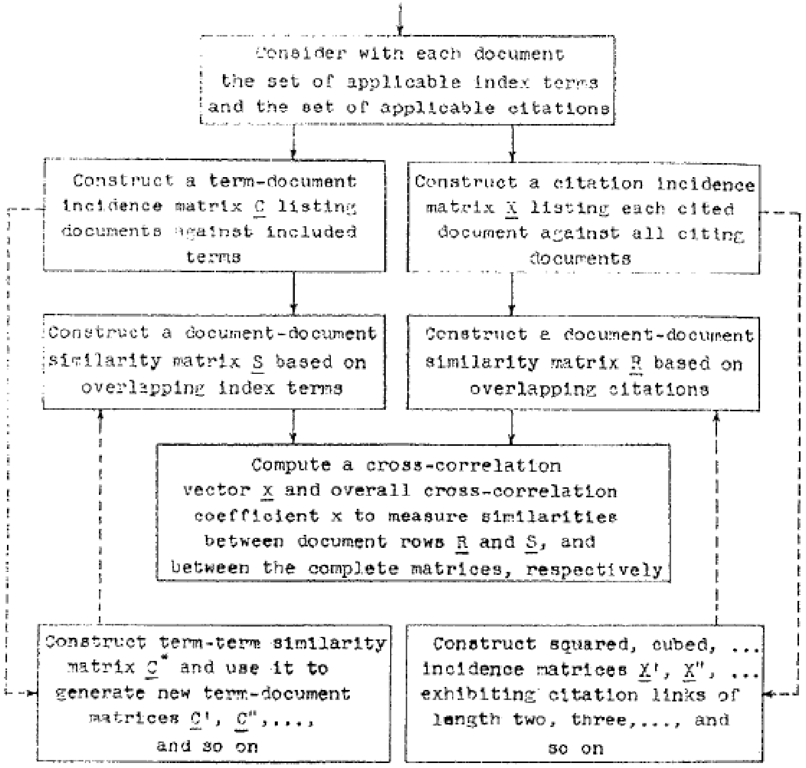
Claim Text from ‘352 Patent	Salton, 1963
	incorporated by reference into this chart.
41. A method for the non-semantical indexing of objects stored in a computer database, the method for use in searching the database for the objects, comprising the steps of: extracting, comprising the steps of:	*See, e.g.,* Salton, 1963, at Abstract, pp. 440-43, 446, 450 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) 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. A typical automatic indexing and abstracting system based on word frequency counts is shown in Figure 1. (Salton, 1963, pp. 440-41) Figure 1

Claim Text from ‘352 Patent	Salton, 1963
	 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 W1, W2, … , Wn, 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
	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) 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) 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)
[41a] labeling objects with a first numerical representation; and	*See, e.g.,* Salton, 1963, at pp. 441, 447 Figure 1

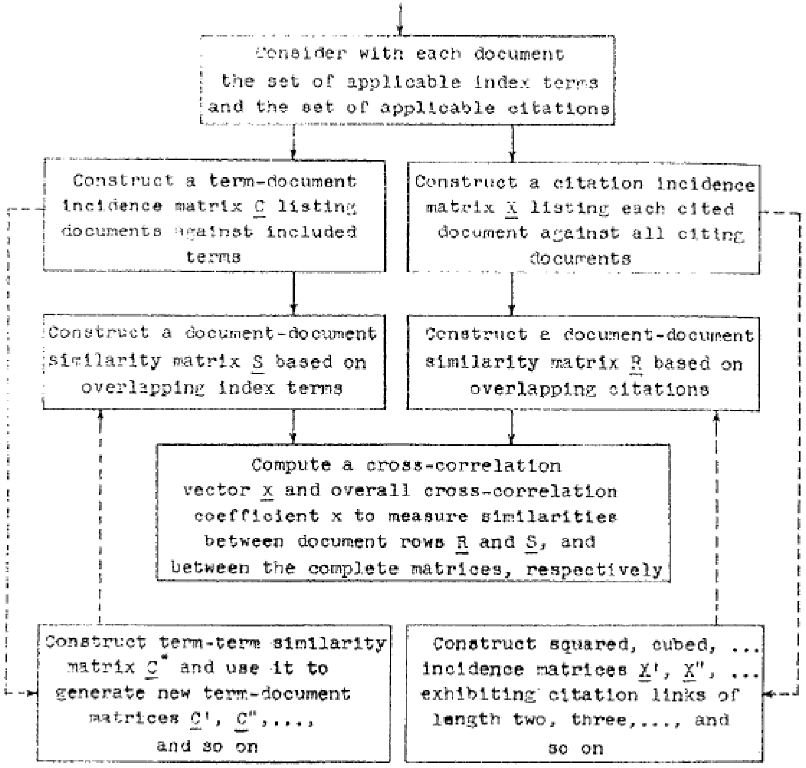
Claim Text from '352 Patent	Salton, 1963
	 FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts. (Salton, 1963, p. 441) Consider a collection of m documents each of which is characterized by the property of being cited by one or more of the other documents in the same collection. Each document can then be represented by an m-dimensional logical vector Xi , where Xji = 1 if and only if document i is cited by document j, and Xji = 1 otherwise. (Salton, 1963, p. 447)
[41b] generating a second numerical representation	*See, e.g.,* Salton, 1963, at pp. 446 n.1, 447, 450

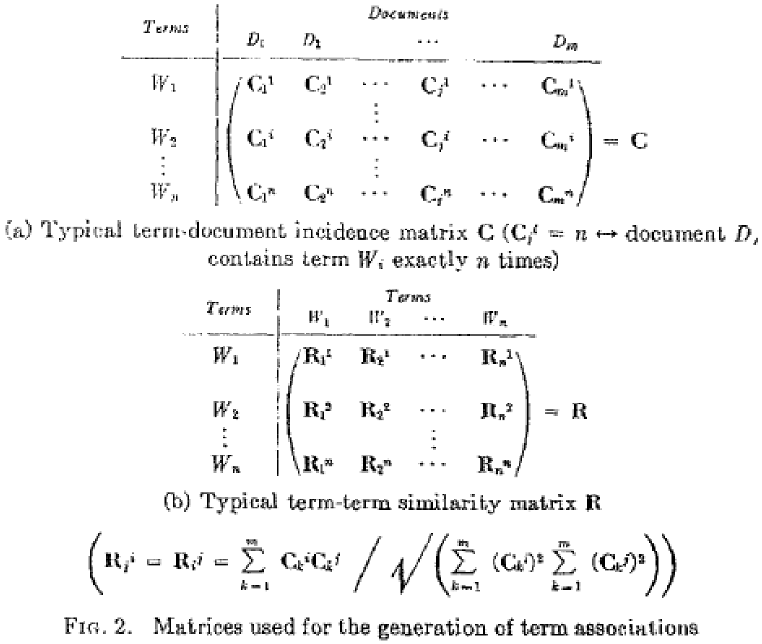
Claim Text from '352 Patent	Salton, 1963
for each object based on each object's references to other objects;	A citation index consists of a set of bibliographic references (the set of cited documents), each followed by a list of all those documents (the citing documents) which include the given cited 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 Xi, where Xij = 1 if and only if document i is cited by document j, and Xij = 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.  $\begin{array}{c\|cccc} \text{Cited} & \multicolumn{4}{c}{\text{Citing documents}} \\ \text{documents} & D_1 & D_2 & \cdots & D_m \\ \hline D_1 & X_1^1 & X_2^1 & \cdots & X_m^1 \\ D_2 & X_1^2 & X_2^2 & \cdots & X_m^2 \\ \vdots & \vdots & & & \vdots \\ D_m & X_1^m & X_2^m & \cdots & X_m^m \end{array} = X$ $(X_j^i = 1 \leftrightarrow \text{document } D_i \text{ is cited by document } D_j)$ FIG. 4. Matrix X exhibiting direct citations (Salton, 1963, p. 447) Figure 5

Claim Text from '352 Patent	Salton, 1963
	 FIG. 5. Comparison of citation similarities with index term similarities (Salton, 1963, p. 450)
[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	*See, e.g.*, Salton, 1963, at pp. 448, 450, 451-52 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,

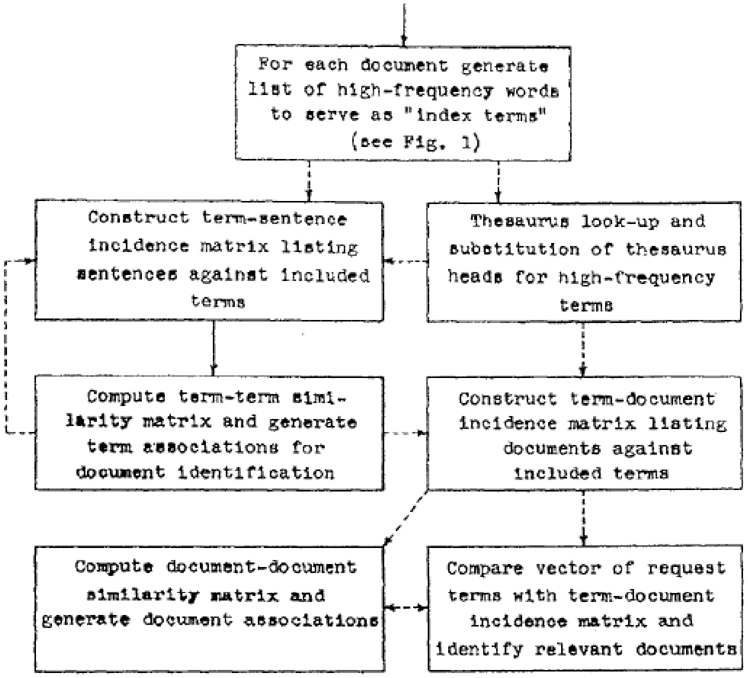
Claim Text from ‘352 Patent	Salton, 1963
among objects;	$[\mathbf{X}']_j^{\,i} = \bigvee_{k=1}^{m} (\mathbf{X}_k^{\,i} \wedge \mathbf{X}_j^{\,k}),$ $[\mathbf{X}'']_j^{\,i} = \bigvee_{k=1}^{m} (\mathbf{X}_k^{\,i} \wedge (\mathbf{X}')_j^{\,k}),$ and so on. Boolean multiplication is used, since the new 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) 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. 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) 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

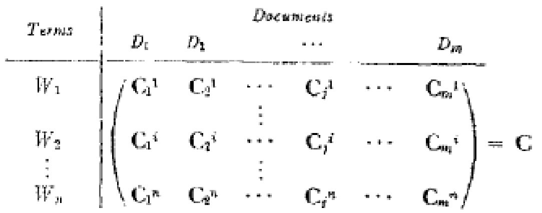
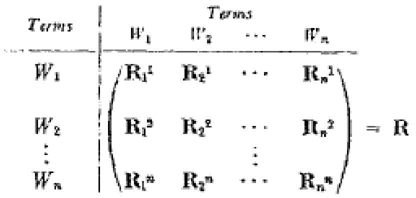
Claim Text from '352 Patent	Salton, 1963
	link exists between A and B. Instead, A and B are then linked by a path of length two, since an extraneous document C exists between documents A and B. Similarly, if the path between two documents includes two extraneous documents, they are linked by a path of length three, and so on. (Salton, 1963, p. 448) 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) Figure 5

Claim Text from '352 Patent	Salton, 1963
	 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

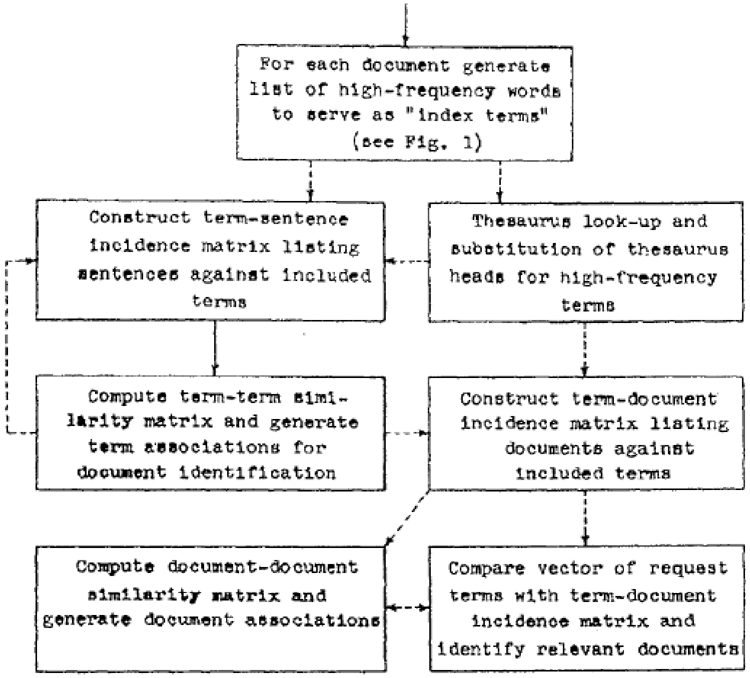
Claim Text from '352 Patent	Salton, 1963
	of length greater than n increases as n becomes larger. Thus more and more documents will exhibit individual similarity coefficients of zero value, thus tending to decrease the value of the overall coefficient. Second, as the length of the links increases and 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)
[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	*See, e.g.,* Salton, 1963, at pp. 443-45, 447-48 Figure 2  $\begin{array}{c\|c} \text{Terms} & \text{Documents} \\ & \begin{array}{cccccc} D_1 & D_2 & & \cdots & & D_m \end{array} \\ \hline \begin{array}{c} W_1 \\ \\ W_2 \\ \vdots \\ W_n \end{array} & \begin{pmatrix} C_1^1 & C_2^1 & \cdots & C_j^1 & \cdots & C_m^1 \\ & & \vdots & & & \\ C_1^i & C_2^i & \cdots & C_j^i & \cdots & C_m^i \\ & & \vdots & & & \\ C_1^n & C_2^n & \cdots & C_j^n & \cdots & C_m^n \end{pmatrix} = \mathbf{C} \end{array}$ (a) Typical term-document incidence matrix C ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times) $\begin{array}{c\|c} \text{Terms} & \text{Terms} \\ & \begin{array}{cccc} W_1 & W_2 & \cdots & W_n \end{array} \\ \hline \begin{array}{c} W_1 \\ \\ W_2 \\ \vdots \\ W_n \end{array} & \begin{pmatrix} R_1^1 & R_2^1 & \cdots & R_n^1 \\ \\ R_1^2 & R_2^2 & \cdots & R_n^2 \\ & & \vdots & \\ R_1^n & R_2^n & \cdots & R_n^n \end{pmatrix} = \mathbf{R} \end{array}$ (b) Typical term-term similarity matrix R $\left(R_j^i = R_i^j = \sum_{k=1}^{m} C_k^i C_k^j \Big/ \sqrt{\left(\sum_{k=1}^{m} (C_k^i)^2 \sum_{k=1}^{m} (C_k^j)^2 \right)} \right)$ FIG. 2. Matrices used for the generation of term associations (Salton, 1963, p. 443)

Claim Text from '352 Patent	Salton, 1963
	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 Rji between term Wi and term Wj 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) 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

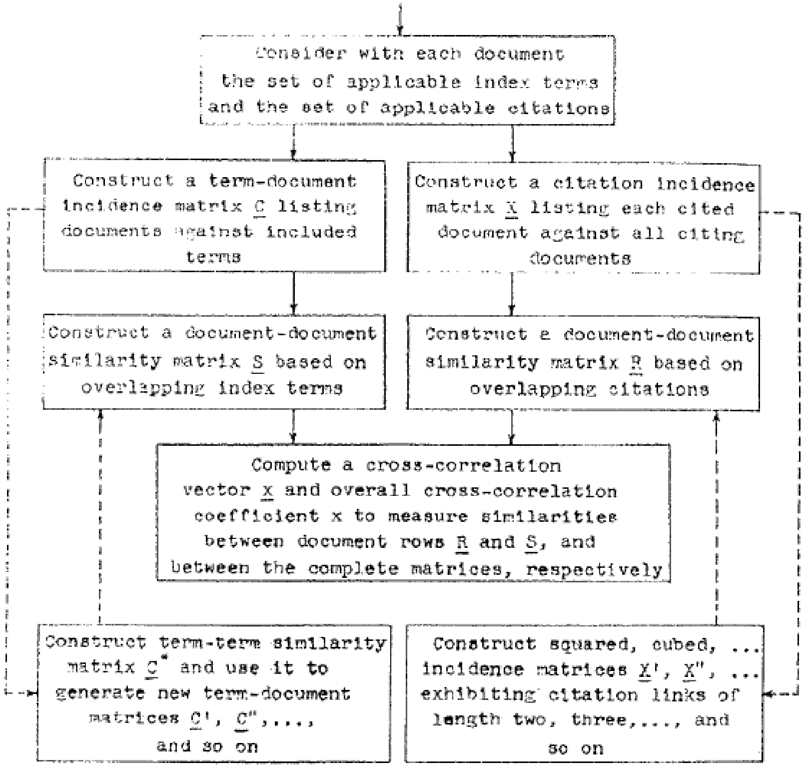
Claim Text from '352 Patent	Salton, 1963
	 FIG. 3. Typical automatic document retrieval system using term and document associations → optional paths → compulsory paths (Salton, 1963, p. 445) To test the significance of bibliographic citations, a comparison was made between citation similarities and index 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 Rij is the value of the similarity coefficient between the ith and jth 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  (a) Typical term-document incidence matrix C ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times)  (b) Typical term-term similarity matrix R $\left(R_j^i = R_i^j = \sum_{k=1}^{m} C_k^i C_k^j \Big/ \sqrt{\left(\sum_{k=1}^{m} (C_k^i)^2 \sum_{k=1}^{m} (C_k^j)^2\right)}\right)$ 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-

Claim Text from ‘352 Patent	Salton, 1963
	dimensional vectors. The similarity coefficients can be displayed in an n x n symmetric term-similarity matrix R, where the coefficient of similarity Rji between term Wi and term Wj is $\displaystyle R_j^i = R_i^j = \frac{\sum_{k=1}^{m} C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^{m} (C_k^i)^2 \sum_{k=1}^{m} (C_k^j)^2\right)}}.$ … 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) A term-term similarity matrix is then constructed by using co-occurrence if 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

Claim Text from ‘352 Patent	Salton, 1963
	 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 Rij is the value of the similarity coefficient between the ith and jth rows (columns) of X. (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 Rij is the value of the similarity coefficient between the ith and jth 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 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) 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) Figure 5

Claim Text from '352 Patent	Salton, 1963
	 FIG. 5. Comparison of citation similarities with index term similarities (Salton, 1963, p. 450)
[41f] storing the fifth numerical representations in the computer database as the index for use in searching for objects in the database.	*See, e.g.,* Salton, 1963, at pp. 440-41, 442, 450 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

Claim Text from ‘352 Patent	Salton, 1963
	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. A typical automatic indexing and abstracting system based on word frequency counts is shown in Figure 1. (Salton, 1963, pp. 440-41) Figure 1

<table>
<tr><th>Claim Text from ‘352 Patent</th><th>Salton, 1963</th></tr>
<tr><td></td><td>

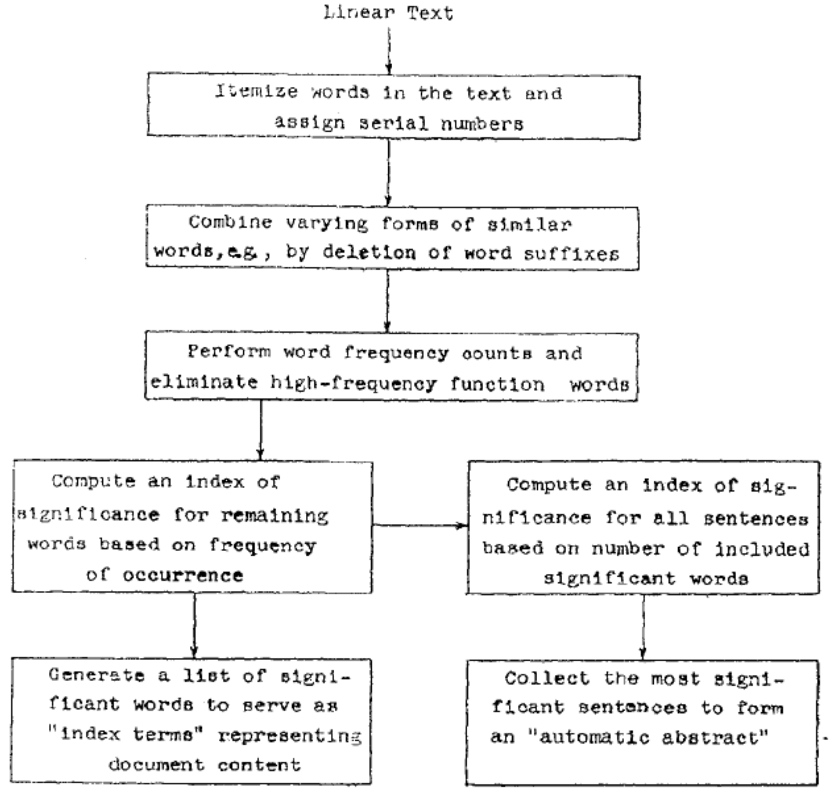


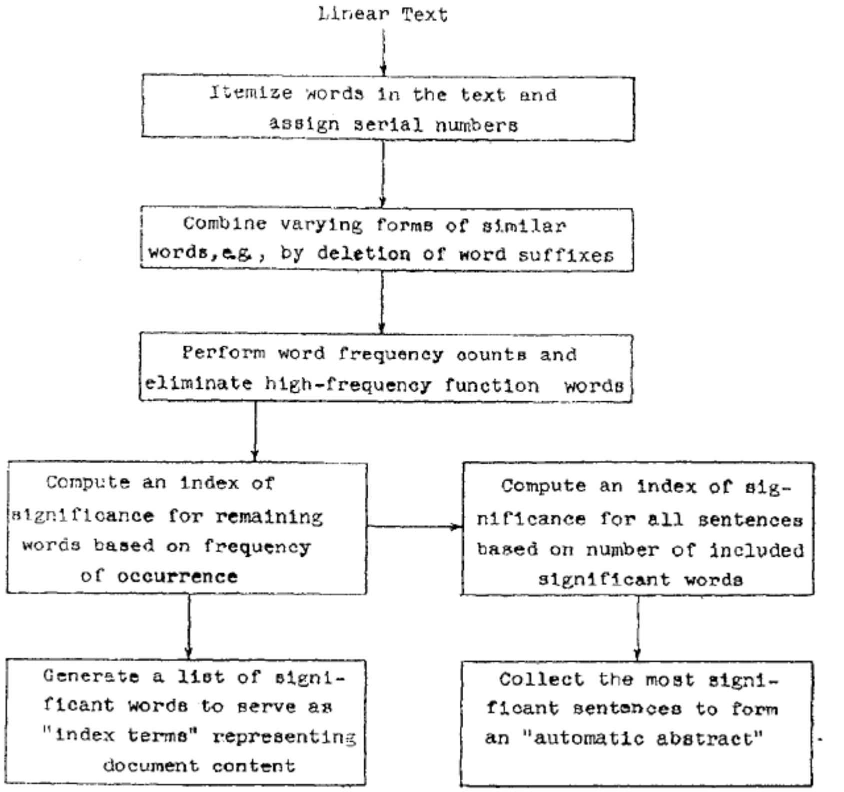
FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts.

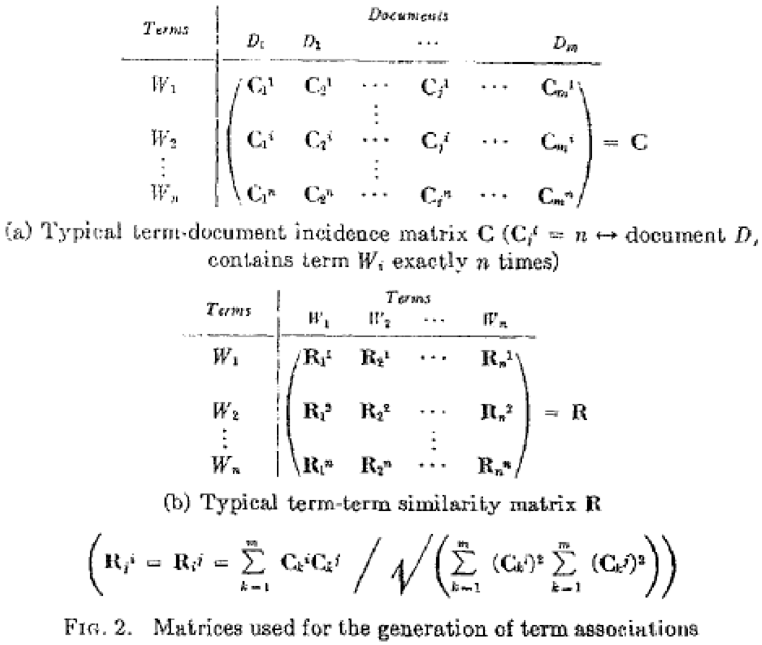
(Salton, 1963, p. 441)

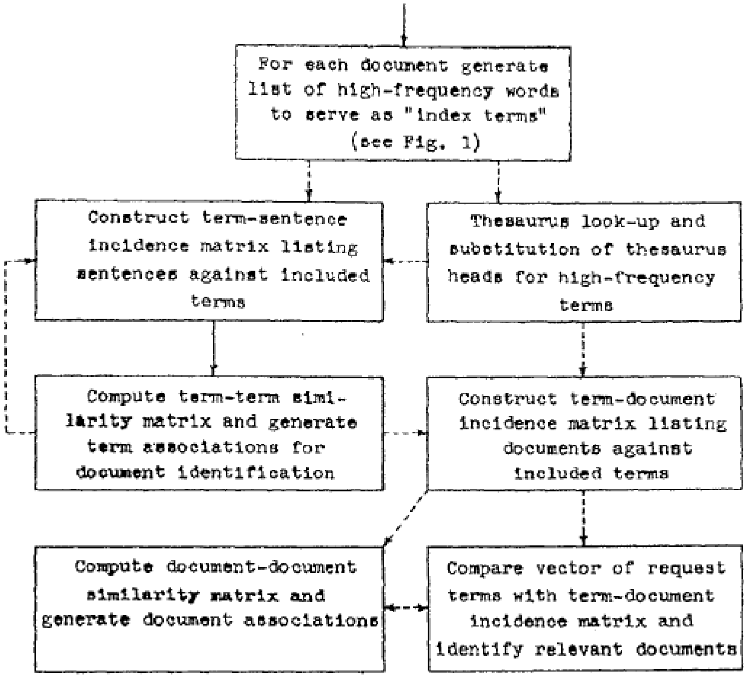
Most associative retrieval systems are based on the statistical word frequency counting procedures previously illustrated in Figure 1. Thus, given a document collection, it is possible to extract a set of n distinct high-frequency words W1, W2, … , Wn, such that each document within the collection is initially identified by some subset of the set of n given words. (Salton, 1963, p. 442)

</td></tr>
</table>

Claim Text from '352 Patent	Salton, 1963
	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)
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.	*See, e.g.,* Salton, 1963, at pp. 441, 443-45, 447, 449-52 Figure 1

Claim Text from ‘352 Patent	Salton, 1963
	 FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts. (Salton, 1963, p. 441) Figure 2

Claim Text from '352 Patent	Salton, 1963
	 $\begin{array}{c\|cccccc} \text{Terms} & & & \text{Documents} & & & \\ & D_1 & D_2 & & \cdots & & D_m \\ \hline W_1 & \mathbf{C}_1^1 & \mathbf{C}_2^1 & \cdots & \mathbf{C}_j^1 & \cdots & \mathbf{C}_m^1 \\ & & & \vdots & & & \\ W_2 & \mathbf{C}_1^i & \mathbf{C}_2^i & \cdots & \mathbf{C}_j^i & \cdots & \mathbf{C}_m^i \\ \vdots & & & \vdots & & & \\ W_n & \mathbf{C}_1^n & \mathbf{C}_2^n & \cdots & \mathbf{C}_j^n & \cdots & \mathbf{C}_m^n \end{array} = \mathbf{C}$ (a) Typical term-document incidence matrix $\mathbf{C}$ ($\mathbf{C}_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times) $\begin{array}{c\|cccc} \text{Terms} & & \text{Terms} & & \\ & W_1 & W_2 & \cdots & W_n \\ \hline W_1 & \mathbf{R}_1^1 & \mathbf{R}_2^1 & \cdots & \mathbf{R}_n^1 \\ W_2 & \mathbf{R}_1^2 & \mathbf{R}_2^2 & \cdots & \mathbf{R}_n^2 \\ \vdots & & & \vdots & \\ W_n & \mathbf{R}_1^n & \mathbf{R}_2^n & \cdots & \mathbf{R}_n^n \end{array} = \mathbf{R}$ (b) Typical term-term similarity matrix $\mathbf{R}$ $\left(\mathbf{R}_j^i = \mathbf{R}_i^j = \sum_{k=1}^{m} \mathbf{C}_k^i \mathbf{C}_k^j \Big/ \sqrt{\left(\sum_{k=1}^{m} (\mathbf{C}_k^i)^2 \sum_{k=1}^{m} (\mathbf{C}_k^j)^2\right)}\right)$ 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 x n symmetric term-similarity matrix R, where the coefficient of similarity Rji between term Wi and term Wj is $\mathbf{R}_j^i = \mathbf{R}_i^j = \frac{\sum_{k=1}^{m} \mathbf{C}_k^i \mathbf{C}_k^j}{\sqrt{\left(\sum_{k=1}^{m} (\mathbf{C}_k^i)^2 \sum_{k=1}^{m} (\mathbf{C}_k^j)^2\right)}}.$

Claim Text from ‘352 Patent	Salton, 1963
	… 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 of more of the other documents in the same collection. Each document can then be represented by an m-dimensional logical vector Xi, where Xij = 1 if and only if document i is cited by document j, and Xij = 0 otherwise. If these m vectors arranged in

Claim Text from ‘352 Patent	Salton, 1963
	rows one below the other a square logical incidence matrix is formed similar to the matrix exhibited in Figure 4. $\begin{array}{c\|cccc} \text{Cited} & \multicolumn{4}{c}{\text{Citing documents}} \\ \text{documents} & D_1 & D_2 & \cdots & D_m \\ \hline D_1 & X_1^1 & X_2^1 & \cdots & X_m^1 \\ D_2 & X_1^2 & X_2^2 & \cdots & X_m^2 \\ \vdots & \vdots & & & \vdots \\ D_m & X_1^m & X_2^m & \cdots & X_m^m \end{array} = X$ ($X_j^i = 1 \leftrightarrow$ document D_i is cited by document D_j) FIG. 4. Matrix X exhibiting direct citations (Salton, 1963, p. 447) 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, $[X']_j^i = \bigvee_{k=1}^{m} (X_k^i \wedge X_j^k),$ $[X'']_j^i = \bigvee_{k=1}^{m} (X_k^i \wedge (X')_j^k),$ 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
	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) 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) 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)
44. The method of claim 41 wherein the step of creating the third numerical representations further comprises the steps of:	*See, e.g.,* Salton, 1963, at pp. 448, 450-52 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

Claim Text from '352 Patent	**Salton, 1963**
	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) 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, $[\mathbf{X}']_j^i = \bigvee_{k=1}^{m} (\mathbf{X}_k^i \wedge \mathbf{X}_j^k),$ $[\mathbf{X}'']_j^i = \bigvee_{k=1}^{m} (\mathbf{X}_k^i \wedge (\mathbf{X}')_j^k),$ 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) 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) 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
	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)
[44a] analyzing the second numerical representation against a plurality of empirically defined patterns, wherein certain patterns are more important than others; and	*See, e.g.,* Salton, 1963, at pp. 444, 448 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) 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)
[44b] weighing the analyzed second numerical representations according to the importance of the patterns.	*See, e.g.,* Salton, 1963, at p. 444 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

Claim Text from ‘352 Patent	Salton, 1963
	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)
45. A method for searching indexed objects, wherein the index is stored, comprising the steps of:	*See, e.g.,* Salton, 1963, at pp. 440-41, 442, 450 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. A typical automatic indexing and abstracting system based on word frequency counts is shown in Figure 1. (Salton, 1963, pp. 440-41) Figure 1

<table>
<tr><th>Claim Text from ‘352 Patent</th><th>Salton, 1963</th></tr>
<tr><td></td><td>

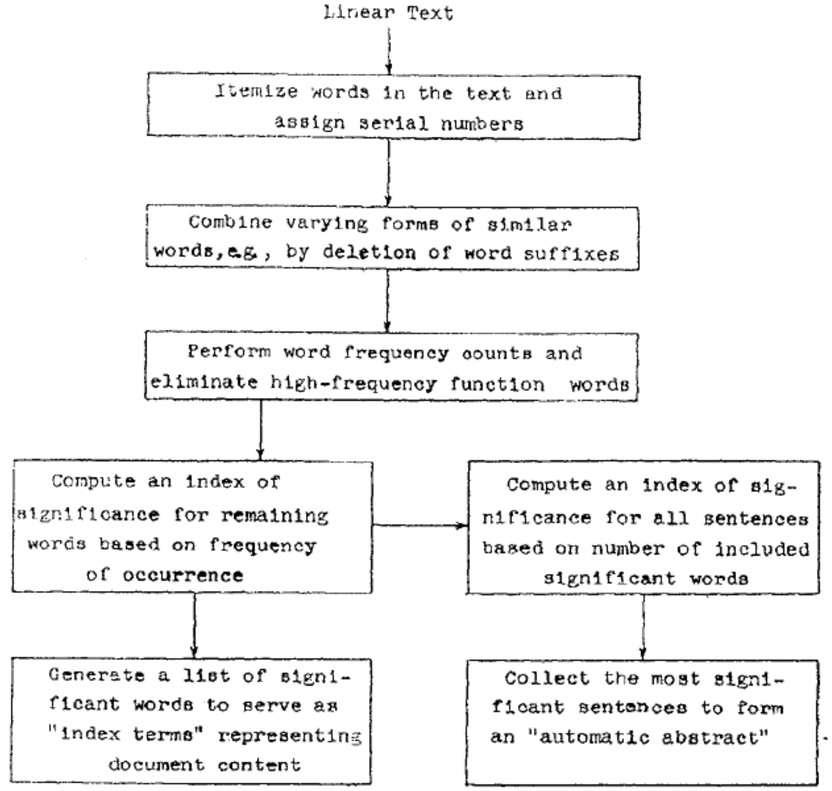


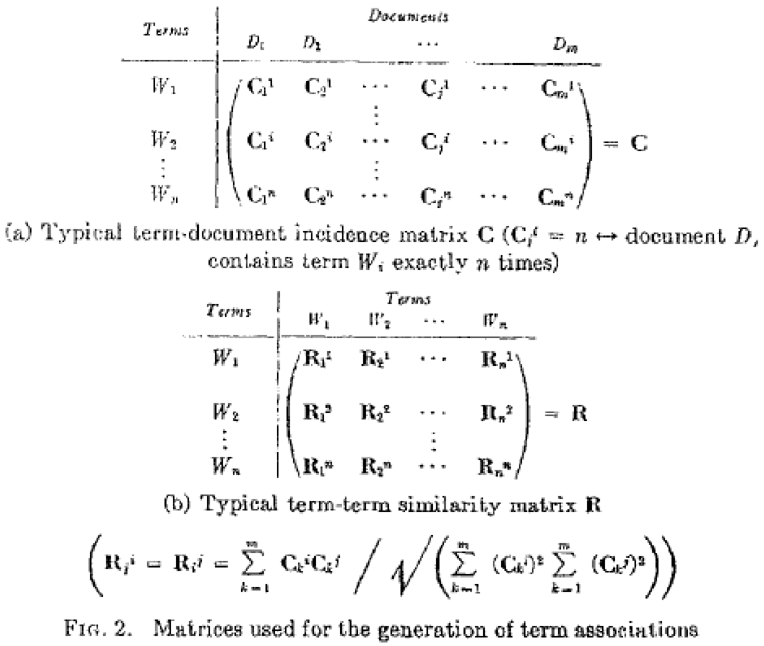
FIG. 1. Typical automatic indexing and abstracting system based on word frequency counts.

(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 W1, W2, … , Wn, such that each document within the collection is initially identified by some subset of the set of n given words. (Salton, 1963, p. 442)

</td></tr>
</table>

Claim Text from ‘352 Patent	**Salton, 1963**
	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)
[45a] entering search commands;	*See, e.g.,* Salton, 1963, at p. 442 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)
[45b] processing the search commands with a processor;	*See, e.g.,* Salton, 1963, at p. 442 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)
[45c] retrieving the stored index using the processor;	*See, e.g.,* Salton, 1963, at pp. 442-46, 449

Claim Text from '352 Patent	Salton, 1963
	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  $\begin{array}{c\|ccccccc} \text{Terms} & & & & \text{Documents} & & & \\ & D_1 & D_2 & & \cdots & & D_m & \\ \hline W_1 & C_1^1 & C_2^1 & \cdots & C_j^1 & \cdots & C_m^1 & \\ W_i & C_1^i & C_2^i & \cdots & C_j^i & \cdots & C_m^i & = \mathbf{C} \\ \vdots & & & \vdots & & & & \\ W_n & C_1^n & C_2^n & \cdots & C_j^n & \cdots & C_m^n & \end{array}$ (a) Typical term-document incidence matrix C ($C_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times) $\begin{array}{c\|cccc} \text{Terms} & & \text{Terms} & & \\ & W_1 & W_2 & \cdots & W_n \\ \hline W_1 & R_1^1 & R_2^1 & \cdots & R_n^1 \\ W_2 & R_1^2 & R_2^2 & \cdots & R_n^2 \\ \vdots & & & \vdots & \\ W_n & R_1^n & R_2^n & \cdots & R_n^n \end{array} = \mathbf{R}$ (b) Typical term-term similarity matrix R $\left(R_j^i = R_i^j = \sum_{k=1}^{m} C_k^i C_k^j \Big/ \sqrt{\left(\sum_{k=1}^{m} (C_k^i)^2 \sum_{k=1}^{m} (C_k^j)^2\right)}\right)$ 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

<table>
<tr><th>Claim Text from ‘352 Patent</th><th>Salton, 1963</th></tr>
<tr><td></td><td>

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

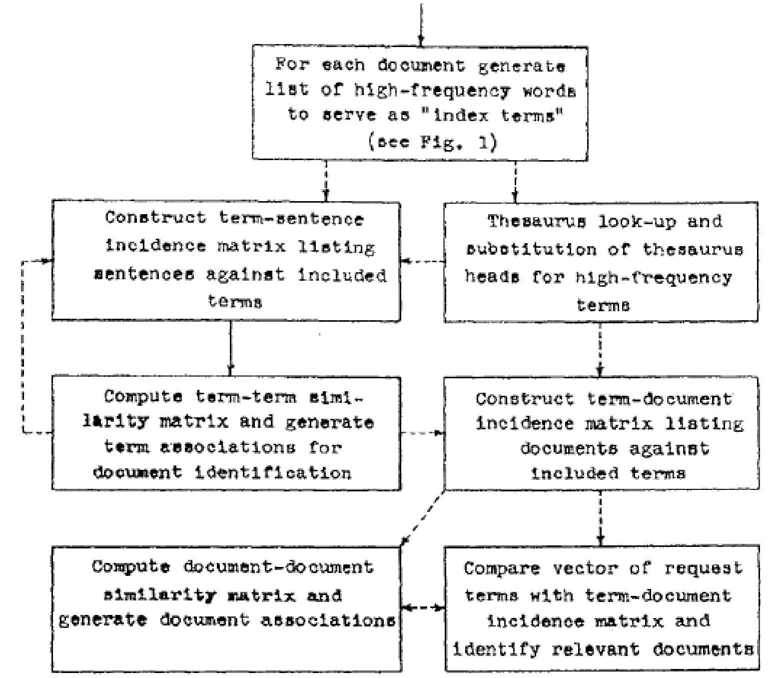


FIG. 3. Typical automatic document retrieval system using term and document associations
→ optional paths → compulsory paths

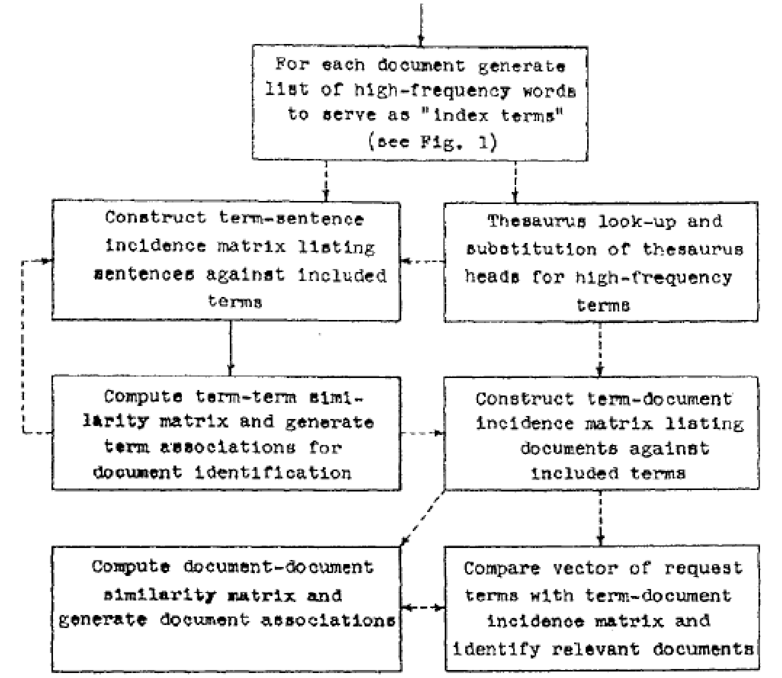
(Salton, 1963, p. 445)

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

</td></tr>
</table>

Claim Text from '352 Patent	Salton, 1963
	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. 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)
[45d] Analyzing the index to identify a pool of objects, comprising the steps of:	*See, e.g.,* Salton, 1963, at pp. 442-46, 449 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

Claim Text from ‘352 Patent	Salton, 1963
	$\begin{array}{ccccccc} & & & & \text{Documents} & & \\ \text{Terms} & D_1 & D_2 & & \cdots & & D_m \\ \hline W_1 & \mathbf{C}_1^1 & \mathbf{C}_2^1 & \cdots & \mathbf{C}_j^1 & \cdots & \mathbf{C}_m^1 \\ & & & \vdots & & & \\ W_2 & \mathbf{C}_1^i & \mathbf{C}_2^i & \cdots & \mathbf{C}_j^i & \cdots & \mathbf{C}_m^i \\ \vdots & & & \vdots & & & \\ W_n & \mathbf{C}_1^n & \mathbf{C}_2^n & \cdots & \mathbf{C}_j^n & \cdots & \mathbf{C}_m^n \end{array} = \mathbf{C}$ (a) Typical term-document incidence matrix $\mathbf{C}$ ($\mathbf{C}_j^i = n \leftrightarrow$ document D_j contains term W_i exactly n times) $\begin{array}{ccccc} & & \text{Terms} & & \\ \text{Terms} & W_1 & W_2 & \cdots & W_n \\ \hline W_1 & \mathbf{R}_1^1 & \mathbf{R}_2^1 & \cdots & \mathbf{R}_n^1 \\ W_2 & \mathbf{R}_1^2 & \mathbf{R}_2^2 & \cdots & \mathbf{R}_n^2 \\ \vdots & & & \vdots & \\ W_n & \mathbf{R}_1^n & \mathbf{R}_2^n & \cdots & \mathbf{R}_n^n \end{array} = \mathbf{R}$ (b) Typical term-term similarity matrix $\mathbf{R}$ $\left(\mathbf{R}_j^i = \mathbf{R}_i^j = \sum_{k=1}^{m} \mathbf{C}_k^i \mathbf{C}_k^j \Big/ \sqrt{\left(\sum_{k=1}^{m} (\mathbf{C}_k^i)^2 \sum_{k=1}^{m} (\mathbf{C}_k^j)^2\right)}\right)$ 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
	 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;	*See, e.g.,* Salton, 1963, at pp. 442, 447 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) 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 Xi , where Xji = 1 if and only if document i is cited by document j, and Xji = 1 otherwise. (Salton, 1963, p. 447)
[45f] identifying a group of objects that have a relationship to the selected object, wherein the step of identifying comprises the steps of:	*See, e.g.,* Salton, 1963, at pp. 443-48, 450 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 Rji between term Wi and term Wj is

Claim Text from '352 Patent	Salton, 1963
	$R_j^i = R_i^j = \frac{\sum_{k=1}^{m} C_k^i C_k^j}{\sqrt{\left(\sum_{k=1}^{m} (C_k^i)^2 \sum_{k=1}^{m} (C_k^j)^2\right)}}.$ … 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) 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) 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) 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 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

Claim Text from '352 Patent	Salton, 1963
	jth rows (columns) of X. (Salton, 1963, p. 448) 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)
[45g] Identifying objects that are referred to by the selected object; and	*See, e.g.*, Salton, 1963, at pp. 445, 446 n.1, 450 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) 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) 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)
[45h] Identifying objects that refer to the selected object	*See, e.g.*, Salton, 1963, at pp. 445, 446 n.1, 450 In particular, it may be conjectured that information associated with the author of a given

Claim Text from '352 Patent	Salton, 1963
	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) 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) 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)
[45i] quantifying the relationship of the selected object to each object in the group of objects; and	*See, e.g.,* Salton, 1963, at pp. 444, 448, 450-52 An estimate of document relevance is then obtained by computing for each document the similarity coefficient between the request column Cm+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) 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) 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,

Claim Text from ‘352 Patent	Salton, 1963
	$[\mathbf{X}']_j^{\,i} = \bigvee_{k=1}^{m} (\mathbf{X}_k^{\,i} \wedge \mathbf{X}_j^{\,k})$, $[\mathbf{X}'']_j^{\,i} = \bigvee_{k=1}^{m} (\mathbf{X}_k^{\,i} \wedge (\mathbf{X}')_j^{\,k})$, 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) 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) 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)

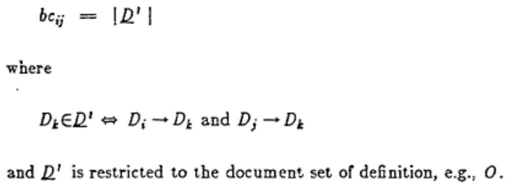
Claim Text from '352 Patent	**Salton, 1963**
[45j] ranking the objects in the group of objects in accordance to the quantified relationship to the selected object; and	*See, e.g.*, Salton, 1963, at p. 444 An estimate of document relevance is then obtained by computing for each document the similarity coefficient between the request column Cm+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)
[45k] presenting one or more objects from the group of objects in ranked order.	*See, e.g.*, Salton, 1963, at p. 444 An estimate of document relevance is then obtained by computing for each document the similarity coefficient between the request column Cm+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)

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:	*See infra; see also, e.g.*, Chapters 1, & 6-9.
Selecting a node for analysis;	*Id.* 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 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; *see also, e.g.*, 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;	*Id.* 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,

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>


$$bc_{ij} = |\underline{D}'|$$

where

$$D_k \in \underline{D}' \Leftrightarrow D_i \to D_k \text{ and } D_j \to D_k$$

and $\underline{D}'$ is restricted to the document set of definition, e.g., O.

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.”,

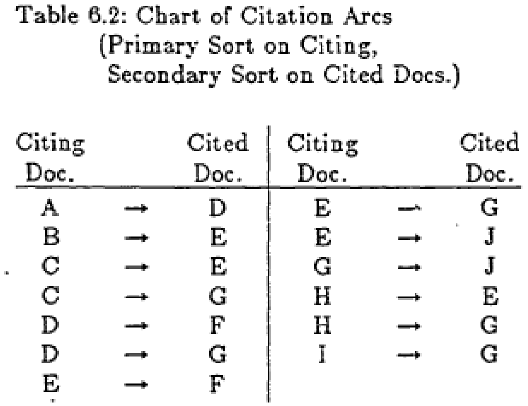
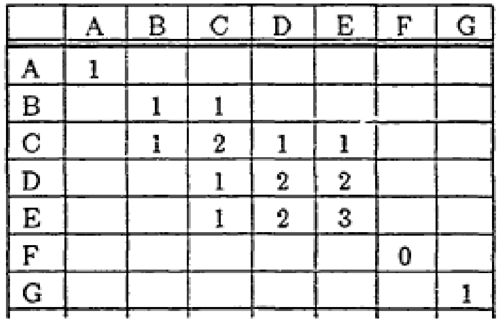


Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<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>
</table>
</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>
Figure 6.3: <u>bc</u> Submatrix

<table>
<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>
<tr><td>A</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>1</td><td>1</td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td>1</td><td>2</td><td>1</td><td>1</td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td>1</td><td>2</td><td>2</td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td>1</td><td>2</td><td>3</td><td></td><td></td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>0</td><td></td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td></tr>
</table>
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

$$cc_{ij} = |\underline{D}''|$$
where
$$\underline{D}'' \subseteq C,$$
the source set of documents considered, and
$$D_k \in \underline{D}'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j .$$
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

$cc_{E,G} = 2$ $cc_{F,G} = 2$ $cc_{F,J} = 1$,”
</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

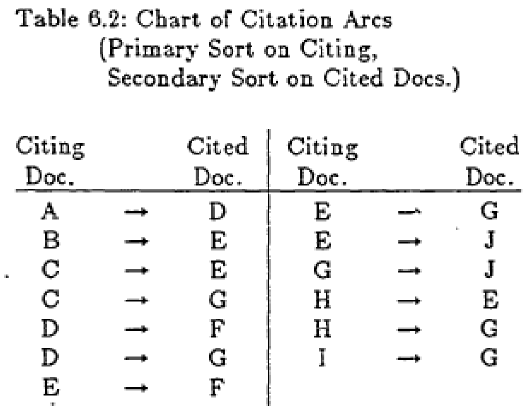



Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)
<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<tr><td>A</td><td>→</td><td>D</td><td>E</td><td>→</td><td>G</td></tr>
<tr><td>B</td><td>→</td><td>E</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>
</table>

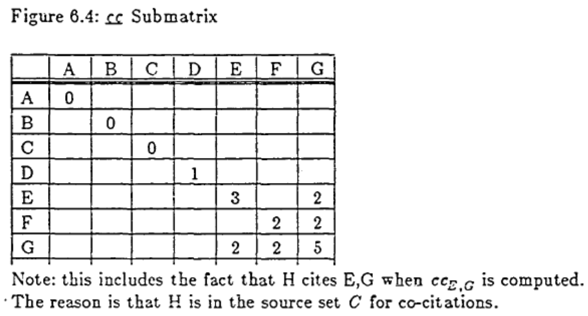
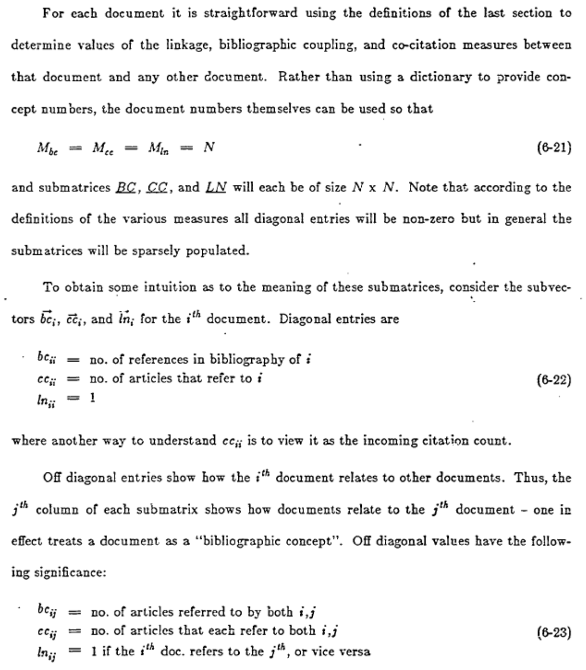



Figure 6.4: <u>cc</u> Submatrix
<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th><th>F</th><th>G</th></tr>
<tr><td>A</td><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>0</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>0</td><td></td><td></td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td></td><td></td><td>3</td><td></td><td>2</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>2</td><td>2</td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td>2</td><td>2</td><td>5</td></tr>
</table>
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:
</td></tr>
</table>

Claim Text for ’494 Patent	Fox Thesis, 1983
	 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 $M_{bc} = M_{cc} = M_{ln} = N$ (6-21) and submatrices $\underline{BC}$, $\underline{CC}$, and $\underline{LN}$ will each be of size N x 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 bc_{ii} = no. of references in bibliography of i cc_{ii} = no. of articles that refer to i (6-22) ln_{ii} = 1 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: bc_{ij} = no. of articles referred to by both i,j cc_{ij} = no. of articles that each refer to both i,j (6-23) ln_{ij} = 1 if the i^{th} doc. refers to the j^{th}, or vice versa ,” pp. 171-182, 205-206, 240 (Figure 8.2, Sample computations of inner products); *see also, e.g.*, Chapter 1, Chapters 6-9.
Deriving actual cluster links from the candidate cluster links;	*Id.* 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; *see also, e.g.*, Chapters 1, & 6-9, Charts for the preceding limitation (including the quotations and descriptions set forth therein, which are incorporated by reference herein).
identifying one or more nodes for display; and	*Id.* 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

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 <u>bc</u>, <u>cc</u>, and <u>ln</u> 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; *see also, e.g.*, 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.	*Id*. 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.” *See* chapters 5 and 8; *see also, e.g.*, 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:	*See infra*; *see also, e.g.*, Chapters 1, & 6-9, Chart for Claim 1, *supra* (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, *supra* (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, *supra* (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	*Id.* 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

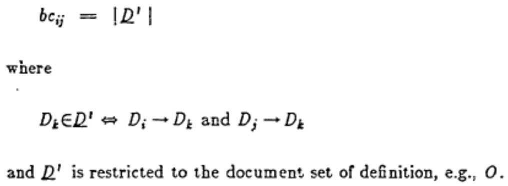
Claim Text for ’494 Patent	Fox Thesis, 1983
under analysis.	more concern here, bibliographic information such as direct references between documents and other derived measures such as those of bibliographic coupling and co-citation strength can be employed . . . The bibliographic measures described here have been useful in both retrieval and clustering applications.”), 160-172, 173: “The use of <u>bc</u>, <u>cc</u>, and <u>ln</u> 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; *see also, e.g.*, Chapters 1, 7-9, Chart for Claim 1, *supra* (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.	*Id.* 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 <u>bc</u>, <u>cc</u>, and <u>ln</u> 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; *see also, e.g.*, Chapters 1, 7-9, Chart for Claim 1, *supra* (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:	*See infra; see also, e.g.*, Chapters 1, & 6-9.
Selecting an object to determine the proximity of other objects to the selected object;	*Id.* 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 <u>bc</u>, <u>cc</u>, and <u>ln</u> 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; *see also, e.g.*, Chapters 1, 7-9.

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td>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;</td><td>

Id. 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,



$$bc_{ij} = |\underline{D}'|$$

where

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

and $\underline{D}'$ is restricted to the document set of definition, e.g., O.

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.”,

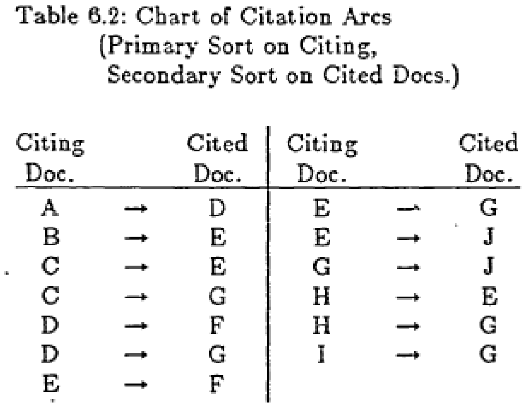
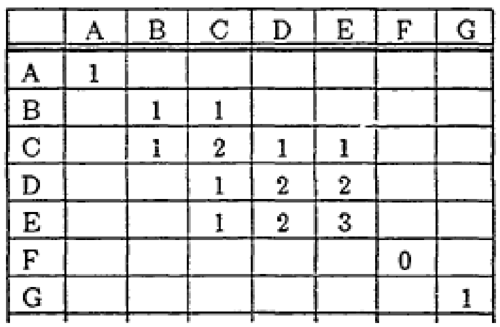


Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<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>
</table>

</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>
Figure 6.3: bc Submatrix

<table>
<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>
<tr><td>A</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>1</td><td>1</td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td>1</td><td>2</td><td>1</td><td>1</td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td>1</td><td>2</td><td>2</td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td>1</td><td>2</td><td>3</td><td></td><td></td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>0</td><td></td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td></tr>
</table>
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
$$cc_{ij} = |\underline{D}''|$$
where
$$\underline{D}'' \subseteq C,$$
the source set of documents considered, and
$$D_k \in \underline{D}'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j.$$
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
$cc_{E,G} = 2$ $cc_{F,G} = 2$ $cc_{F,J} = 1$,”
</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

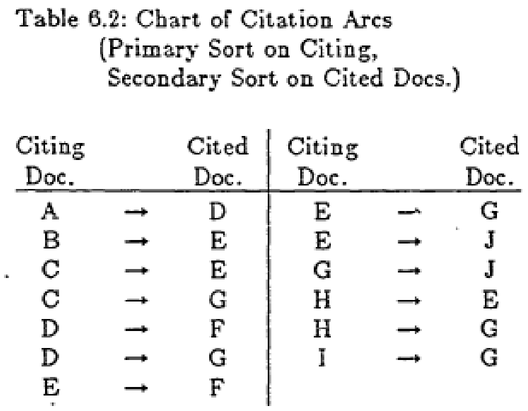



Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)
<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<tr><td>A</td><td>→</td><td>D</td><td>E</td><td>→</td><td>G</td></tr>
<tr><td>B</td><td>→</td><td>E</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>
</table>

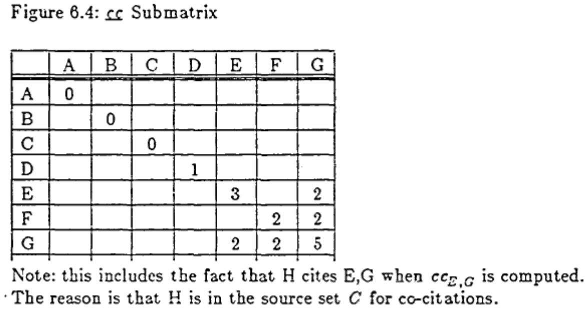
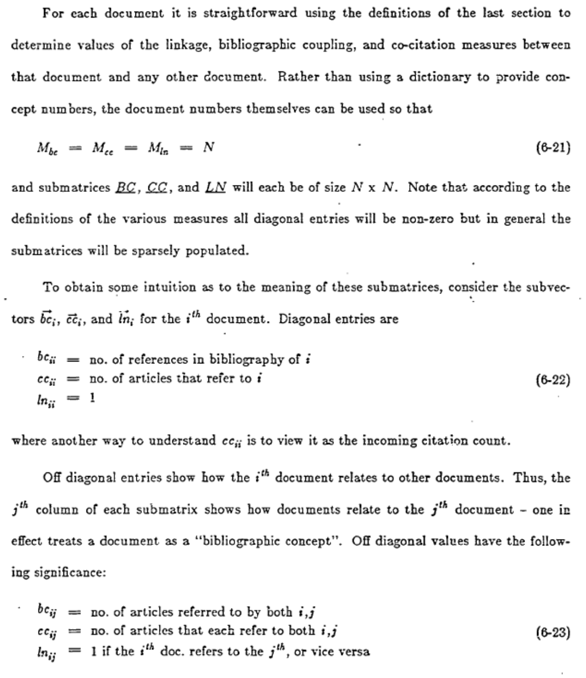
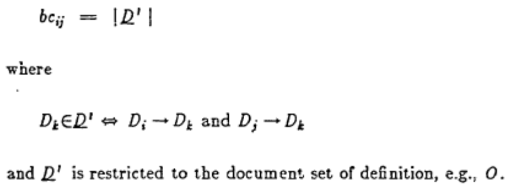



Figure 6.4: cc Submatrix
<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th><th>F</th><th>G</th></tr>
<tr><td>A</td><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>0</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>0</td><td></td><td></td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td></td><td></td><td>3</td><td></td><td>2</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>2</td><td>2</td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td>2</td><td>2</td><td>5</td></tr>
</table>
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:
</td></tr>
</table>

Claim Text for ’494 Patent	Fox Thesis, 1983
	 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 $M_{bc} = M_{cc} = M_{ln} = N$ (6-21) and submatrices $\underline{BC}$, $\underline{CC}$, and $\underline{LN}$ will each be of size N x 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 bc_{ii} = no. of references in bibliography of i cc_{ii} = no. of articles that refer to i (6-22) ln_{ii} = 1 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: bc_{ij} = no. of articles referred to by both i,j cc_{ij} = no. of articles that each refer to both i,j (6-23) ln_{ij} = 1 if the i^{th} doc. refers to the j^{th}, or vice versa ,” pp. 171-182, 205-206, p. 240 (Figure 8.2, Sample computations of inner products); *see also, e.g.*, Chapter 1, Chapters 6-9.
Deriving an actual cluster link set for the selected object using the generated candidate cluster link set; and	*Id.* 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 <u>bc</u>, <u>cc</u>, and <u>ln</u> 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

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; *see also, e.g.*, 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.	*Id.* 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.” *See* chapters 5 and 8; *see also, e.g.*, 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.	*Id.* 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,  $bc_{ij} = \lvert \underline{D}' \rvert$ where $D_k \in \underline{D}' \Leftrightarrow D_i \rightarrow D_k$ and $D_j \rightarrow D_k$ and $\underline{D}'$ is restricted to the document set of definition, e.g., O. 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.”,

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

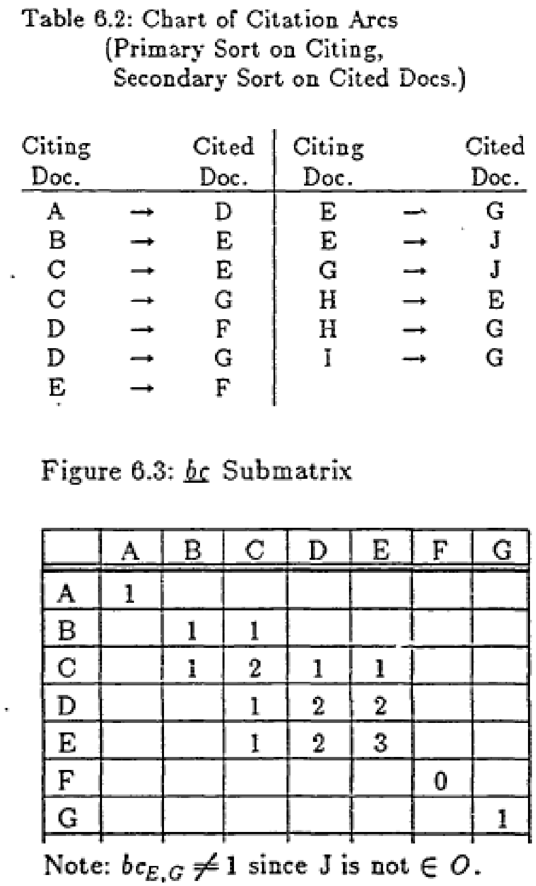
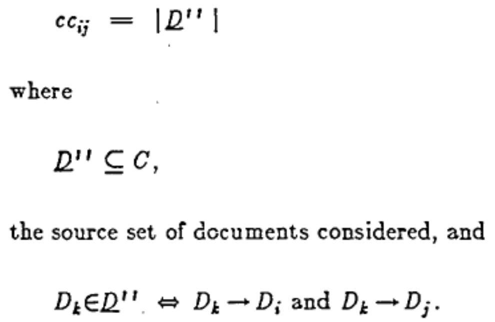



Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)
<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<tr><td>A</td><td>→</td><td>D</td></tr>
<tr><td>B</td><td>→</td><td>E</td></tr>
<tr><td>C</td><td>→</td><td>E</td></tr>
<tr><td>C</td><td>→</td><td>G</td></tr>
<tr><td>D</td><td>→</td><td>F</td></tr>
<tr><td>D</td><td>→</td><td>G</td></tr>
<tr><td>E</td><td>→</td><td>F</td></tr>
<tr><td>E</td><td>→</td><td>G</td></tr>
<tr><td>E</td><td>→</td><td>J</td></tr>
<tr><td>G</td><td>→</td><td>J</td></tr>
<tr><td>H</td><td>→</td><td>E</td></tr>
<tr><td>H</td><td>→</td><td>G</td></tr>
<tr><td>I</td><td>→</td><td>G</td></tr>
</table>
Figure 6.3: <u>bc</u> Submatrix
<table>
<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>
<tr><td>A</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>1</td><td>1</td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td>1</td><td>2</td><td>1</td><td>1</td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td>1</td><td>2</td><td>2</td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td>1</td><td>2</td><td>3</td><td></td><td></td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>0</td><td></td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td></tr>
</table>
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
</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>



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

where

$$\underline{D}'' \subseteq C,$$

the source set of documents considered, and

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

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

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

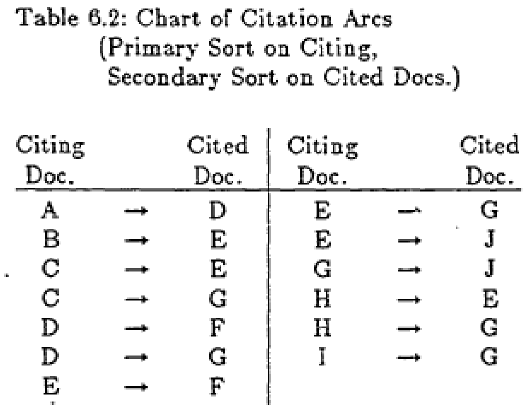


Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<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>
</table>

</td></tr>
</table>

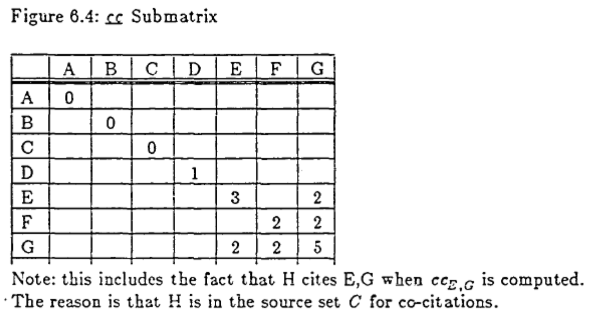
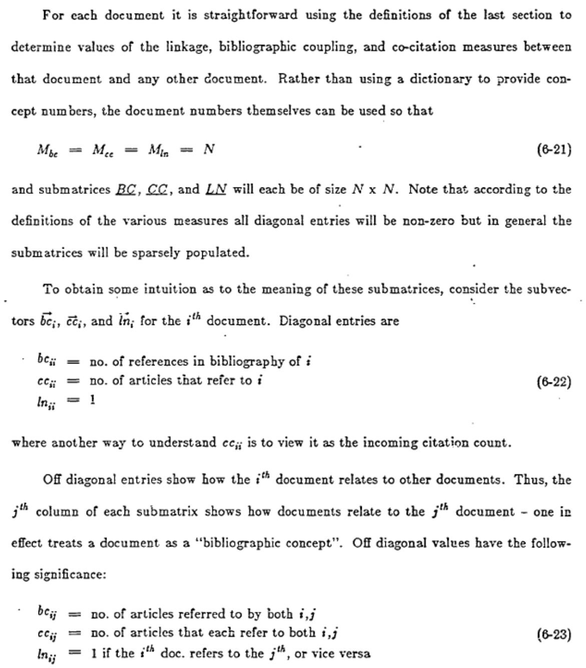
<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>


Figure 6.4: $\underline{cc}$ Submatrix

<table>
<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>
<tr><td>A</td><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>0</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>0</td><td></td><td></td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td></td><td></td><td>3</td><td></td><td>2</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>2</td><td>2</td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td>2</td><td>2</td><td>5</td></tr>
</table>

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

$$M_{bc} = M_{cc} = M_{ln} = N \qquad (6\text{-}21)$$

and submatrices $\underline{BC}$, $\underline{CC}$, and $\underline{LN}$ will each be of size N x 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

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

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

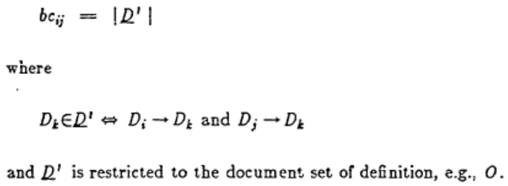
Off diagonal entries show how the i^{th} document relates to other documents. Thus, the j^{th} column of each submatrix shows how documents relate to the j^{th} document – one in effect treats a document as a “bibliographic concept”. Off diagonal values have the following significance:

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

, 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.
</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>199-201 (“Clustering Process”)); *see also, e.g.*, Chapters 1, 6, 8-9.</td></tr>
<tr><td>14. A method for representing the relationship between nodes using stored direct links, paths, and candidate cluster links, comprising the steps of:</td><td>*See infra*; *see also, e.g.*, Chapters 1, & 6-9.</td></tr>
<tr><td>initializing a set of candidate cluster links;</td><td>

Id. 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,



$$bc_{ij} = |\underline{D}'|$$

where

$$D_k \in \underline{D}' \Leftrightarrow D_i \to D_k \text{ and } D_j \to D_k$$

and $\underline{D}'$ is restricted to the document set of definition, e.g., O.

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.”,

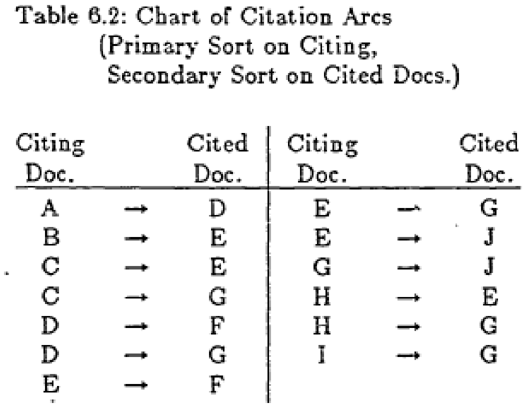
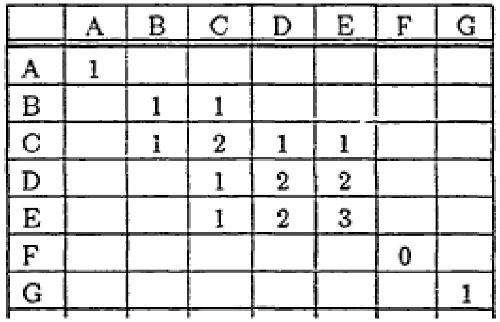


Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<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>
</table>

</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>
Figure 6.3: $\underline{bc}$ Submatrix



<table>
<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>
<tr><td>A</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>1</td><td>1</td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td>1</td><td>2</td><td>1</td><td>1</td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td>1</td><td>2</td><td>2</td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td>1</td><td>2</td><td>3</td><td></td><td></td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>0</td><td></td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td></tr>
</table>

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

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

where

$$\underline{D}'' \subseteq C,$$

the source set of documents considered, and

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

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

$cc_{E,G} = 2$ $cc_{F,G} = 2$ $cc_{F,J} = 1$,”
</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

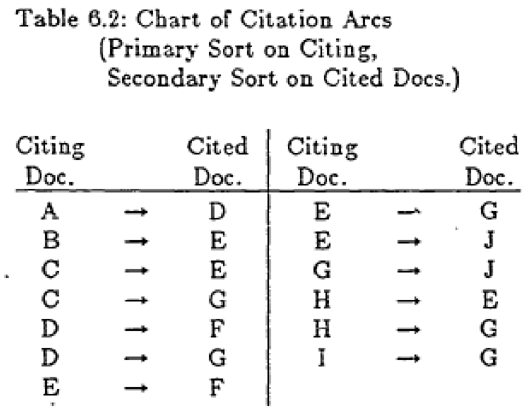



Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)
<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<tr><td>A</td><td>→</td><td>D</td><td>E</td><td>→</td><td>G</td></tr>
<tr><td>B</td><td>→</td><td>E</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>
</table>

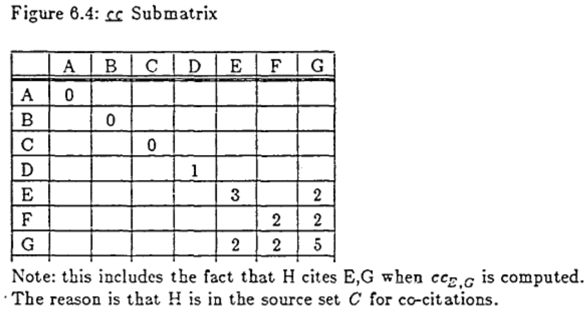
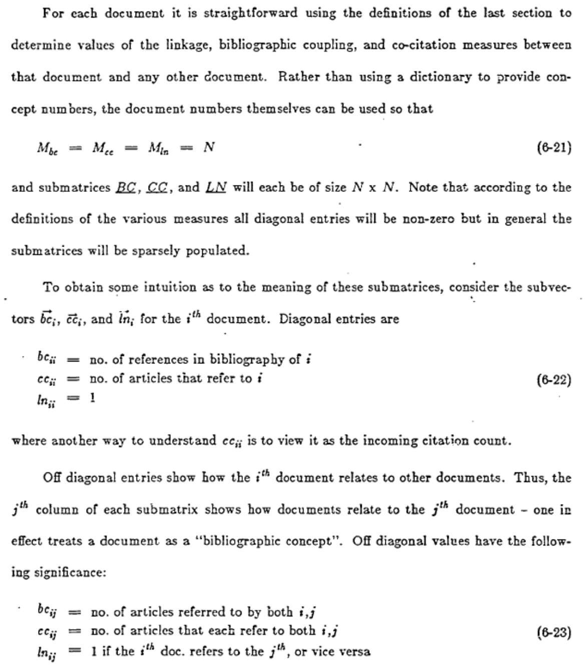
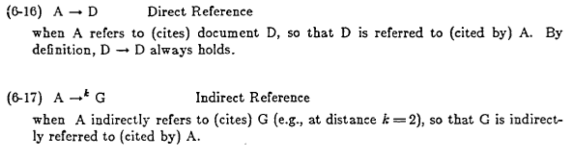



Figure 6.4: <u>cc</u> Submatrix
<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th><th>F</th><th>G</th></tr>
<tr><td>A</td><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>0</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>0</td><td></td><td></td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td></td><td></td><td>3</td><td></td><td>2</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>2</td><td>2</td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td>2</td><td>2</td><td>5</td></tr>
</table>
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:
</td></tr>
</table>

Claim Text for ’494 Patent	Fox Thesis, 1983
	 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 $M_{bc} = M_{cc} = M_{ln} = N$ (6-21) and submatrices $\underline{BC}$, $\underline{CC}$, and $\underline{LN}$ will each be of size N x 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 bc_{ii} = no. of references in bibliography of i cc_{ii} = no. of articles that refer to i (6-22) ln_{ii} = 1 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: bc_{ij} = no. of articles referred to by both i,j cc_{ij} = no. of articles that each refer to both i,j (6-23) ln_{ij} = 1 if the i^{th} doc. refers to the j^{th}, or vice versa ,” pp. 171-182, 205-206, p. 240 (Figure 8.2, Sample computations of inner products); *see also, e.g.*, Chapter 1, Chapters 6-9.
Selecting the destination node of a path as the selected node to analyze;	*Id*. 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:

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>


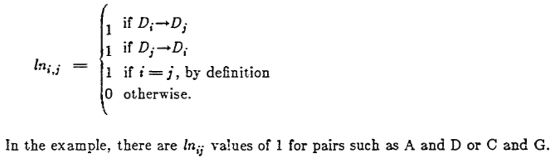
(6-16) A → D Direct Reference

when A refers to (cites) document D, so that D is referred to (cited by) A. By definition, D → D always holds.

(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->D or D-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,



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

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

,” Figure 6.5:

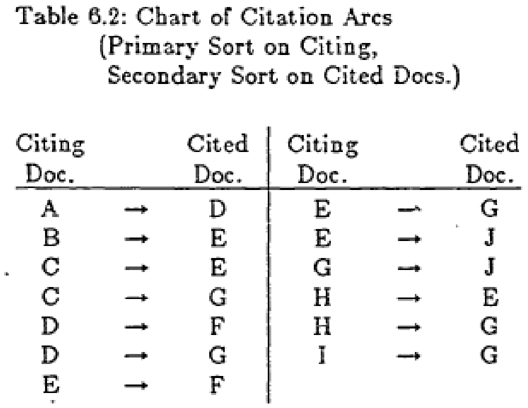
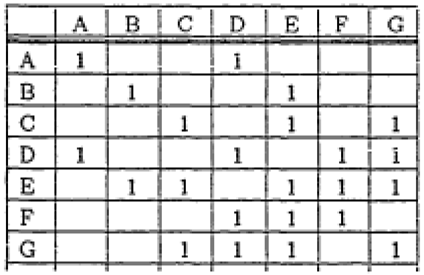
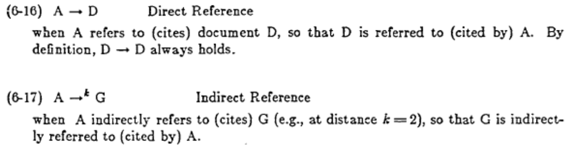
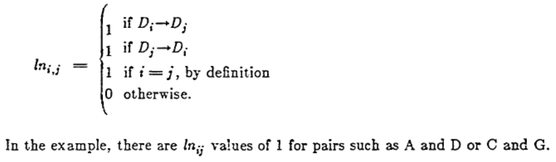


Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<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>
</table>
</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>
Figure 6.5: <u>ln</u> Submatrix

<table>
<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>
<tr><td>A</td><td>1</td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>1</td><td></td><td></td><td>1</td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>1</td><td></td><td>1</td><td></td><td>1</td></tr>
<tr><td>D</td><td>1</td><td></td><td></td><td>1</td><td></td><td>1</td><td>1</td></tr>
<tr><td>E</td><td></td><td>1</td><td>1</td><td></td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td></tr>
<tr><td>G</td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td><td>1</td></tr>
</table>
, 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.
</td></tr>
<tr><td>retrieving the set of direct links from the selected node to any other node in the database;</td><td>
<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:

(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.
(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->D or D-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,
</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>


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

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

,” Figure 6.5:

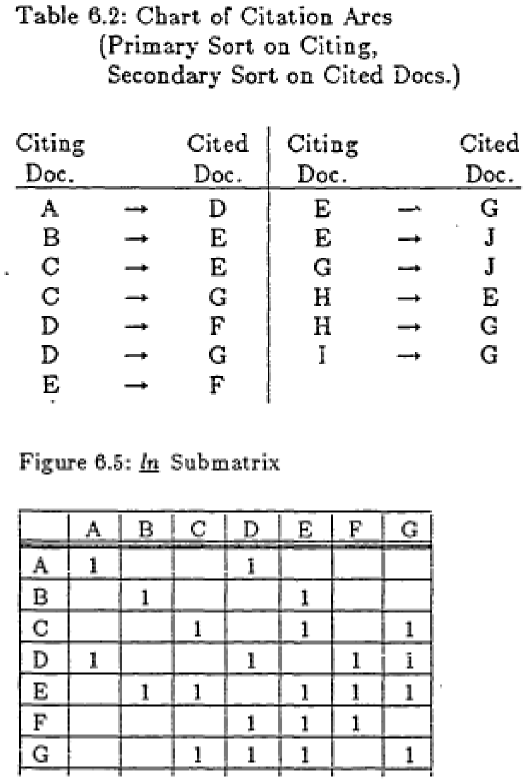


Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

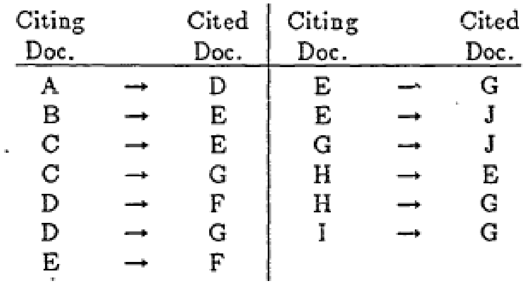
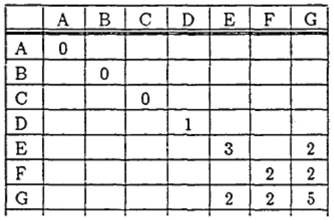
<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<tr><td>A</td><td>→</td><td>D</td><td>E</td><td>→</td><td>G</td></tr>
<tr><td>B</td><td>→</td><td>E</td><td>E</td><td>→</td><td>J</td></tr>
<tr><td>C</td><td>→</td><td>E</td><td>G</td><td>→</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>
</table>

Figure 6.5: <u>ln</u> Submatrix

<table>
<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>
<tr><td>A</td><td>1</td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>1</td><td></td><td></td><td>1</td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>1</td><td></td><td>1</td><td></td><td>1</td></tr>
<tr><td>D</td><td>1</td><td></td><td></td><td>1</td><td></td><td>1</td><td>1</td></tr>
<tr><td>E</td><td></td><td>1</td><td>1</td><td></td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td></tr>
<tr><td>G</td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td><td>1</td></tr>
</table>

, 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.
</td></tr>
<tr><td>Determining the weight of the path using the retrieved direct links;</td><td>*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,</td></tr>
</table>

Claim Text for ’494 Patent	Fox Thesis, 1983
	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. 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 $cc_{ij} = \lvert \underline{D}'' \rvert$ where $\underline{D}'' \subseteq C,$ the source set of documents considered, and $D_k \in \underline{D}'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j .$ 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>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)
<table><tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr><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></table>Figure 6.4: <u>cc</u> Submatrix
<table><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><tr><td>A</td><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>B</td><td></td><td>0</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>C</td><td></td><td></td><td>0</td><td></td><td></td><td></td><td></td></tr><tr><td>D</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr><tr><td>E</td><td></td><td></td><td></td><td></td><td>3</td><td></td><td>2</td></tr><tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>2</td><td>2</td></tr><tr><td>G</td><td></td><td></td><td></td><td></td><td>2</td><td>2</td><td>5</td></tr></table>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.
See also, e.g., Chapters 1, & 6-9.</td></tr>
<tr><td>repeating steps b through d for each path; and</td><td>*See* Charts for previous limitations.</td></tr>
<tr><td>Storing the determined weights as candidate cluster links.</td><td>*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, 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</td></tr>
</table>

Claim Text for ’494 Patent	Fox Thesis, 1983
	E, is referred to by both B and C. 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 $$cc_{ij} = \lvert \underline{D}'' \rvert$$ where $$\underline{D}'' \subseteq C,$$ the source set of documents considered, and $$D_k \in \underline{D}'' \Leftrightarrow D_k \rightarrow D_i \text{ and } D_k \rightarrow D_j .$$ 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 $cc_{E,G} = 2$ $cc_{F,G} = 2$ $cc_{F,J} =1$,”

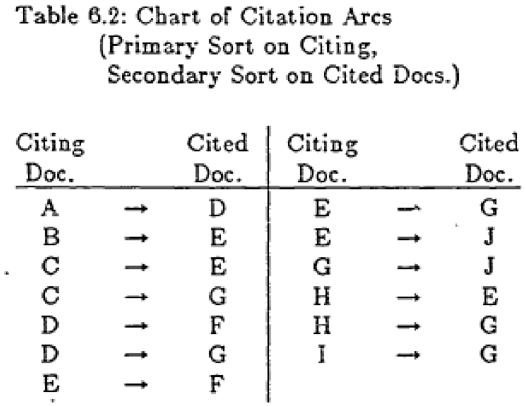
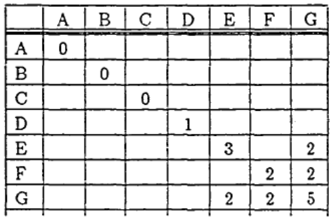
<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>


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

<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<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>
</table>

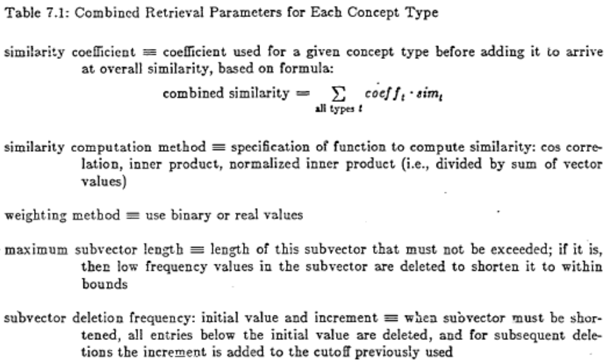
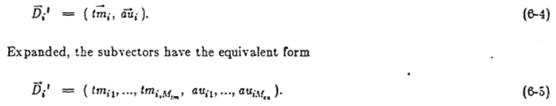
Figure 6.4: $\underline{cc}$ Submatrix



<table>
<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>
<tr><td>A</td><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>0</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>0</td><td></td><td></td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td></td><td></td><td>3</td><td></td><td>2</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>2</td><td>2</td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td>2</td><td>2</td><td>5</td></tr>
</table>

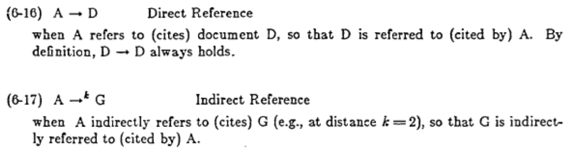
Note: this includes the fact that H cites E,G when $cc_{E,G}$ is computed. The reason is that H is in the source set C for co-citations.

See also, e.g., Chapters 1, & 6-9).
</td></tr>
<tr><td>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.</td><td><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).</td></tr>
<tr><td>16. The method of claim 15 wherein the step of deriving comprises the step of choosing the top rated candidate cluster links.</td><td><i>Id.</i> at 193: “The algorithm produces a hierarchical clustering where all <i>N</i> 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</td></tr>
</table>

Claim Text for ’494 Patent	Fox Thesis, 1983
	the children immediately below regardless of whether those are documents or other centroids. 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. 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.  Table 7.1: Combined Retrieval Parameters for Each Concept Type similarity coefficient ≡ coefficient used for a given concept type before adding it to arrive at overall similarity, based on formula: $$\text{combined similarity} = \sum_{\text{all types } t} coeff_t \cdot sim_t$$ similarity computation method ≡ specification of function to compute similarity: cos correlation, inner product, normalized inner product (i.e., divided by sum of vector values) weighting method ≡ use binary or real values 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 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 ”; *see also, e.g.*, Chapters 1, & 6-9.
18. A method of analyzing a database having objects and a first numerical representation of direct relationships in the database, comprising the steps of:	*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  $$\vec{D}_i' = (\vec{tm}_i, \vec{au}_i). \quad (6\text{-}4)$$ Expanded, the subvectors have the equivalent form $$\vec{D}_i' = (tm_{i1}, \ldots, tm_{i,M_{im}}, au_{i1}, \ldots, au_{iM_{ia}}). \quad (6\text{-}5)$$,” p. 159: “In addition to terms and authors, other types of information are available in many collections. Dates and

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

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:



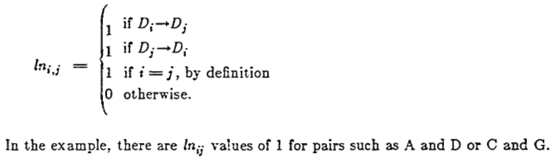
(6-16) A → D Direct Reference

when A refers to (cites) document D, so that D is referred to (cited by) A. By definition, D → D always holds.

(6-17) $A \to^k G$ 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->D or D-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,



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

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

,” Figure 6.5:

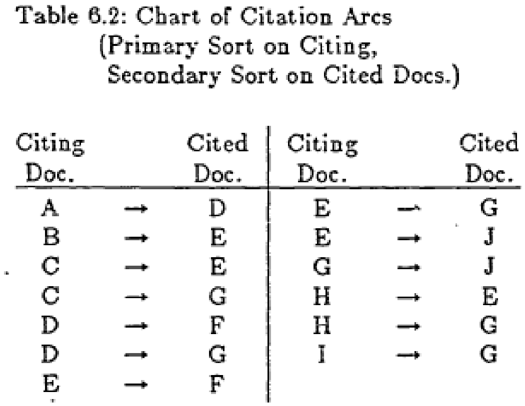


Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

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			

</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

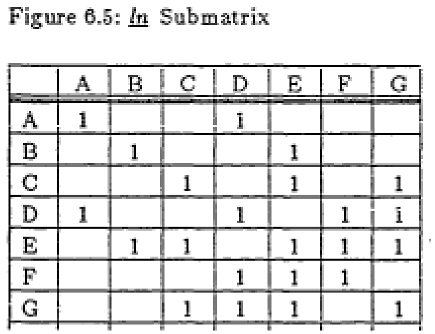
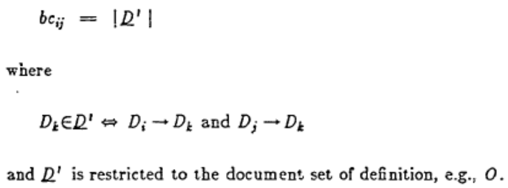



Figure 6.5: <u><i>ln</i></u> Submatrix
<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th><th>F</th><th>G</th></tr>
<tr><td>A</td><td>1</td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>1</td><td></td><td></td><td>1</td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>1</td><td></td><td>1</td><td></td><td>1</td></tr>
<tr><td>D</td><td>1</td><td></td><td></td><td>1</td><td></td><td>1</td><td>1</td></tr>
<tr><td>E</td><td></td><td>1</td><td>1</td><td></td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td></tr>
<tr><td>G</td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td><td>1</td></tr>
</table>
, 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, 7-9, Appendix C.
</td></tr>
<tr><td>generating a second numerical representation using the first numerical representation, wherein the second numerical representation accounts for indirect relationships in the database;</td><td>
<i>Id.</i> at Chapter 6 (<i>e.g., pp. 159-164,</i> 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,


$$bc_{ij} = |\underline{D}'|$$

where

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

and $\underline{D}'$ is restricted to the document set of definition, e.g., O.

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.”,
</td></tr>
</table>

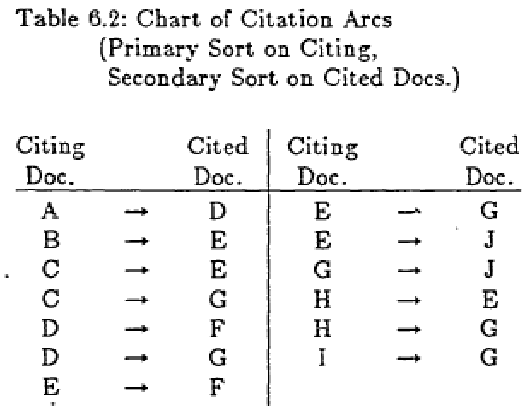
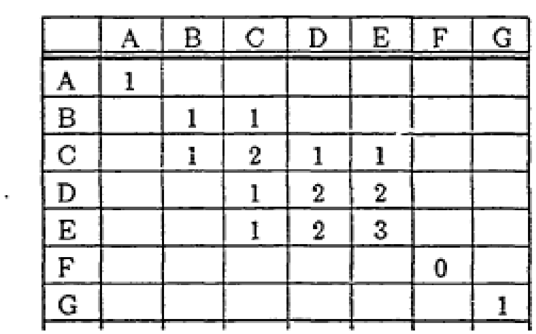
<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>


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

<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<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>
</table>

Figure 6.3: <u>bc</u> Submatrix



<table>
<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>
<tr><td>A</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>1</td><td>1</td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td>1</td><td>2</td><td>1</td><td>1</td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td>1</td><td>2</td><td>2</td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td>1</td><td>2</td><td>3</td><td></td><td></td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>0</td><td></td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td></tr>
</table>

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
</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

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

where

$$\underline{D}'' \subseteq C,$$

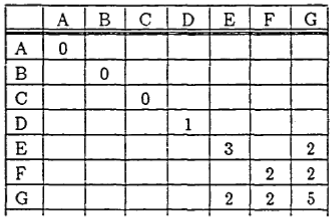
the source set of documents considered, and

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

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,”

Figure 6.4: cc Submatrix

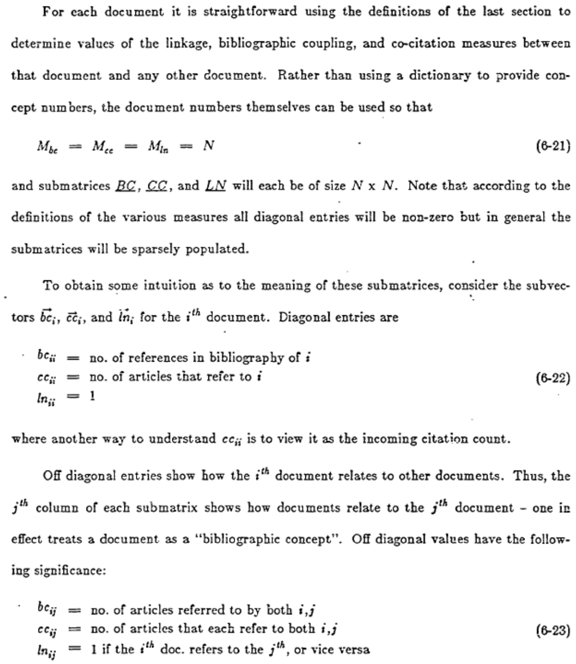
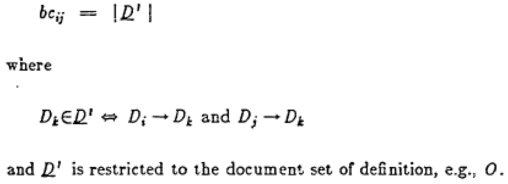


<table>
<tr><th></th><th>A</th><th>B</th><th>C</th><th>D</th><th>E</th><th>F</th><th>G</th></tr>
<tr><td>A</td><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>0</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>0</td><td></td><td></td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td></td><td></td><td>3</td><td></td><td>2</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>2</td><td>2</td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td>2</td><td>2</td><td>5</td></tr>
</table>

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:

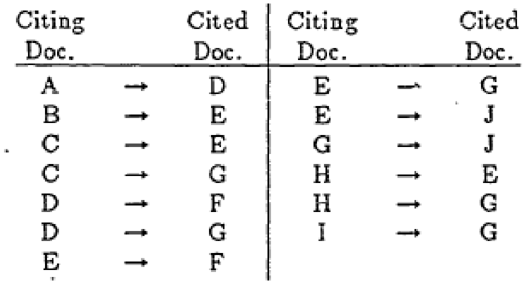
</td></tr>
</table>

Claim Text for ’494 Patent	Fox Thesis, 1983
	 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 $M_{bc} = M_{cc} = M_{ln} = N$ (6-21) and submatrices $\underline{BC}$, $\underline{CC}$, and $\underline{LN}$ will each be of size N x 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 bc_{ii} = no. of references in bibliography of i cc_{ii} = no. of articles that refer to i (6-22) ln_{ii} = 1 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: bc_{ij} = no. of articles referred to by both i,j cc_{ij} = no. of articles that each refer to both i,j (6-23) ln_{ij} = 1 if the i^{th} doc. refers to the j^{th}, or vice versa , pp. 171-182, 205-206, p. 240 (Figure 8.2, Sample computations of inner products); see also*, e.g.*, Chapters 1, & 6-9.
storing the second numerical representation;	*Id.* 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,  $bc_{ij} = \lvert \underline{D}' \rvert$ where $D_k \in \underline{D}' \Leftrightarrow D_i \rightarrow D_k$ and $D_j \rightarrow D_k$ and $\underline{D}'$ is restricted to the document set of definition, e.g., O. In the example of Figure 6.3,

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

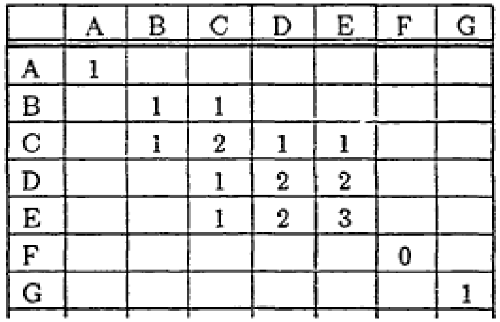
$bc_{B,C}$ = 1 and $bc_{D,E}$ = 2 since one document, E, is referred to by both B and C, while two documents, F and G, are each referred to by both D and E. Thus, B->E, C->E and D->F, E->F, D->G, E->G.”,

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



<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<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>
</table>

Figure 6.3: <u>bc</u> Submatrix



<table>
<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>
<tr><td>A</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>1</td><td>1</td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td>1</td><td>2</td><td>1</td><td>1</td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td>1</td><td>2</td><td>2</td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td>1</td><td>2</td><td>3</td><td></td><td></td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>0</td><td></td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td></tr>
</table>

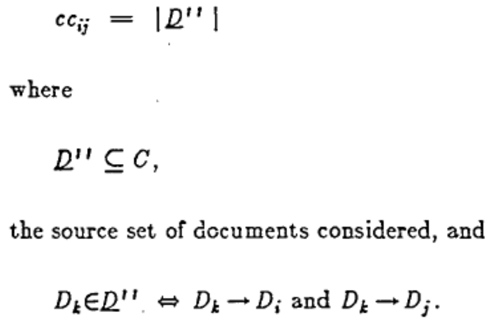
Note: $bc_{E,G} \neq 1$ since J is not $\in O$.

, 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

</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

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



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

where

$$\underline{D}'' \subseteq C,$$

the source set of documents considered, and

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

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

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

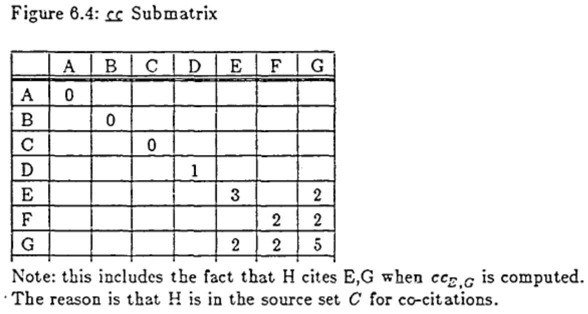


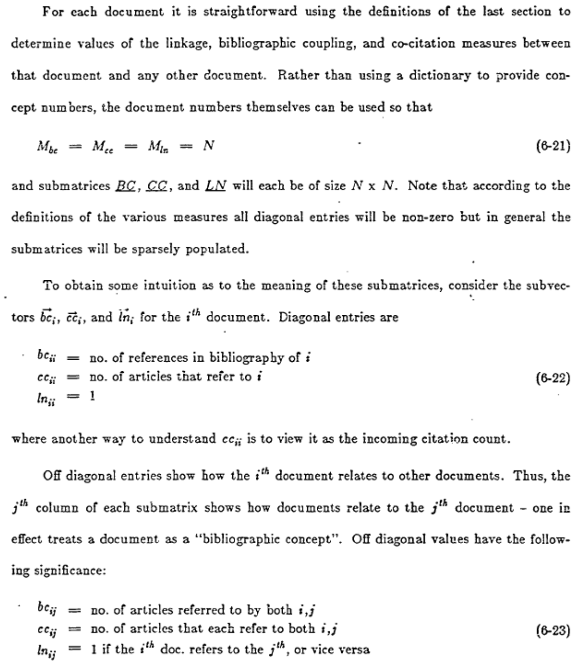
Figure 6.4: <u>cc</u> Submatrix

<table>
<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>
<tr><td>A</td><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>0</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>0</td><td></td><td></td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td></td><td></td><td>3</td><td></td><td>2</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>2</td><td>2</td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td>2</td><td>2</td><td>5</td></tr>
</table>

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.

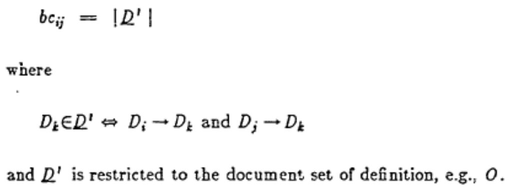
, p. 170:

</td></tr>
</table>

Claim Text for ’494 Patent	Fox Thesis, 1983
	 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 $M_{bc} = M_{cc} = M_{ln} = N$ (6-21) and submatrices $\underline{BC}$, $\underline{CC}$, and $\underline{LN}$ will each be of size N x 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 bc_{ii} = no. of references in bibliography of i cc_{ii} = no. of articles that refer to i (6-22) ln_{ii} = 1 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: bc_{ij} = no. of articles referred to by both i,j cc_{ij} = no. of articles that each refer to both i,j (6-23) ln_{ij} = 1 if the i^{th} doc. refers to the j^{th}, or vice versa , pp. 171-182, 205-206, p. 240 (Figure 8.2, Sample computations of inner products); see also*, e.g.*, Chapters 1, & 6-9.

Claim Text for ’494 Patent	Fox Thesis, 1983
identifying at least one object in the database, wherein the stored numerical representation is used to identify objects; and	*Id.* 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 <u>bc</u>, <u>cc</u>, and <u>ln</u> 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; *see also, e.g.*, Chapters 1, 7-9.
displaying one or more identified objects from the database.	*Id*. 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.” *See* chapters 5 and 8; *see also, e.g.*, Chapters 1, & 6-9.
19. The method of claim 18 wherein the step of generating a second numerical representation comprises: selecting an object in the database for analysis;	*Id.* 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 <u>bc</u>, <u>cc</u>, and <u>ln</u> 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

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>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; see also, e.g., Chapters 1, 7-9.</td></tr>
<tr><td>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.</td><td>Id. 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,



$$bc_{ij} = |\underline{D}'|$$

where

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

and $\underline{D}'$ is restricted to the document set of definition, e.g., O.

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.”,

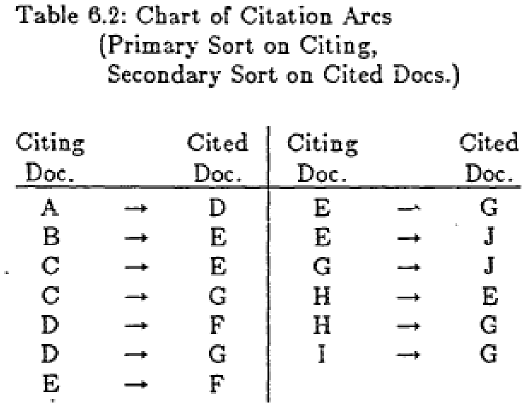
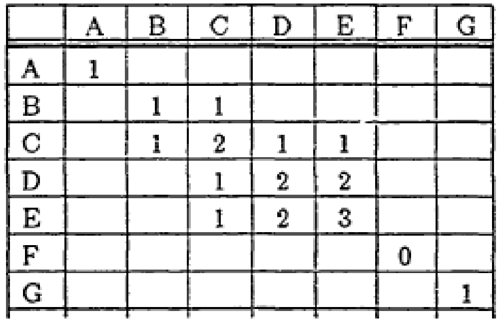


Table 6.2: Chart of Citation Arcs (Primary Sort on Citing, Secondary Sort on Cited Docs.)

<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<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>
</table>
</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

Figure 6.3: $\underline{bc}$ Submatrix



<table>
<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>
<tr><td>A</td><td>1</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>1</td><td>1</td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td>1</td><td>2</td><td>1</td><td>1</td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td>1</td><td>2</td><td>2</td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td>1</td><td>2</td><td>3</td><td></td><td></td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>0</td><td></td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td></td><td></td><td>1</td></tr>
</table>

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

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

where

$$\underline{D}'' \subseteq C,$$

the source set of documents considered, and

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

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

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

</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

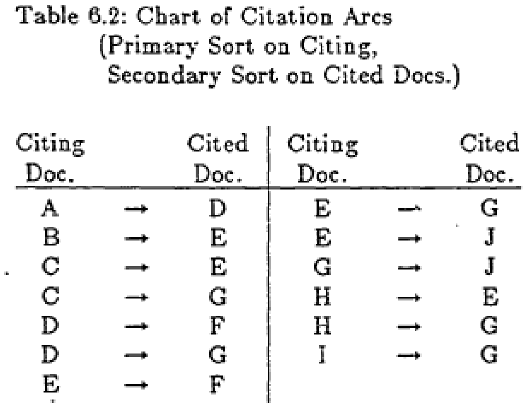
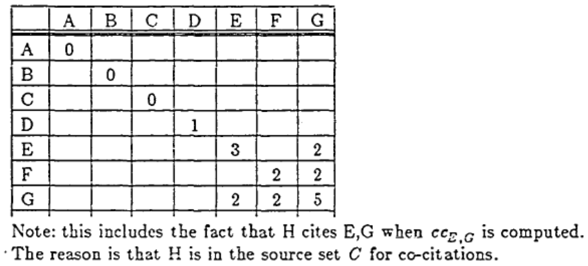
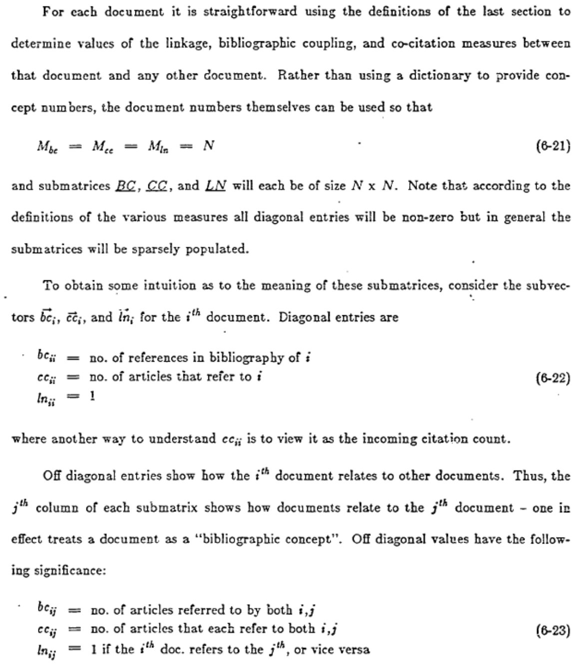



Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)
<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<tr><td>A</td><td>→</td><td>D</td><td>E</td><td>→</td><td>G</td></tr>
<tr><td>B</td><td>→</td><td>E</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>
</table>
Figure 6.4: $\underline{cc}$ Submatrix

<table>
<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>
<tr><td>A</td><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>0</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>0</td><td></td><td></td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td></td><td></td><td>3</td><td></td><td>2</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>2</td><td>2</td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td>2</td><td>2</td><td>5</td></tr>
</table>
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:
</td></tr>
</table>

Claim Text for ’494 Patent	Fox Thesis, 1983
	 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 $M_{bc} = M_{cc} = M_{ln} = N$ (6-21) and submatrices $\underline{BC}$, $\underline{CC}$, and $\underline{LN}$ will each be of size N x 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 bc_{ii} = no. of references in bibliography of i cc_{ii} = no. of articles that refer to i (6-22) ln_{ii} = 1 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: bc_{ij} = no. of articles referred to by both i,j cc_{ij} = no. of articles that each refer to both i,j (6-23) ln_{ij} = 1 if the i^{th} doc. refers to the j^{th}, or vice versa ,” pp. 171-182, 205-206, p. 240 (Figure 8.2, Sample computations of inner products); *see also, e.g.*, Chapter 1, Chapters 6-9.
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.	*Id.* 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 <u>bc</u>, <u>cc</u>, and <u>ln</u> 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

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; *see also, e.g.*, 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.	*Id*. 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.” *See* chapters 5 and 8; *see also, e.g.*, Chapters 1, & 6-9.
23. A method of representing data in a computer database with relationships, comprising the steps of:	*See infra*; *see also, e.g.*, Chapters 1, & 6-9.
assigning nodes node identifications;	*Id*. at 153: “Consider a collection, *C*, containing *N* documents, that is processed by automatic indexing routines which first eliminate stop words and reduce remaining words to their respective stems,” p. 196:

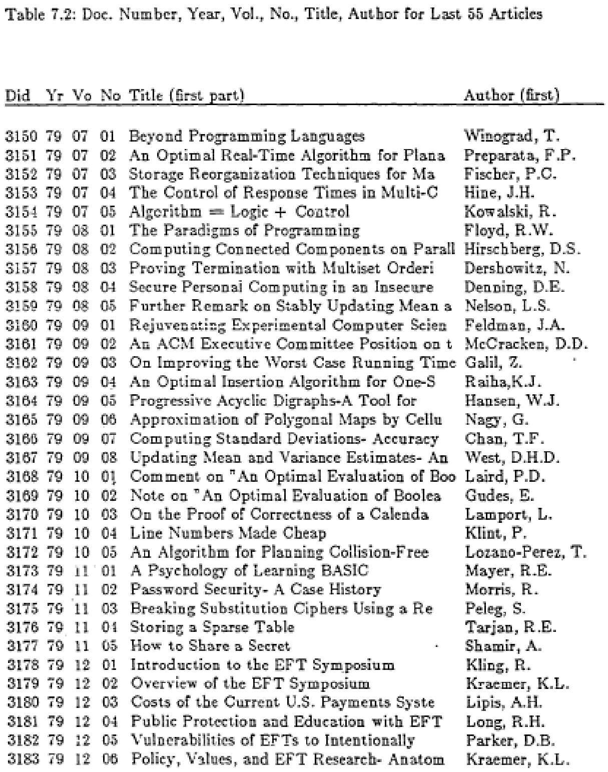
<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>


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

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

p. 198:
</td></tr>
</table>

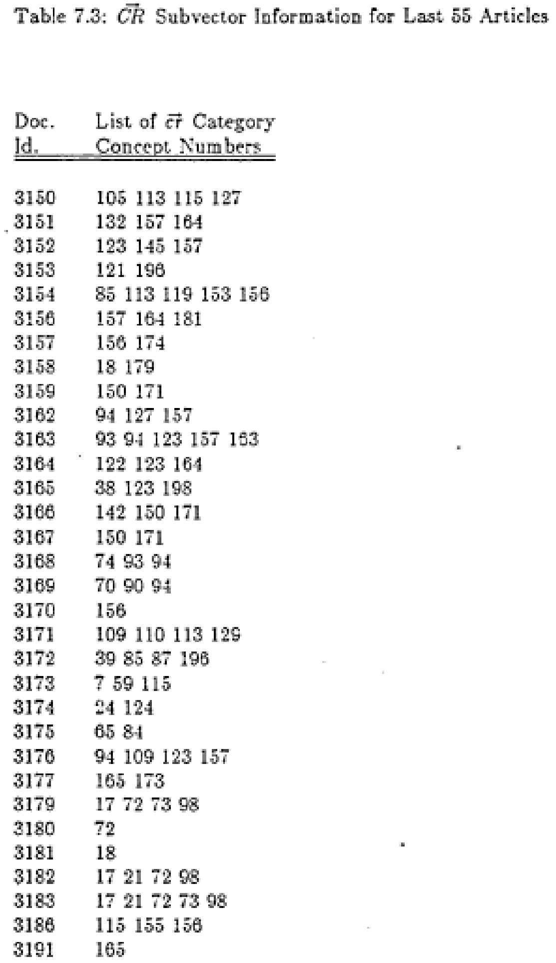
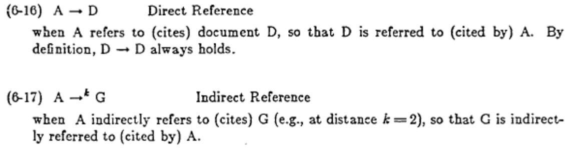
<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>


Table 7.3: $\vec{CR}$ Subvector Information for Last 55 Articles

<table>
<tr><th>Doc. Id.</th><th>List of $\vec{cr}$ Category Concept Numbers</th></tr>
<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 94 123 157 153</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>
</table>

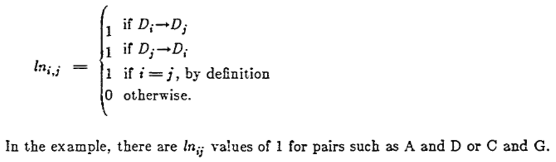
See also, e.g., *id*. at 27, 195, 203, 207, 211-13, 225-26, 229, 230, Tables 7.2, 7.9, Section 6.5.1.5 “Indexing,” Chapters 1, & 6-9.
</td></tr>
<tr><td>generating links, wherein each link represents a relationship between two nodes and is identified by</td><td><i>Id</i>. at 159: “In addition to terms and authors, other types of information are available in many collections. Dates and controlled vocabulary terms may be properly</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td>the two nodes in which the relationship exists;</td><td>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:



(6-16) A → D Direct Reference
when A refers to (cites) document D, so that D is referred to (cited by) A. By definition, D → D always holds.
(6-17) $A \to^k G$ 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->D or D-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,



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

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

,” Figure 6.5:

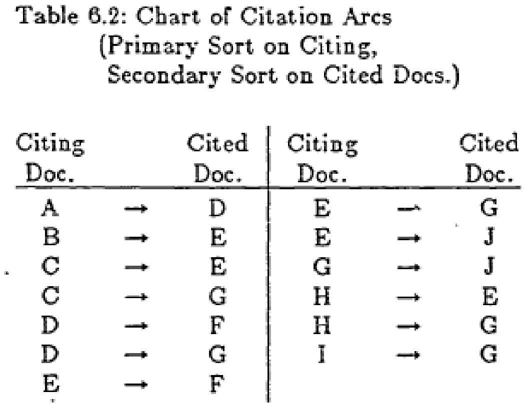
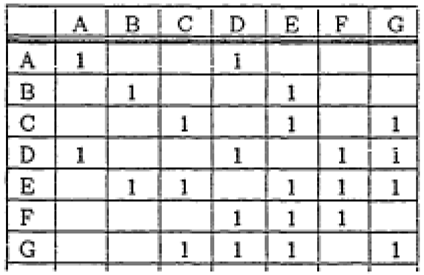
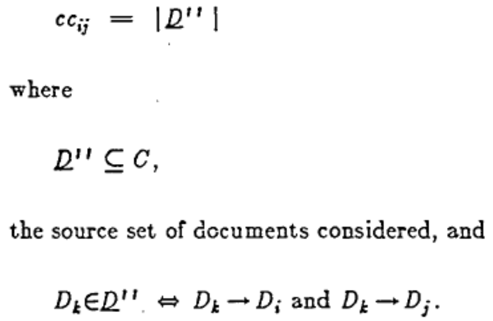


Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<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>
</table>
</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>Figure 6.5: <u><i>ln</i></u> Submatrix
<table><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><tr><td>A</td><td>1</td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr><tr><td>B</td><td></td><td>1</td><td></td><td></td><td>1</td><td></td><td></td></tr><tr><td>C</td><td></td><td></td><td>1</td><td></td><td>1</td><td></td><td>1</td></tr><tr><td>D</td><td>1</td><td></td><td></td><td>1</td><td></td><td>1</td><td>1</td></tr><tr><td>E</td><td></td><td>1</td><td>1</td><td></td><td>1</td><td>1</td><td>1</td></tr><tr><td>F</td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td></tr><tr><td>G</td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td><td>1</td></tr></table>, 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.</td></tr>
<tr><td>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</td><td><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.
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</td></tr>
</table>

<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>


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

where

$$\underline{D}'' \subseteq C,$$

the source set of documents considered, and

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

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

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

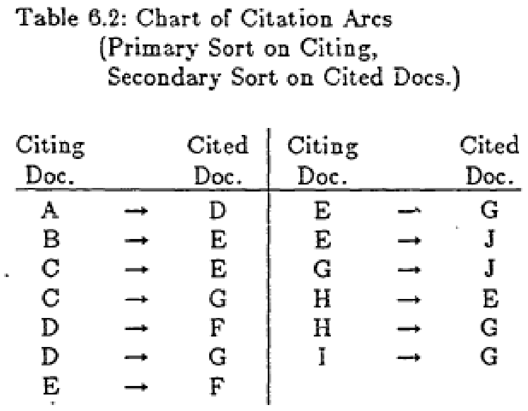


Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)

<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<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>
</table>
</td></tr>
</table>

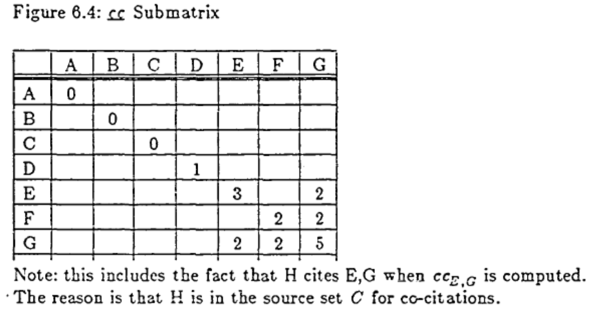
<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

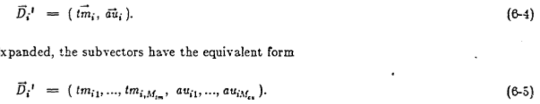
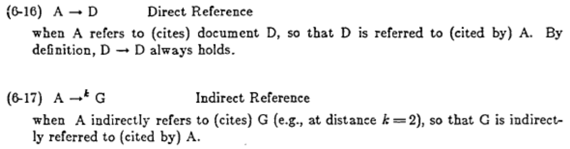
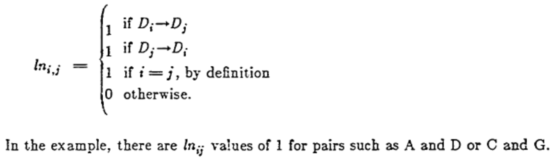
Figure 6.4: <u>cc</u> Submatrix

<table>
<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>
<tr><td>A</td><td>0</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>0</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>0</td><td></td><td></td><td></td><td></td></tr>
<tr><td>D</td><td></td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>E</td><td></td><td></td><td></td><td></td><td>3</td><td></td><td>2</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td></td><td></td><td>2</td><td>2</td></tr>
<tr><td>G</td><td></td><td></td><td></td><td></td><td>2</td><td>2</td><td>5</td></tr>
</table>

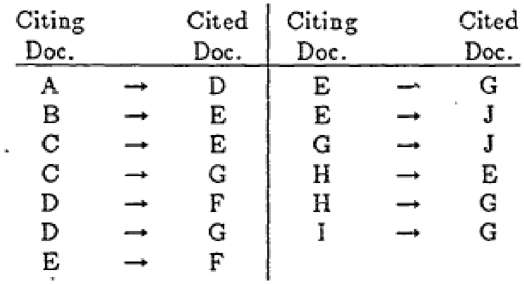
Note: this includes the fact that H cites E,G when $cc_{E,G}$ is computed.
The reason is that H is in the source set C for co-citations.

See also, e.g., Chapters 1, & 6-9.</td></tr>
<tr><td>displaying a node identification.</td><td><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.</td></tr>
<tr><td>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.</td><td><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.</td></tr>
<tr><td>25. The method of claim 23 further comprising the step of: generating link sub-types, comprising the steps of:</td><td><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).”).</td></tr>
</table>

Claim Text for ’494 Patent	Fox Thesis, 1983
identifying each link sub-type with a name; and	*Id*. at 214; *see also, e.g.*, Chapters 1, & 6-9, Chart for Claim 1, *supra* (including the quotations and descriptions set forth therein, which are incorporated by reference herein); *see also*, 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.	*Id*. at 214; *see also, e.g.*, Chapters 1, & 6-9, Chart for Claim 1, *supra* (including the quotations and descriptions set forth therein, which are incorporated by reference herein); *see also*, 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.	*Id.* 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”); *see also, e.g.*, 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.	*Id.* 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”); *see also, e.g.*, 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:	*See infra*; *see also, e.g.*, Chapters 1, & 6-9.
assigning links to represent relationships in the database;	*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 ith subvector as tm_i, and the author subvector as au_i, the ith document is described as

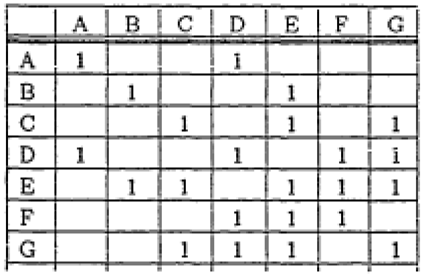
Claim Text for ’494 Patent	Fox Thesis, 1983
	 $\vec{D}_i{}' = (\vec{tm}_i, \vec{au}_i).$ (6-4) Expanded, the subvectors have the equivalent form $\vec{D}_i{}' = (tm_{i1}, ..., tm_{i,M_{tm}}, au_{i1}, ..., au_{iM_{au}}).$ (6-5) ,” p. 159: “In addition to terms and authors, other types of information are 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:  (6-16) A → D Direct Reference when A refers to (cites) document D, so that D is referred to (cited by) A. By definition, D → D always holds. (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->D or D-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,  $ln_{i,j} = \begin{cases} 1 & \text{if } D_i \rightarrow D_j \\ 1 & \text{if } D_j \rightarrow D_i \\ 1 & \text{if } i = j \text{, by definition} \\ 0 & \text{otherwise.} \end{cases}$ 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>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>
Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)



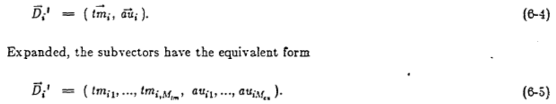
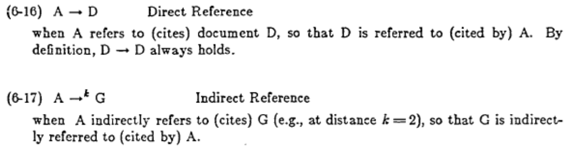
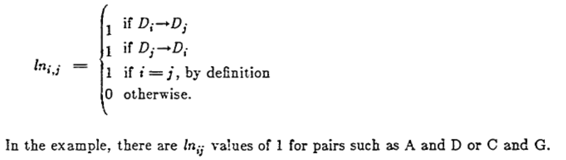
<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<tr><td>A</td><td>→</td><td>D</td><td>E</td><td>→</td><td>G</td></tr>
<tr><td>B</td><td>→</td><td>E</td><td>E</td><td>→</td><td>J</td></tr>
<tr><td>C</td><td>→</td><td>E</td><td>G</td><td>→</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>
</table>

Figure 6.5: <u>ln</u> Submatrix



<table>
<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>
<tr><td>A</td><td>1</td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>1</td><td></td><td></td><td>1</td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>1</td><td></td><td>1</td><td></td><td>1</td></tr>
<tr><td>D</td><td>1</td><td></td><td></td><td>1</td><td></td><td>1</td><td>1</td></tr>
<tr><td>E</td><td></td><td>1</td><td>1</td><td></td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td></tr>
<tr><td>G</td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td><td>1</td></tr>
</table>

, 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.
</td></tr>
<tr><td>generating node identifications based upon the assigned links, wherein node identifications are generated so that each link represents a relationship between two identified nodes;</td><td>*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</td></tr>
</table>

Claim Text for ’494 Patent	Fox Thesis, 1983
	 $\vec{D}_i' = (\vec{tm}_i, \vec{au}_i).$ (6-4) Expanded, the subvectors have the equivalent form $\vec{D}_i' = (tm_{i1}, ..., tm_{i,M_{tm}}, au_{i1}, ..., au_{iM_{au}}).$ (6-5) ,” 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:  (6-16) A → D Direct Reference when A refers to (cites) document D, so that D is referred to (cited by) A. By definition, D → D always holds. (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->D or D-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,  $ln_{i,j} = \begin{cases} 1 & \text{if } D_i \rightarrow D_j \\ 1 & \text{if } D_j \rightarrow D_i \\ 1 & \text{if } i = j, \text{ by definition} \\ 0 & \text{otherwise.} \end{cases}$ 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>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

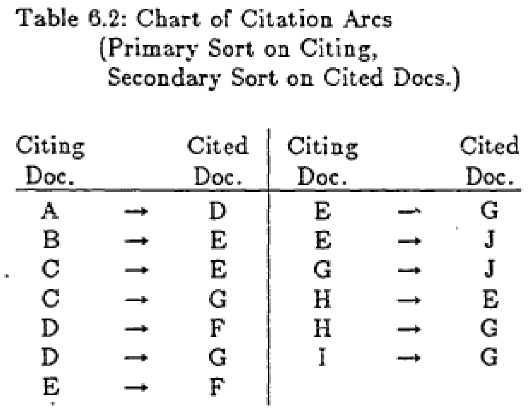
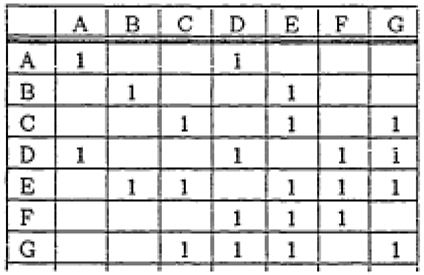
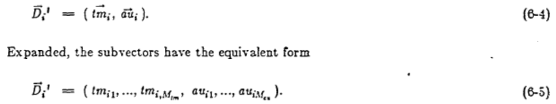
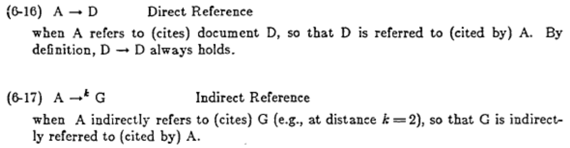
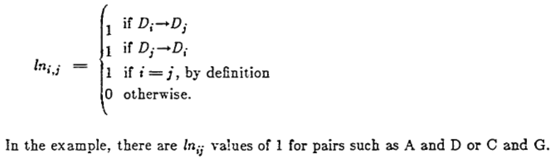
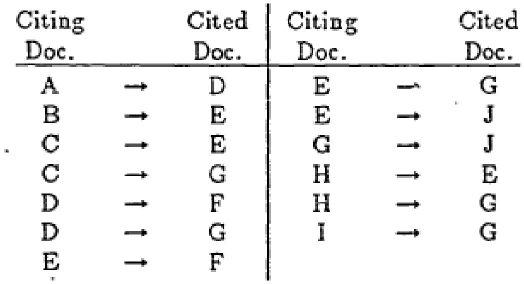



Table 6.2: Chart of Citation Arcs
(Primary Sort on Citing,
Secondary Sort on Cited Docs.)
<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<tr><td>A</td><td>→</td><td>D</td><td>E</td><td>→</td><td>G</td></tr>
<tr><td>B</td><td>→</td><td>E</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>
</table>
Figure 6.5: <u>ln</u> Submatrix

<table>
<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>
<tr><td>A</td><td>1</td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>1</td><td></td><td></td><td>1</td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>1</td><td></td><td>1</td><td></td><td>1</td></tr>
<tr><td>D</td><td>1</td><td></td><td></td><td>1</td><td></td><td>1</td><td>1</td></tr>
<tr><td>E</td><td></td><td>1</td><td>1</td><td></td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td></tr>
<tr><td>G</td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td><td>1</td></tr>
</table>
, 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.
</td></tr>
<tr><td>storing the links and node identifications, wherein the links and nodes may be retrieved;</td><td><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</td></tr>
</table>

Claim Text for ’494 Patent	Fox Thesis, 1983
	 $\vec{D}_i' = (\vec{tm}_i, \vec{au}_i).$ (6-4) Expanded, the subvectors have the equivalent form $\vec{D}_i' = (tm_{i1}, \ldots, tm_{i,M_{tm}}, au_{i1}, \ldots, au_{iM_{au}}).$ (6-5) ,” 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:  (6-16) A → D Direct Reference when A refers to (cites) document D, so that D is referred to (cited by) A. By definition, D → D always holds. (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->D or D-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,  $ln_{i,j} = \begin{cases} 1 & \text{if } D_i \rightarrow D_j \\ 1 & \text{if } D_j \rightarrow D_i \\ 1 & \text{if } i = j \text{, by definition} \\ 0 & \text{otherwise.} \end{cases}$ In the example, there are ln_{ij} values of 1 for pairs such as A and D or C and G. ,” Figure 6.5:

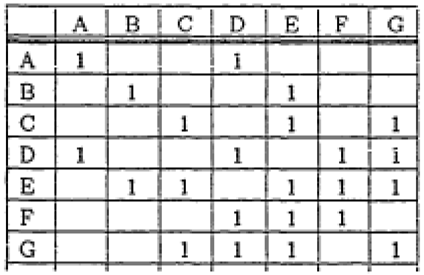
<table>
<tr><th>Claim Text for ’494 Patent</th><th>Fox Thesis, 1983</th></tr>
<tr><td></td><td>

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



<table>
<tr><th>Citing Doc.</th><th></th><th>Cited Doc.</th><th>Citing Doc.</th><th></th><th>Cited Doc.</th></tr>
<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>
</table>

Figure 6.5: *ln* Submatrix



<table>
<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>
<tr><td>A</td><td>1</td><td></td><td></td><td>1</td><td></td><td></td><td></td></tr>
<tr><td>B</td><td></td><td>1</td><td></td><td></td><td>1</td><td></td><td></td></tr>
<tr><td>C</td><td></td><td></td><td>1</td><td></td><td>1</td><td></td><td>1</td></tr>
<tr><td>D</td><td>1</td><td></td><td></td><td>1</td><td></td><td>1</td><td>1</td></tr>
<tr><td>E</td><td></td><td>1</td><td>1</td><td></td><td>1</td><td>1</td><td>1</td></tr>
<tr><td>F</td><td></td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td></tr>
<tr><td>G</td><td></td><td></td><td>1</td><td>1</td><td>1</td><td></td><td>1</td></tr>
</table>

, 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.</td></tr>
<tr><td>searching for node identifications using the stored links; and</td><td>*Id.* at Chapter 6, *e.g.*, 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 *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 combined to aid</td></tr>
</table>

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; *see also, e.g.*, Chapters 1,6 & 9.
displaying node identifications, wherein the displayed node identifications are located in the searching step.	*Id*. 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"; *see also* Chapters 5 and 8; *see also, e.g.*, 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.