NIELSEN PATENTS
STEVEN NIELSEN (133864)
steve@NielsenPatents.com
100 Larkspur Landing Circle, Suite 21
Larkspur, CA 94939-1743
Telephone: (415) 272-8210

HARDY PARRISH YANG LLP
VICTOR G. HARDY (pro hac vice)
vhardy@hpylegal.com
MINGHUI YANG (pro hac vice)
myang@hpylegal.com
4412 Spicewood Springs Rd., Suite 202
Austin, TX 78759
Telephone: (512) 520-9407

Attorneys for Plaintiff
SOFTWARE RIGHTS ARCHIVE, LLC

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
OAKLAND DIVISION

SOFTWARE RIGHTS ARCHIVE, LLC,

Plaintiff,

v.

TWITTER, INC.,

Defendant.

Case No. 4:12-cv-03972-HSG

PLAINTIFF'S FIRST AMENDED COMPLAINT FOR PATENT INFRINGEMENT

Judge: Hon Haywood S. Gilliam

Jury Trial Demanded

FIRST AMENDED COMPLAINT

For its Complaint, Software Rights Archive, LLC ("SRA") alleges as follows:

I. NATURE OF THE ACTION

1. This is a patent infringement action to end Defendant's direct, joint, contributory, and/or induced infringement of Plaintiff SRA's patented inventions, including but not limited to Defendant's unauthorized and infringing use, sale, offering for sale, manufacture, and/or importation of products and/or methods incorporating Plaintiff's inventions.

2. SRA has obtained all rights and interest to United States Patent No. 5,544,352 (the "'352 Patent"), United States Patent No. 5,832,494 (the "'494 Patent"), and United States Patent No. 6,233,571 (the "'571 Patent").

3. Defendant provides, uses, puts to use, sells, offers for sale, distributes, manufactures, and/or imports infringing products and services, and encourages others, including its customers, to use Defendant's products and services in an infringing manner.

4. Plaintiff SRA seeks damages from Defendant's infringement of Plaintiff's patent rights. Plaintiff further seeks past damages and prejudgment and postjudgment interest for Defendant's past infringement of Plaintiff's patents.

II. THE PARTIES

5. Software Rights Archive, LLC is a limited liability company organized and existing under the laws of Delaware.

6. Upon information and belief, Defendant Twitter, Inc. ("Twitter" or "Defendant") is a corporation organized and existing under the laws of the State of Delaware with a principal place of business at 795 Folsom Street, San Francisco, CA 94107. Twitter can be served with process by serving its registered agent for service of process in the State of California by serving Alexander Macgillivray,795 Folsom Street, Suite 600, San Francisco, California 94107.

III. JURISDICTION AND VENUE

7. This action arises under the United States Patent Act, codified at 35 U.S.C. § 1 *et seq*. This Court has exclusive subject matter jurisdiction under 28 U.S.C. §§ 1331 and 1338(a).

8. This Court has personal jurisdiction over Twitter, Inc. because, upon information and belief, Twitter resides in this District, has transacted business in this District, has committed acts of infringement in this District and continues to commit acts of infringement in this District.

9. Venue is proper in the Northern District of California pursuant to 28 U.S.C. §§ 1391(b), (c) and 1400(b), because Defendant resides in this District, has committed acts of direct and indirect infringement in this District, has transacted business in this District, and has established minimum contacts with this District.

IV. INTRADISTRICT ASSIGNMENT

10. This is an intellectual property action and, therefore, under Civil Local Rules 3-5(b) and 3-2(c), may be assigned to any division in this District.

V. PLAINTIFF'S PATENTS

11. On August 6, 1996, the '352 Patent, entitled "Method and Apparatus for Indexing, Searching and Displaying Data" was duly and lawfully issued by the United States Patent and Trademark Office, naming Daniel Egger as sole inventor and Libertech, Inc. as assignee. A true and correct copy of the '352 Patent is attached as **Exhibit A**. The '352 Patent was subject to *ex parte* reexamination by the United States Patent Office, and an *Ex Parte* Reexamination Certificate was issued for the '352 Patent on September 20, 2011, a true and correct copy of which is attached as **Exhibit B**. SRA is the assignee of all right, title and interest in and to the '352 Patent, and holds the right to sue and recover for past, present, and future infringement thereof.

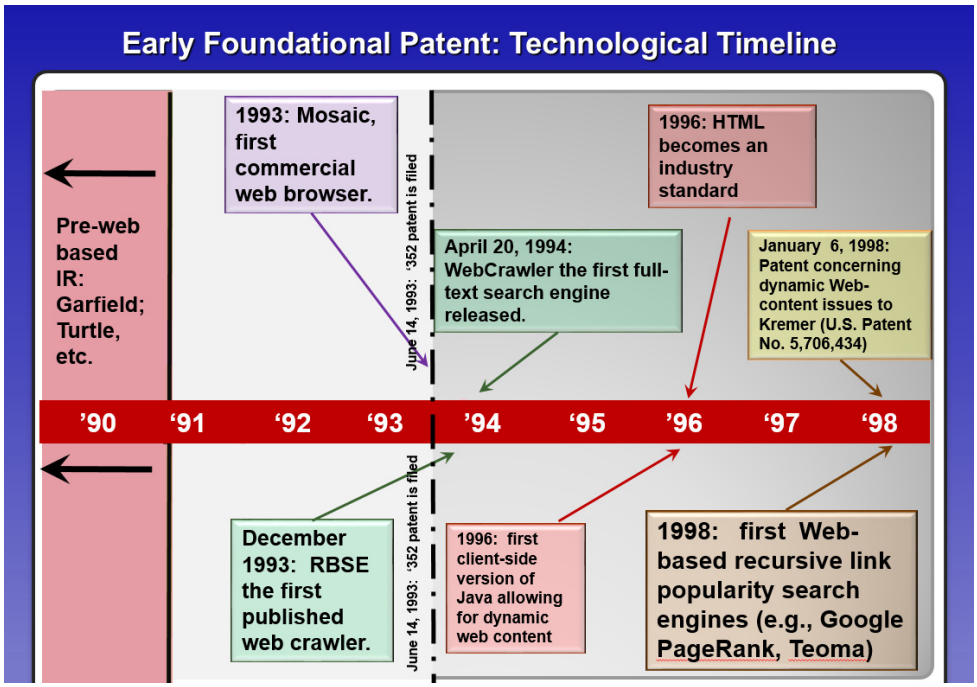
12. On November 3, 1998, the '494 Patent, entitled "Method and Apparatus for Indexing, Searching and Displaying Data" was duly and lawfully issued by the United States Patent and Trademark Office naming Daniel Egger, Shawn Cannon, and Ronald D. Sauers as inventors, and Libertech, Inc. as assignee. A true and correct copy of the '494 Patent is attached as **Exhibit C**. The '494 Patent was subject to *ex parte* reexamination by the United States Patent Office, and an *Ex Parte* Reexamination Certificate was issued for the '494 Patent on September 27, 2011, a true and correct copy of which is attached as **Exhibit D**. SRA is the assignee of the '494 Patent and holds the right to sue and recover for past, present, and future infringement thereof.

13. On May 15, 2001, the '571 Patent, entitled "Method and Apparatus for Indexing, Searching and Displaying Data" was duly and lawfully issued by the United States Patent and Trademark Office to Daniel Egger. A true and correct copy of the '571 Patent is attached hereto as **Exhibit E**. The '571 Patent was subject to *ex parte* reexamination by the United States Patent Office, and an *Ex Parte* Reexamination Certificate was issued for the '571 Patent on October 4, 2011, a true and correct copy of which is attached as **Exhibit F**. SRA is the assignee of the '571 Patent and holds the right to sue and recover for past, present, and future infringement thereof.

VI. OVERVIEW OF THE PATENTED TECHNOLOGY IN VIEW OF PATENT ELIGIBILITY UNDER 35 U.S.C. § 101

14. The Patents-in-Suit relate to the use of non-semantic link analysis (i.e., the analysis of citation and hyperlink relationships between records) to enhance computerized searching of electronic databases such as those related to the World Wide Web. Inventor Daniel Egger is a pioneer in the field of electronic database searching. His development of the cutting edge technology at issue in this case traces to the early 1990s, years before the accused products in this case were developed and commercialized. Indeed, the Patents-In-Suit were the basis on which the Patent and Trademark Office rejected several of Defendant Google Inc.'s patent claims directed to the accused PageRank® algorithm, self-described as the "heart of [Google's] software". (*Our Search: Google Technology*, http://www.google.com/technology/ (July 23, 2008)). No less than nine patents from Google founder Lawrence Page, who created the PageRank algorithm for Google, and Amit Singhal, head of Google's Search Quality Group that implements the PageRank algorithm, cite at least one of the Patents-In-Suit.

15. The inventions of the '352, '494, and '571 Patents are directed to, among other things, the use of non-semantic indirect relationships in search to solve problems present in electronic information retrieval systems of the early 1990s. The Patents recognize the importance of efficient search of electronic databases and attempt to improve the ability of the then state of the art search systems and the data structures used by these systems. The patents were developed at a time when electronic searching of large databases was early in its development. At the time of filing of the first patent in 1993, the internet was in its infancy and research in information retrieval was focused on smaller databases outside of the web. As shown below, the first commercial web browser had just been released, and the first published web crawler to index the web would not be released for another few months:



It would be years before HTML would even become the industry standard. Indeed, the '352 patent was filed five years and the '494 patent was filed two years before the first commercial search engine that analyzed non-semantical indirect relationships (i.e., Google Search Engine) other than Daniel Egger's own search system (V-Search). It was nine years before Facebook was founded in 2004.

A. The Patents Identify Problems with Prior Art Computer Search Systems

16. Egger's early work primary focused on legal research systems—the largest available computer databases at that time—and later migrated to the World Wide Web. Egger struggled with the challenges of locating relevant case law amid the multitude of electronic records made available by subscription services such as Westlaw® and Lexis-Nexis®. At the time—when the Internet was in its infancy—Egger recognized and solved several fundamental limitations of conventional search techniques. The searching of records (or "objects") in a database is customarily performed using semantic techniques that query the database for records containing a particular key word or group of words of interest to the database user. In a very large database, however, a large number of records—many of which are irrelevant—may be returned from even a restrictive word search. Egger experienced the frustration of long hours spent locating the wheat among the sea of chaff returned from such searches.

17. Egger's novel solution involved implementing computerized indexing and searching techniques that use citation data to enhance semantic word-based search techniques to better pinpoint relevant and important records in a large database. It would rank order the results so that the most relevant and most significant records would be returned first to the user. Today, Egger's patented techniques have become the standard approach and method for Internet search.

18. Systems such as Westlaw and Lexis predated the search engines of the World Wide Web by a number of years. Through his work, he was able identify problems inherent in existing computer search technology—the same problems that would later be encountered when applying those search methods to the World Wide Web. The patents describe search technology in the early 1990s as being reliant upon semantically text analysis using Boolean search terms:

> Our society is in the information age. Computers maintaining databases of information have become an everyday part of our lives. The ability to efficiently perform computer research has become increasingly more important. Recent efforts in the art of computer research have been aimed at reducing the time required to accomplish research. Computer research on non-textual objects is very limited. *Current computer search programs use a text-by-text analysis procedure (Boolean Search) to scan a database and retrieve items from a database*. The user must input a string of text, and the computer evaluates this string of text. Then the computer retrieves items from the database that match the string of text. The two popular systems for computerized searching of data used in the legal profession are Westlaw™, a service sold by West Publishing Company, 50 W. Kellogg Blvd., P.O.

> Box 64526, St. Paul, Minn. 55164-0526, and Lexis™, a service sold by Mead Data Central, P.O. Box 933, Dayton, Ohio 45401.

(Ex. E, '571 patent, 1:27-45).

19. These type of systems relying on semantically analysis had a host of problems that prevented them from presenting a select number of the most relevant results out of millions of possibly responsive items to the user in an easily accessible manner. Among the problems described by the patents is that, unless a request is precisely phrased, searches were unlikely to return the precise desired result:

> However, Boolean searches of textual material are not very efficient. Boolean searches only retrieve exactly what the computer interprets the attorney to have requested. **If the attorney does not phrase his or her request in the exact manner in which the database represents the textual object, the Boolean search will not retrieve the desired textual object. Therefore, the researcher may effectively by denied access to significant textual objects that may be crucial to the project on which the researcher is working**.

(Ex. E, '571 patent, 1:46-54).

20. Another problem of existing systems is that they included too many search results, including large numbers of irrelevant or low relevance results, without presenting them in the order of the document's significance or importance:

> A second problem encountered with Boolean searches is that the search retrieves a significant amount of irrelevant textual objects. (It should be noted that in the context of research, a textual object could be any type of written material. The term textual object is used to stress the fact that the present invention applies to all types of databases. The only requirement that a textual object must satisfy in order to be selected by a Boolean search program is that part of the textual object match the particular request of the researcher. Since the researcher cannot possibly know all of the groupings of text within all the textual objects in the database, the researcher is unable to phrase his request to only retrieve the textual objects that are relevant. … Even if one assumes that all the textual objects retrieved from a Boolean search are relevant, the listing of the textual objects as done by any currently available systems does not convey some important and necessary information to the researcher. The researcher does not know which textual objects are the most significant (i.e., which textual object is referred to the most by another textual object) or which textual objects are considered essential precedent (i.e., which textual objects describe an important doctrine).

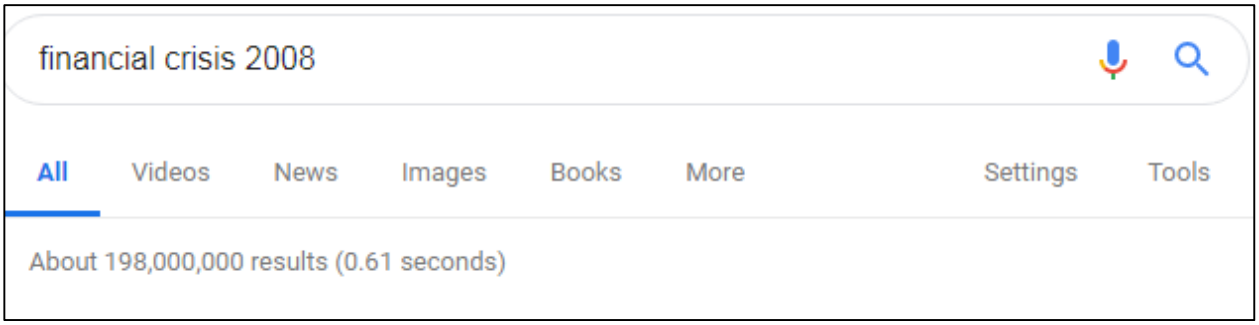
(Ex. E, '571 patent, 1:54-2:13). Thus, the researcher was forced to sift through large amounts of irrelevant information before finding the document of most importance or interest. In a case of a large database, this make makes the search results meaningless. See section VI.C, *infra*

21. The Patents describe these problems and inefficiencies as pervasive through all existing computerized research systems:

> Computerized research tools for legal opinions and related documents are probably the most sophisticated computer research tools available and therefore form the background for this invention. However, the same or similar computer research tools are used in many other areas. For example, computer research tools are used for locating prior art for a patent application. **The same problems of inefficiency discussed above exist for computer research tools in many areas of our society**.

(Ex. E, '571 patent, 2:38-46).

22. The problems identified above with respect to semantically based search systems in the 1990s are even more severe with Web search systems. For example, as shown in the screenshot, if a user wished to search for articles describing the 2007-2008 financial crisis and used the search terms "financial" "crisis," "2008" he would receive about 198 million webpages with these terms: Thus, the then state of the art method using Boolean search terms on databases as large as the World Wide Web would produce results so large as to render the search results meaningless. If the user wanted a specific result, prior art systems had no way of identifying that result from all other responsive results and typically required the user to know the specific, narrowly tailored search parameters before the search in order to obtain just that result.



B. The Patents Are Directed to Improvements to Computerized Search Systems by Describing Unconventional Search Methods Using Non Semantic Indirect Relationships

23. The patented invention solves the above problems with improvements to search methods, databases, and data structures using a proximity index that identify, represent analyze referential indirect relationships (i.e., chains of citations including hyper link relationships). It is important to note that the terms "direct relationship" and "indirect relationships" in the claims have

been given a very specific structural meaning in the patents. The PTO has construed them to refer to when one object cites to another object or when two objects are connected by a chain of citations:

> We, therefore, construe *direct relationships* as "relationships where one object cites to another object," and *indirect relationships* as "relationships where at least one intermediate object exists between two objects and where the intermediate objects connect the two objects through a chain of citations."

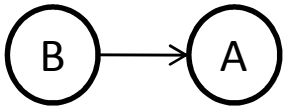
(Ex. G, IPR2013-00479 Institution Decision, p 11). Fundamentally, the patents recognize that certain citation relationships including hyperlink pointers on the World Wide Web contain useful information concerning an objects "importance" that could be used to identify the most relevant objects among a pool of objects. (Ex. A, '352 Patent, 5:41-50, 7:63-19:15, 19:62–20:7). In this way, a more relevant pool objects can be further located from a pool of otherwise responsive objects containing particular textual terms using direct and indirect citations to the objects. Similarly, ranks may be developed to order the position objects on screen by relevance or importance so that the most important objects are displayed first or in an easily acceded manner.

24. Returning to our previous example involving the Google Search Engine identification of 198 million words with the semantically terms "financial crisis 2008", one can locate the top 10 most important objects to be displayed from the 198 million search results containing the terms "financial crisis 2008" and present them to the user on the first page of results. The recognition by Egger of the importance of non-semantic relationships expressed in hyperlinks would later be considered a revolutionary idea by industry years after the patents in suit and change the way Web searches are conducted**.** See Section VI.C, *infra*.

25. One of the improvements of Egger's invention over the prior art methods of the day was Egger's use of a proximity index of indirect relationships to enhance conventional word-based searching. Egger's "Proximity Indexing" is a method of preparing data in a database for *subsequent* searching by advanced data searching programs." (Ex. A, '352 Pat., 4:5-14) (emphasis added). Prior to Egger's invention, conventional electronic search systems such as Westlaw and Lexis/Nexis focused exclusively on semantic word-based text matching and did not use a proximity index. Egger's approach marked a fundamental shift by harnessing the value of non-semantic

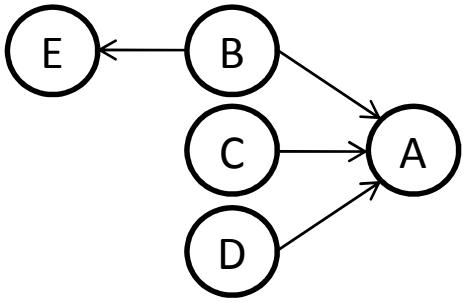
relationships, which are referential relationships (*e.g.*, citations or hyperlinks, where one record may point to another record) to further enhance a word-based search.

26. Egger made several key observations concerning legal research that were important to his invention. First, cases that cite to each other tend to be more likely to discuss similar subject matter and/or reflect the importance of an object. If B directly cites to A (*i.e.*, a direct relationship), then A and B are likely to discuss related subject matter. *Id.*, 5:15–24.



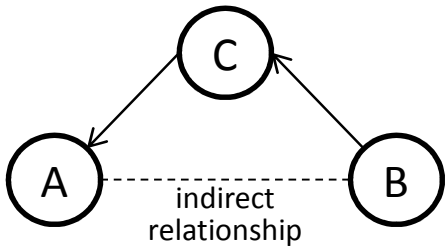
Thus, searching direct citations tends to find related subject matter.

27. Second, the more legal citations that cite a given case, the more important or significant the case is likely to be. *Id.*, 5: 25–31. If B, C, and D cite A, and only B cites E (all other things being equal), then A is likely to be more important than E.

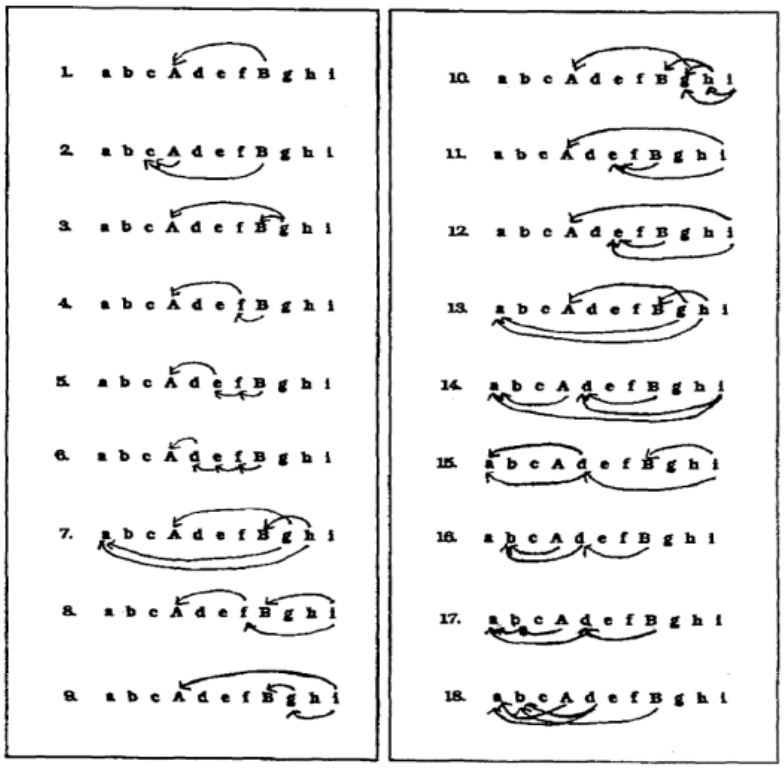


Thus, citation data can be used to rank a pool of objects by order of importance.

28. Third, Egger noted that indirect relationships between objects are also useful for determining the similarity or importance of objects. For example, if B cites C and C cites A, B is indirectly related to A.



Egger's algorithms make extensive use of multiple indirect citation relationships. *Id.*, 12:31–46. Indeed, as shown in Fig. 6, the preferred embodiment of the '352 Patent employs up to eighteen different "patterns" of relationships, seventeen of which involve indirect relationships:



See id., Fig. 6. All of the asserted claims contemplate analyzing non-semantic indirect relationships and other factors to enhance the search for objects. Embodiments disclose the use of both semantic and non-semantic factors to search for objects.

29. Egger uses the analysis of non-semantic citation relationships in a proximity index to support seven types of search routines. *See id.*, Figs. 4A-B. There are four pool search subroutines and three query-by-example search subroutines. (Ex. A, '352 Pat., 19:62–20:7). The CDSPM search routine that is most important to this litigation is the Pool Importance routine.

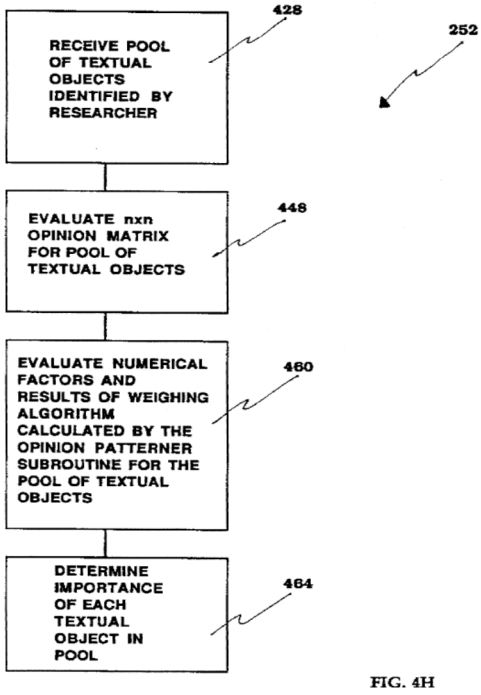


FIG. 4H

30. The Pool Importance search routine begins by selecting a pool of objects via a semantic keyword search. *See id.*, Fig. 4H, 5:41–50, 7:63–19:15. Those objects that contain the keywords form a pool of objects. The search system, Computer Search Program for Data represented in Matrices ("CSPDM"), then evaluates the indexes and matrices of the proximity indexer and other factors to rank each object's importance within the pool; the objects are then displayed using their rank. *Id.*, 21:21–32.

31. Still Fig. 14 A and B describe two additional index, search and display routines with

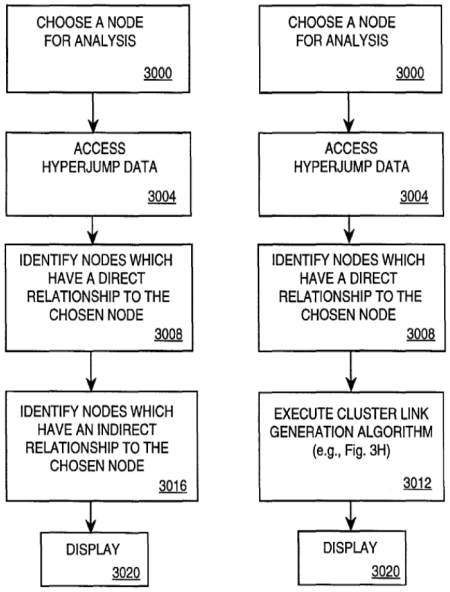


Fig. 14A *Fig. 14B*

14B describing the use of clusters formed by cluster links to locate objects for display.

All of the claims asserted in this litigation define at least a search method that uses indirect relationships.

32. The patents set forth specific data structures representing indirect relationships and search methods that can rank order objects based upon importance. In this way, a relevant pool objects can be further located from a pool of objects containing particular textual terms using direct and indirect citations to the objects.

33. The patents also recognize that hyper link relationships in a network such as the Web also contained useful information for search. These type of referential relationships differed significantly than bibliographic citations that were experimented with in the prior art. (Ex. H, Declaration of Amy Langville, ¶ 50). The patents analyzed hyperlink relationships in the same way it analyzed textual bibliographic citations and considered chains of hyperlink relationships to define indirect relationships as discussed in the patents:

> In the preferred embodiment, this system uses the cluster link generation algorithm described in FIG. 3H to search and identify closely associated documents located on the Internet in the same manner as described above. The system treats hyperlinks 2004 on the Web in the same manner as it treats links 2004 in a database, and it treats web pages on the Web in the same manner as it treats nodes 2008 in a database 54. Source links 2004 on the Web link a source node 2008 (or source web page) to a second node (or second web page). Influence links 2004 perform the same function in reverse. Direct links 2032 (as described above) are the same as hyperlinks 2004, which use URLs, in the World Wide Web, and they directly link one web page (or node) to another. Indirect links 2036 link two web pages or nodes 2008 through more than one path.

(Ex. E, '571 patent, 48:63-49:10). The inventive concepts of creating data structures representing and analyzing indirect hyperlink relationships on the Web for purposes of for improved search was unconventional, non-routine and not well understood. As discussed more in section VI.C, at the time of the filing of the patents in suit, no search engine analyzed indirect hyperlink relationships for purposes on enhancing search. Nor would experiments even be conducted on such relationships. Indeed, the very idea of collecting and conducting "any meaningful" analysis of hyperlinks of the Web for search was considered a "revolutionary" idea in 1998, years after the patents in suit. See ¶ 37, *infra*. Daniel Egger disclosed the usefulness of these relationships for search two years before Larry Page was hailed as a major innovator for making this recognition. *Id.*

34. Furthermore, as discussed at length in the next section, the inventive concepts of Daniel Egger differed from prior experiments in the field which largely focused only the presence of a co-citation relationships and bibliographic couplings found in the bibliographies of research papers, the simplest of citation patterns. Daniel Egger's numerical representations were directed to many different unique and novel type of patterns of citation relations (*e.g.*, the 18 patterns) and combined them to form a single numerical representations with other unique weighting factors to produce a weight that could be used in searching. Similarly, the cluster link generator was capable of identifying important non-semantic relationships of any link length or pattern by its unique link by link weighting statistical analysis. Thus, Egger's methods obtained used different types of indirect relationships and obtained far more useful information from the network than just the presence bc and cc patterns. (Ex. H, Langville Decl., ¶ 50). Consequently, Daniel Egger's technology was developed in a commercial product (Libertech V-Search) and eventually deployed by major commercial search engines while the prior art experiments largely produced failed or meager results.

C. The Use of Non-Semantic Indirect Relationships as Claimed to Improve Search Was Not Conventional, Well-Understood, or Routine at the Filing of the Patents in Suit

35. The patents in suit and claims are directed to specific search and display methods that use indirect relationships to improve computerized search systems. Search and display methods that use indirect relationships (e.g., the ordered combinations of '571 claims 26, 28, and 31), and the specific data structures described and claimed by the patents representing these relationships, were not conventional, well-understood, or routine in the art at the time of filing of the patents. Prior to the patents, the study of the use of indirect relationships to improve search was confined to handful of experiments and papers that largely produced negative or meager results that did not justify incorporation into an automated retrieval machine. These few experiments largely took place in the 1980s and then link analysis was ignored by the search engines until the late 1990s. The experts in the field did not appreciate indirect relationships could be used as claimed to improve search in the manner claimed, and no commercial search engine actually used these relationships prior to Google and Daniel Egger's V-search systems. The prior art experiments primarily focused

on the mere presence of indirect relationships (bc and cc) and consequently only achieved negative or meager results that were unreliable.

36. The use of indirect relationships in automated retrieval was not in common use or was well understood by the field. In the few experiments that were conducted, leaders in the field largely concluded that the use of citation relationship was ineffectual for use in actual systems and that a much better understanding of them was needed:

> [1986] Overall, the [citation analysis] procedure is not sufficiently reliable to warrant incorporation into operational automatic retrieval systems.
>
> None of the proposed methods for the improvement of document representation [including Fox's bc and cc indirect relationship vectors] has proved to be generally useful when applied to a variety of different retrieval environments.
>
> Since no obvious way exists for distinguishing the positive from the negative effects, the citation methodology cannot be recommended for inclusion in practical retrieval environments.
>
> [1988] "Other recent attempts to supply expanded document representations **using citations and other bibliographic indicators** attached to texts and documents have also led to the conclusion that **effective term expansion methods** valid for a variety of different collections **are difficult to generate**." [(Ex. M, Salton & Buckley, On the Use of Spreading Activation Methods in Automatic Information Retrieval, pp. 147-148)].
>
> [1992] Retrieval experiments in a collection of bibliographic references showed that following citations – a kind of referential links— produces ambiguous results ….The hope is that our semantic links contain the information necessary to decide whether a further nodes should be visited by the retrieval algorithm or not.
>
> [1993] However, despite the significant efforts to explore and develop these models, there remain concerns about the models' utility for the searching of large scientific databases. Using the p-norm retrieval experiment described in Fox (1983) as an example, I will present my three major concerns…the reliability of extrapolating the performance of research systems that use the collection to a system to search a file over 750 times larger than the collection is highly questionable…"
>
> [1982] What they [his experimental results] do not, and cannot, demonstrate is …whether or not it [his methodology] can be developed into a component part of an operational commercial system.
>
> [1994] Unfortunately, … viable methods for automatically building large hypertext structures and for using such structures in a sophisticated way have not been available.
>
> [1990] There is a great deal of research to be done. . . . **We need a great deal more understanding** of how to model users and their information needs, as well as how to collect unobtrusively the required data to build these models. We need better ways to forms links automatically between citation markers and cited works and between related discussions.

(Ex. I, Declaration of Paul Jacobs, ¶¶ 172-198). These experiments demonstrate a lack of appreciation of the importance of indirect relationships in search and understanding of how to effectively use indirect relationships in the manner claimed. They further show why the use of indirect relationships for computer search never became of conventional or routine use prior to the patents in suit but was confined to a set of a few experiments by researchers. Attached as Appendix A is a timeline summarizing in more detail the few experiments have been done direct and indirect relationships. This is further discussed in detail in the Expert Reports of Dr. Amy Langville and Dr. Paul Jacobs, incorporated by reference here and attached as Exhibits H and I.

37. No commercial search engine used any analysis of indirect relationships prior to the filing of the patents in suit. Upon release, Google's search engine was hailed as a major innovation. Sergey Brin, the founder of Google, describes how the notion that you could do "anything meaningful" using the hyperlinks for search was a "revolutionary idea":

> We originally developed PageRank kind of playing around with all the links on the web and that too was a pretty **revolutionary idea**, though it seems very simple, the fact you can even just collect [the links] and **do anything meaningful with them** …

(Ex. H, Langville Decl., ¶ 63). Brin further noted that their discovery of studying hyperlinks was unexpected rather than predictable or in conventional use:

> And we sort of stumbled upon a way to do that by studying links. . . . But what we found was we-- kind of by accident almost-- we found that this processing of the link structure of the web, we could create a search that was better in important ways. In ways that these search engines had ignored.

(Ex. H, Langville Decl., ¶ 62). Google quickly overtook all other existing search engines which did not analyze indirect relationships for search or employ the specifically claimed data structures and algorithms described in the patents in suit. The notion that the other major search engines in 1998 (years after the patents in suit) did not know that "anything meaningful" could be done with analyzing the links on the web and "ignored" them further supports the lack of conventionality, routineness and well understanding of the use of indirect relationship for automated retrieval, particularly those relationships involving hyperlinks.

38. Similarly, the fact that Google used indirect link analysis to take over a search industry that did not use the technology after the patents in suit demonstrates that at the time of the

patents in suit link analysis of indirect relationships of web links was not well understood or in conventional or routine use by the industry. Google's Search Engine was highly successful, quickly dominated the search engine market, and rendered all previous search engines that did not use the patented technology obsolete. The success of the Google Search Engine was largely attributed to its analysis direct and indirect non-semantic relationships using the PageRank algorithm:

> "Last week, after months of testing, the two 26-year-old PhD candidates from Stanford University formally launched what many experts consider to be the most powerful search tool on the Net, called Google. **Its edge over other search engines lies in sophisticated mathematics for analyzing links among hundreds of millions of Web pages, and then ranking the pages by relative importance**." Google peers into a future where there will be so many Web pages that conventional search engines will be useless, thus raising the value of software that can deliver.
>
> **the main reason** [Google] was so successful there are technologies that we developed initially that made it work really well and having to with using the web as a whole [(*i.e.*, link structure)] rather than just what words appear on each page."

Neil Gross, Movers & Shakers: "Can Google's Prodigies Make a Search Tool Pay?"

(http://www.businessweek.com/ebiz/9909/em0929.htm)

> In 1998, Google didn't exist; Yahoo and Alta Vista were leading the young search industry, and there was no place for a late comer. By bringing to the market a major innovation (***the "page rank" technology***), Google put the previous order of competitors upside down.
>
> . . . .
>
> As Page stated in the provisional patent application for Google:
>
> *The reason why my system works so well* is that it decides which documents to return, and in what order, by using an approximation to how well cited or `important' the matching documents are.
>
> . . . .
>
> A Google research director, Monika Henzinger, has described that:
>
> The biggest "success story" is certainly the PageRank algorithm … It led to significant improvement in search quality and gave rise to the creation of the Google search engine … the PageRank algorithm ***initiated*** research ***in hyperlink analysis on the web***, which has become a flourishing area of research.

(Ex. H, Langville Decl., ¶¶ 90, 92, 94). Although Google was a late comer, it use of link analysis allowed it to take over the search market. *Id.* ¶¶ 43-65, 66-114. Companies that did not deploy link analysis in their search engines could not compete and only those companies that remained were ones that used adopted this technology. *Id.*

39. Now major search engines use link analysis embodied in the claims of the patents in suit.

> Perhaps the best known innovation in Google is link popularity. All major search engines now use this technique in somewhat different forms.

http://innovations.ziffdavisenterprise.com/2007/01/tips_and_tricks_for_raising_yo.html (archived at web.archive.org). Indeed, SRA has licensed over ninety percent of the search engine market under the patents in suit. See section VI.E, *infra* regarding how the Patents-in-Suit describe and specifically claim the major computational features of the PageRank algorithm.

D. The Patents in Suit Are Directed to Specific Improvements in Computer Technology by Describing Unconventional Database Representation and Data Structures Representing Non-Semantic Indirect Relationships that Can be used to Efficiently Search Databases and Display Results

40. As part of the solution of using non-semantic indirect relationships the patents describe several unconventional data structures (described below) representing indirect relationships to allow the computer to efficiently locate and display objects of interest and thereby improve the functioning of the computer itself. These claimed data structures that analyze and represent indirect relationships for search and the search and display methods utilizing these data structures represent inventive concepts that are unconventional, non-routine and was not well understood. These data structures and their claimed use define specific technological improvements or implementation to the functioning of the computer itself and represent technological solutions the technical problems identified in the specification concerning semantic search: (1) Proximity Index; (2) Cluster Links (3) Clusters of Indirect Relationships; and (4) Patterner, Patterner, Patterned Vectors, Opinion Pattern Matrix and Scalar F.

Proximity Index

41. One objective of the patents in suit is to create a "proximity index" to allow for efficient search of a database:

> It is an object of the invention to utilize statistical techniques along with empirically generated algorithms to reorganize, re-index and reformat data in a database into a more efficient model for searching [i.e., improved efficient data structures for search].
>
> It is an object of the invention to utilize statistical techniques along with empirically generated methods to increase the efficiency of a computerized research tool [improved efficient search methods using statistical analysis].

> It is an object of the invention to create a system of computerized searching of data that **significantly reduces the number of irrelevant objects retrieved**.

(Ex. E, '571 patent, 7:41-50). Proximity Indexing" is a method of preparing data in a database for *subsequent* searching by advanced data searching programs." The proximity index is a representation of the database and its relationships. *Id.* 4:5-9 ("The Proximity Indexing Application Program indexes (or represents) the database in a more useful format to enable the Computer Search Program for Data Represented by Matrices (CSPDM) to efficiently search the database."). Complex link analysis concerning non-semantic relationships and other significant data concerning a database or network can be calculated prior to the search and placed into a proximity index. The proximity index serves as a quick-reference" so that search merely has to retrieve a value from the proximity index, rather than calculate complex values during the pendency of a search request:

> The invention can be used with an existing database by indexing the data and creating a numerical representation of the data. This indexing technique called proximity indexing generates a **quick-reference** of the relations, patterns, and similarity found among the data in the database. Using this proximity index, **an efficient search for pools of data having a particular relation**, pattern or characteristic can be effectuated. This relationship can then be graphically displayed.

(Ex. E, '571 patent, 3:31-39). A Proximity Index data structure containing representations of indirect relationships for the purpose of search was not in conventional or routine practice, nor was it well understood by the industry at the time of the patents in suit. See ¶¶ 25-28, *supra*, regarding the use of indirect relationships for search.

42. The proximity indexes of the patents-in-suit and their claims contain at least three types of unconventional data structures representing (and analyzing) the indirect relationships between objects in a database that further constituted an inventive concepts: (1) cluster links; (2) clusters; and (3) Pattern Matrix and Scalar F.

Cluster Links Representing Indirect Relationships

43. One improvement to the functioning of a computer is an improved data structure representing a relationship between two indirectly linked objects in a database. As shown in Fig. 3 of the V-Search Manual incorporated by reference to the specification, the Cluster Link Generator mathematically analyzes "clusters" of "links" between two nodes (hence, a "cluster link") to define the statistical relationship between the two nodes:

V-Search Publisher's Toolkit User's Manual

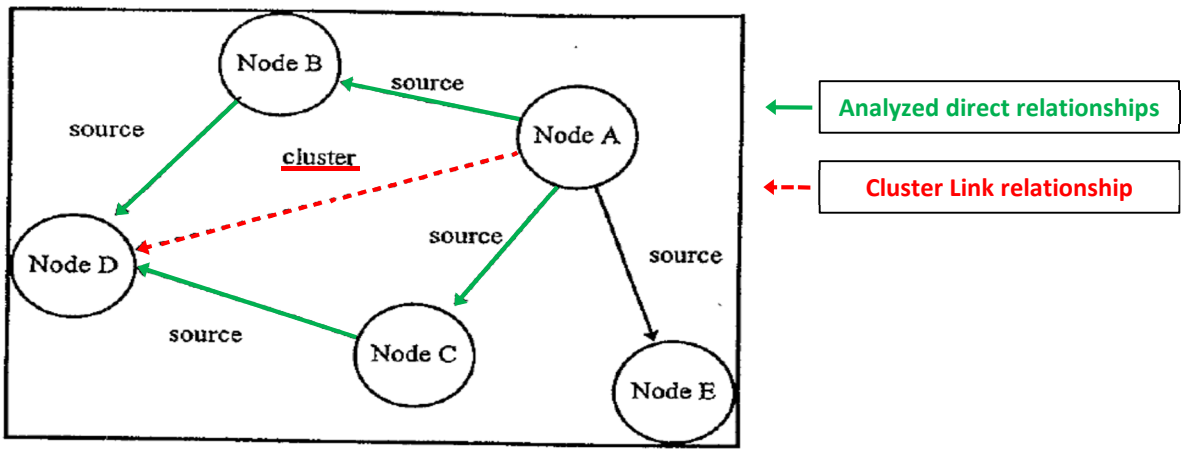
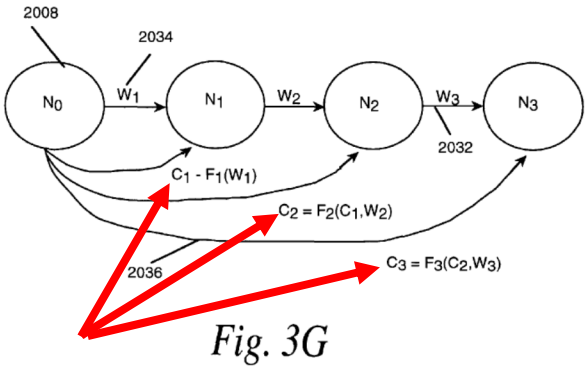


Figure 3: A Cluster link from Node A to Node D.

Ex. J, V-SEARCH PUBLISHER'S TOOLKIT USER'S MANUAL 4, fig.3 (1995) (emphasis added) (hereinafter "V-SEARCH MANUAL").[1]

44. As set forth in Figure 3G below, the cluster link is a value obtained from statistical analysis of direct links in a set of paths between the nodes. As depicted below, "[t]he set of cluster links are also shown in the figure [3G] as functions of the weights associated with the direct links…" '494 Pat., 8:52–56. The three values C1, C2 and C3 are cluster links for the node pairs (N0-N1); (N0-N2); and (N0-N3) and are produced by the statistical analysis of the path between the two nodes of the pair. In Fig. 3G, the value is determined as a function of the weight of the direct links between start node N0-and one of the destination nodes N1, N2, or N3.

[1] V-Search was one of Mr. Egger's commercial embodiments and its corresponding V-SEARCH PUBLISHER'S TOOLKIT USER'S MANUAL was included in the prosecution history.

 *Fig. 3G* **Cluster links**	Consider a set of nodes 2008 N_0 . . . N_3 connected by a ***sequence of direct links 2004 whose weights 2034 are given by W_1 . . . W_3,*** as shown in FIGS. 3F. . . . Each path provides some evidence that the start node 2008 (N_0) and destination node 2008 (N_1, N_2, or N_3) are related to some extent. ***The strength of the implied relationship depends on the length of the path, and on the weights 2034 of the individual direct links 2004 along that path.*** (Ex. C, '494 Pat., 22:5–15) (emphasis added).

Thus, relationship is expressed as a value or number in an index, which is used to represent the strength of relationships of node pairs as defined by set of paths of direct link between the node pairs. *See* Figs. 3G (below) and 3 (above). In this way, the interconnectivity of many link paths and other factors between two nodes may be expressed by a single value or number indicating the strength of the relationship. The previous construction of cluster link by this Court is:

> a relationship between two nodes based upon a statistical analysis of multiple relationships between nodes in a database

(Ex. K, Claim Construction Order at 15-18).

45. The claims of the patents in suit explicitly claim the cluster link data structure for use in search or display. For example, claim 16 of the '571 patent provides wherein Universal Resource Locators which have an indirect relationship to the chosen document are located, wherein the step of analyzing further comprises analyzing the Universal Resource Locators for indirect relationships *using cluster links;* and displaying a located document." Similarly, the claims reflect limitations that are directed to generating the cluster links or other similar data structures concerning indirect relationships. For example claim 23 of the 571 patent provides: *the step of proximity analyzing comprises: analyzing indirect relationships by scoring one or more paths of direct links between two indirectly related nodes by analyzing weights associated with direct links that make up the path between the nodes.*

46. Cluster links and the disclosed steps for generating them (e.g., scoring one or more paths of direct links between two indirectly related nodes by analyzing weights associated with direct links that make up the path between the nodes) are specific technological improvements to the functioning of the computer itself to address problems in the prior art with sematic search. This improved data structure/data representation and the ordered combination of the methods using and creating them represent inventive concepts that are unconventional, not well understood or in routine use. First, it pertains to representing non semantic relationships and in the case of '494 claim 52 by way of example that are related to hyperlinks in a network. The use of indirect relationships at all, much less these specific data structures representing these relationships, were not conventional, routine, or well understood at the time of the filing of the patents in suit. The use of the relationships greatly improve search methods by allowing for a more accurate search that has less irrelevant objects and more efficient display of the most relevant objects. The data structure "Cluster link" allows for a more efficient and accurate search of a database to provide the most relevant results over systems that do not use them. The cluster link is superior to other data structures in that it can be used to represent any indirect relationship of any length or pattern. The vector models used in prior art experimentation were limited to relationships of a certain pattern. Since the cluster link uses a statistical analysis, it can be used to judge the strength of the indirect relationship of a given node pair verses other relationships between two nodes, thereby allowing for differentiation between types of indirect relationships. The cluster link also accounts for multiple indirect paths of differing length between two nodes which prior art structures did not. A cluster link can be calculated prior to a search and is disclosed as part of the patterner index. (*E.g.*, Ex. E, '571 Patent, 21:30-33). All the computer needs to do is retrieve the cluster link value rather than make complex calculations at search time. The single value representing multiple relationships that can be retrieved also reduces search time processing. This allows the quick processing of complex analyses and ranking of search results for purposes of importance which allows the system to identify efficiently the most important results among a pool of otherwise relevant objects that have the search terms. Similarly. the cluster link scores can be used to identify clusters of only the strongest links. These clusters can then be used to locate nodes for purposes of

display.. This allows the quick processing and ranking of search results for purposes of importance which allows the system to identify the most important results among a pool of otherwise relevant objects that have the search terms. Similarly, the cluster link scores can be used to identify clusters of only the strongest links. These clusters can then be used to locate nodes for purposes of display.

Clusters of Indirectly Related Nodes

47. The patents describe the creation of clusters of links indirectly related to a chosen node that bear the strongest relation to the chosen webpage. These clusters of indirect links are then used to locate nodes for display in the search routines.

48. The specification describes for a selected node, the preferred embodiment cluster link generator "classifies" a "set" of nodes as being indirectly related to the selected node, i.e. a cluster of indirectly related nodes for a selected node:

> FIG. 14B describes the embodiment of the invention which executes 3020 the cluster link generator algorithm 2044 to generate direct and indirect links 2004 to find the set of candidate cluster links. After identifying 3008 all of the URLs referenced in the source web page, in the preferred embodiment, the cluster link generation algorithm 2044 retrieves 2056 a list of URLs and classifies them as the direct links 2032 to be analyzed. ***The cluster link generator 2044 traces the links 2032 to their destination nodes 2008 (a web site or web page) and performs a web crawl to retrieve 2056 a list of URLs referenced by the source nodes 2008. The generator 2044 classifies the second set of nodes 2008 as being indirectly linked to the source node 2004, and the links 2036 to these nodes 2008 are added 2072 to the list of candidate cluster links***. In order to find the set of candidate cluster links, the cluster link generator 2044 repeats the above steps 2052. In the more general method described in FIG. 14A, the system identifies 3012 the links 2036 which have an indirect relationship and then displays 3020 the direct 2032 and indirect 2036 links.

(Ex. E, '571 Patent, 49:36-56)

49. Since cluster links may contain values relevant to the strength of an indirect relationship, they can be used to identify the strongest indirect links for a selected webpage called the "actual cluster links" from a list of generated "candidate" cluster links:

> In this embodiment, only a subset of the candidate cluster links 2004, the actual cluster links 2004, which meet a specified criteria are used to identify nodes (2008) for display 38.

(Ex. C, '494 Patent, 22:1-4)

> The set of all candidate cluster links 2004 is then sorted by weight 2034. A subset of the candidate links 2004 is chosen as actual cluster links 2004. The number of cluster links 2004 chosen may vary, depending on the number of direct links 2004 from N_0, and on the total number of candidate cluster links 2004 available to choose from.

Id., 22:40-45

> Once the candidate cluster link 2004 set has been generated, deriving the actual cluster links 2004 is a simple matter of selecting or choosing the T top rated candidate links 2004, and eliminating the rest.

Id., 24:1-4

> Following weighting, the generator 2044 sorts the set of candidate cluster links 2004 by weight, and a subset of these links 2004 (those links 2004 above a specified cut-off weight) are retained for display 3020 to the end user.

Id., 50:15–22. Thus, the set of chosen actual links from the candidate set represents a cluster of the strongest indirectly linked nodes to a selected node.

50. The data structure of "clusters" of indirectly linked nodes used for search and display and the ordered combination of processing steps creating and using them represent improved data structures and methods that were not conventional, routine or well understood at the time of the filing of the patents in suit. This inventive concept improves the functioning of the computer itself over the art and are used to create more efficient search and remedy problems in prior art searching.

Patterner, Pattern Vectors, Opinion Pattern Matrix and Scalar F

51. The patents in suit describe a patterner which creates a matrix or table as part of a proximity index that contains numerical representations of indirect relationships and other factors to efficiently search the database. The cluster link generator is disclosed as a type of patterner and is used in the search routines that refer to patterners. (Ex. E. '571 Patent, 21:30-33).

52. The patents in suit captured these useful indirect relationships of Fig 5, among other ways, in specific types of improved data structures in the form of weighted pattern vectors created as follows:

> For purposes of explaining how patterns are used to generate the Proximity Index, only the two simplest patterns are illustrated.
>
> The simplest, Pattern #1, is "B cites A." See FIG. 6. In the notation developed, this can be diagramed: a b c A d e f B g h i where the letters designate textual objects in chronological order, the most recent being on the right, arrows above the text designate citations to A or B, and arrows below the text designate all other citations. The next simplest pattern between A and B, Pattern #2, is "B cites c and A cites c," which can also be expressed as "there exists c, such that c is an element of (A intersect B)." See Appendix #1. This can be diagramed: a b c A d e f B g h i. For every textual object c from 0 to (A-1), the existence of Pattern #2 on A and B is signified by 1, its absence by 0. This function is represented as P#2AB(c)=1 or P#2AB(c)=0. The complete results of P#1AB and P#2AB can be represented by an (A)x(1) citation vector designated X.

> The functions of some Patterns require an (n).times.(1) matrix, a pattern vector. Therefore it is simplest to conceive of every Pattern function generating an (n).times.(1) vector for every ordered pair of full textual objects in the database 54, with "missing" arrays filled in by 0s. Pattern Vectors can be created for Pattern #1 through Pattern #4 by just using the relationships among textual object A and the other textual objects in the database 54 and among textual object B and the other textual objects in the database 54. Pattern Vectors for Patterns #5 through #18 can only be created if the relationship of every textual object to every other textual object is known.

(Ex. E, '571 Patent, 14:61-15:22)

53. As the patent discloses, one or more pattern vectors can be used to form an opinion pattern matrix representing the indirect relationships in the database. The values contained within this matrix are used to create ranks for or to determine the most relevant results of the database. The Opinion Pattern Matrix took multiple pattern vectors of different types of indirect relationships to create values that reflected the relationship between two indirectly related nodes:

> **Process the Opinion Citation Matrix through each of the pattern algorithms described above and in FIG. 6 for each ordered pair of full textual objects to create opinion pattern vectors for each pattern and for each pair of full textual objects**. The pattern algorithms determine relationships which exist between the ordered pair of textual objects. The first four pattern algorithms can be run utilizing just the Opinion Citation Vector for the two subject full textual objects. Each pattern algorithm produces a opinion pattern vector as a result. The fifth through eighteenth pattern algorithms require the whole Opinion Citation Matrix to be run through the Opinion Patterner Subroutine 100.

(Ex. E, '571 Patent, 17:38-49). The pattern vectors are used to generate a weighted number called "scalar f" that represented a relationship between two nodes and which were entries in the Opinion Pattern Matrix:

> Calculate a weighted number F(A,B) which represents the relationship between full textual object A and full textual object B. The weighted number is calculated using … each of the 18 patterns. The weighing algorithm uses empirical data or loading factors to calculate the resulting weighted number.
>
> . . . .
>
> Compile the Opinion Pattern Matrix by entering the appropriate resulting numbers from the weighing algorithm into the appropriate cell locations to form an n [(times)] n Opinion Pattern Matrix

(Ex. E, '571 Patent, 18:5-20). The Pattern Vectors, Pattern Matrix, and Scalar F all represent unconventional data structures and representations describing indirect relationships in a proximity matrix. These features including the systems, matrices, scalars, and other data structures are

inventive concepts that are not conventional routine or well understood. See section above about indirect relationships and cluster links. These inventive concepts represent technological improvements to prior art search methods and data structures and remedy deficiencies and problems of prior art systems.

E. Additional Data Structures and Search Method Technological Improvements

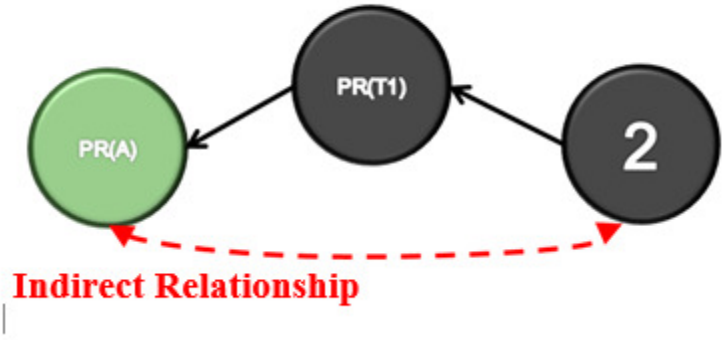
54. The patents in suit describe and claim additional unconventional improvements to the data structures and methods of search ranking analysis that improve the ability to efficiently search. The patents disclose numerical representations of indirect relationships (*e.g.*, cluster links and scalar f) and search methods that use ranks or values involving (1) recursive analysis of direct links; (2) damping weighting factors; (3) the number of hyperlinks on a page weighting factor; and (4) visits or views of web objects weighting factor. Each of the limitations are directed to improvements in the data structures representing indirect relationships used for search and the analyses used by search methods.

55. It should be noted that the first three of these additional factors: (1) recursive analysis of hyperlinks; (2) damping weighting factors; and (3) the number of hyperlinks on a page weighting factor) are explicit elements of the PageRank algorithm used by Google that revolutionized Web Search. First, PageRank analyzes indirect relationships by a recursive analysis of direct links on the web: "PageRank handles both these cases and everything in between by recursively propagating weights through the link structure of the web." (The Anatomy of a Large-Scale Hypertextual Web Search Engine, available at http://infolab.stanford.edu/~backrub/google.html).

56. The formula for calculating the PageRank of the page, as provided by *The Anatomy of a Large Scale Search Engine,* by Brin and Page (1998) (attached as Exhibit N) is as follows:

> We assume page A has pages T1....Tn which point to it (*i.e.*, are citations). ***The parameter d is a damping factor*** which can be set between 0 and 1. We usually set d to 0.85. Also ***C(A) is defined as the number of links going out of page A***. The PageRank of a page A is given as follows:
>
> $$PR(A) = (1-d) + d(PR(T1)/C(T1) + ... + PR(Tn)/C(Tn))$$

(Ex. N at 4). As shown above, the variable "c" of PageRank is the number of hyperlinks on a page, and variable "d" is a damping factor. Furthermore, the PageRank algorithm scores paths of direct links between two nodes:



"PageRank is a link analysis algorithm and it assigns a numerical weighting to each element of a hyperlinked set of documents, such as the World Wide Web, with the purpose of "measuring" its relative importance within the set. The algorithm may be applied to any collection of entities with reciprocal quotations and references." (https://en.wikipedia.org/wiki/PageRank) (archived at web.archive.org).

Using Recursive Analysis and/or a Damping Factor to Represent Indirect Relationships of Cluster Links or Scalars

57. Additional dependent claims describe further improvements to the data structure and search methods that represent independent grounds for patent eligibility. These claims, including for example, claim 46 of the '494 patent:

> 46. The method of claim 18, wherein the direct relationships are hyperlink relationships between objects on the world wide web and wherein generation of the second numerical representation uses a **recursive analysis** of a set of direct links between two objects and a **damping factor**; and said direct link weights are calculated using a quantity of direct relationships of an object.

58. The recursive analysis of direct links is described by the patent, including for example Fig. 3G. Fig. 3G shows how an indirect relationship (i.e., between nodes N0 and N3) is analyzed by considering the path from N0 to N3 and assigning weights W1, W2, W3 to each link in the path. The overall value of the indirect relationships is determined based on these weights and not just the path between the nodes.

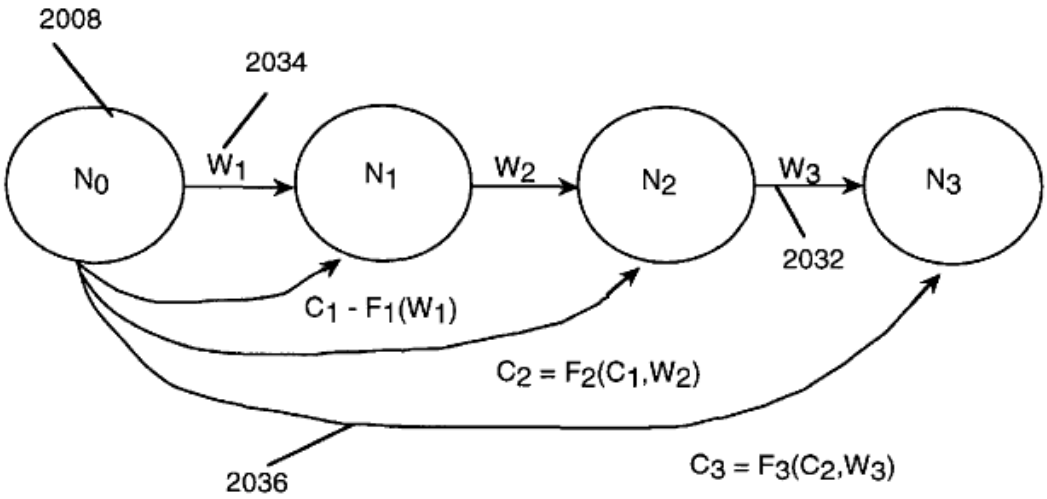


Fig. 3G

The algorithm of Fig. 3G evaluates indirect relationships by recursively analyzing the direct link between the nodes as shown by the depicted function C2=F2(C1,W2). This recursive analysis of the path of direct links that use the prior calculation of C1 in generating C2 results in a faster computation time because of not having to repeat the prior calculation. It also accurately reflects the analysis of direct links for any type of relationship which improves the data representation and the usefulness of the data structure for search. Consideration of the impact of any particular direct link weight is enabled by this analyses, thereby, thereby improving over prior art.

59. The patents also disclose using a damping or decay factor to differentiate certain link relationships from others when generating the values that constitute cluster links:

> The individual functions F1 . . . F3 describe how to combine the weights 2034 of the direct links 2004 to determine the weight 2034 of an implied link. Selecting appropriate functions is the key to making cluster link generation work well. A

> preferred definition of F_N is as follows: . . . where D_N is a ***damping factor*** that decreases rapidly as N increases.

(Ex. E, ‘’571 Patent, 22:23-32).

> wherein the combined weight, WC_{i+1}, is computed from the weight of the path P (WC_i), **a dampening factor (D_{i+1})** and the weight of direct Link L (W_{i+1}), and wherein the combined weight is computed using the following formula: $WC_{i+1}=\min(WC_i, D_{i+1}*W_{i+1})$;

(Ex. C, ’494 Patent, Claim 17, 53:20-25). Damping or decay factors involve using weights to weaken or lower the contribution of a particular link to a rank or score. For example, the algorithm weakens the contribution to the score using a damping factor based upon length of relationship (number of direct links) of two indirectly related nodes. Similarly the algorithm can weaken the contribution of a link based upon time or type. Thus, the use of a damping factor improve the data representation of an indirect relationship by allowing for differentiation of links based upon important characteristics. The use of a damping factor in the ordered combination of steps generating data structures or search methods pertaining to indirect relationships is an inventive concept that improves search methods and data structures relating to indirect relationships. This use in connection with indirect relationships was not conventional, routine or well understood at the time of the patents in suit. While highly experimental prior art systems investigated the use of indirect relationships, they focused on the mere presence of certain relationships such as bibliographic coupling (bc) and co-citation (cc) but did not further evaluate more complicated indirect relationships between objects or derive different values for those indirect relationships based on a value assigned to direct relationships between objects in the relationship. (Ex. H, Declaration of Amy Langville, ¶ 50).

60. The claimed invention improved upon the prior art experiments that tested the efficacy of using indirect relationships for search and these specific algorithms and improvements were directly responsible for why the invention is able to achieve improved search results whereas leaders in the field determined indirect relationships degraded search. As Dr. Langville testified, the prior art experiments of Fox would not have found bc and cc to improve search results as he was testing the wrong indirect relationships. *Id*.

61. Bc and cc alone does not provide the necessary improvement over semantical search. Rather, the recursive analysis of higher order indirect relationship is required to produce the improved results. Furthermore, weights must be assigned to each link in the path to determine the most important indirect relationships to further refine a search. *Id*.

62. Thus, the patents are directed to this specific improvement to the shortcomings of the prior art. Because the leaders in the field did not appreciate the use of indirect relationships for search, the specific method described in the patent of using recursive analysis and weighted links must be unconventional and not routine. Indeed, the balance of all evidence shows experts in the field dismissed the use of indirect relationships and determined the ones they did investigate could not be effectively used for search.

63. More importantly, both the patents and the claims themselves describe how recursive analysis and weighted links are used to improve searching using indirect relationships rather than merely reciting the abstract idea of searching using recursion or weights. The claims specifically generate a second numerical representation based on a set of direct links between two objects, and those direct links are assigned weights which influence the value of the final second numerical representation.

F. Use of Weights That Considers the Number of Outbound Links or Views by Users of a Website

64. The patent discloses the number of hyperlinks on a webpage and visits by users to a website) as a weighting factors in the searching ranking mechanism and data structures of the preferred embodiment.

> The Program 62 weighs the patterns by importance, giving one type of data document more importance than another type. For example, it may give more importance to a web site than to a single document which has no other links. The Program 62 may use other factors to weigh the data documents, such as the number of "hits" (***visits by other end users to the site***, a number which is available to web users) a data document receives in a specific time frame or the ***number of hyperlinks within a page***. The Program 62 then forms a matrix based on ordered pairs of documents, and the matrix calculations discussed before of this specification can be carried out. The Program 62 generates a coefficient of similarity which will determine the relatedness of web pages to each other and to the source web page. The Program 62 displays the most similar web pages to the user.

The ordered combination of using these weighting factors in combination with indirect relationships in search methods and data structures is a technological improvement that was unconventional, not well understood, and non-routine at the time of the filing of the patents.

65. Several claims claim using the number of hyperlinks on a page as a weighting factor. Similarly, other claims describe algorithms that consider the number of outbound links and further represent another independent ground for patent eligibility. These claims include, for example, claim 39 of the '494 patent:

> 39. The method of claim 38, wherein one or more of said direct links includes a weight based upon a quantity of direct references from an object to other objects.

66. By considering the number of outbound links, the patents provides a technological solution to the technological problem of searching large databases by weighing objects based on the number of outbound links in connection with a use of indirect relationships. This process identifies objects that have more outbound links than other objects, which then used to rank or create data structures using indirect relationships. This weighting factor can be used to differentiate the strength of different indirect relationships for a more accurate data representation of the indirect link. This improves upon prior systems which could not rank results in this manner or display them to the user is any order. The importance of this weighting factor is shown by the fact that this factor was a key component of the tremendously successful PageRank algorithm.

67. The inventive concept of combining this weighting factor with an analysis of indirect relationships was not conventional, well understood and/or routine at the time of the filing of the patents in suit. The prior art used semantical algorithms for search, and the few investigations into non-semantical search did not use the number of outbound links to order search results.

68. The patents also disclose and claim algorithms that consider the number of times an object is visited and further represent another independent ground for patent eligibility. Claims embodying this feature include, for example, Claim 28 of the '571 patent:

> 28. A method for visually displaying data related to a web having identifiable web pages and Universal Resource Locators with pointers, comprising:
>
> choosing an identifiable web page;

identifying Universal Resource Locators for the web pages, wherein the identified Universal Resource Locators either point to or point away from the chosen webpage;

analyzing Universal Resource Locators, including the identified Universal Resource Locators,

wherein Universal Resource Locators which have an indirect relationship to the chosen web page are located,

wherein the step of analyzing further comprises cluster analyzing the Universal Resource Locators for indirect relationships; and

displaying identities of web pages, wherein the located Universal Resource Locators are used to identify web pages,

wherein the step of displaying is influenced by a number of times a web object is visited and wherein the cluster analysis uses a damping factor.

(Ex. F. '571 Pat. Reexam Cert., 4:11-33).

69. The patents specifically disclose using a number of visits to websites to weigh data documents is a specific way of improving the accuracy of search results obtained through importance searching. (Ex. E, '571 Patent, 50:12-27) ("The Program 62 weighs the patterns by importance, giving one type of data document more importance than another type. … The Program 62 may use other factors to weigh the data documents, such as the number of "hits" (***visits by other end users to the site***, a number which is available to web users)…). The combination of a factor based upon visits to a web object with an analysis of indirect relationships is an inventive concept that is not conventional, routine or well understood. The number of visits to a web page is an indicator of how important or "of interest" a particular website may be. This factor may be used to increase the contribution o f a link relationship involving a highly visited objects when generating the values for cluster links and scalars for indirect relationships. Similarly this factor can be used in conjunction with cluster links and scalars of indirect relationships for determining the importance of an object and if and what position such object will be displayed to the user.

70. The prior art used semantical algorithms for search, and the few investigations into non-semantical search did not use the number of visits to order search results. No commercial search engines used web link analysis of indirect relationships in combination with this weighting factor for search until well after the filing of the patents. No experimental systems incorporated

this element. It should be noted that the examiner specifically relied upon the presence of this factor in claims analyzing indirect relationships in finding the patentability of claims 26, 28 and 31. (Ex. L, IPR2013-00481 Institution Decision at 21-25)

G. Improved Interface and Methods for Displaying Retrieved Information

71. The patented invention also reflect improvements to the display of a computerized search system. The patents describe and claim an improved interface and methods for displaying retrieved information. The patents disclose a user interface that significantly enhances the presentation of search results to the user:

> A computer research tool for indexing, searching and displaying data is disclosed. Specifically, a computer research tool for performing computerized research of data including textual objects in a database or a network and for providing a **user interface that significantly enhances data presentation is described**. …

(Ex. E, '571 Patent, Abstract).

> The invention simplifies the research task by improving upon methods of searching for data including textual objects and by implementing a user interface **that significantly enhances the presentation of the data**.

(Ex. E, '571 Patent, 3:27-30).

72. As described by the patents, a serious problem of existing semantical based search algorithms was their inability to present the most important search results in a pool of hundreds to millions of results to the user in an easily accessible manner. As the patents improved upon existing semantical methods by ranking results in importance by analyzing the indirect relationships of all objects in the database, the patents also improve upon presentation of those results to the user in various manners.

> The Internet can be viewed as an immense collection of linked documents providing varied information to the public via an elaborate electronic distribution channel. In the past, the end user's ability to search, find, index, and navigate through relevant documents of interest has been primarily limited to word based queries which primarily rely on the target document's text indexing. Instead of relying on textual searching, this method and apparatus for indexing, searching, and displaying data analyzes hyperlinks which connect web pages to other web pages in order to help the end user to search, find, and navigate through the relevant documents of interest. This system analyzes hyperlinks using proximity indexing or clustering technology discussed previously. Once identified, the system displays the results in a variety of ways and end users are able to navigate directly to the documents identified by this system's analyzation technology.

(Ex. E, '571 Patent, 48:46-62).

73. The improved display and display methods solves the problem of overwhelming a user with millions of potentially relevant results by displaying only objects as determined by an analysis of indirect relationships to allow the user to quickly navigate or identify the desired result:

> The Computer Search program, called the Computer Search Program for Data represented in Matrices (CSPDM), provides efficient computer search methods. The CSPDM rank orders data in accordance with the data's relationship to time, a paradigm datum, or any similar reference. An alternative embodiment of the invention employs a cluster link generation algorithm which uses links and nodes to index and search a database or network. The algorithm searches for direct and indirect links to a search node and retrieves the nodes which are most closely related to the search node. The user interface program, called the Graphical User Interface (GUI), provides a user friendly method of interacting with the CSPDM program and prepares and presents a visual graphical display.

(Ex. E, '571 Patent, Abstract)

> The remaining two programs in the present invention are the CSPDM and the GUI Program. The CSPDM has seven subroutines that each search for different pools of objects. The GUI Program also has seven subroutines. Each CSPDM subroutine performs a different type of search. Each of the subroutines of the GUI uses the results of the corresponding subroutine of the CSPDM to create the proper display on the display.

(Ex. E, '571 Patent, 4:42-50).

74. The patented display can consist of a list of the most important items or produce a map of the relationships of the objects in the database. The patents also show an improved display that uses indirect relationships to display a list of only important or related objects, thereby preventing information overload to the user of every object that matches the search results:

> The graphical user interface program 70 displays a list of the most related web pages to the source web page. This list includes documents, web sites, and pages which are directly or indirectly linked to the subject document or the subject topic. . . . The GUI program 70 displays all of this information either in the list format or in the text box 1032.

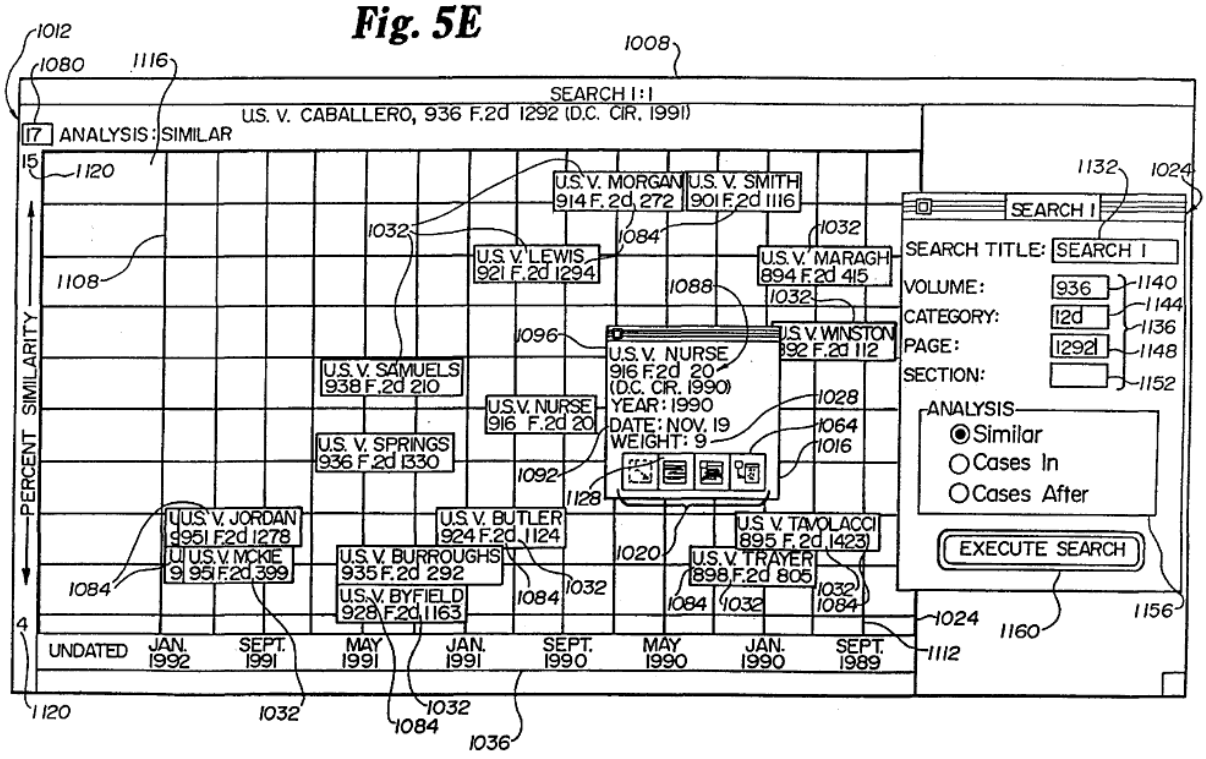
(Ex. E, '571 Patent, 50:50-64).

75. The patent further describes specific innovative features of the user interface. For example, the interface may include a map of a particular object in the database and its relationship to other database objects:

> By using the graphical display, the researcher can view immediately a visual representation of trends in the data (for example, trends developing in the law and current and past legal doctrines). In addition, the researcher can immediately identify important data or important precedent and which object serving as the precedent is most important to the project on which the researcher is working. **This visual representation is a vast improvement over the current computerized research tools.**

(Ex. E, '571 Patent, 7:15-23).

Fig. 5E



76. For example, the interface may include a map of a particular object in the database and its relationship to other database objects:

> **An innovative feature of the preferred embodiment is the ability to call up a search screen or map while viewing the data of a particular object in the** database 54. This feature is implemented through the use of embedded active links 2004. By using embedded icons that are active within the data of an object being viewed or by using embedded text which is active within the data of an object in the database 54, this feature allows the user to jump from viewing data to a search screen, menu, map or the like. The search screen or map can be one which has been previously generated or can be generated at the time of selecting the embedded active icon or active text. The preferred method of using this feature is with text documents. Active icons or active text are embedded within the text documents and the user is alerted to these active icons or text through the use of highlighting or different coloring of the active icon or text. When the user sees an active icon or active text while viewing an object in the database 54, the user may choose to jump out of the object and into a map, search screen, or the like.

(Ex. E, '571 Patent, 45:58-46:10).

77. The patented display is an improvement over the prior art in several ways. First, the display uses link analysis of indirect relationships to identify the most relevant objects and only presents a subset of the most relevant objects to the user, reducing the amount of information

presented and information overload. Further, the displayed identities of the most relevant objects have embedded active links that allow users to easily navigate between the most relevant objects:

> The preferred embodiment of the network application of this system uses the graphical user interface program 70 to display the results of the algorithm as a list showing the selected links 2004 and the various data associated with the links 2004. The links 2004 shown on the screen to the end user are active links 2004, similar to the active comments used in the text boxes 1032 described previously in this application. The end user may instantaneously link to the destination node 2008 that the user selects. The list format provides link information in a style familiar to user of the Internet. However, this system is also capable of displaying the results in the user-friendly graphical format as described above. The graphical user interface program 70 described previously uses box coloring and sizing to communicate large amounts of information quickly and intelligibly to the user. In a preferred embodiment, different colors for boxes 1032 are assigned depending on what type of node 2008 they represent (e.g., a web page, web site, a document, a file transfer protocol (FTP) (a common internet designation for news sites)). Preferably, the box 1032 is given depth. The amount of URL links a node 2008 contains may determine the amount of depth.

(Ex. E, '571 Patent, 50:28-49). This list of active links to the most relevant objects allows improved navigation through the network.

78. Each of these improvements represent an unconventional patent-eligible inventive concept. Using link analysis to identify a subset of relevant objects for display was not conventional or routine in the art. Displaying the objects identified by the link analysis of indirect relationships was further unconventional and not routine. And incorporating hyperlinks in the display to allow users to navigate between relevant displayed objects was certainly unique to the patents and not found in the prior art. The art simply did not understand or appreciate using an analysis of indirect hyperlink relationships to create a display of hyperlinks to objects in any way.

79. These features are present in the claims, for example, claim 23 of the '571 patent, which includes the specific steps of identifying URLs for web pages, analyzing the URLs for indirect relationships, and then displaying the web pages using the URLs and the analysis:

> 28. A method for visually displaying data related to a web having identifiable web pages and Universal Resource Locators with pointers, comprising:
>
> choosing an identifiable web page;
>
> identifying Universal Resource Locators for the web pages, wherein the identified Universal Resource Locators either point to or point away from the chosen web page;

analyzing Universal Resource Locators, including the identified Universal Resource Locators, wherein Universal Resource Locators which have an indirect relationship to the chosen web page are located, wherein the step of analyzing further comprises cluster analyzing the Universal Resource Locators for indirect relationships; and

displaying identities of web pages, wherein the located Universal Resource Locators are used to identify web pages, wherein the step of displaying is influenced by a number of times a web object is visited and wherein the cluster analysis uses a damping factor.

80. These features are also embodied in, for example, claim 1 of the '352 patent and claim 26 of the '571 patent:

1. A research system for computerized searching of textual objects, wherein textual objects are stored in a database, comprising:

a computer processor for processing commands and manipulating the textual objects stored in the database;

a means, coupled to the computer processor, for entering the commands to be processed by the computer processor;

a means for indexing the textual objects using the computer processor and the entered commands comprising:

a means for creating vectors representing the textual objects wherein the vectors are created using non-semantical relationships that exist among or between the textual objects;

a means for searching the indexed textual objects using the vectors to obtain a pool of textual objects comprising a means for vector searching of the indexed textual objects using the vectors;

a graphical user interface means for converting the pool of textual objects into a graphical view comprising:

a means for forming a box to graphically represent one or more of the textual objects in the pool; and

a display, operably coupled to the graphical user interface means, for showing the graphical view including any of the boxes formed.

(Ex A, '352 Patent, 31:4-28).

26. The method of claim 23, wherein **the step of displaying is influenced by a number of times a web object is visited**.

(Ex F, '571 Patent Reexamination Certificate, 3:47-48)

81. Indeed, the PTO has already specifically found that claim 26 and the step of displaying based on the number of times a web object is visited is novel and unobvious over the prior art. Based on the PTO's finding this claim was never before practiced by the art, it cannot be said that the claim is conventional or routine in view of the art.

82. These features are a specific interface design that allows users to quickly identify the most relevant results from a database with potentially billions of entries and millions of entries relevant to a particular search.

83. Accordingly, the above interface features used in combination with an analysis of indirect relationships to determine whether a relevant search result will be included in search results and the position on the screen that they will be returned represents an inventive concept that was not conventional, routine, or well understood at the filing of the patents. The above discussed features are improvements to the prior art display routines and data structures that did not incorporate the use of indirect relationships or the other weighing factors discussed above.

VII. DEFENDANT'S INFRINGING ACTS

84. Twitter is a social networking service and website located on the World Wide Web at the URL www.twitter.com (and other related URLs). In addition, Twitter also provides applications and other software for mobile and other electronic devices. Users must register before using the site, after which they may create a personal profile, follow other users, and exchange messages, including automatic notifications when they update their profile. Additionally, users may categorize the users they follow by adding them to lists such as "People From Work" or "Close Friends." Twitter provides many features for searching, serving, locating, recommending, analyzing, and displaying content and other information (including but not limited to profiles, advertisements, software, products, media, apps, status updates (*e.g.*, tweets), entities, places, stories, activity, etc.) that analyze or use indirect relationships.

85. In addition to the above features, Twitter's systems and functionality include the following:

(a) Processes for search on or by Twitter, including but not limited to: searching for Tweets, People, Entities and Locations, each of which Twitter may be designated previously using different feature names (*see*

https://dev.twitter.com/docs/platform-objects) and other objects. Twitter publicly refers to these processes on "How to Search on Twitter" (https://support.twitter.com/groups/31-twitter-basics/topics/110-search/articles/132700-how-to-search-on-twitter). Such functionality is available on the World Wide Web at the URL https://twitter.com/#!/search-home, through Twitter API and other facilities;

(b) Processes for searching for, recommending and otherwise locating and displaying content and other information within the Twitter community, including but not limited to: "Stories" (*see* https://twitter.com/i/discover); "Activity" (*see* https://twitter.com/#!/activity); and "Who to Follow" (*see* https://twitter.com/#!/who_to_follow/suggestions);

(c) Link analysis used and developed by Twitter to prioritize the display of content, including but not limited to, Tweets, People, Entities, and Locations; and

(d) Supporting infrastructure may include, but is not necessarily limited to, Hadoop, Pig, Hbase, Scalding, Cascading, Cassovary, Early Bird, and Blender.

(collectively, Section II is hereinafter "Infringing Methods and Systems").

86. The actual claims that will be asserted in this litigation will be governed according to the infringement contentions served in this case and the local rules governing amendment. The prior served infringement contentions serve as notice as to the nature of the infringement of the Defendant.

VIII. PATENT INFRINGEMENT

A. COUNT I – INFRINGEMENT OF THE '352 PATENT

87. Defendant has infringed and continues to infringe, without the permission of SRA, the '352 Patent because it makes uses, offers for sale, and sells Infringing Methods and Systems and related services covered by the claims of the '352 Patent.

88. Examples of infringement of these claims for purposes of giving notice to Defendant of the nature of its infringement is set forth in the infringement contentions previously served in this case and attached as Exhibit O.

89. Moreover, at least since the filing of SRA's Original Complaint (Dkt. No. 1) on July 27, 2012, Defendant, without the permission of SRA, has been and is presently indirectly infringing the '352 Patent through the provision of the Infringing Methods and Systems, including actively

inducing infringement of the '352 Patent under 35 U.S.C. § 271(b) and contributing to infringement of the '352 Patent under 35 U.S.C. § 271(c). To the extent it does not immediately cease its infringing activities, its infringement is and continues to be willful and deliberate. Such indirect infringements include, without limitation, with specific intent to encourage the infringement, knowingly inducing customers—including but not limited to users of www.twitter.com—to use, or knowingly contributing to customers' infringing uses of, without any substantial noninfringing use, Infringing Methods and Systems that Defendant knew infringed or demonstrated willful blindness with respect to infringement of one or more claims of the '352 Patent.

90. Twitter induces its users to give Twitter direction and control over what is displayed on the screens of their electronic devices, including, but not limited to, computers and mobile devices.

91. Twitter induces its users to display content identified by Twitter in accordance with instructions provided by Twitter on their electronic devices, including, but not limited to, computers and mobile devices.

92. On information and belief, Twitter continued to induce its users to allow Twitter to direct and control the displays on their electronic devices despite its knowledge that its usage would likely infringe claims of the '352 Patent.

93. Twitter provides its users with customized lists of URL links based on that user's preferences. Twitter provides this list with the intention and expectation that its users will select on the URL links and display the webpage referenced by said URL link. On information and belief, this is a core aspect of Twitter's business model and is essential to its operation.

94. On information and believe, Twitter sells access to its website and use of Twitter's software to its customers in return for the valuable consideration of providing Twitter with that customer's personal data and agreeing to the associated terms of service. Twitter's website and software are especially adapted to practice the claims of the '352 Patent and lack substantial non-infringing uses.

95. Defendant's acts of infringement have caused damage to SRA. SRA is entitled to recover from Defendant the damages sustained by SRA as a result of Defendant's wrongful acts in

an amount subject to proof at trial, but not less than a reasonable royalty. Plaintiff is seeking damages for past infringement of the '352 Patent and does not seek damages for infringement beyond the expiration date of the '352 Patent.

B. COUNT II – INFRINGEMENT OF THE '494 PATENT

96. Defendant has infringed and continues to infringe, without the permission of SRA, the '494 Patent because it makes uses, offers for sale, and sells Infringing Methods and Systems and related services covered by the claims of the '494 Patent.

97. Examples of infringement of these claims for purposes of giving notice to Defendant of the nature of its infringement is set forth in the infringement contentions previously served in this case and attached as Exhibit P.

98. Moreover, at least since the filing of SRA's Original Complaint (Dkt. No. 1) on July 27, 2012, Defendant, without the permission of SRA, has been and is presently indirectly infringing the '494 Patent through the provision of the Infringing Methods and Systems, including actively inducing infringement of the '494 Patent under 35 U.S.C. § 271(b) and contributing to infringement of the '494 Patent under 35 U.S.C. § 271(c). To the extent it does not immediately cease its infringing activities, its infringement is and continues to be willful and deliberate. Such indirect infringements include, without limitation, with specific intent to encourage the infringement, knowingly inducing customers—including but not limited to users of www.twitter.com—to use, or knowingly contributing to customers' infringing uses of, without any substantial noninfringing use, Infringing Methods and Systems that Defendant knew infringed or demonstrated willful blindness with respect to infringement of one or more claims of the '494 Patent.

99. Twitter induces its users to give Twitter direction and control over what is displayed on the screens of their electronic devices, including, but not limited to, computers and mobile devices.

100. Twitter induces its users to display content identified by Twitter in accordance with instructions provided by Twitter on their electronic devices, including, but not limited to, computers and mobile devices.

101. On information and belief, Twitter continued to induce its users to allow Twitter to direct and control the displays on their electronic devices despite its knowledge that its usage would likely infringe claims of the '494 Patent.

102. Twitter provides its users with customized lists of URL links based on that user's preferences. Twitter provides this list with the intention and expectation that its users will select on the URL links and display the webpage referenced by said URL link. On information and belief, this is a core aspect of Twitter's business model and is essential to its operation.

103. On information and believe, Twitter sells access to its website and use of Twitter's software to its customers in return for the valuable consideration of providing Twitter with that customer's personal data and agreeing to the associated terms of service. Twitter's website and software are especially adapted to practice the claims of the '494 Patent and lack substantial non-infringing uses.

104. Defendant's acts of infringement have caused damage to SRA. SRA is entitled to recover from Defendant the damages sustained by SRA as a result of Defendant's wrongful acts in an amount subject to proof at trial, but not less than a reasonable royalty. Plaintiff is seeking damages for past infringement of the '494 Patent and does not seek damages for infringement beyond the expiration date of the '494 Patent.

C. COUNT III – INFRINGEMENT OF THE '571 PATENT

105. Defendant has infringed and continues to infringe, without the permission of SRA, the '571 Patent because it makes uses, offers for sale, and sells Infringing Methods and Systems and related services covered by the claims of the '571 Patent.

106. Examples of infringement of these claims for purposes of giving notice to Defendant of the nature of its infringement is set forth in the infringement contentions previously served in this case and attached as Exhibit Q.

107. Moreover, at least since the filing of SRA's Original Complaint (Dkt. No. 1) on July 27, 2012, Defendant, without the permission of SRA, has been and is presently indirectly infringing the '571 Patent through the provision of the Infringing Methods and Systems, including actively inducing infringement of the '571 Patent under 35 U.S.C. § 271(b) and contributing to infringement

of the '571 Patent under 35 U.S.C. § 271(c). To the extent it does not immediately cease its infringing activities, its infringement is and continues to be willful and deliberate. Such indirect infringements include, without limitation, with specific intent to encourage the infringement, knowingly inducing customers—including but not limited to users of www.twitter.com—to use, or knowingly contributing to customers' infringing uses of, without any substantial noninfringing use, Infringing Methods and Systems that Defendant knew infringed or demonstrated willful blindness with respect to infringement of one or more claims of the '571 Patent.

108. Twitter induces its users to give Twitter direction and control over what is displayed on the screens of their electronic devices, including, but not limited to, computers and mobile devices.

109. Twitter induces its users to display content identified by Twitter in accordance with instructions provided by Twitter on their electronic devices, including, but not limited to, computers and mobile devices.

110. On information and belief, Twitter continued to induce its users to allow Twitter to direct and control the displays on their electronic devices despite its knowledge that its usage would likely infringe claims of the '571 Patent.

111. Twitter provides its users with customized lists of URL links based on that user's preferences. Twitter provides this list with the intention and expectation that its users will select on the URL links and display the webpage referenced by said URL link. On information and belief, this is a core aspect of Twitter's business model and is essential to its operation.

112. On information and believe, Twitter sells access to its website and use of Twitter's software to its customers in return for the valuable consideration of providing Twitter with that customer's personal data and agreeing to the associated terms of service. Twitter's website and software are especially adapted to practice the claims of the '571 Patent and lack substantial non-infringing uses.

113. Defendant's acts of infringement have caused damage to SRA. SRA is entitled to recover from Defendant the damages sustained by SRA as a result of Defendant's wrongful acts in an amount subject to proof at trial, but in no event less than a reasonable royalty. Plaintiff is

seeking damages for past infringement of the '571 Patent and does not seek damages for infringement beyond the expiration date of the '571 Patent.

IX. PRAYER FOR RELIEF

WHEREFORE, SRA prays for relief against Defendant as follows:

a. Judgment that Defendant has infringed, induced others to infringe, and/or committed acts of contributory infringement with respect to the claims of the '352, '494 and '571 Patents;

b. Awarding SRA damages adequate to compensate for the infringement by Defendant, but in no event less than a reasonable royalty for the use made of the inventions by Defendant, together with interests and costs under 35 U.S.C. § 284;

c. Awarding pre- and post-judgment interest on the damages assessed;

d. Declaring this case exceptional pursuant to 35 U.S.C. § 285, and awarding SRA its reasonable attorney fees;

e. SRA's costs of court; and

f. Awarding to SRA such other and further relief as the Court deems just.

X. JURY DEMAND

SRA demands a trial by jury.

Respectfully submitted

Dated: July 30, 2019

By: */s/ Victor Hardy*
Victor G. Hardy
Minghui Yang
HARDY PARRISH YANG LLP

Attorneys for Plaintiff
SOFTWARE RIGHTS ARCHIVE, LLC