

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

**DECLARATION OF BENJAMIN B. BEDERSON, PH.D.
IN SUPPORT OF PETITION FOR INTER PARTES REVIEW OF
U.S. PATENT 7,698,329**

Table of Contents

I.	Introduction and Background.....	1
II.	Qualifications	1
III.	Legal Standards	11
	A. Overview.....	11
	B. Claim Construction	11
	C. Person of Ordinary Skill in the Art.....	12
	D. Anticipation.....	13
	E. Obviousness	13
	F. Secondary Considerations of Nonobviousness.....	16
IV.	Level of Ordinary Skill in the Art.....	17
V.	Overview of the '329 Patent	18
	A. Overview of the Technology	18
	B. Overview of the Prosecution History.....	27
VI.	Interpretation of Claim Terms.....	30
VII.	Overview of the Prior Art	32
	A. Background and State of the Art for the '329 Patent.....	32
	B. Overview of Page.....	39
	C. Overview of FusionBot.....	44
	D. Overview of Emens.....	49
	E. Overview of Li.....	51
	F. Overview of Cucerzan	53
VIII.	Ground 1: Claims 1-3, 5, 8-10, and 12 are obvious in view of Page in combination with FusionBot.	56
	A. Claim 1	56
	1. [1p] A method, comprising;.....	56
	2. [1a] ranking a plurality of documents recalled by a search engine for a query;	56
	3. [1b] wherein the plurality of documents contain certain documents, each document of said certain documents	

	containing at least one section that is not used by said search engine for recall	57
4.	[1c] and one or more sections that are used by said search engine for recall;	63
5.	[1d] wherein ranking a plurality of documents includes ranking said plurality of documents based, at least in part, on the at least one section of said certain documents not used by said search engine to recall documents; and	64
6.	[1e] wherein the method is performed by one or more computing devices.	70
B.	Claim 2: The method of claim 1, wherein said at least one section of said certain documents not used by said search engine to recall documents are demarcated by a tag of an element having a name from a set of one or more names.	71
C.	Claim 3: The method of claim 1, wherein at least one document of said certain documents does not comprise data that contains element tags.	72
D.	Claim 5.....	73
1.	[5a] The method of claim 1, said search engine using a search engine index to recall said plurality of documents; and	73
2.	[5b] wherein said search engine index does not index any term in said certain documents that is found only in one or more sections of said certain documents established as a no-recall section.	73
E.	Claim 8.....	74
F.	Claim 9.....	75
G.	Claim 10.....	75
H.	Claim 12.....	76
IX.	Ground 2: Claims 4 and 11 are obvious in view of Page in combination with FusionBot and Emens.	76
A.	Claim 4.....	78
1.	[4a] The method of claim 1, further including said search engine generating data representing results of said query, wherein: said data includes an abstract describing each document of said plurality of documents; and.....	78

2.	[4b] for each respective abstract of each document of said certain documents, said abstract excludes terms from the respective at least one section not used by said search engine to recall said each document.	78
B.	Claim 11	80
X.	Ground 3: Claims 6 and 13 are obvious in view of Page in combination with Fusionbot and Li.	80
A.	Claim 6: The method of claim 1, wherein ranking a plurality of documents recalled by a search engine for a query depends in part on a vector space model associated with each of said documents.	82
B.	Claim 13	84
XI.	Ground 4: Claims 7 and 14 are obvious in view of Page in combination with FusionBot and Cucerzan.	84
A.	Claim 7	86
B.	Claim 14	87
XII.	Conclusion.....	87

I. INTRODUCTION AND BACKGROUND

I, Benjamin B. Bederson, hereby declare as follows:

1. I submit this declaration in support of American Airlines' and Hilton's ("Petitioners'") Petition for Inter Partes Review of U.S. Patent 7,698,329 ("the '329 patent").

2. I understand that my curriculum vitae has been submitted as Exhibit 1028.

3. In preparing this declaration, in addition to my knowledge and experience, I have reviewed and am familiar with the following references:

- The '329 patent (Ex. 1001)
- File History of the '329 patent (Ex. 1002)
- U.S. Patent 6,285,999 to Page ("Page," Ex. 1004)
- FusionBot Webpages ("FusionBot," Ex. 1005)
- U.S. Patent 6,295,559 to Emens ("Emens," Ex. 1006)
- International Patent Publication WO97/49048 to Li ("Li," Ex. 1007)
- U.S. Patent Application Publication 2007/0214131 to Cucerzan ("Cucerzan," Ex. 1008).

I have also considered all other materials cited herein.

II. QUALIFICATIONS

4. Since 1998, I have been a Professor of Computer Science at the University of Maryland ("UMD"), where I have joint appointments at the Institute for Advanced Computer Studies and the College of Information Studies

(Maryland's "iSchool") and am currently Professor Emeritus. I was also Associate Provost of Learning Initiatives and Executive Director of UMD's Teaching and Learning Transformation Center from 2014 to 2018. I am a member and previous director of the Human-Computer Interaction Lab ("HCIL"), the oldest Human-Computer Interaction research group in the country, and one of the best-known. From 2006 to 2014, I was also co-founder and Chief Scientist of Zumobi, Inc., a Seattle-based startup that is a publisher of content applications and advertising platforms for smartphones. I am a co-founder and co-director of the International Children's Digital Library ("ICDL"), a website launched in 2002 that provides the world's largest collection of freely available online children's books from around the world and has an interface aimed to make it easy for children and adults to search and read children's books online. I am also co-founder and prior Chief Technology Officer of Hazel Analytics, a data analytics company whose product sends alerts in warranted circumstances. In addition, for more than 25 years I have consulted in the area of user interfaces for numerous companies, including Logitech, Microsoft, the Palo Alto Research Center, Sony, Lockheed Martin, and the NASA Goddard Space Flight Center.

5. I received a Bachelor of Science degree in Computer Science with a minor in Electrical Engineering from Rensselaer Polytechnic Institute ("RPI") in

1986. I received a Master of Science degree and a Ph.D. in Computer Science from New York University ("NYU") in 1989 and 1992, respectively.

6. For more than 30 years, I have studied, designed, and worked in the field of computer science and human-computer interaction. My experience includes 30 years of teaching and research, with research interests in human-computer interaction and the software and technology underlying today's interactive computing systems. This includes the design and implementation of user interfaces and software applications, including web browsers and search engines as described below.

7. At UMD, my research is in the area of Human-Computer Interaction ("HCI"), a field that relates to the development and understanding of computing systems to serve users' needs. Researchers in this field are focused on making universally usable, useful, efficient, and appealing systems to support people in a wide range of activities. My approach is to balance the development of innovative technology that serves people's practical needs. Following this approach, I have built the following systems: Pad++ (1994 software for zoomable user interfaces), KidPad (1997 software to help children collaboratively write stories), PadPrints (1998 software to graphically show interactive web browsing histories), PhotoMesa (2001 software to help end users browse personal photos), Pocket PhotoMesa (2004 software for browsing personal photos on mobile devices),

DateLens (2002 software to help end users use their mobile devices to efficiently access calendar information), LaunchTile (2005 "home screen" software for mobile devices to allow users to navigate apps in a zoomable environment), SpaceTree (2001 software to help end users efficiently browse very large hierarchies), ICDL (2002 International Children's Digital Library), and Story Kit (2009 iPhone application to help children create stories).

8. Pad++, which I wrote in 1994, included a number of techniques to support high performance interactive visualization.¹ Pad++ included a custom web browser that I built from scratch and distributed as open source software at the Pad++ website. The web browser was described in numerous publications, including "Pad++: A Zoomable Graphical Interface for Exploring Alternate Interface Physics."² For the web browser, I implemented from scratch a rendering engine that supported the version of HTML, available at the time (HTML, 1.0) as well as an embedded scripting language. The web browser supported displaying embedded webpages that could be zoomed along with the rest of the pages. The following images are from that custom web browser that I built, which used a

¹ Bederson, B.B. & Meyer, J. (1998). Implementing a Zooming User Interface: Experience Building Pad++. *Software: Practice and Experience*, 28(10), 1101-1135.

² Bederson, B. B. & Hollan, J. D. (1994). "Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics," *ACM Conference on User Interface Software and Technology*, 17-26.

"zoomable user interface" so one page could be shown larger while other recently visited pages would be displayed smaller:

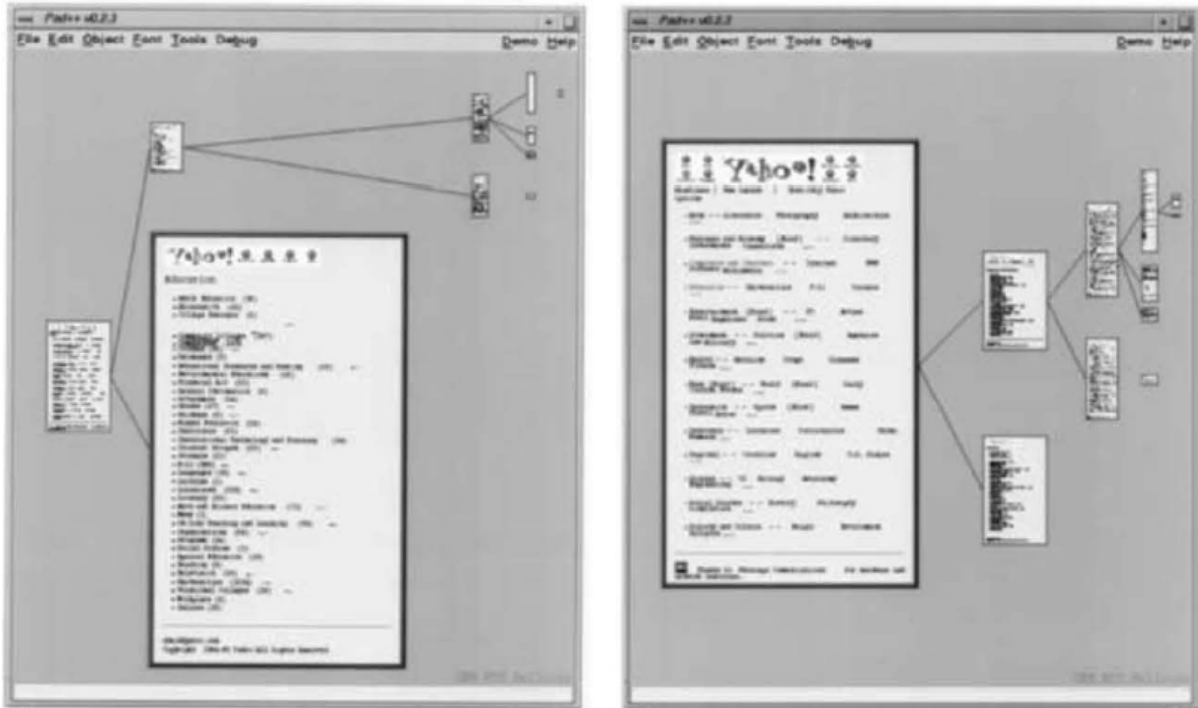


Figure 5 from "Pad++: A Zoomable Graphical Sketchpad for Exploring Alternate Interface Physics"³

9. By 1998, my team had combined Pad++ with Netscape Navigator (a standard web browser of the time) to offer a graphical web history, visually showing what pages a user had visited with an application we called "PadPrints."⁴

³ Bederson, B. B., Hollan, J. D., Perlin, K., Meyer, J., Bacon, D. & Furnas, G. (1996). Pad++: A Zoomable Graphical Sketchpad for Exploring Alternate Interface Physics. Journal of Visual Languages & Computing, 7(1), 3-32.

⁴ Ron R. Hightower, Laura T. Ring, Jonathan I. Helfman, Benjamin B. Bederson & James D. Hollan. 1998. PadPrints: Graphical Multiscale Web Histories. In Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology (UIST '98). Association for Computing Machinery, New York, NY, USA, 121-122. <https://doi.org/10.1145/288392.288582>

We did this by implementing a web "proxy," which was a standard way of modifying the built-in behavior of web browsers at the time. Some proxies work by intercepting all web traffic coming from and going to a web browser. Once intercepted, the proxy can do whatever it wants with the traffic, such as caching it for better performance, simply allowing the traffic through and monitoring it, modifying the data itself, or holding that connection and opening new connections. PadPrints' proxy worked by monitoring web traffic and letting PadPrints know what the user was doing so PadPrints could display the web history.

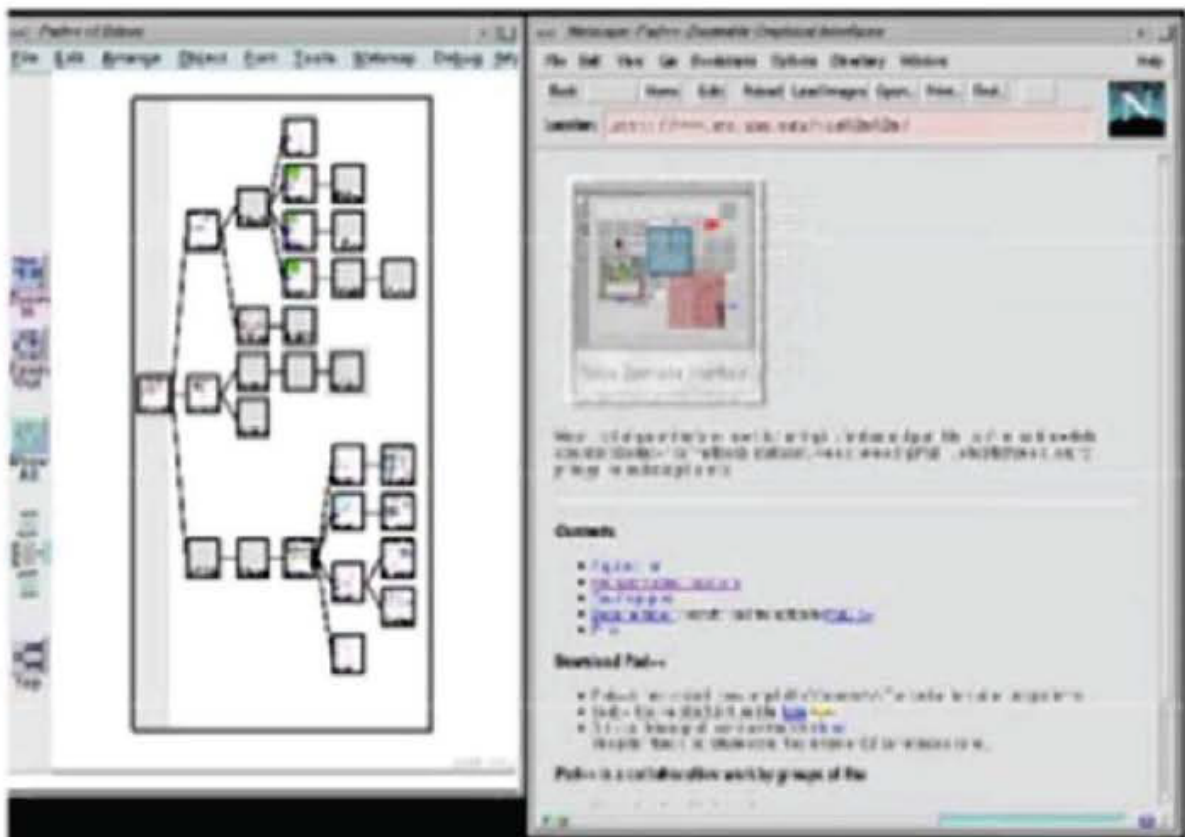


Figure 1 from PadPrints: Graphical Multiscale Web Histories

10. I have since 2002 been the technical director responsible for the design and implementation of the International Children's Digital Library⁵ that can be found at www.childrenslibrary.org (originally at www.icdlbooks.org). In particular, I have been closely involved in designing the user interface and software architecture for the website since its inception in 2002. The ICDL is the world's largest collection of freely available children's books from around the world. From the very beginning, it included a search engine that supported searching for books in the library based on book metadata such as title, genre, and color of the corresponding book cover. I was part of the team that designed this search engine functionality. The search engine could recall the full content of each book, but did not consider the full content of the book in ranking search results.



Screen Capture Taken February 19, 2017 by Benjamin B. Bederson from <http://www.childrenslibrary.org/icdl/SimpleSearchCategory?ids=25&langid=&pnum=1&cnum=1&text=&lang=English&ilang=English>.

11. Since the mid-1 990s, I have been responsible for the design and implementation of numerous other websites in addition to the ICDL. For example, I designed and built my own professional website when I was an Assistant Professor of Computer Science at the University of New Mexico in 1995, and I have continued to design, write code for, and update both that site (which I moved to the University of Maryland in 1998, currently at <http://www.cs.umd.edu/~bederson/>) and numerous project websites, such as Pad++

(located at <http://www.cs.umd.edu/hcil/pad++/>). I received the Janet Fabri Memorial Award for Outstanding Doctoral Dissertation for my Ph.D. work in robotics and computer vision. I have combined my hardware and software skills throughout my career in human-computer interaction ("HCI") research, building various interactive electrical and mechanical systems that couple with software to provide an innovative user experience.

12. My work has been published extensively in more than 160 technical publications, and I have given about 100 invited talks, including 9 keynote lectures. I have won a number of awards, including the Brian Shackel Award for Outstanding Contribution with International Impact in the Field of HCI in 2007 and the Social Impact Award from Association for Computing Machinery's ("ACM") Special Interest Group on Computer Human Interaction ("SIGCHI") in 2010. ACM is the primary international professional community of computer scientists, and SIGCHI is the primary international professional HCI community. I have been honored by both professional organizations. I am an "ACM Distinguished Scientist," which "recognizes those ACM members with at least 15 years of professional experience and 5 years of continuous Professional Membership who have achieved significant accomplishments or have made a significant impact on the computing field." I am a member of the "CHI Academy," which is described as follows: "The CHI Academy is an honorary group of

individuals who have made substantial contributions to the field of human-computer interaction. These are the principal leaders of the field, whose efforts have shaped the disciplines and/or industry, and led the research and/or innovation in human-computer interaction." The criteria for election to the CHI Academy are: (1) cumulative contributions to the field; (2) impact on the field through development of new research directions and/or innovations; and (3) influence on the work of others.

13. I have appeared on radio shows numerous times to discuss issues relating to user interface design and people's use and frustration with common technologies, websites, and mobile devices. My work has been discussed, and I have been quoted, by mainstream media around the world over 120 times, including by *The New York Times*, *The Wall Street Journal*, *The Washington Post*, *Newsweek*, the *Seattle Post-Intelligencer*, *The Independent*, *Le Monde*, NPR's *All Things Considered*, *New Scientist Magazine*, and MIT's *Technology Review*.

14. I have designed, programmed, and publicly deployed dozens of user-facing software products that have cumulatively had millions of users. I am the co-inventor of 14 U.S. patents and 20 U.S. patent applications, which are generally directed to user interfaces and experience.

III. LEGAL STANDARDS

A. Overview

15. I am not an attorney or a legal expert, and I do not opine on matters of law. All the legal standards listed in this section, and which I applied in reaching the conclusions expressed throughout this declaration, were provided to me by counsel. In preparing this declaration, I have endeavored to faithfully apply these legal standards. I do not intend to offer opinions on the law. Below, I describe the law that I applied in analyzing the claims of the '386 patent.

16. I also understand that an unpatentability analysis is conducted in two steps. First, the claims of the patent are construed to determine their meaning and scope. Second, after the claims are construed, the content of the prior art is compared to the construed claims.

17. I understand that a claimed invention is only patentable when it is new, useful, and non-obvious in light of the prior art. That is, the invention, as defined by the claims of the patent, must not be anticipated or rendered obvious by the prior art.

B. Claim Construction

18. I have been informed by counsel that the first step in determining validity is to properly construe the claims. I have further been informed by counsel that, in this proceeding, the Board must determine the scope of the claims by using the same claim construction standard that United States District Courts use. Under

this standard, words of a claim are given their “ordinary” and “customary” meaning as would have been understood by a person of skill in the art (“POSITA”) at the time of the invention. I have been informed that such a person looks to the claims, specification, and file history in understanding the meaning of the claim terms.

19. I understand that in addition to the claims, specification, and prosecution history, other evidence may be considered to ascertain the meaning of claim terms, including textbooks, encyclopedias, articles, and dictionaries. I have been informed that this other evidence is often less significant and less reliable than the claims, specification, and prosecution history.

C. Person of Ordinary Skill in the Art

20. I have been informed that the patentability of a claimed invention must be analyzed from the perspective of "one of ordinary skill in the art" in the same field as the patent-in-suit at the time of the invention. I have also been informed that several factors are considered in assessing the level of ordinary skill in the art, including (1) the types of problems encountered in the art; (2) the prior art solutions to those problems; (3) the rapidity with which innovations are made; (4) the sophistication of the technology; and (5) the education level of active workers in the field.

D. Anticipation

21. I understand that a patent claim is invalid as anticipated if a single piece of prior art teaches every element of the claims, viewed from the perspective of a person of ordinary skill in the art. I also understand that an anticipatory reference does not have to recite word for word what is in the anticipated claims. Anticipation can also occur when a claimed limitation is inherent in the relevant reference. I have been advised that if the prior art necessarily functions in accordance with, or includes, a claimed limitation, it can anticipate even though the limitation is not expressly disclosed.

E. Obviousness

22. I understand that a patent claim is invalid as obvious if the claimed invention would have been obvious to a POSITA at the time the claimed invention was made. In this context, "obvious" means that even if all of the elements of a claimed invention cannot be found in a single prior art reference that would anticipate the invention, a POSITA who knows about all the prior art would have come up with the invention. I understand that in an obviousness determination, a POSITA is presumed to have knowledge of all material prior art. I understand that whether a claim is obvious is based upon a determination of several factual issues.

23. In considering obviousness, I understand that one must determine the scope and content of the prior art. I understand that, in order to be considered

"prior art" to a patent under consideration, a reference must be reasonably related to the claimed invention of that patent. A reference is reasonably related if it originates from the same field as the claimed invention or from another field to which a POSITA would look to solve a known problem.

24. I understand that an assessment of obviousness requires a determination as to what differences, if any, existed between the claimed invention and the prior art.

25. I understand that a patent claim composed of several elements is not proved obvious merely by demonstrating that each of its elements was independently known in the prior art. In evaluating whether a claim would have been obvious, one may consider whether one or more reasons existed that would have prompted a POSITA to combine the elements or concepts from the prior art in the same way as in the claimed invention. There is no single way to define the line between true inventiveness on the one hand (which is patentable) and the application of common sense and ordinary skill to solve a problem on the other hand (which is not patentable). For example, market forces or other design incentives, rather than true inventiveness, may inspire a given change.

26. I understand that one may consider whether (1) the change was merely the predictable result of using prior art elements according to their known functions or the result of true inventiveness; (2) there was some teaching or

suggestion in the prior art to make the modification or combination of elements claimed in the patent; (3) the innovation applied a known technique that had been used to improve a similar device or method in a similar way; (4) the claimed invention would have been obvious to try, meaning that the claimed innovation was one of a relatively small number of possible approaches to the problem that a POSITA could have implemented with a reasonable expectation of success; (5) the invention merely substituted one known element for another known element to obtain predictable results; (6) the invention merely applied a known technique to a known device, method, or product to yield predictable results; or (7) known work in the field prompted variations in use of the same inventions in the same or different fields due to market forces or design incentives that would have been predictable to a POSITA.

27. I understand that one must be careful not to determine obviousness using the benefit of hindsight, as many true inventions might seem obvious after the fact. I understand that the determination should be based on the position of a POSITA at the time the claimed invention was made and should not account for what is known today or what is learned from the teachings of the patent under consideration.

F. Secondary Considerations of Nonobviousness

28. I understand that certain secondary considerations may be considered in evaluating obviousness in order to prevent hindsight bias. These secondary considerations include commercial success of products that practice the patent, any long-felt need for the patented technology, failure by others to solve the problem addressed by the patent, initial skepticism by others in the industry, industry recognition and praise of the patented products, and efforts by others to copy the patented technology.

29. I have been informed that these secondary considerations of nonobviousness must be balanced against the strength of the prima facie case of obviousness. Where the invention represents no more than the predictable use of prior art elements according to their established functions, secondary considerations are inadequate to establish nonobviousness.

30. I have also been informed that the patentee must establish a nexus between any secondary consideration factors and the claimed invention. Where a secondary consideration factor results from something other than what is both claimed and novel in the patent, there is no nexus to the merits of the claimed invention.

31. I have been informed that, for example, the need for a nexus as it applies to the evidence-of-commercial-success factor requires that the patentee

show that the sales of a product that allegedly practices the claimed invention are or were a direct result of the unique characteristics of the claimed invention, as opposed to other economic and commercial factors unrelated to the patented invention. I understand that this means that if the product's commercial success is due to an element that existed in the prior art or that is unclaimed by the patent, then no nexus exists.

32. I have been further informed that, for example, the nexus requirement as it applies to the long-felt-need factor must likewise be supported by evidence. Where differences between the prior art and the claimed invention are minimal, it cannot be said that any long-felt need was satisfied by the patented invention or that the patent solved any technological problems left unaddressed by others.

33. I have been informed that evidence of industry praise must also show a nexus with the claimed invention, and that self-serving statements by the patentee regarding the merits of the claimed invention are generally given no weight.

IV. LEVEL OF ORDINARY SKILL IN THE ART

34. In my opinion, a POSITA pertinent to the '329 patent as of its earliest priority date would have had at least a four-year degree in computer science, computer engineering, or a closely related field and at least one year of experience in researching, designing, or developing computerized search systems such as search engines. Additional formal education could substitute for professional

experience, and significant professional experience could substitute for formal education. I meet this definition of a POSITA now, and I did so as of the priority date of the '329 patent.

35. I am also familiar with the knowledge of a POSITA as of the priority date of the '329 patent. I can opine on how the person of ordinary skill in the art would have understood the disclosure and claims of the '329 patent, the disclosures of the prior art, motivations to combine the prior art,⁵ and the combinations that would have been obvious to a POSITA.

V. OVERVIEW OF THE '329 PATENT

A. Overview of the Technology

36. The '329 patent describes "[a] method and apparatus for improving search results" by "delineating sections of a document that are not relevant to the main content" of the document, but performing a ranking analysis based on the entire content of the document. '329 patent, Abstract. When the document is included in results recalled by a search engine in response to a query, terms in the document's so-called "no-recall sections" are omitted from the results. *Id.* However, "[t]he results are ordered at least in part by the rankings attributed to the identified no-recall sections." *Id.* The '329 patent explains that its embodiments are

⁵ As used herein, the phrases "found it obvious" and "motivated to combine" are used as shorthand to refer to a POSITA's motivation to combine with a reasonable expectation of success.

implemented with generic hardware and software, *id.*, 4:62-6:52, and it concedes that the embodiments "are not limited to any specific combination of circuitry and software," *id.*, 5:38-40.

37. The '329 patent explains that search engines comprise "combination[s] of ... software and computational resources [that] are dedicated to searching a set of information resources," typically in response to "queries submitted to the search engine." *Id.*, 1:18-25. These search engines typically perform their work in three major steps: First, because "[s]earch engines use a search engine index ... to search for documents," an index for the search engine must be built. *Id.*, 1:45-51. While this can be done manually by humans, typically, this function is performed by "crawlers," which "explore information over the Internet" by following links between documents, index the contents of the documents they find in search engine indexes, and "summarize [the documents] in databases." *Id.*, 1:47-57. Second, when a query is entered into the search engine, the search engine examines the indexes and databases it built for responsive documents. *Id.*, 1:57-59. Finally, after identifying responsive documents, the search engine generates a "list of documents," which "may contain summary information about the document[s]." *Id.*, 1:60-62. As the '329 patent notes, search engines such as Yahoo™ and Google™ were "widely used on the Internet, the World Wide Web ... and other

large internetworks and information resource webs" well before the priority date of the '329 patent. *Id.*, 1:25-32.

38. The '329 patent also notes that a list of documents generated by a search engine may be ordered by rank after the search engine "assign[s] a rank to each document in the list." *Id.*, 1:62-67. A rank may be assigned to a document based on its "relevance to a search query," which the '329 patent explains "is a measure of how closely the subject matter of [the] document matches a search query's terms." *Id.*, 1:67-2:3. Ranking a document is therefore different from recalling a document, which simply refers to "[t]he inclusion of a document within the search engine results generated by a search engine for a search engine query." *Id.*, 2:3-5.

39. The '329 patent explains that search engines are susceptible to certain "nefarious techniques," called search engine spamming, that "trick search engines into recalling documents and inflating their rank." *Id.*, 2:6-8. The patent notes that the problems posed by search engine spamming techniques can be exacerbated when webmasters—i.e., the creators of certain web-based documents—seek to avoid indexing certain content in their documents, as the content that is not indexed can be used as a vehicle for "hiding search engine spam." *Id.*, 2:28-34. To counter these problems, the '329 patent presents mechanisms for "affect[ing] the recall and ranking of documents to more accurately reflect [their] relevance ... to

search engine queries." *Id.*, 3:1-3. These mechanisms take advantage of the structure of each document to identify sections that should not be indexed by the search engine-and therefore should not be used to recall the document in response to a query-but should be used to rank the document when it has been so recalled. *See id.*, 3:7-31.

40. The '329 patent explains that a "document" is a "unit of information that may be indexed by search engine indexes" and may include "plain or formatted text, inline graphics, and other multimedia data, and hyperlinks to other documents." *Id.*, 1:33-38. Documents may conform to certain mark-up standards like Hypertext Markup Language ("HTML") or Extensible Mark-Up Language ("XML") or fall within any number of file formats, such as Microsoft Word™ or Adobe™ Portable Document Format ("PDF") formats. *Id.*, 1:38-43. The '329 patent uses webpage 100 as an exemplary document. *See id.*, Fig. 1, 3:32-39.

41. The structure of webpage 100 can be divided into sections, which include the following: a title section 101, a navigation pane 102, a main content section 103, a copyright notice section 104, a section populated by content that regularly changes (i.e., "related blogs" section) 105, and an advertising section 106. *Id.*, 3:33-37. As discussed *infra* in Section VII.A, dividing webpages into similar sections was a common practice of both webmasters (to improve their websites' informativeness and functionality and optimize webpage development) and search

engine developers (to maximize their engines' storage capacity and analytical efficiency). This structure as it might present itself to a user viewing the webpage through a browser is reflected in the following figure:

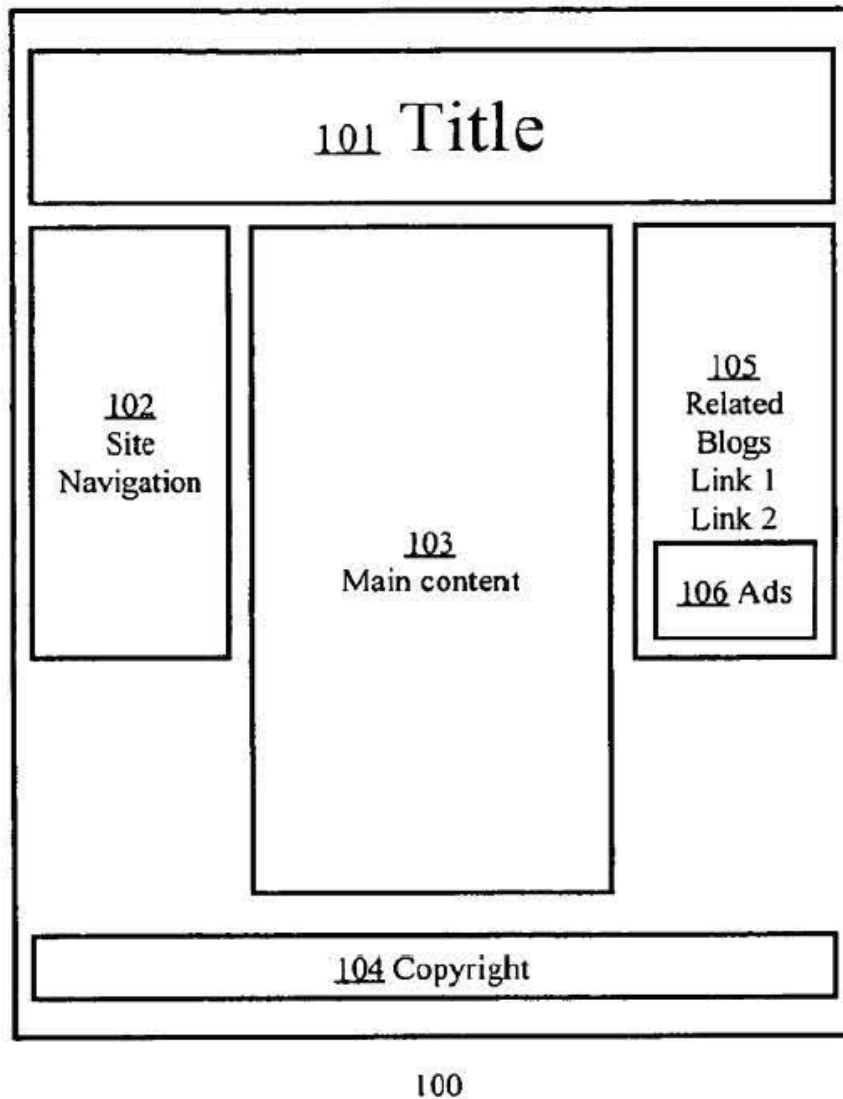


Figure 1.

Id., Fig. 1. The '329 patent explains that the navigation pane 102, the copyright notice section 104, the related blogs section 105, and the advertising section 106

are unrelated to the main content of the webpage. *Id.*, 3:37-39. This structure may be implemented by the following exemplary HTML code, which incorporates tags and nested tables that identify (1) the type of HTML code used to generate the document; (2) where the document's content is located; and (3) how the various sections of the webpage are positioned:

```

200 <HTML>
201 <HEAD>
202 ...
203 </HEAD>
204 <BODY>
205     <TABLE>
206         <TR>
207             <TD> Title </TD>
208         </TR>
209         <TR>
210             <TD>
211                 <TABLE>
212                     <TR>
213                         <div class="robots-noindex">
214                             <TD> Navigation </TD>
215                         </div class="robots-noindex">
216                     </TR>
217                     <TR>
218                         <TD> Main Content </TD>
219                     </TR>
220                     <TR>
221                         <div class="robots-noindex">
222                             <TD> Related Blogs, Ads</TD>
223                         </div class="robots-noindex">
224                     </TR>
225                 </TABLE>
226             </TD>
227         </TR>
228         <TR>
229             <div class="robots-noindex">
230                 <TD> Copyright </TD>
231             </div class="robots-noindex">
232         </TR>
233     </TABLE>
234 </BODY>
235 </HTML>

```

Figure 2.

Id., Fig. 2, 3:40-50.

42. The '329 patent describes embodiments in which a search engine crawler, upon encountering the document reflected in Figures 1 and 2 above, "identifies certain sections to ignore, that is, to not index in search engine indexes and[/]or otherwise use for recalling the document." *Id.*, 3:7-11. The '329 patent calls these ignored sections "no-recall sections." *Id.*, 3:11-12. The remaining sections of the document, which are indexed by the search engine, are called "recall sections." *Id.*, 3:12-14. The '329 patent provides two embodiments for identifying no-recall sections.

43. First, the patent describes "demarcat[ing]" or "delimit[ing]" no-recall sections in the document's HTML code by "a tag," such as a `<div class="robotsnoindex">` tag. *Id.*, 3:14-15, 3:51-52. For example, with reference to the exemplary embodiment in Figure 2, reproduced above, the '329 patent explains that the tag has been used to delimit the navigation pane 102, the copyright notice section 104, the related blogs section 105, and the ad section 106. *Id.*, 3:14-15, 3:51-54. The `<div class="robots-noindex">` tag causes any content it encloses "to be ignored for purposes of recalling the document." *Id.*, 3:55-57. As a result, even if a query term can be found within a section demarcated by a `<div class="robots-noindex">` tag, the document will not be recalled based on that particular instance of the query term. *Id.*, 3:55-62. However, while "a `<div class="robots-noindex">`

element ... is not indexed by a search engine index," the tag does not preclude the section from being "considered for spam and relevancy analysis" once the document has been recalled by a search engine for a query. *Id.*, 4:7-16. Instead, "[a]ll the attributes in all of the sections of [the] document such as 'links', frequency of terms, coloring, font, etc. are ... considered for spam and relevancy analysis." *Id.*, 4:13-16.

44. In the second disclosed embodiment, "a no-recall section may be identified by analyzing section content" and assigning "[s]ection rating scores" indicative of the relevance of the section to "the main topic of the page." *Id.*, 3:15-17; 4:17-53. These scores may be assigned based on "various attributes" of each section, "such as the number, frequency and order of appearance of terms, fonts, and colors" or how frequently the content of a section changes as compared to the content of the rest of the document, as well as the characteristics of outgoing links in each section, such as "link text and link quantity and quality." *Id.*, 4:25-29, 4:43-50. Identification of no-recall sections in this manner may incorporate the use of a document model, such as a vector space model, a standard or extended Boolean model, or a natural language model. *Id.*, 4:31-43. An exemplary process by which this might be achieved is outlined in Figure 3. *See id.*, Fig. 3; 4:17-53.

45. As previously mentioned, the '329 patent discloses that the identified no-recall sections are not indexed by the search engine, so the terms they contain

"do not contribute to . . . document term frequency counts and are not used for recalling the documents in response to search engine queries." *Id.*, 3:17-20. Despite this, the content of no-recall sections-including the quality and quantity of links or advertisements they contain and/or other features-can impact document ranking. *Id.*, 3:20-31.

46. After recalling and ranking the documents in response to a query, the search engine can return a list of documents that has been ordered by document rank. *See id.*, 1:60-67. This list may "contain short descriptive paragraphs or abstracts" which provide concise descriptions of the documents-for the recalled documents. *Id.*, 4:54-57. However, according to the '329 patent, should a recalled document contain one or more no-recall sections, "the contents of [those] no-recall sections are excluded from the [document's] abstract." *Id.*, 4:57-60.

47. The '329 patent includes two independent claims (claims 1 and 8). Claim 1, reproduced below, is representative and is, on its face, broader than the disclosed embodiments. For example, claim 1 does not expressly recite using an "index" or tags to demarcate sections of documents. Instead, claim 1 recites the following:

1. A method, comprising:

ranking a plurality of documents recalled by a search engine for a query;

wherein the plurality of documents contain certain documents, each document of said certain documents containing at least one section that is not used by said search engine for recall and one or more sections that are used by said search engine for recall;

wherein ranking a plurality of documents includes ranking said plurality of documents based, at least in part, on the at least one section of said certain documents not used by said search engine to recall documents;

and

wherein the method is performed by one or more computing devices.

B. Overview of the Prosecution History

48. I have reviewed the prosecution history of the application that led to issuance of the '329 patent. Ex-1002. During prosecution of the '329 patent, the Examiner twice rejected the claims under 35 U.S.C. § 103 as obvious over U.S. Patent Publication No. 2003/0120654 to Edlund et al. ("Edlund") in view of Matt Cutts, "Handling noindex meta tags" (Aug. 30, 2006) ("Cutts") and, for two dependent claims, further in view of U.S. Patent Publication No. 2005/0091580 to Kamholz et al. ("Kamholz"). *See id.* at 35-43, 85-93.

49. In response to the first of these rejections, Applicant amended certain claims (then claims 1, 2, and 4), canceled other claims, and added new claims. *Id.*, 72-76. The newly added claims were nearly identical to the existing claims, except

that they recited a "computer-readable storage medium" instead of a "method." *Id.* Applicant argued that the following element of the independent claims overcame the Examiner's rejections:

wherein ranking a plurality of documents includes ranking said plurality of documents based, at least in part, on the one or more sections of said certain documents not used by said search engine to recall documents

See id. at 77-81.

50. When the Examiner once again rejected the claims on the same grounds, Applicant and the Examiner conducted an Examiner Interview, *see id.* at 114-22, after which Applicant amended the claims "exclusively [to] improve readability and clarity of the claims and not for the purpose of overcoming alleged prior art," *id.* at 102-10. As for its arguments against the Examiner's rejections under § 103, Applicant initially noted that during the Examiner Interview, the Examiner "indicat[ed] that the use of data within noindex metatags for recall or ranking is contrary to industry standards and is generally frowned upon." *Id.* at 107. It then argued that Edlund and Cutts "fail[ed] to suggest or describe[] 'ranking a plurality of documents includes ranking said plurality of documents based, at least in part, on the one or more sections of said certain documents **not used by said search engine** to recall documents,'" noting particularly that "Cutts [made] no distinction between the use of noindex data for recall and ranking." *Id.* at 108 (emphasis in original). Applicant pointed out that, in contrast, "Claim 1 teaches

ignoring the data contained within a noindex metatag for the purposes of recall but using the data contained within a noindex metatag for the purposes of ranking." *Id.* To reinforce its position, Applicant included the following table, which it asserted "illustrate[d] the differences in the treatment of the *data within a noindex tag* for the search engines tested by Cutts and the treatment of the *noindex sections* recited by Claim 1":

Search Engine	Use noindex data for recall?	Use noindex data for ranking?	Difference Comparison
Google	No	No	No/No
Ask	No	No	No/No
MSN	Yes	Yes	Yes/Yes
Yahoo!	Yes	Yes	Yes/Yes
Claim 1	No	Yes	No/Yes

Id. at 109 (emphases added). Based on this table, Applicant concluded that "none of the search engines tested by Cutts *ignore (do not use) the noindex data for recall and use the noindex data for ranking.*" *Id.* (emphasis added). For these reasons, it asserted that each of the claims was nonobvious and allowable over Edlund in view of Cutts and Kamholz. *Id.* at 109-10.

51. The Examiner subsequently issued a Notice of Allowance. *See id.* at 123-29. In his Statement of Reasons for Allowance, the Examiner stated that the claims of the '329 patent were allowable *specifically because* they claimed "reading the ignored (or the 'noindex') section and using that section for purposes of ranking the web page." *Id.* at 127.

[T]he Claims do not index the section *containing the command to ignore the specified section*. It only uses that section for ranking purposes. This is *demonstrated in Applicant Remarks in at least the table on Page 8*. The reason for allowance *lies specifically in the concept of using an ignored section only for purposes of ranking and not indexing*.

Id. (emphasis added).

52. Applicant's chart above implicitly concedes that there are only *four* discrete options for resolving the problem of how search engines should account for content within noindex sections in their recall and ranking processes. Each of these choices incorporates binary decisions as to whether the data should be used for recall and/or ranking. These options are reflected in the table below.

	Noindex Data Used for Recall	Noindex Data Used for Ranking
Option 1	No	No
Option 2	No	Yes
Option 3	Yes	No
Option 4	Yes	Yes

VI. INTERPRETATION OF CLAIM TERMS

53. As discussed above, I have been informed that for the purposes of inter partes review, the standard for claim construction of terms within the claims

of the patent is the same as that applied in federal district court litigation. I understand that a district court has previously construed terms of this patent, and I have been informed of those constructions. Specifically, I understand the district court to have issued the following constructions:

Claim Term	Claim(s)	District Court's Construction
"document"/"documents"	Claims 1-14	"any unit of information that may be indexed by search engine indexes"
"each document of said certain documents containing at least one section that is not used by said search engine for recall and one or more sections that are used by said search engine for recall"	Claims 1 and 8	Plain and ordinary meaning except that "recall" means "generating results for a search engine query" and "section" means "defined portion within the structure of a document"
"wherein ranking a plurality of documents includes ranking said plurality of documents based, at least in part, on the at least one section of said certain documents not used by said search engine to recall documents"	Claims 1 and 8	Plain and ordinary meaning except that "recall" means "generating results for a search engine query" and "section" means "defined portion within the structure of a document"
"said data includes an abstract describing each document of said plurality of documents"	Claims 4 and 11	Plain and ordinary meaning
"for each respective abstract of each	Claims 4 and 11	Plain and ordinary meaning except that "recall" means

document of said certain documents, said abstract excludes terms from the respective at least one section not used by said search engine to recall said each document"		"generating results for a search engine query" and "section" means "defined portion within the structure of a document"
--	--	---

Ex.-1003, 42-53.

54. I have applied these constructions in my analysis. For terms that the district court did not previously construe, I have generally applied the ordinary and customary meaning as understood by a POSITA in light of the specification and the prosecution history. However, in light of the prosecution history summarized above, I have also considered whether the prior art discloses "ignoring the data contained within a noindex metatag for the purposes of recall but using the data contained within a noindex metatag for the purposes of ranking," as Applicant alleged was taught by Claim 1. Ex-1002 at 108.

VII. OVERVIEW OF THE PRIOR ART

A. Background and State of the Art for the '329 Patent

55. At the time the '329 patent was filed, search engines and their functionalities were well known to a POSITA. In fact, all of the concepts disclosed in the '329 patent were known well before the patent's priority date. For example, it was known that a search engine could identify and collate documents, such as webpages, into an index or database so that they could later be recalled in response

to search queries. *See, e.g.*, Kamvar (Ex-1009), 1:21-24; Kamholz (Ex-1010), ¶¶5-6. It was well known that this operation would be performed by a "crawler" (also known as a "spider" or "robot"), which would follow hyperlinks between documents, review the content of each document it found, and build an index or database using portions of that content. *See, e.g.*, Kamvar, 1:24-31; Kamholz, ¶¶6-7. The search engine, upon receiving a query from a user, would examine the contents of the index or database it constructed to identify and recall documents responsive to a query. *See, e.g.*, Kamvar, 1:31-37; Kamholz, ¶6; Google Technology Overview (Ex-1011), 2. The search engine would then generate results identifying the recalled documents. *See, e.g.*, Kamvar, 1:31-37; Nicholson (Ex-1012), 6-14 (describing indexing and abstracting mechanisms for six search engines). Accordingly, it was well known in the art that the contents of a search engine's index generally restricted the number of documents that could be recalled by said search engine in response to a query and produced as search engine results.

56. Before the priority date of the '329 patent, a POSITA would also have been well aware that documents such as webpages could be-and often were divided into sections. *See, e.g.*, Lin (Ex-1013), 590 ("Based on [the] HTML tag <TABLE>, a page is partitioned into several content blocks Obviously, <TABLE> is not the only way to partition a page into blocks. If a content block includes too [much] text[], ... it can be partitioned into several smaller blocks according to tags, such as

the title, headings, <P>, or <TR> and <TD> embedded in <TABLE>.").

Webmasters often split webpages into sections to improve the informativeness and functionality of their websites and optimize webpage development processes. *See, e.g., id.*, 588-89 (noting that "almost all dot-com Web sites present their service channels, navigation panels, copyright and privacy announcements and advertisements [o]n every page for business purposes" and to make their pages "easy [to] access" and more "user-friendly," and that many used <TABLE> tags as page layout templates because they made it "easy and convenient to modularize an HTML page" that needed to present information on each of these different topics). Additionally, search engine developers eventually began to program their search engines to account for sections in webpages because doing so would "reduc[e] the size of index[es] and increas[e] the precision of [document] retrieval." *Id.*, 583; *see also* Ma (Ex-1014), 2-3 (using "table text chunks" to identify and remove templated content to improve the accuracy of search engine results).

57. Moreover, it was well understood in the art that search engines could be designed to exclude either whole documents or sections of documents from their indexes. *See, e.g.,* FBFAQ #29, 1 ("You can prevent a particular section of your page from being indexed by surrounding the content you wish to exclude with the <noindex> ... </noindex> tags."); Milosevic (Ex-1015), 1 ("This help topic describes how to prevent *sections* of a document from being indexed. To prevent

an *entire* document from being indexed, see How to prevent your pages from being indexed." (emphasis in original)); Janes (Ex-1016), 1 ("The Robots 11ETA tag ... is used to provide page-specific direction for web crawlers [I]ts page-specific nature means it cannot be used to restrict crawlers from indexing only certain sections of a document. ... [T]he Robot Exclusion Profile defines a microformat that can be applied to any element or set of elements in a page."); Kamholz, ¶¶ 9-11.

58. It was known, too, that excluding certain sections of documents from search engine indexes could be done to further improve the relevance and quality of search engine results. *See, e.g.*, Lin, 588 (noting that "intra-page redundancy"- i.e., the inclusion of redundant content blocks like "company logos, navigation panels or advertisement banners"-affects "the precision of search"); FBFAQ #29, 1 (explaining that not indexing "common footer bar[s] or navigation bar[s] on every page" could avoid "skew[ing] search results"); Milosevic, 1 (explaining that a customer may "need [to] prevent[] certain parts of a document - such as a navigational sidebar - from being included in the search"). Various mechanisms existed in the art by which this could be achieved, including, but not limited to, recognizing meta tags or attributes webmasters incorporated into the documents' HTML code and excluding document sections designated by those tags or attributes, *see* FBFAQ #29; Milosevic; Janes; Kamholz, ¶¶ 35-45, and designing

search engines to recognize templated documents and exclude document sections that conformed to those templates, *see* Lin; Ma.

59. Particularly well understood in the art was the use of HTML tags to identify content to be excluded from a search engine's index. For example, POSITAs understood that a webmaster could include tags in the HTML code of a document to direct one or more search engine robots to act in particular ways with respect to the content designated by the tags. *See, e.g.*, FBFAQ #29, 1; Milosevic, 1-2; Janes, 1-3; Kamholz, ¶¶35-41. Relevant here, the prior art explained that a webmaster could include <noindex> tags to preclude one or more search engine robots from indexing the content identified by the tags. *See, e.g.*, FBFAQ #29, 1; Milosevic, 1-2; Janes, 1-3; Kamholz, ¶¶ 35-41. Similarly, a webmaster could include <nofollow> tags to preclude one or more search engine robots from following hyperlinks within the content identified by the tags. *See, e.g.*, FBFAQ #47; Milosevic, 1-2; Janes, 1-2; Kamholz, ¶¶ 35-41.

60. It was additionally well known in the art that search engines could be, and often were, designed to rank results they generated in response to a query. *See, e.g.*, Kamvar, 1:37-40; Kamholz, ¶ 6. Various techniques existed in the art to rank search engine results-particularly by relevance to a query. *See, e.g.*, Kamholz, ¶ 6; Page; Acharya (Ex-1017); Pant (Ex-1018); Bharat (Ex-1019). Several of these techniques incorporated the analysis of hyperlinks linking documents to one

another to identify "authoritative" or "trusted" documents. *See, e.g.*, Page; Bharat; PageRank Paper (Ex-1020); Anatomy Paper (Ex-1021). Moreover, it was well known that relevance rankings could be improved by incorporating more than one of these techniques into any given search engine. *See, e.g.*, Page, 8:42-46 (noting that ranking may incorporate many "factors such as ... objective ranking and textual matching"); Li, 22:23-23:15 (recommending making a "final relevance ranking for [a] query" depend on a combination of "link-based ranking" and "conventional" or "other means of relevance ranking"); Cucerzan, ¶ 71 (stating that the disclosed relevance ranking method could be overlaid onto other "internal relevance algorithm(s) of [a] search engine" to "re-rank the search result(s)").

61. Additionally, the concept of search engine optimization ("SEO"), or the means by which webmasters could modify documents' content to improve the documents' relevance to various search engine queries-and therefore their rankings in search engine results responsive to those queries-was well understood by POSITAs. *See, e.g.*, Gross (Ex-1022), ¶ 10; Acharya, ¶ 8. In fact, certain types of SEO, called search engine spamming, bore a particularly negative connotation in the art. *See, e.g.*, Gross, ¶ 10. It was known that a webmaster could "spam" a search engine by, for example, incorporating into his or her document keywords that were invisible to users but recognizable by robots. *See, e.g., id.* The robots, upon encountering these keywords, which may or may not have been relevant to

the document's main content, would nevertheless boost the relevance ranking of the document with respect to queries incorporating those keywords. *See id.*

Another mechanism by which a webmaster could "spam" a search engine is by, for example, incorporating a number of hyperlinks into "dummy" documents that linked to those documents for which the webmaster sought to improve relevance rankings. *See, e.g., id.*; Kamvar, 4:59-5:6. The interconnected document structures created for this purpose were commonly known as link farms or spam farms. *See* Gross, ¶ 10; Kamvar, 4:59-5:6.

62. In view of the above, POSITAs were also well informed as to the ways in which search engines could be designed to account for various kinds of search engine spam. For example, it was well known that some search engines would exclude certain documents from their indexes upon identifying spam within said documents. *See, e.g.,* Gross, ¶ 27. It was also well-understood that other search engines would reduce the rank of (i.e., downrank) documents upon encountering spam within said documents. *See, e.g.,* Acharya, ¶ 39. However, POSITAs were very familiar with the notion that efforts to spam search engines and to eliminate spam altogether were in constant tension with one another and interacted in what could properly be considered a cat-and-mouse game. *See, e.g.,* Gross, ¶ 11.

B. Overview of Page

63. Page is a patent assigned to Google- the invention disclosed therein was developed by a Google co-founder, Larry Page. Page, Cover. Page was filed on January 9, 1998 and claims priority to a provisional application filed January 10, 1997, exactly ten years before the '329 patent's filing date. *Id.* Page describes search and ranking functionalities that, by the time the '329 patent was filed, were extremely familiar to POSITAs. For example, much of Page's disclosure refers to one of Google's search engine ranking methods, called "PageRank," that was as of the priority date of the '329 patent one of the most influential forces in the fields of search engine development, webpage development, and SEO. This was because Google used PageRank and variations thereof as a major component of its search engine result ranking algorithms. *See, e.g.,* Google PageRank Information; Google Technology Overview, 1. Consequently, to improve the relevance of their websites in Google's search engine results, webmasters and SEO companies continuously sought better ways to acquire (and avoid losing) what they referred to as "PageRank" or "link juice." *See, e.g.,* Enge (Ex-1024), 1; Digital Point (Ex-1025), 1-2. Likewise, search engine developers were keenly aware of the effectiveness of PageRank and Google's web search engine methodology.

64. Page discloses methods of "assign[ing] importance ranks to nodes in a linked database," which may be "any database of documents containing citations,

the world wide web, or any other hypermedia database." Page, Abstract. Page explains that its disclosed methods were designed to combat a problem common to search engines collating and searching large document databases like the web: because such databases often "contain many low-quality documents," "searches typically return hundreds of irrelevant or unwanted documents which camouflage the few relevant ones." *Id.*, 1:36-39.

65. Page describes some of the prior art techniques for improving the relevance of search results that existed in the late '90s. For example, limiting the scope of searches to particular subsets of the relevant databases or adding additional search terms could, for searches in large, heterogeneous databases like the web, "increase[] the chances that the desired information w[ould] be inadvertently eliminated from the search results." *Id.*, 1:39-50. As another example, Page explains that ranking documents using only "variations of a standard vector space model" improved the relevance of results little and was "vulnerable to 'spamming' techniques that authors can use to artificially inflate their document's relevance." *Id.*, 1:50-67.

66. Page therefore presents a method based on "the well known idea of citation counting," by which a document's importance is determined by counting the number of citations (or backlinks) to said document. *Id.*, 2:20-25. The idea is that, generally, the more a document is cited, the more relevant it will be to a

search query. *See id.*, 2:26-29. Accordingly, Page's method "tak[es] advantage of the linked structure of a database to assign a rank to each document in the database, where the document rank is a measure of [its] importance." *Id.*, 2:51-54. To account for the fact that databases like the web can "have extreme variations in the quality and importance of documents," *id.*, 2:29-35, the method assigns ranks to documents based not only on the number of citations to them, but also on the importance of the documents in which those citations are found, *id.*, 2:61-3:2. The ranks are calculated using an iterative procedure. *Id.*, 3:2-3.

67. The method of Page generates a "directed graph of N nodes" from the contents of a linked database such as the web, wherein "each node corresponds to a webpage document and ... the directed connections between nodes correspond to links from one document to another." *Id.*, 3:56-63. Each node in the graph has a set of *forward links* and *backlinks* connecting the node to other nodes (or documents) in the database. *Id.*, 3:63-4:4. As shown in Figure 1 below, a forward link is typically identified by the inclusion of hyperlink text pointing to another document. *See id.*, Fig. 1, 3:67-4:3. For example, hyperlink text in each of Documents B and C points to Document A, so Document A would be considered a forward link of Documents B and C. *Id.* Documents B and C, on the other hand, would be considered backlinks of Document A. *Id.*

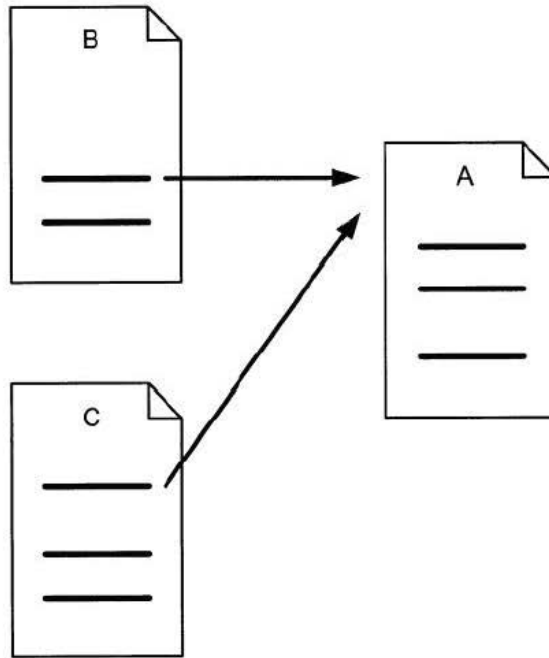


FIG. 1

Id., Fig. 1.

68. Through a process that iterates until convergence, *id.*, 5:20-59, the method of Page calculates a rank for each document based on both (1) the number of backlinks the document has; and (2) the rank of each document to which it backlinks, *id.*, 4:25-37. When added together across all documents in the database, "[t]he ranks form a probability distribution over webpages, so that the sum of ranks over all webpages is unity." *Id.*, 4:38-40. Any given document's rank represents the probability that a typical surfer of the database will arrive at the document after following a bunch of forward links. *Id.*, 4:40-42.

69. Page describes a number of potential modifications to the formulas and calculations disclosed therein that it posits may improve the method's performance for certain purposes. *Id.*, 6:61-7:36. For instance, it notes that the method's performance may be improved for particular purposes by weighting the relative importance of links based on their positioning or formatting within documents. *Id.*, 7:28-32. As an example, "highly visible links that are near the top of a document"—such as near a document title—"can be given more weight." *Id.*, 7:29-30.

70. Further, Page explains that the method may be used for myriad purposes, *id.*, 7:36-8:48, including and especially for "enhancing the quality of results from web search engines," *id.*, 8:6-8. In this vein, Page discloses a "ranking method" by which a search engine crawler, while exploring the web and indexing the content of documents, also creates "a directed graph of nodes corresponding to the structure of hyperlinks" between the web documents. *Id.*, 8:12-16. "The nodes of the graph (i.e.,] pages of the web) are then ranked according to importance" using the calculations discussed above. *Id.*, 8:16-20.

71. Likewise, Page explains that when recalling documents to be ranked by its method, the search engine can search the entire text of a document or can search only certain sections. For example, Page discloses that its ranking method is used in conjunction with search engines that locate responsive documents "either

by searching full text, or by searching titles only." *Id.*, 8:21-23. Once the documents have been recalled, the search engine can then return an ordered "list of documents," "with high ranking documents first and low ranking documents last." *Id.*, 8:42-44. Page intimates that ranking may also be based on other factors like "objective ranking" and "textual matching." *Id.*, 8:44-47.

C. Overview of FusionBot

72. FusionBot (Ex-1005) is a collection of webpages published on the FusionBot product website hosted by LOGIKA Corporation, a company that offered web search solutions. See generally FB What's New. The FusionBot website included webpages answering "Frequently Asked Questions" ("FAQs") on the various capabilities of the FusionBot search engine and web development platform products. *See, e.g.*, FBFAQ #26; FBFAQ #29; FBFAQ #47.

73. Two of the four cited FusionBot webpages-those for FAQs #29 and #47-explain that the FusionBot platform allowed users to selectively tag content on webpages that would instruct search engine crawlers as to (1) whether the content should be indexed, FBFAQ #29; or (2) whether hyperlinks within the content should be followed to other webpages, FBFAQ #47. FusionBot explains that these tagging features had been incorporated to offer webmasters the ability to control what content search engine crawlers could discover and/or index and improve the

relevance of search results returned by search engines on their websites. See FB FAQ #29, 1; FBFAQ #47.

74. Notably, FusionBot explains that, as early as January 16, 2001, a user of the FusionBot platform could prevent a crawler from indexing a section of the body of a webpage "by surrounding the content ... with the <noindex> ... </noindex tags>." FB What's New, 4; FBFAQ #29, 1. The tags could be implemented as shown below:

<noindex>
This is the content you wish
to be excluded from your index
within a particular page.
</noindex>

FBFAQ #29, 1. Upon encountering the tags within the body of the webpage's HTML code, a crawler would identify and ignore the material sought to be excluded as content that should not be indexed by the corresponding search engine. *See id.* Consequently, this non-indexed content would not be searched by the search engine in response to an eventual search query. *Id.*

75. This designation of non-indexable content within a page could even be done on a crawler-specific basis. *Id.* Specifically, a user could use the tags <noindexfb> ... </noindexfb> to instruct the FusionBot crawler to not index the

content while making the content indexable for all other search engine crawlers. *Id.* Additionally, a similar function could be achieved by using comment tags (e.g.,`<!--noindex begin--> ... <!--noindex end-->` or `<!--begin noindex--> ... <!--end noindex-->`) to identify the content to be ignored by crawlers. *Id.* FusionBot explains that a webmaster may wish to include these kinds of tags to exclude from a search engine index webpage elements like "common footer bar[s] or navigation bar[s]" that appear on every page of a website. *Id.* Should these elements not be omitted, terms that appear in those elements, when included in a search engine query, could "cause[] every page [on a website] to be returned" and thereby produce results containing a number of documents irrelevant to the query. *Id.*

76. FusionBot also notes that including noindex-type tags would "only prevent the indexing of actual words on [the] page\ it would "NOT prevent [the] spider from following a specific link or links ... included between the <NOINDEX> tags." *Id.* FusionBot explains that a user could prevent a crawler from following specific links on a webpage by using nofollow-type tags, as shown below:

```
<NOFOLLOW><A HREF="http://www.mysite.com/private.htm">Click Here</a></NOFOLLOW>
```

- OR -

```
<NOFOLLOWFB><A HREF="http://www.mysite.com/private.htm">Click Here</a>  
</NOFOLLOWFB>
```

FBFAQ #47. Consequently, FusionBot explains that enclosing particular content in noindex-type tags would preclude a crawler from indexing the content *without* preventing the crawler from following hyperlinks within that content and discovering how webpages were interconnected. *See* FBFAQ #29, 1; FBFAQ #47.

77. Another webpage on the FusionBot website, discussing FAQ #26, remarks that the FusionBot platform also allowed users to control the elements included in results output by a search engine in response to a query. FBFAQ #26. The FusionBot platform allowed the search results to present up to "seven basic elements within the text content": (1) the title of the document, which was extracted from the <TITLE> tag for the page; (2) a description of the document, if any, which would be extracted from the <MET A Description> tag for the page; (3) an excerpt from the body of the document where a query term was found; (4) a rank value (indicative of whether the document was a "good match" for the query); (5) the Uniform Resource Locator ("URL") of the document; (6) the size of the document; and (7) the date the document was created and/or last modified. *Id.* However, the FusionBot platform permitted users to customize their results displays by excluding any of elements (2)-(7) from the results; only the title of each document had to be displayed. *Id.*

78. In my opinion, a POSITA would have considered the four cited FusionBot webpages to be a single publication because they were all located on the

same product website and elaborated upon the functionalities of the same FusionBot product line. Specifically, three of the webpages (FBFAQ #26, FBFAQ #29, and FBFAQ #47) are directed to answering FAQs about related capabilities of the FusionBot software. And the fourth (FB What's New) documents when certain of these features (among others) were added to the FusionBot product line.

79. If it is determined that FusionBot is not a single printed publication, but is instead four separate publications, then in my opinion it would have been readily apparent and obvious to a POSITA to combine their teachings. Again, all four webpages are from the same website and describe the same FusionBot software. The three FAQ pages describe related content by explaining (1) how FusionBot uses tags to prevent certain sections of a document from being indexed (FBF AQ #29); (2) how FusionBot uses related tags to prevent hyperlinks in certain sections from being followed (FBFAQ #47); and (3) how search results displays may be customized using the FusionBot product (FBFAQ #26). Likewise, the FB What's New page provides links to some of these same FAQs and discusses additional features and functionalities of the FusionBot platform. These documents would have been readily available to a POSITA who simply visited and clicked through the FusionBot website to learn more about the FusionBot product line. A POSITA would have been motivated to combine the teachings of these four webpages on the same website and expected success in doing so because the

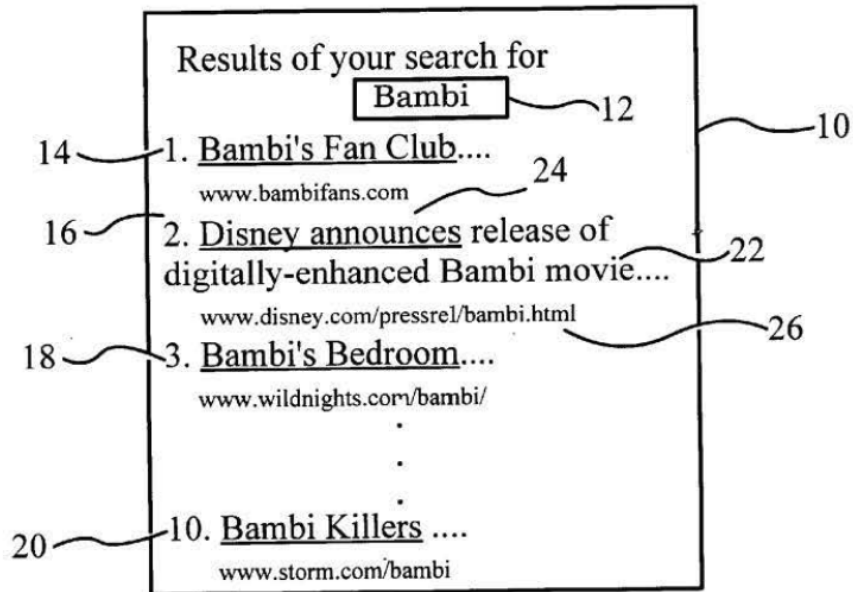
FusionBot website expressly states that they already related to and were combined in the same preexisting system.

D. Overview of Emens

80. Emens was filed on August 26, 1999, was issued September 25, 2001, and discloses "a method for providing rated hypermedia content," wherein the rating indicates the degree to which the hypermedia content might be considered objectionable. Emens, Cover, Abstract. The disclosed method seeks to target problems related to the exposure of immature populations, like children, to overly violent, pornographic, or other forms of offensive or objectionable content because of their unfettered access to the breadth of information on the web via web search engines. *Id.*, 1:20-30. Accordingly, an embodiment disclosed by Emens uses "raw search result page[s] created automatically by a search engine in response to a query" to produce and transmit to users rated versions of the search results. *Id.*, Abstract. In this way, Emens presents a method for improving the quality and utility of search engine results.

81. Emens explains that when a search engine receives a query, it will "search an index to find applicable documents" and automatically "create a search result page listing a number of the located documents." *Id.*, 1:56-58, 3:60-4:1, 4:59-67, 6:57-62. This search result page will often present, for each document,

"summary information ... , which may include a title and short abstract or extract," and a URL, as shown in the following figure:



Prior Art

FIG. 1

Id., Fig. 1, 2:59-3:7, 4:1-7, 1:58-61.

82. Additionally, Emens discloses that a search engine's index, which is built by a crawler associated with the search engine, will "contain[] Web site meta-information, including name, keywords, URL, content rating vectors, and other relevant information" for each indexed document. *Id.*, 4:61-63, 5: 15-18. This "meta- information" may encompass the document's "title, summary information, keywords, and URL." *Id.*, 5:18-20. In other words, Emens discloses that when the search engine crawler adds a document to its index, summary information for that document, which may include an abstract or short extract, is also stored in said

index until it needs to be presented with the document in the search engine's results returned for one or more queries. *See id.*, 1:56-61, 3:60-4:7, 4:59-67, 5:15-20, 6:57-62. Emens further provides a means of "block[ing] access" to both recalled objectionable documents and summary information for these documents on search results pages. *Id.*, 4:25-28, 6:54-7:25.

E. Overview of Li

83. Li, a patent publication dated December 24, 1997, discloses methods directed to "searching databases distributed over wide-area networks such as the World Wide Web." Li, Cover, 1:3-5; *see also id.*, Abstract. Li explains that database systems, like the web, that incorporate forms of hypertext or hypermedia can be organized into non-linear arrays of nodes connected by hyperlinks. *Id.*, 1:7-14. Li defines a hyperlink as "a relationship between two anchors" (i.e., a linking document and a linked document), *id.*, 1:15-16, and explains that hyperlinks "are generally identified by underscoring or highlighting certain text or graphics" associated with the linking document/anchor, *id.*, 1:16-2:4. Li aimed to improve upon prior art ranking methodologies by taking advantage of anchor text. *See id.*, Abstract.

84. Li explains that almost all prior art search engine ranking techniques based on relevance at the time "depend[ed] on the frequency of query terms in a given document," though it notes that additional factors like "how many documents

contain the [query] term[s]" could also be considered. *Id.*, 3:18-4:1. These factors, once determined, could then be input into one of "various models[,] such as the vector space model, probab[i]listic model, fuzzy logic models, etc.," to rank documents. *Id.*, 4:1-6.

85. Li provides an example of how a vector space model might have been used to rank a document based on the frequency with which a query term appears in said document:

[I]n the vector space model, a user query Q is represented as a vector where each query term (qt) is represented as a dimension of a query vector.

$$Q = < qt_1, qt_2, \dots, qt_m >$$

Documents in the database are also represented by vectors with each term or key word (dt) in the document represented as a dimension in the vector.

$$D = < dt_1, dt_2, \dots, dt_m >$$

The relevance score is then calculated as the dot product of Q and D.

Id., 4:7-14. Li explains that the relevance rankings of search engine results could be improved by weighting the query terms and/or vectors in various ways or by normalizing query term weighting based on the lengths of documents in which query terms were found. *Id.*, 4:15-5:11. However, it noted that the improvements

these mechanisms afforded were limited, and it recognized a need to further enhance relevance ranking strategies for search engines. *See id.*, 5:11-7:8.

86. Consequently, Li sets forth a "method of ranking documents ... based on the document[s'] relevance to a query," wherein a document's relevance ranking is determined by comparing the query terms to the terms in the anchor text of all hyperlinks pointing to said document. *Id.*, 8:17-9:2. Li discloses that this determination may be achieved through use of a vector space model in which (1) "[t]he query may be represented by a query vector ... contain[ing] a dimension of each term in the query"; (2) the "document may be represented by document link vectors for each hyperlink pointing to the document," with each document link vector "contain[ing] a dimension for each term in the corresponding hyperlink pointing to that document"; and (3) "[c]omparing the words in the query to the words in the hyperlinks includes calculating [a] dot product of the query vector with the document link vector for [each] hyper link" and "summing the dot products." *Id.*, 9:3-15; *see id.*, 19:21-22:7. Li further discloses that results returned for a given query may be "sorted so that the documents having higher summed relevance rankings are listed above those with lower rankings." *Id.*, 22:8-10.

F. Overview of Cucerzan

87. Cucerzan is a September 13, 2007 publication of a U.S. patent application filed March 13, 2006. Cucerzan, Cover. Cucerzan discloses methods

for "improving the relevance of search results through utilization of a query log."
Id., Abstract.

88. Cucerzan explains that search engines operate by using indexes to identify web documents responsive to user queries, ranking the identified documents "based on measures of how good the match is (such as term frequency, inverse document frequency, and term proximity) in conjunction with measures of Website/document popularity," and producing "ordered list[s]" in response to the queries. *Id.*, ¶ 2. However, it observes that the relevance of search engine results was hamstrung by users' inability to precisely convey their needs and/or intent through search queries and the fact that users often were too busy or unwilling to set up user profiles that could facilitate more relevant searches. *See id.*, ¶¶ 3-7. It further notes that, for apparently ambiguous searches, measures of "absolute relevance of web pages or web sites (such as page rank) may not be indicative of how relevant those pages or sites are." *Id.*, ¶ 7. Consequently, Cucerzan sets forth techniques for ranking or re-ranking search results by relevance, wherein "[t]he relevance of the search results for a target query can be judged based on one or more queries in [a query] log that are related to the target query temporally and/or lexically." *Id.*, Abstract; *see also id.*, ¶ 11.

89. In embodiments of the method Cucerzan discloses, a "query log component" receives a list of search results generated by a search engine in

response to a target query. *Id.*, ¶¶33, 41, 50, 58, 62, 76, 85. This list may include associated relevance rankings. *Id.*, ¶¶41, 58. The query log component can then build a language model-which may be "a vector of terms" or a probability distribution for the target query using stored or retrieved information about the target query and temporally or lexically related queries stored in a query log. *Id.*, ¶¶35, 41-42, 62-63, 77, 86, Claim 1. The newly constructed language model is compared with the "lexical properties of the search results for the target query" by a matching component to determine a "match score" from either all or a subset of the search results. *Id.*, ¶¶36-37, 43, 51, 63, 77, Claim 1. This comparison may be accomplished by, for example, tokenizing a "segment of a document or web page" and creating a word vector, *id.*, 143; *see id.* ¶¶65-66, or by generating "a language model for every search result" using properties obtained after "processing a URL, page title(s), page content (body, text), snippets of the results of a document query pair, etc.," *id.*, ¶¶37, 50, 62-63. A ranking component of the search engine can then "rank or re-rank the search results based on the original ranking" the search engine provided and/or the match score, at which point the newly ranked or re-ranked search results can be provided to the user. *Id.*, ¶¶37, 43, 52, 63, 77, Claim 1; *see also id.*, ¶71.

VIII. GROUND 1: CLAIMS 1-3, 5, 8-10, AND 12 ARE OBVIOUS IN VIEW OF PAGE IN COMBINATION WITH FUSIONBOT.

A. Claim 1

1. [1p] A method, comprising;

90. To the extent the preamble is limiting, Page discloses it. For example, Page discloses "a computer implemented method . . . for scoring linked database documents." Page, 3:17-19; *see also id.*, 2:39-50 (summarizing "systems and methods for ranking documents in a linked database"). As discussed below, Page's disclosed method also performs all of the steps recited in claim 1.

2. [1a] ranking a plurality of documents recalled by a search engine for a query;

91. Page discloses "ranking a plurality of documents." For example, Page explains that "[t]his invention ... relates to methods for assigning ranks to nodes in a linked database, such as any database of documents containing citations, the world wide web or any other hypermedia database." Page, 1:18-22; *see also id.*, 2:39-3:33 (summarizing "various aspects" of Page's disclosures as "provid[ing] systems and methods for ranking documents in a linked database"). Page's disclosure of rank ing "documents" (plural) teaches ranking a "plurality of documents," as claimed. Moreover, a POSITA would have understood that by applying its document ranking methods to "extremely large databases such as the world wide web," *id.*, 2:45-50, Page discloses ranking a plurality of documents.

92. Page also discloses that the ranked documents have been "recalled by a search engine for a query." For example, Page explains that an "important application and embodiment of the ... invention is directed to enhancing the quality of results from web search engines." *Id.*, 8:6-8. "A search engine employing [the] ranking method" of Page includes a web crawler that "explores the web and creates an index of the web content, as well as a directed graph of nodes corresponding to the structure of hyperlinks. The nodes of the graph (*i.e.*, pages of the web) are then ranked according to importance." *Id.*, 8:13-20. Page discloses that when "[t]he search engine is used to locate documents that match ... specified search criteria," "a set of documents is identified that match the search terms." *Id.*, 8:21-22, 8:42-43. Page further discloses that, at that point, "the list of documents is ... sorted with high ranking documents first and low ranking documents last." *Id.*, 8:42-44.

93. Page's disclosures related to a search engine locating documents are consistent with the '329 patent's description of "document recall" as "[t]he inclusion of a document within ... search engine results generated by a search engine for a search engine query." '329 patent, 2:3-5. They are also consistent with the prior district court construction of the word "recall" as "generating results for a search engine query." *See supra* Section VI; Ex-1003, 49, 51.

3. [1b] wherein the plurality of documents contain certain documents, each document of said certain documents

containing at least one section that is not used by said search engine for recall

94. A POSITA would have understood that a plurality of documents would necessarily "contain certain documents." Specifically, a "plurality of documents" must contain at least two documents. Moreover, Page identifies groups of certain "high ranking documents" and certain "low ranking documents" within the entire "set of documents." Page, 8:42-44.

95. Page also explains that each document of the certain documents contains at least one section that is not used by said search engine for recall under the ordinary and customary meaning of that phrase. For example, Page discloses that a search engine may exclude at least one section of each document in the certain documents from use in recall by explaining that, in an exemplary search engine, "[t]he search engine is used to locate documents that match specified search criteria, either by searching full text, *or by searching titles only.*" *Id.*, 8:21-23 (emphasis added). As explained above in Section VII.A, a POSITA would have understood that webpages typically comprise various sections, such as title sections, navigation sections, copyright and privacy announcement sections, advertisement sections, and main content sections, which may or may not be partitioned through the use of tags. *See, e.g.*, Lin, 588-90; Ma, 2-3. As a result, by explaining that the search engine searches using only titles of the documents or webpages, Page discloses that the search engine recalls documents using only

content found in their title sections. Page therefore simultaneously discloses to a POSITA the inverse: that the search engine uses no content found in other, non-title sections of the documents (e.g., navigation sections, copyright and privacy announcement sections, advertisement sections, main content sections) for recall.

96. Thus, by disclosing that Page's exemplary search engine searches only by title, this embodiment of Page discloses the ordinary and customary meaning of limitation [1b].

97. To the extent the Board determines that the Applicant disclaimed limitation [1b], limiting it to documents having <noindex> tags, a POSITA would have found it obvious to implement Page's disclosure of selectively excluding sections of documents from recall by enclosing those sections within <noindex> tags, as was known in the art prior to the '329 patent's earliest priority date.

98. Page explains that a search engine may include a "web crawler" that "explores the web and creates an index of the web content, as well as a directed graph of nodes corresponding to the structure of hyperlinks." Page, 8:11-16. As explained above in Section VII.A, a POSITA would have understood that, to respond to a search query, the search engine would use the index it created to recall documents matching the query. *See supra* Section VII.A; *see also* Kamvar, 1:31-37 ("When a user makes a search query having one or more terms, the search engine searches the index for documents that satisfy the query, and provides a list of

matching documents "); Kamholz, ¶ 6 ("Everything [a] crawler finds goes into an index The search engine software, which is yet another part of the search engine, is a program that sifts through the pages recorded in the index to find documents fulfilling a search query submitted by a user."). Thus, a POSITA would have understood that any sections of a given document that were not indexed would not be searchable by the search engine-i.e., would not be used by the search engine to recall the document in response to a search query-as expressly contemplated by Page.

99. FusionBot describes the <noindex> tags that were known in the art. Specifically, FusionBot explains that "[y]ou can prevent a particular section of your page from being indexed by surrounding the content you wish to exclude with the <noindex> ... </noindex> tags." FBFAQ #29, 1. FusionBot includes an example demonstrating how to implement <noindex> tags to exclude a particular section of a document from a search engine index-and therefore preclude it from being used by a search engine for recall:

<noindex>

**This is the content you wish
to be excluded from your index
within a particular page.**

</noindex>

Id.

100. A POSITA would have been motivated and found it obvious to combine Page's and FusionBot's teachings. Doing so would have combined two known elements (FusionBot's `<noindex>` tags for excluding parts of a document from an index and Page's teachings of a search engine that excludes sections of a document from recall) to achieve the predictable result of an efficient web crawler that indexed only those portions of documents to be used in recall. The `<noindex>` tag was well known to POSITAs, and its functions were taught not only by FusionBot, but also by other references. *See supra* Section VII.A; *see also* Milosevic, 1-2 (using `<!-- robots content="noindex" --> ... <!-- /robots -->` comment tags to "prevent *sections* of a document from being indexed" (emphasis in original)); Janes, 1 (using `<dt id="robots-noindex">robots-noindex</dt>;<dd> ... </dd>` tags to "[i]nform[] robots that the content of [an] element is not to be included as part of the page"); Kamholz, ¶¶ 37-41 ("designat[ing] specific content that is prevented from being indexed ... by a bot" by enclosing the content in `<robot="noindex"> ... </robot>` tags).

101. Given Page's disclosure that in one embodiment a search engine only uses the title section of each document for recall, a POSITA would have been motivated and found it obvious to implement systems and methods that accounted for FusionBot's `<noindex>` tags placed around non-title sections in documents.

Doing so would enable a crawler to quickly and efficiently index web content because the web crawler would only index the portions of those documents it would use for recall-e.g., the titles of those documents. It would therefore be able to (1) significantly reduce the time necessary to index volumes of documents, and (2) lessen the storage requirements and computational burdens imposed by the generation of search engine indexes as compared to search engines designed to index each document in its entirety. The resulting time and cost savings would have motivated a POSITA to implement such a design.

102. FusionBot also provides express motivation for combining Page and FusionBot. For example, whereas Page suggests excluding certain sections of a document from recall (e.g., non-title sections), FusionBot explains that one may wish to exclude from an index (and therefore from recall) only "common footer bar[s]" or "navigation bar[s]" that "may skew the search results." FBFAQ #29, 1. Based on this disclosure, a POSITA would have been motivated and found it obvious to modify Page's embodiment of a search engine designed to search on "full text" to exclude document navigation bar and/or common footer bar sections from the search engine's index-and therefore from recall. Page already contemplates excluding certain sections of documents from searches (by excluding said sections from search engine indexes), and FusionBot teaches that the navigation bar and common footer bar sections are specific examples of sections

that one may wish to exclude to prevent search results from being "skew[ed]." *Id.* Again, this would have been a straightforward combination of two known elements (Page's search engine, which contemplates excluding sections of a document from recall, and FusionBot's <noindex> tags and express teaching that the navigation bar and/or common footer bar should be excluded) to achieve the predictable result of improving search results, as expressly taught by FusionBot.

4. [1c] and one or more sections that are used by said search engine for recall;

103. The combination of Page and FusionBot as explained in element [1b] also teaches this limitation. As explained with reference to element [1b] above, *see supra* Section VIII.A.3, Page discloses that one or more sections (e.g., a title section) of each document of said certain documents are used by said search engine for recall under the ordinary and customary meaning of that phrase because Page discloses that "[t]he search engine is used to locate documents that match the specified search criteria, either by searching full text, or by searching titles only," Page, 8:21-23.

104. Moreover, a POSITA would have found it obvious to use FusionBot's <noindex> tags to exclude certain sections of documents, e.g., the main content of the documents or navigation bars and common footer bars, from a search engine's index (and thus from recall). However, in each example, the remaining sections of

the document (i.e., those not enclosed within the <noindex> tags) would be indexed and thus used by the search engine for recall.

5. [1d] wherein ranking a plurality of documents includes ranking said plurality of documents based, at least in part, on the at least one section of said certain documents not used by said search engine to recall documents; and

105. Page discloses that "ranking a plurality of documents includes ranking said plurality of documents based, at least in part, on the at least one section of said certain documents not used by said search engine to recall documents."

106. As explained above regarding elements [1b] and [1e], Page discloses an embodiment where the search engine uses only the title section of each document for recall. *See supra* Sections VIII.A.3-4. Thus, in this embodiment, the rest of the document, including the main content section, is not used for recall. Page, however, discloses using the entire document for ranking and therefore teaches ranking the documents "based, at least in part, on the at least one section of said certain documents not used by said search engine to recall documents."

107. Specifically, Page describes embodiments of a method in which the recalled documents are ranked using an algorithm "where the document rank is a measure of the importance of a document." Page, 2:51-54. In this sense, Page recognizes that documents, such as webpages, may contain hyperlinks to other documents, and that the importance of a particular page (and hence its rank) depends on both "the number of citations it has" and "the importance of the citing

documents." *Id.*, 2:51-3:16, 3:56-4:4. Consequently, Page's method recursively ranks each webpage in a set of search engine results based on (1) the number of backlinked webpages it connects to, and (2) the ranks of each of those backlinked webpages, the latter of which may be determined by the number of forward links from still other pages to each of the backlinked webpages. *Id.*, 3:56-4:48.

108. Page's link-based relevance ranking is similar to those link-based "forms of analysis" for ranking that are disclosed in the '329 patent. '329 patent, 3:20-31. Notably, an embodiment of the method disclosed in the '329 patent incorporating these forms of analysis assigns document ranks based on "[o]utgoing links" and "link characteristics," such as "where the links lead, the link text and link quantity and quality," *id.*, 4:17-31, and can "readily identif[y]" "low quality links" and rank documents accordingly, *id.*, 3:28-31. This is unsurprising because at the time of the '329 patent's filing, Google's link-based relevance ranking methods were well known and widely implemented, as evidenced not only by Page but also by other papers and patent applications published by Google co-founder Larry Page (Page's named inventor), others at Google, and beyond. *See, e.g.*, Page; Anatomy Paper; PageRank Paper; *see also* Kamvar, 1:59-2:32; Gross, ¶¶24-25, 31-33; Acharya, ¶7.

109. Page's link-based ranking method accounts for hyperlinks found throughout entire documents (*i.e.*, all sections of all documents)—not just their title

sections. For example, Page discloses a simplified embodiment involving just three search results (Documents A, B, and C). Page, Fig. 2, 4:49-6:11. In this example, the links used for ranking are clearly not located at the tops of the pages, and thus are clearly not in the pages' title sections:

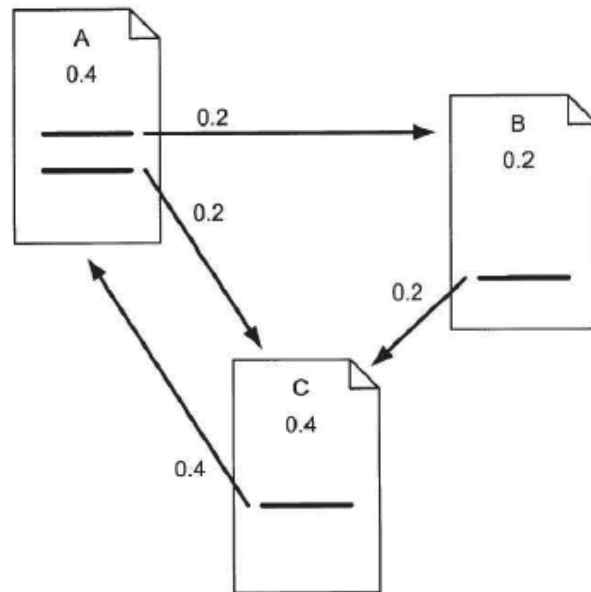


FIG. 2

Page, Fig. 2. When describing this example, Page clearly refers to the entirety of each document and not just its respective title section, explaining that "Document A has a single backlink to document C, and this is the only forward link of document C," and that "Document B has a single backlink to document A, but this is one of two forward links of document A." *Id.*, 4:49-5:8.

110. Moreover, Page confirms elsewhere that the entirety of each document—and not just its title section—is used for ranking because Page discloses that in certain embodiments "highly visible links that are near the top of a document can be given more weight." *Id.*, 7:28-30; *see also id.*, 7:30-32 ("Also, links that are in large fonts or emphasized in other ways can be given more weight."). This disclosure would have conveyed to a POSITA that hyperlinks that were not "near the top of a document," *e.g.*, that were not in the title section, would also have been used in ranking, even if in one embodiment their contributions to the ranking would be given less weight. Page therefore discloses to a POSITA a search engine capable of using the contents of non-title sections of documents for ranking, even when the contents of such sections were not used for recall.

111. Moreover, the combination of Page and FusionBot teaches this element, and a POSITA would have found it obvious to implement Page's link-based ranking method to rank documents based at least in part on the content of sections in those documents enclosed within <noindex> tags.

112. As explained above, Page discloses ranking search results based on the quantity and quality of backlinks to each document in a search engine's results. Page discloses that, with this method, "[t]he rank of a page can be interpreted as the probability that a surfer will be at the page after **following** a large number of forward links." Page, 4:39-42 (emphasis added). Moreover, and as discussed

above, Page discloses that "a web crawler explores the web and creates an index of the web content, *as well as* a directed graph of nodes corresponding to the structure of hyperlinks. The nodes of the graph (i.e. pages of the web) are then ranked according to importance" using Page's ranking method. *Id.*, 8:12-20. In this manner, the method Page discloses ranks documents not "only from the intrinsic content of [the] document[s]" (i.e., what is found in the index), but also based on "the extrinsic relationships between documents" (i.e., what is represented by the directed graph of nodes). *Id.*, 2:51-59. Notably, "implementations of the invention have the advantage that the convergence is very fast ... and it is much less expensive than building a full-text index." *Id.*, 7:37-40. Based on these disclosures, a POSITA would have understood that Page's ranking method relies on the ability of the web crawler to follow hyperlinks on various pages, construct a directed graph of nodes in view of the interconnectedness of the hyperlinks it follows, and rank the pages based on that directed graph. A POSITA would have further understood that the process of constructing the directed graph of nodes is one that is contemporaneously yet independently performed by a crawler as it follows hyperlinks to explore the web. *Cf.* Ex-1026, Cutts Interview, 2-3 (explaining that "a NoIndex page can accumulate PageRank" and pass PageRank "because the links are still followed outwards from a NoIndex page").

113. As discussed regarding element [1b] above, a POSITA would have found it obvious to enclose within FusionBot's <noindex> tags certain sections of each webpage (e.g., non-title sections to search by title only or the navigation bar and/or common footer bar sections to avoid skewing search results). *See supra* Section VIII.A.3. FusionBot expressly discloses that <noindex> tags "will only prevent the indexing of actual words on a page, but will **NOT** prevent [a] spider from following a specific link or links that may be included between the <NO INDEX> tags." FBFAQ #29, 1 (emphasis in original). Thus, in the combination of Page and FusionBot, Page's web crawler would still follow any hyperlinks within the sections of documents enclosed within <noindex> tags and account for them in the directed graph of nodes on which its ranking method is based. The search engine would therefore incorporate backlink information associated with those hyperlinks into its document rankings, even though the sections in which the hyperlinks were found were not indexed and thus not considered (i.e., ignored) for search/recall.

114. This combination thus renders obvious element [1d] and is the same as the '329 patent's disclosure that "[t]he terms inside no-recall sections ... are not used for recalling the documents," but that "the no-recall sections are included as input to forms of analysis of the document that affect, for example, the document's ranking" because "[l]inks inside the no-recall sections as well as the rest of the

document may be followed," and "low quality links will be readily identified and ranked accordingly." '329 patent, 3:17-31.

115. Moreover, a POSITA would have found it obvious to implement Page's link-based ranking method to rank documents based at least in part on the content of document sections enclosed within <noindex> tags because doing so would have been one of four discrete, predictable solutions to the problem of how a search engine should account for content within noindex sections during search engine recall and ranking processes. *See supra* Section V.B. In fact, this implementation would have particularly appealed to a POSITA, who would have been incentivized by the need to further "reduc[e] the size of [search engine] index[es]" while simultaneously "increasing the precision of [document] retrieval" and improving the relevance of search engine results. Lin, 6.

6. [1e] wherein the method is performed by one or more computing devices.

116. Page discloses that the above method is performed by one or more computing devices because Page's entire disclosure relates to "a computer implemented method ... for scoring linked database documents" that includes "obtaining" documents and ranking them. Page, 3:17-30; *see also id.*, Claims 1, 8, 9, 10, 18, 19.

117. Moreover, Page explains that the above method is performed by a "web search engine," *Id.*, 8:6-48, and a POSITA would have known that a web

search engine is implemented on one or more computing devices. Likewise, Page discloses that its embodiments are performed "using current processors," *id.*, 7:37-40, which a POSITA would know referred to computer processors. Page thus discloses to a POSITA that its methods are performed by one or more computing devices.

B. Claim 2: The method of claim 1, wherein said at least one section of said certain documents not used by said search engine to recall documents are demarcated by a tag of an element having a name from a set of one or more names.

118. The combination of Page and FusionBot teaches this element. For example, FusionBot explains that "[y]ou can prevent a particular section of your page from being indexed by surrounding the content you wish to exclude with the <noindex> ... </noindex> tags." FBFAQ #29, 1. FusionBot provides an example demonstrating the implementation of <noindex> tags to exclude a particular section of a document from a search engine index:

<noindex>
This is the content you wish
to be excluded from your index
within a particular page.
</noindex>

Id. Here, <noindex> tags demarcate a section of a document that is not indexed by the search engine and therefore not used by the search engine for recall. The tag

bears the name "<noindex>" from a set of one or more names (e.g., "<noindex>," "<noindexfb>," "<nofollow>," "<noarchive>"). *Id.* Notably, the '329 patent's illustrative "tag ... used to delimit the no-recall sections of [a] webpage" carries a strikingly similar name: "<div class =\"robots-noindex\">." '329 patent, 3:51-67.

C. Claim 3: The method of claim 1, wherein at least one document of said certain documents does not comprise data that contains element tags.

119. The combination of Page and FusionBot teaches this element. As discussed above, Page discloses an embodiment in which the search engine searches using only document titles. *See supra* VIII.A.1.c. Moreover, FusionBot teaches that a webmaster may optionally add <noindex> tags to sections of a document that the webmaster does not want indexed and used for recall. FBFAQ #29, 1 (explaining that "[y]ou can," but do not have to, "prevent a particular section of your page from being indexed by surrounding the content you wish to exclude with the <noindex> ... </noindex> tags"). A POSITA would have understood that, if a search engine searched only on document titles, a webmaster could—but need not—enclose non-title sections of his or her webpages within <noindex> tags to preclude those sections from being used for recall. *See supra* Section VIII.a.3. A POSITA also would have understood that, because each webmaster had the option to include <noindex> tags on his or her webpages, for various reasons specific to each webmaster, some documents the search engine indexed would contain

<noindex> tags, and others would not. Regardless of whether such tags were included, the search engine would continue to use only the title of each document for recall. However, a POSITA would have been motivated to generate systems and methods that accounted for these <noindex> tags where they were included, as noted above, because doing so would help optimize the storage capacity and analytical efficiency of a search engine that searched on titles only. *See supra* VIII.B.1.c. The combination of Page and FusionBot thus teaches the limitation of claim 3.

120. Moreover, whether the documents contain or do not contain <no index> tags is a binary choice. It thus presents only two possible approaches to this problem, each of which a POSITA could have implemented with a reasonable expectation of success. Because of this, the combination would have at least been obvious to try.

D. Claim 5

- 1. [5a] The method of claim 1, said search engine using a search engine index to recall said plurality of documents; and**

121. As discussed above in Sections VIII.A, the combination of Page and FusionBot presents a search engine that uses an index to recall a plurality of documents.

- 2. [5b] wherein said search engine index does not index any term in said certain documents that is found only in one or**

more sections of said certain documents established as a no-recall section.

122. As discussed above in Sections VIII.A.3 and VIII.B, the combination of Page and FusionBot renders obvious excluding one or more sections of documents from a search engine index by demarcating those sections with `<noindex>` tags. A POSITA would have understood that tagging a section in this manner establishes it as a "no-recall section," as claimed. Indeed, this is virtually identical to the embodiment discussed in the '329 patent with reference to Figure 2, which uses a `<div class="robots-noindex">` tag to "delimit the no-recall sections of [a] webpage." *See* '329 patent, Fig. 2, 3:51-4:16.

123. A POSITA would have understood that any term that appeared only in a section demarcated by `<noindex>` tags-i.e., any term "found only in one or more [no-recall] sections"—would not be indexed.

E. Claim 8

124. Claim 8 recites "[a] computer-readable storage medium that stores instructions which, when executed by one or more processors, cause performance of steps identical to elements [la]-[ld] of claim 1, which are analyzed above. *See supra* Sections VIII.A.2-5. The combination of Page and FusionBot discloses the elements of claim 8 for the same reasons articulated above regarding claim 1. Page likewise discloses the claimed "computer-readable storage medium that stores instructions which, when executed by one or more processors, cause performance

of the claimed method for at least the reasons discussed above regarding element [1e], which similarly requires that claim 1 is performed by one or more computing devices. *See supra* Section VIII.A.6. A POSITA would have understood that, because Page's methods were performed by computing devices and designed to enhance web search engines, they would necessarily be performed by one or more processors executing programs stored on a computer-readable storage medium. Moreover, Page expressly confirms this notion by reciting in multiple claims "[a] computer-readable medium that stores instructions executable by one or more processors to perform" its methods. Page, Claims 18-19.

F. Claim 9

125. Dependent claim 9 is substantively identical to dependent claim 2, except that it recites "[t]he computer-readable storage medium of claim 8" instead of "[t]he method of claim 1." Page and FusionBot teach the limitations of claim 9 for the same reasons discussed above regarding claims 2 and 8.

G. Claim 10

126. Dependent claim 10 is substantively identical to dependent claim 3, except that it recites "[t]he computer-readable storage medium of claim 8" instead of "[t]he method of claim 1." Page and FusionBot teach the limitations of claim 10 for the same reasons discussed above regarding claims 3 and 8.

H. Claim 12

127. Dependent claim 12 is substantively identical to dependent claim 5, except that it recites "[t]he computer-readable storage medium of claim 8" instead of "[t]he method of claim 1." Page and FusionBot teach the limitations of claim 12 for the same reasons discussed above regarding claims 5 and 8.

IX. GROUND 2: CLAIMS 4 AND 11 ARE OBVIOUS IN VIEW OF PAGE IN COMBINATION WITH FUSIONBOT AND EMENS.

128. As I explain below, the combination of Page and FusionBot in further view of Emens renders obvious claims 4 and 11.

129. As explained above in Section VII.D, Emens sets forth "a method for providing rated hypermedia content," wherein the rating indicates the degree to which the hypermedia content might be considered objectionable. Emens, Abstract. In this manner, Emens combats problems associated with exposure of children and other immature user populations to overly violent, pornographic, or other forms of offensive or objectionable content available to them via web search engines. *Id.*, 1:20-30. Accordingly, an embodiment disclosed by Emens uses "raw search result page[s] created automatically by a search engine in response to a query" to produce and transmit to users rated versions of the search results. *Id.*, Abstract.

130. A POSITA would have understood, based on its disclosures, that Emens presents a method for improving the quality and utility of search engine

results for certain segments of users. Similarly, an "important application" of Page's ranking method "is directed to enhancing the quality of results from web search engines." Page, 8:6-8. FusionBot likewise presents a means of customizing search results so as to maximize their utility for various search engine users. FBFAQ #26. Accordingly, a POSITA would have naturally looked to Emens in addition to Page and FusionBot and accounted for its disclosures so as to optimize search engine results for all user populations.

131. In addition, combining Emens with the combination of Page and FusionBot would have been nothing more than the application of known techniques (i.e., Emens's techniques for rating hypermedia content and presenting rated versions of search results) to known systems and/or methods (i.e., Page's search engine and related ranking methodology and FusionBot's techniques for customizing search results) to yield predictable results (i.e., improved methods for displaying relevant and useful search results to users).

132. Further, a POSITA would have found it obvious to combine Emens's teachings regarding the use of titles and abstracts in search engine indexes and search results and the customization of search results with Page's and FusionBot's teachings of a search engine that permits customization of the types of content displayed in search results. *See supra* Section VII.D. Doing so would have been nothing more than the combination of known elements (Emens's user-configurable

display of titles and abstracts and FusionBot's search results display options) to yield the predictable result of a more customizable system for displaying content to users. A POSITA would have recognized that this combination would be useful both in achieving Emens's stated goal of protecting certain user populations and FusionBot's broader goal of ensuring all users are presented with the most relevant and useful search results displays.

A. Claim 4

- 1. [4a] The method of claim 1, further including said search engine generating data representing results of said query, wherein: said data includes an abstract describing each document of said plurality of documents; and**

133. Emens discloses this element. Emens explains that a "search result page typically contains summary information ... , which may include a title and short abstract or extract," "along with the Uniform Resource Locator (URL), for each retrieved document." Emens, 1:59-61, 4:1-10. A POSITA would have recognized that both the "title" and "short abstract" in Emens teach the claimed "abstract" as that term is used in the '329 patent. A POSITA would have been motivated to combine Emens' teachings of an abstract with Page-FusionBot's teachings to provide a user with a useful and concise summary of the recalled results.

- 2. [4b] for each respective abstract of each document of said certain documents, said abstract excludes terms from the**

respective at least one section not used by said search engine to recall said each document.

134. Emens discloses this element. As explained above, Emens notes that search results typically include summary information such as a title and an abstract or extract. Emens, 1:59-61, 4:1-10. Emens also explains that this summary information is stored, along with other "meta-information," in the search engine's index. *Id.*, 2:61-64, 5:15-20. Emens explains that, when a search engine receives a search query, the search engine "locates entries in [the] index ... that match [the] search query ... to create a search result set." *Id.*, 2:64-67. A POSITA would have understood that, by building search results from summary information that is stored in the index, the search engine would exclude from any titles, abstracts, or extracts in the search results terms that are not included in the index—e.g., terms found only in sections of documents that are not indexed and therefore not used for recall.

135. Emens therefore confirms Page's and FusionBot's disclosures that document titles are included in the search engine's index. And based on the disclosures of each of Page, FusionBot, and Emens, a POSITA would understand that the combination of Page, FusionBot, and Emens would generate abstracts that exclude terms from the no-recall sections of each document.

B. Claim 11

136. Dependent claim 11 is substantively identical to dependent claim 4, except that it recites "[t]he computer-readable storage medium of claim 8" instead of "[t]he method of claim 1." The combination of Page and FusionBot in further view of Emens teaches the limitations of claim 11 for the same reasons discussed above regarding claims 4 and 8.

X. GROUND 3: CLAIMS 6 AND 13 ARE OBVIOUS IN VIEW OF PAGE IN COMBINATION WITH FUSIONBOT AND LI.

137. As I explain below, the combination of Page and FusionBot in further view of Li renders obvious claims 6 and 13.

138. As discussed above in Section VII.E, Li describes methods directed to "searching databases distributed over wide-area networks[,] such as the World Wide Web," that can be organized into non-linear arrays of nodes connected by hyperlinks. Li, Abstract, 1:3-14. Specifically, it presents a "method of ranking documents ... based on the document[s'] relevance to a query" by using vector space models to compare query terms to terms in the anchor text of hyperlinks pointing to each document. *Id.*, 8:17-9:2. In this vein, Li presents a strategy that combats a problem almost identical to the one identified by Page and that, like Page's method, relies on the interconnectedness of databases like the web. *See* Page, 1:18-22.

139. With this knowledge and the understanding that relevance rankings could be improved by incorporating more than one ranking technique into a search engine, *see supra* Section VII.A, a POSITA would have been motivated and found it obvious to combine the ranking techniques Li discloses with the disclosures of a combination of Page and FusionBot to further improve relevance rankings for search engine results. Indeed, this fact is underscored by the fact that Google used both PageRank and what it referred to as "hypertext-matching analysis"—a ranking methodology that resembles Li's method-in its ranking algorithms. Google Technology Overview, 1 ("Google also analyzes the content of neighboring web pages to ensure the results returned are the most relevant to a user's query."). Importantly, Page expressly contemplates the possibility that the ranking method it discloses can be combined with other "factors such as ... objective ranking and textual matching" to further improve its relevance rankings. Page, 8:44-46. Thus, the prior art references themselves provide an express motivation for the combination of Li with the combination of Page and FusionBot. In addition, combining Li with the combination of Page and FusionBot is nothing more than the application of known techniques (i.e., Li's ranking techniques) to known systems and/or methods (i.e., Page's search engine and related ranking methodology) to yield predictable results (i.e., improved search engine ranking methods).

A. Claim 6: The method of claim 1, wherein ranking a plurality of documents recalled by a search engine for a query depends in part on a vector space model associated with each of said documents.

140. Claim 6 recites a conventional method in the search engine art.

POSITAs knew of "vector space models" and implemented them in search engine relevance ranking methodology long before the earliest filing date of the '329 patent. *See, e.g.*, Page, 1:51-67; Li, 3:18-5:11. Notably, the '329 patent provides no details on the vector space model, stating only that "[i]n an algebraic approach, terms inside a section or document can be represented by multidimensional vectors having direction and magnitude, [and] the relevancy can be ascertained by the magnitude of vector deviations." '329 patent, 4:32-36.

141. Page explains that using vector space models to rank documents by relevance was prior art at least as early as Page's filing date-nine years before the '329 patent was filed. Specifically, Page discloses that "variations of a standard vector space model"—which could account for "how recently the document was updated" or "how close the search terms [we]re to the beginning of the document" were known and used by prior art search engines to rank documents. Page, 1:51-67. While Page explains that using vector space model-based methods alone would produce results of "relatively low quality," it also notes that results ranked using such models were "better than [results] with no ranking at all." *Id.*, 1:56-58. Thus, Page explains that this method was well known in the art and that it would have

been obvious to combine the method with Page's ranking protocol to influence a page's ranking based at least in part on "how recently the document was updated" or "how close the search terms [we]re to the beginning of the document." Page, 1:51-58.

142. Li also discloses this feature. Li explains that, at the time the application giving rise to Li was filed, "[a]lmost all ranking techniques of search engines depend[ed] on the frequency of query terms in a given document," and that factors like "document frequency, i.e.,] how many documents contain the term, [could] also be taken into account in determining a relevance score." Li, 3:18-4:1. Li further discloses that, "[o]nce the various factors ... have been determined for a particular query, various models such as the vector space model" could be "used to develop a numerical relevance ranking." *Id.*, 4:1-5. Li then outlines how a vector space model of the kind known in the art could be used to calculate the rank of a document, explaining in detail how the query and documents would be represented as vectors and how the dimensions for the vectors could be weighted and normalized. *Id.*, 4:8-5:11. Moreover, Li discloses "a method of ranking documents ... based on the document' s relevance to a query" using query terms and the anchor text of hyperlinks associated with each document in conjunction with a vector space model. *Id.*, 8:17-9:21; *see id.*, 19:22-22:13. In sum, Li emphasizes that methods of ranking pluralities of documents recalled by search engines for queries

that depended in part on vector space models associated with each of the recalled documents was well known in the art.

B. Claim 13

143. Dependent claim 13 is substantively identical to dependent claim 6, except that it recites "[t]he computer-readable storage medium of claim 8" instead of "[t]he method of claim 1." Page alone or Page (alone or with FusionBot) in view of Li teaches the limitations of claim 13 for the same reasons discussed above regarding claims 6 and 8.

XI. GROUND 4: CLAIMS 7 AND 14 ARE OBVIOUS IN VIEW OF PAGE IN COMBINATION WITH FUSIONBOT AND CUCERZAN.

144. As I explain below, the combination of Page and FusionBot in further view of Cucerzan renders obvious claims 7 and 14.

145. As explained above in Section VII.F, Cucerzan discloses methods for "improving the relevance of search results." Cucerzan, Abstract. Cucerzan observes that the relevance of search engine results was necessarily limited by users' inability to precisely convey their needs and/or intent through search queries, *see* Cucerzan, ¶¶ 3-7, and it notes that, for ambiguous queries (*i.e.*, queries "for which a very large number of web pages contain the queried terms"), measures of "absolute relevance of web pages or web sites (such as page rank) may not be indicative of how relevant those pages or sites are," *id.*, ¶ 7. Consequently, Cucerzan presents techniques for ranking or re-ranking search results by relevance

to a target query, wherein relevance is "judged based on one or more queries in [a query] log that are related to the target query temporally and/or lexically." *Id.*, Abstract; *see also id.*, ¶ 11.

146. As noted previously, a POSITA would have understood that relevance rankings could be improved by incorporating more than one ranking technique into a search engine. *See supra* Section VII.A. This understanding is explicitly confirmed by Cucerzan, which proposes a method for "re-ranking the top search results returned by a search engine," Cucerzan, ¶ 10, using a "ranking component" that "can be configured to rank or re-rank the search results based on the original ranking provided by the search engine ... or based on [a] match score," *id.*, ¶ 37; *see also id.*, ¶¶ 41, 58. And Cucerzan even acknowledges that the initial ranking provided by the search engine may be based on "page rank." *Id.*, ¶ 7. Cucerzan therefore ties directly into Page's express disclosure that the ranking method Page sets forth can be combined with other factors to further improve its relevance rankings, *see* Page, 8:44-46, and actively encourages POSITAs to combine Cucerzan's teachings with other search engine ranking methods like those disclosed in Page.

147. In addition, combining Cucerzan with the combination of Page and FusionBot is nothing more than the application of known techniques (i.e., Cucerzan's ranking techniques) to known systems and/or methods (i.e., Page's

search engine and related ranking methodology) to yield predictable results (i.e., improved search engine ranking methods). A POSITA therefore would have been motivated and found it obvious to combine the ranking techniques of Cucerzan with the disclosures of Page or the combination of Page and FusionBot to further improve relevance rankings for search engine results.

A. Claim 7

148. Cucerzan discloses this feature. Cucerzan sets forth a method for "improving the relevance of search results" using query log information in conjunction with language models. Cucerzan, Abstract, Fig. 3. In the disclosed method, a "system ... for re-ranking search results retrieved by a search engine" "build[s] a language model ... for a target query" and compares that language model "with a language model for every search result," which may be generated based on "one or more properties of the search results" like "URL[s], page title(s), [and] page content (body, text)." *Id.*, ¶¶ 33, 35, 37, 41, 43, 50-51, 58, 62-63, 76, 85. The language models used for comparison "can be, for example, a vector of terms, an associated probability distribution, and/or a probability distribution over query terms extracted." *Id.*, ¶¶ 42, 77. The comparison of language models is used to "determine a match score" that can then be communicated to another part of the system and used to "rank or re-rank the search results," potentially in conjunction

with an "original ranking provided by the search engine." *Id.*, ¶¶ 36-37, 43, 51-52, 63, 77, Claim 1.

B. Claim 14

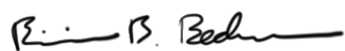
149. Dependent claim 14 is substantively identical to dependent claim 7, except that it recites "[t]he computer-readable storage medium of claim 8" instead of "[t]he method of claim 1." Page in view of FusionBot and Cucerzan teaches the limitations of claim 14 for the same reasons discussed above regarding claims 7 and 8.

XII. CONCLUSION

150. This declaration represents the opinions that I have formed up to date based on the material currently available to me. I reserve the right to revise, supplement, and/or amend my opinions stated in this declaration based on new information that becomes available to me and on my continuing analysis of the materials available to me.

151. I hereby declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct, and that all statements made of my own knowledge are true and correct, and that all statements made on information and belief are believed to be true. I understand that willful false statements and the like are punishable by fine or imprisonment, or both (18 U.S.C. ¶ 1001).

Executed on February 26, 2023, at Windsor, MA.

A handwritten signature in black ink, appearing to read "Ben B. Bederson". The signature is fluid and cursive, with a long horizontal stroke at the end.

Benjamin B. Bederson, Ph.D.