EXHIBIT A

IN THE UNITED STATES DISTRICT COURT

FOR THE DISTRICT OF DELAWARE

PERSONALIZED USER MODEL, L.L.P., Plaintiff,	)	
v.	)	C.A. No. 09-525-LPS
GOOGLE INC., Defendant.	)	**JURY TRIAL DEMANDED**

REPORT OF DEFENDANTS' EXPERT
MICHAEL I. JORDAN, PH.D., CONCERNING INVALIDITY OF

CLAIMS 1, 11, 22, 32, AND 34
OF U.S. PATENT NO. 6,981,040

AND

CLAIMS 1, 3, 5, 6, 7, 21, AND 22
OF U.S. PATENT NO. 7,685,276

TABLE OF CONTENTS

I. INTRODUCTION

1. My name is Michael I. Jordan. I have been retained by Defendant Google Inc. ("Google") to give my expert opinion as to the validity of the patent claims asserted by Personalized User Model, L.L.P. ("PUM") in the above-captioned matter. Below, I set forth the reasons that I believe the asserted patent claims to be invalid.

2. My analysis covers claims 1, 11, 22, 32,[1] and 34 of U.S. Patent No. 6,981,040 (hereinafter "the '040 Patent") and claims 1, 3, 5, 6, 7, 21, and 22 of U.S. Patent No. 7,685,276 (hereinafter "the '276 Patent"). It is my opinion that each of the asserted claims are invalid at least for anticipation and/or obviousness in light of the prior art.

3. I receive $550 per hour for my work. My compensation is not dependent upon the outcome of this case.

4. The matters referenced in this report are based upon my personal knowledge, and if called upon as a witness I could testify completely as to these matters.

II. QUALIFICATIONS

5. I am the Pehong Chen Distinguished Professor in the Department of Electrical Engineering and Computer Science and the Department of Statistics at the University of California, Berkeley. I obtained my PhD in 1985 at the University of California, San Diego, my MA in 1980 at Arizona State University and my BS in 1978 at Louisiana State University. I was a professor at MIT from 1988 to 1998 and joined the University of California, Berkeley in 1998.

6. I am a member of the National Academy of Sciences, a member of the National Academy of Engineering and a member of the American Academy of Arts and Sciences. I was the President of the International Society for Bayesian Analysis in 2011. I am a Fellow of the American Association for the Advancement of Science. I am a Fellow of the Association for

[1] While PUM is not asserting infringement of claim of claim 32 of the '040 patent directly, asserted claim 34 depends from claim 32.

Computing Machinery, the Institute of Mathematical Statistics, the Institute of Electrical and Electronics Engineers, the Association for the Advancement of Artificial Intelligence, the American Statistical Association and the Cognitive Science Society. I received the ACM/AAAI Allen Newell Award in 2009, the SIAM Activity Group on Optimization Prize in 2008 and the IEEE Neural Networks Pioneer Award in 2006. I have been named a Neyman Lecturer and a Medallion Lecturer by the Institute of Mathematical Statistics.

7. I am known internationally for my work on statistical machine learning and Bayesian statistics. I have made seminal contributions to the areas of neural networks, Bayesian nonparametric analysis, probabilistic graphical models, spectral methods, and kernel machines. I am also well known for my work on applying machine learning in the fields of natural language processing, information retrieval, statistical genetics, signal processing, computational biology and computer systems.

8. My citation in the National Academies is for "contributions to the foundations and applications of machine learning." I am the only researcher in machine learning who has been elected to membership in the National Academy of Sciences.

9. I have published over 350 peer-reviewed articles on my research, in publications such as *Science*, *Proceedings of the National Academy of Sciences*, *Nature*, *IEEE Transactions on Information Theory*, *Annals of Statistics*, *Journal of the ACM*, *SIAM Review*, *Journal of the American Statistical Association*, *American Journal of Human Genetics* and *Journal of Machine Learning Research*. Eighty of my articles have been cited more than 100 times in the literature (according to Google Scholar) and six have been cited more than 1000 times. These latter six articles are the following (with the citation counts in brackets):

> Blei, D., Ng, A., and Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993-1022. [2836]

Jacobs, R. A., Jordan, M. I., Nowlan, S., and Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation, 3,* 1--12. [2095]

Jordan, M. I., and Jacobs, R. A. (1994). Hierarchical mixtures of experts and the EM algorithm. *Neural Computation, 6,* 181--214. [1880]

Ng, A., Jordan, M. I., and Weiss, Y. (2002). On Spectral Clustering: Analysis and an algorithm. In T. Dietterich, S. Becker & Z. Ghahramani (Eds.), *Advances in Neural Information Processing (NIPS) 14*, Cambridge, MA: MIT Press. [1856]

Jordan, M. I. (Ed.). (1999). *Learning in Graphical Models*, Cambridge, MA: MIT Press. [1109]

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning, 37(2),* 183--233. [1171]

10. Other recent articles of particular relevance include the following:

Ng, A. Y., Zheng, A. X., and Jordan, M. I. (2001). Stable algorithms for link analysis. *Proceedings of the 24th International Conference on Research and Development in Information Retrieval (SIGIR),* New York, NY: ACM Press.

Blei, D. M., Jordan, M. I. and Ng, A. Y. Hierarchical Bayesian models for applications in information retrieval. (2003). In: J. M. Bernardo, M. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West (Eds.), *Bayesian Statistics 7.*

Blei, D. M. and Jordan, M. I. Modeling annotated data. (2003). *Proceedings of the 26th International Conference on Research and Development in Information Retrieval (SIGIR),* New York: ACM Press.

Barnard, K., Duygulu, P., De Freitas, N., Forsyth, D. A., Blei, D. M., and Jordan, M. I. Matching words and pictures. (2003). *Journal of Machine Learning Research, 3,* 1107-1135.

Teh, Y. W., Jordan, M. I., Beal, M. J. and Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association, 101,* 1566-1581.

Liang, P., Jordan, M. I., and Klein, D. (2009). Learning semantic correspondences with less supervision. *47th Annual Meeting of the Association for Computational Linguistics (ACL)*, Singapore.

Blei, D. M., Griffiths, T. and Jordan, M. I. (2010). The nested Chinese restaurant process and Bayesian inference of topic hierarchies. *Journal of the ACM, 57,* 1-30.

Duchi, J., Mackey, L. and Jordan, M. I. (2010). On the consistency of ranking algorithms. In T. Joachims and J. Fuernkranz (Eds.), *International Conference on Machine Learning (ICML)*, New York: ACM Press.

Liang, P., Jordan, M. I. and Klein, D. (2011). Learning dependency-based compositional semantics. *49th Annual Meeting of the Association for Computational Linguistics (ACL)*, Portland, OR.

11. A full list of my qualifications and experience is contained in my CV, which I attached as Exhibit 2 to this report.

12. I have reviewed extensive materials relating to this case including the asserted patents, the patent histories, the ongoing re-examinations, the claim construction briefs and order, and numerous technical papers and articles discussing the scope and content of the prior art in the timeframe relevant for the asserted patent. In all cases, I have applied the claim constructions propounded by the Court in its Order and Opinion dated January 25, 2012 or constructions agreed by the parties for terms not expressly construed by the Court. The materials relied upon are listed in Exhibit 1.

13. In this report, where I have cited a reference as prior art, either the reference predates the priority date of the Patents or I have been informed by counsel for Defendant that Defendant will be able to prove at trial that the reference is prior art as to the Patents.

14. I may present my opinions in the form of a tutorial or otherwise and reserve the right to respond to any evidence PUM may present concerning the subject matter of this report.

15. It may be necessary for me to supplement this report based on material that subsequently comes to light in this case, and I reserve the right to do so. I may be asked to present demonstrative evidence at trial, and I reserve the right to do so.

16. It may be necessary for me to revise or supplement this report, or submit a supplemental or responsive report, based on any supplemental or responsive report of PUM, and I reserve the right to do so.

III. LEGAL PRINCIPLES

17. As an expert assisting the Court in determining invalidity, I am obliged to follow existing law. I have therefore been asked to apply the following legal principles to my analysis, and I have done so:

a. For a claim to be anticipated, every limitation of the claimed invention must be found in a single prior art reference, either expressly or inherently, arranged as in the claim.

b. When a claim covers several alternative structures or compositions of elements, either generically or as alternatives, the claim is deemed anticipated if any of the structures or compositions within the scope of the claim is disclosed or practiced in a single prior art reference.

c. For a claim element to be inherently present in a prior art reference, the element must be "necessarily present" in the disclosed apparatus, system or method, not merely probably or possibly present.

d. A claim is invalid for obviousness if differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. To be properly applied as an obviousness or anticipation reference, the reference must predate the invention of the subject matter of the claim, unless a statutory bar applies.

e. In determining whether a claimed invention is obvious, one should consider the scope and content of the prior art, the level of ordinary skill in the relevant art, the differences between the claimed invention and the prior art, and whether the claimed invention would have been obvious to one of ordinary skill in the art in light of those differences.

f. If one of ordinary skill in the art can implement a predictable variation prompted by market forces or design incentives, such a variation is obvious. If a technique has been used to improve one device, and one of ordinary skill in the art would recognize that it would improve similar devices in the same way, using the technique is obvious unless its actual application is beyond ordinary skill. Stated differently, the proper question is whether one of ordinary skill, facing the wide range of needs created by developments in the field of endeavor, would have seen a benefit to combining the teachings of the prior art.

g. Where there is a design need or market pressure to solve a problem and there are a finite number of identified, predictable solutions, it is obvious to pursue the known options within the grasp of one of ordinary skill.

h. Contemporaneous development of similar variations of a device or method by other parties is indicative of obviousness.

i. In establishing obviousness, one must avoid the " temptation to read into the prior art the teachings of the invention in issue" and "guard against slipping into the use of hindsight." The prior art itself, and not the applicant's alleged achievement, must establish the obviousness of the combination.

j. I understand that certain objective factors, sometimes known as "secondary considerations" may also be taken into account in determining whether a claimed invention would have been obvious. Such secondary considerations as "commercial success, long felt but unsolved needs, [and] failures of others" may be evidence of non-obviousness. If such factors are present, they must be considered in determining obviousness.

k. The person of ordinary skill is a hypothetical person who is presumed to be aware of all of the pertinent art. The person of ordinary skill is not an automaton, and may be able to fit together the teachings of multiple prior art references employing ordinary creativity and the common sense that familiar items may have obvious uses beyond their primary purposes. It is not necessary to demonstrate precise teachings directed to the specific subject matter of the challenged claim, for a court can take account of the inferences and creative steps that a person of ordinary skill in the art would employ. A patent which merely claims predictable uses of old elements according to their established functions to achieve predicable results may be found invalid as obvious.

l. Art that is analogous to the subject matter of the patent may properly be used as an obviousness reference. I understand that a reference is reasonably pertinent if, even though it may be in a different field from that of the inventor's endeavor, it is one which, because of the matter with which it deals, logically would have commended itself to an inventor's attention in considering his problem.

m. An invention is obvious if one of ordinary skill in the art, faced with the wide range of needs created by developments in the field, would have found it obvious to employ the solution tried by the applicant to meet such needs.

IV. OVERVIEW OF THE ASSERTED PATENTS

18. The patents-in-suit are U.S. Patent Nos. 6,981,040 and 7,685,276, both entitled "Automatic, Personalized Online Information and Product Services." The '040 Patent was filed on June 20, 2000 and issued on December 27, 2005. The '276 Patent was filed on January 8, 2008 and issued on March 23, 2010. There are no substantive differences in the content of the specifications. Both patents claim priority to a provisional application filed on December 28, 1999. Additionally, the '276 Patent claims priority to the '040 Patent.[2]

[2] The '276 patent also claims priority to U.S. Patent No. 7,320,031, which it a continuation of the '040 Patent. I understand that the '031 patent is no longer in this case.

19. I understand that PUM has asserted infringement of claims 1, 11, 22, and 34 of the '040 Patent. Claim 1 is an independent claim, and both claims 11 and 22 depend directly from it. Claim 34 depends from unasserted claim 32, which is an apparatus version of method claim 1.

20. I further understand that PUM has asserted infringement of claims 1, 3, 5, 6, 7, 21, and 22 of the '276 Patent. Claim 1 of the '276 Patent is an independent claim with substantially similar elements as Claim 1 of the '040 Patent, save that it requires monitoring browser use rather than general computer use, and it also requires receiving and processing a search query. Claims 3, 5, 6, 7, 21, and 22 all depend directly from Claim 1.

21. Broadly speaking, the asserted patents describe a method for estimating the probability of a user's interest in a document and using the estimated probability to deliver personalized information to that user. The Abstract[3] reads:

> A method for providing automatic, personalized information services to a computer user includes the following steps: [a] transparently monitoring user interactions with data during normal use of the computer; [b] updating user-specific data files including a set of user-related documents; [c] estimating parameters of a learning machine that define a User Model specific to the user, using the user-specific data files; [d] analyzing a document to identify its properties; [e] estimating the probability that the user is interested in the document by applying the document properties to the parameters of the User Model; and [f] providing personalized services based on the estimated probability. Personalized services include personalized searches that return only documents of interest to the user, personalized crawling for maintaining an index of documents of interest to the user; personalized navigation that recommends interesting documents that are hyperlinked to documents currently being viewed; and personalized news, in which a third party server customized its interaction with the user. The User Model includes continually-updated measures of user interest in words or phrases, web sites, topics, products, and product features. The measures are updated based on both positive examples, such as documents the user bookmarks, and negative examples, such as search results that the user does not follow. Users are clustered into groups of similar users by calculating the

[3] As noted above, the '040 and '276 Patents have substantially similar specifications. Unless otherwise noted, all quotations and citations to the specification in this report refer to the '040 Patent.

distance between User Models.

22. The bracketed letters were added for clarity. As we will see, all asserted claims require the following six elements or slight variations thereof:

a. Transparently monitoring user interactions

b. Updating user-specific data files

c. Estimating parameters of a learning machine that define a User Model specific to the user

d. Analyzing a document to identify its properties

e. Estimating the probability that the user is interested in the document by applying the document properties to the parameters of the User Model

f. Providing personalized services based on the estimated probability

A. <u>The Asserted Patents Generally</u>

23. The asserted patents describe a method for developing a user profile and using that profile to filter documents presented to the user. According to the specification, the amount of information available through the Internet is "staggering" and growing at an exponential rate. (1:22-24.) Since users may be overwhelmed by the sheer amount of data, a variety of techniques have arisen to filter and otherwise manage the set of available information. One such filtering technique is to develop a profile of the user's activities and/or interests and use that profile to personalize the documents that are retrieved or presented to a particular user. (1:24-29.)

24. In the "Background Art" section, the asserted patents describe two types of prior art personalization techniques. The first type, termed "information filtering techniques," create a mathematical representation of each document that had been selected or designated by the user. (*See generally* 1:32 – 2:15.) The mathematical representation often tracks keywords within the documents, *i.e.* words or phrases considered significant like "United States" and "supernova" and unlike "the" and "two." For instance, a CNN.com article about the Philadelphia Eagles would

probably contain words like "Eagles," "Philadelphia," and "Andy Reid," and thus that document's representation will similarly note the occurrence of those words. These mathematical representations—often termed "vectors"—can be used to estimate parameters of a learning machine that serves as a model or profile for the specific user. In one embodiment of the idea, known as the naïve Bayes method, the vectors representing the documents selected or designated by the user are averaged into a single vector that represents the documents selected by the user. Thus a user who selected a lot of Eagles documents will likely have higher scores for words like "Andy Reid" and "Philadelphia."

25. After the machine learning system develops the user representation, it uses vector analysis techniques to analyze prospective documents. A prospective document is converted into a keyword vector using similar techniques as those used for the user documents. That new keyword vector is then compared with the user vector—*e.g.,* in the case of naïve Bayes, the average of all of the user's previously selected documents. Documents with vectors which are "close" to the user vector are accordingly similar to documents that the user selected or designated, and thus potentially of interest to the user. Documents with vectors which are not "close" to the user vector are dissimilar to the user's previous documents, and thus potentially not of interest to the user. (*See generally* 1:32-60.)

26. For example, the user vector of an Eagles fan is likely to contain higher scores for words like "Andy Reid" and "Eagles," because those words are likely to appear frequently in his selected or designated documents. Thus if a prospective document also contains words like "Andy Reid" and "Eagles," its vector is likely to be "close" to the user vector and thus that document is more likely to be recommended to the user. Similarly, if the document contains many words that are not present in the user's vector and are thus not in documents that were

preferred by the user, *e.g.* "Immanuel Kant" or "the Federal Reserve," the document's vector is not likely to be close to the user vector and thus may not be recommended to the user.

27. Column 1, lines 55-60 of the '040 Patent succinctly summarizes the characteristics of the prior art information filtering techniques:

> The user is represented as a vector of the most informative words in a set of user-associated documents. New documents are parsed to obtain a list of the most informative words, and this list is compared to the user's vector to determine the user's interest in the new document.

28. The patents assert that there are several drawbacks to existing information filtering techniques. First, the patents claim that existing systems only personalize the filtration of documents, and do not personalize the initial collection of documents. (1:62-65.) The patents assert this is impractical, as it would require applying the filter to every potential document on the Internet. (1:65 – 2:6.) Second, the patents claim that the user representations are relatively limited, as they typically include only keywords or parameters from a single mode of interaction. (2:6-10.) Finally, the patents assert that existing information filtering systems do a poor job updating the user profile as the user's interests change. (2:10-15.)

29. The patents label the second type of prior art personalization techniques "collaborative filtering methods," which compare user models to each other rather than to documents. (*See generally* 2:16 – 3:21.) As with content-based filtering, collaborative filtering systems track documents selected implicitly or explicitly by the user. (2:23-26.) Rather than analyzing the content of those documents so as to compare previous documents with future documents, collaborative filtering systems compare user profiles to each other and recommend documents based on shared tastes. For example, if a user selected several documents about the Philadelphia Eagles, e.g. http://www.philadelphiaeagles.com/ or http://www.nfl.com/teams/philadelphiaeagles/profile?team=PHI, the system would look for other users that selected those <u>same</u> documents. Once a group of "similar" users has been located, the system determines

whether those other users selected the prospective document, e.g. http://sports.yahoo.com/nfl/ teams/phi. If so, the document is shown to the user. Put more plainly, collaborative filtering systems locate other users with similar tastes as the current user, then recommend documents that those other users liked.

30. For example, the patents list Amazon.com's "Customer's Who Bought" feature as a collaborative filtering system. (2:40:47.) This system is conjectured to keep track of which products are purchased by which customers, then to recommend products to a user that similar users also purchased. A variation of this system still exists on Amazon.com today.

Figure 1: Screen capture from an Amazon.com product page

31. The patents assert that the main drawback of collaborative filtering systems is that they do not consider content. Because recommendations are based on user opinions, a purely collaborative filtering system is unable to analyze a document that has never been seen (and thus was never selected by) any user.[4] Collaborative filtering systems are also putatively unable to determine whether documents are similar to each other, save in that they may have been selected by the same group of users. Accordingly, such systems maintain a record of <u>every</u> potential

[4] Recall that collaborative filtering systems recommend documents that other users liked. If no one has seen a particular document, no one could have liked the document, and the system would be unable to recommend it to anyone.

document—a storage cost that putatively becomes prohibitive as the number of available documents enters the billions and trillions. (3:9-21.)

32. The alleged invention, termed "Personal Web," is meant to address the perceived shortcomings of the aforementioned prior art systems by "provid[ing] automatic, personalized information and product services to a computer network user." (7:4-6.) Personal Web operates in three modes: initialization, updating, and application. (8:54-56.) In the initialization stage, Personal Web analyzes a user's documents and creates a user-specific User Model based on those user documents. (8:56-58.) In the update stage, Personal Web transparently monitors a user's interactions with a computer, updates a set of user-specific data files, and uses those data files to update the parameters of the User Model. (8:59-67.) In the application phase, Personal Web applies the User Model to unseen documents, estimates the probability of the document being interesting to the user, and provides personalized services based on that probability. (9:2-6.)

33. The "updating" and "application" stages of the alleged invention are depicted in Figure 2, which the patents refer to as "a block diagram of a method of the present invention for providing personalized product and information services to a user." (6:8-10.) The figure tracks the six steps present in the Abstract, which are present in every claim:

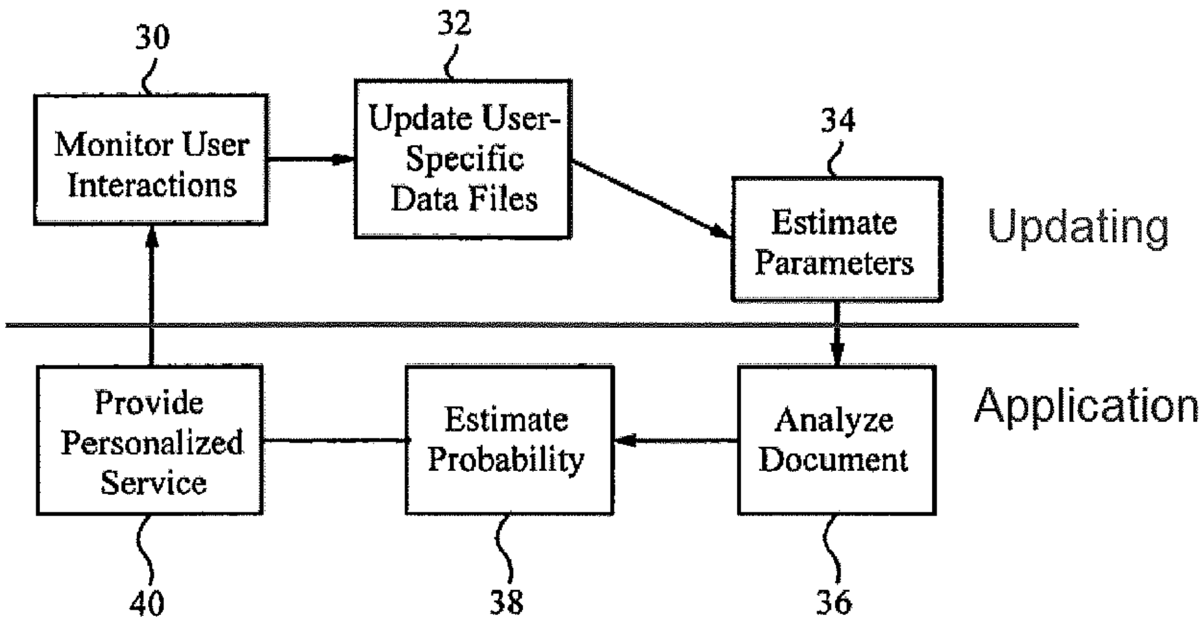

Figure 2: Asserted Patents, Fig. 2 (annotations added)

These steps are described in further detail below.

1. Transparently Monitoring User Interactions

34. The updating phase begins by monitoring user activity. As described in the patents, user interactions may be monitored during multiple distinct modes of user interaction with the network, such as searching, navigation, browsing, email reading, email writing, etc. (4:57-63.) These interactions are transparent, *i.e.* they do not require any additional activity by the user. The patents distinguish this type of monitoring from requiring explicit feedback from the user in the form of ratings. (2:23-28.)

2. Updating User-Specific Data Files

35. The patents define the "user-specific data files" as including both "a set of documents and products associated with the user" as well as "monitored user interactions with data." (8:67 – 9:2.) An exemplary structure for "record[ing] all user interactions with documents" (6:39-40) is depicted in Figure 14:

User Recently Accessed Butter

Document ID	Access Time	Interaction Type	Context	Degree of Interest
www.herring.com/insider	5/12/2000 14:37:21	Navigation	bookmark access	positive, followed 3 links 12 minutes
www.m-w.com	5/12/2000 15:08:21	Search	query "dictionary"	positive, followed 5 links bookmarked, 21 minutes

Figure 3: Asserted Patents, Fig. 14

As the patents explain,

> Through his or her actions, the user creates positive and negative patterns. Positive examples are documents of interest to a user: search results that are visited following a search query, documents saved in the user favorites or bookmarks file, web sites that the user visits independently of search queries, etc. Negative examples are documents that are not of interest to the user, and include search results that are ignored although appear at the top of the search result, deleted bookmarks, and ignored pushed news or email. **Conceptually, positive and negative examples can be viewed as additions to and subtractions from the user data** and resources.

(22:15-26.)

3. <u>Estimating Parameters of a User-Specific User Model</u>

36. Personal Web then updates the parameters of the user-specific User Model based on the identified documents. The User Model is detailed in Figs. 4A-4E of the patent, which depict tables for informative words and phrases, web sites, topics, products, and product features. (*See generally* 10:29 – 14:24.) Figure 4A depicts an embodiment of the word/phrase portion of the User Model, while Figure 15B depicts candidate words and phrases that may later be used to update the User Model:

Informative Word/Phrase List

Word ID	Word Grade	Last Access Time	Number of Accesses
Vegan	0.86	3/6/2000 12:22:41	173
Parasail	0.72	4/15/2000 18:51:27	220

Fig. 4A

Figure 4: Asserted Patents, Fig. 4A

User Word Candidate Table

Word ID	Word Spelling	Word Spelling	Word Grade	Last Access Time
Cytochrome	Cytochrome	Cytocrome	0.67	4/16/200 7:10:01
Hyperbilirubinemia	Hyperbilirubinemia	Hyperbilirubenema	0.58	4/27/2000 12:18:42

Fig. 15B

Figure 5: Asserted Patents, Fig. 15B

37. When updating the User Model, the documents stored in the recently accessed buffer are parsed and their keywords extracted. Those keywords are then added to the user word candidate table depicted in Figure 15B, above. The word candidate table is then used to update the informative word/phrase list shown in Figure 4A through "incremental learning techniques," which changes the Word Grade for existing keyword and adds new keywords as well. (22:64 – 23:9; 23:55-63.) The Word Grade may correspond to the frequency of the word in user documents (11:1-4), the term frequency / inverse document frequency ("TFIDF") of the word (11:12-20), or a measurement of mutual information (11:44 – 12:24).

4. <u>Analyzing Documents</u>

38. The application phase of the invention refers to the process of using the User Model to evaluate documents so as to provide personalized services to the user. Prospective documents are analyzed with the same parsing process employed when updating the User Model. (24:60-64.) That is, the system extracts keywords, location information, products, etc. from the document. *See* Figure 17, below:

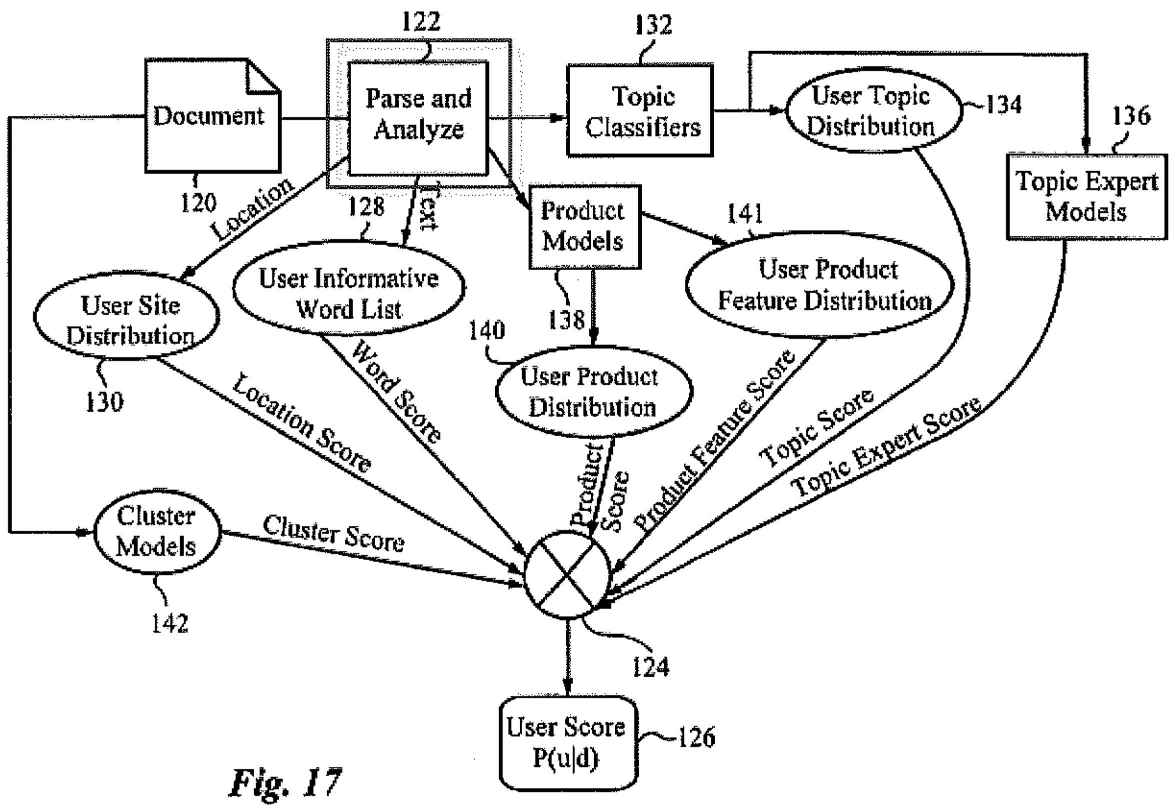

Fig. 17

Figure 6: Asserted Patents, Fig. 17 (notation added)

5. Estimating Probability of User Interest

39. Once the document has been parsed and analyzed, Personal Web applies the user-specific User Model to the extracted features to estimate the probability that the user is interested in that document. The entries within the User Model contains corresponding grades as shown in Figure 4A below:

Informative Word/Phrase List

Word ID	Word Grade	Last Access Time	Number of Accesses
Vegan	0.86	3/6/2000 12:22:41	173
Parasail	0.72	4/15/2000 18:51:27	220

Fig. 4A

Figure 7: Asserted Patents, Fig. 4A (notations added)

40. Personal Web matches the entries (e.g., keywords) in the document with entries in the User Model, then takes the corresponding grades to generate scores for each aspect of the User Model, e.g. a Word Score from the Informative Word/Phrase List, a Location Score from the Web Site Distribution, etc. These separate scores are in turn combined to compute a final probability that the user is interested in the document. (24:65 – 25:3; *see generally* 25:4 – 26:3.) Figure 17 again shows this process:

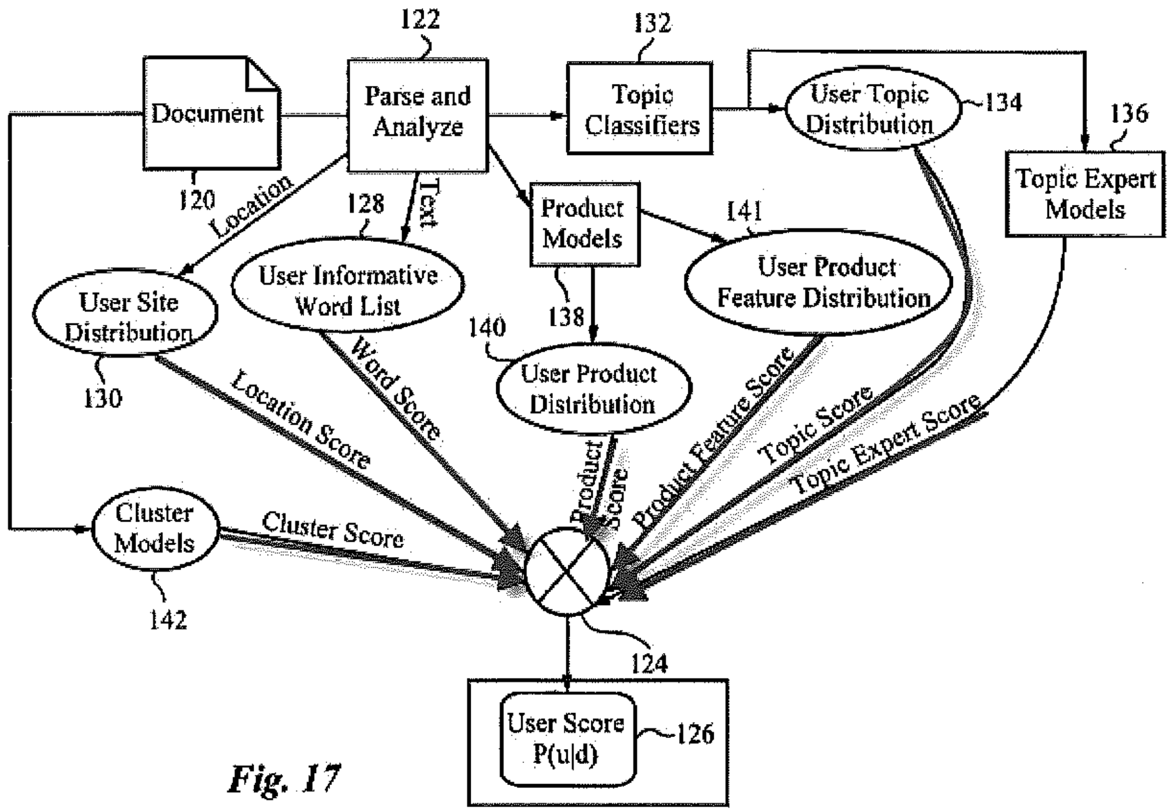

Figure 8: Asserted Patents, Fig. 17 (notations added)

6. <u>Providing Personalized Services</u>

41. Finally, the application phase uses the probability to provide personalized services to the user. The patents list several such services, including Personal Search, Personal Crawler, and Personal Browsing. (*See generally* 27:1 – 32:17.)

B. The Prosecution Histories

1. The '040 Patent

42. The Applicants filed the application that would become the '040 Patent on June 20, 2000. On June 2, 2003, the Examiner rejected all claims in light of U.S. Patent No. 6,006,218 to Breese. Breese discloses a system that transparently monitors user interactions with documents and constructs a user representation based on those interactions. That representation is then used to compute the probability that a user had previously seen a prospective document within a candidate set of results. Documents that are judged less likely to have been seen are given higher priority than documents that are more likely to have been seen, all else being equal.

43. In their September 10, 2003 response, the Applicants distinguished Breese based on its computation of a probability that the document was seen, rather than a probability of a document being interesting to the user: "[P]ositive examples of interest as understood in the current application would be indicators that the user knows of the document and therefore would reduce the score in Breese, et al. Hence the basic premises and goals of Breese, et al. are not consistent with those of the current application." (September 10, 2003 Response, pp. 3-4.) The Applicants also asserted that Breese lacks a User Model as it is allegedly incapable of computing a probability of interest that is *independent* of the user's current information request.

> Column 9, lines 11-15 of Breese, et al. state, "The score can correspond to the entry's estimated, e.g. calculated, probability of relevance to a user's information retrieval request. Alternatively, the score may be an estimate of the value to the user of reviewing the entry." Note that the claims specifically require the estimation of a probability, and that this probability be estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model. The current application on page 11 states, "All of the above features of Personal Web 12 are based on a User Model 13 that represents user interests in a document or product independently of any specific user information need, i.e., not related to a specific query." Hence the "user's information retrieval request" of Breese, et al. from column 9, lines 11-15, is definitely not a User Model, and the "estimate of the value to the user of a reviewing the entry" of Breese, et al. is not a probability and is too vague to be considered a User Model. Hence, this element of claims 1 and 32 is not found in the cited reference.

(*Id.* at 4 (emphasis added).) Regarding the "transparently monitoring" limitation, the Applicants argued that Breese's "obtaining information from questionnaire results" is "not transparently obtained when the user in engages in normal use of a computer." (*Id.* at 5.)

44. On December 3, 2003, the Examiner again rejected all claims as anticipated by Breese. The Examiner rejected the Applicants' seen/interesting distinction regarding the probabilities, stating that "not only does Breese disclose the probability that a document is of interest to a user, it takes into consideration whether the user knows of the document so that the user has new and useful documents and not stale, outdated, not useful information." (December 3, 2003 Office Action, p. 14.) The Examiner similarly rejected the Applicants' argument that Breese lacks a User Model, stating that "Breese clearly discloses a User Model specific to the user. Breese discloses a database that has information about the user and the user's interests." (*Id.*) Regarding transparent monitoring, the Examiner noted that "Breese specifically discloses monitoring a user's actions while accessing Internet sites." (*Id.* at 15.)

45. On December 16, 2003, the Applicants submitted a Request to Withdraw Finality of the Office Action. In it, the applicants argued that a user database alone was not sufficient to meet the "User Model" limitation

> Applicants respectfully submit that the finality was premature inasmuch as there **remain outstanding grounds of rejection of record <u>not</u> clearly developed to such an extent that applicants may readily judge the advisability of an appeal.** For example, independent claim 1 recites "estimating parameters of a learning machine, wherein the parameters define a User Model..." There are three limitations here, "a learning machine," "parameters," and "a User Model." All three limitations, as well as the deterministic relationship among them (i.e., the User Model is defined by the parameters of the learning model) *must* be present in Breese for an anticipatory type of rejection to stand. The cited columns of Breese refer to a database (storage) that has information (stored data) about the user and the user's interests [Office action, page 14, 2nd para.]. It is not clear at all how such a database *anticipates* or is *identical* to the claimed "User Model," which, according to the particular teaching of the present application, is a function defined by a set of parameters of a learning machine [Spec. page 14, 2nd para.; Fig. 3].

(Request to Withdraw Finality of the Office Action, p. 3.)

46. On January 29, 2004, the Examiner withdrew the finality of the claims, finding all claims invalid under Breese in view of U.S. Patent No. 5,754,939 to Herz.[5] The Examiner stated that while Breese met the "transparently monitoring," "updating user-specific files," and "analyzing documents" limitations, it did not meet the "estimating parameters," "estimating probability," or "providing personalized services" limitations. (January 29, 2004 Office Action, pp. 2-3.) However, the Examiner found those limitations present in the Herz reference, which discloses a system for matching "target profiles" with "user target profile interest summaries." (*Id.* at 3-4; *see also* Herz at 4:36-58.)[6]

---

[5] The Action incorrectly identifies the first inventor of the '939 patent as "Hertz."

[6] The Examiner did not include an analysis as to whether Herz disclosed the other elements of the asserted patents. For example, I note that Herz also discloses transparently monitoring user interactions. (Herz at 5:28-30; 7:25-29.)

47. In their March 8, 2004 reply, the Applicants again argued that Breese's probability that a user has seen a document is unlike the probability that a user will find the document interesting. (March 8, 2004 Reply at 3.) The Applicants further argued that Breese did not teach generalization, which the Applicants stated went beyond "keeping score or tracking what happened in the past." (*Id.* at 5.) The Applicants further argued that Herz also did not teach generalization (*Id.* at 5), and that Herz failed to assign probabilities to the ordered articles. *Id.* at 6 (*see also id.* at 9-10):

> Furthermore, *Hertz* (Col. 5, lines 4-21) teaches ordering articles. The question arises what the importance is of the ordered articles. For instance, is it important enough to drag your boss out of a meeting to show the article? *Hertz* does not have a solution for this problem. Ordering articles could be useless if on one day the article is of high importance and the next day is of low importance. This is in contrast to the present invention, which determines for every document an absolute score of importance, e.g. 0.9 probability that a document is of interest to a user, independent what the other documents on today's list were. This aspect is clearly claimed in element 1(e) and 1(f) (vice versa in claim 32) of the present application.

48. On June 4, 2004, the Examiner again rejected all claims in light of *Breese*. As in the January rejection, the Examiner stated that Breese disclosed the "transparently monitoring," "updating user-specific files," and "analyzing documents" steps but not the "estimating parameters," "estimating probability," or "providing personalized services" steps. (June 4, 2004 Office Action at 2-3.) Rather than finding the remaining steps in Herz, the Examiner gave a more detailed account of Breese's probability computation, then argued that one of skill in the art would be motivated to adapt that computation toward the claimed invention:

> According to Breese, if the user already knows the document, it is considered to be of little or no interest. Known documents may be thought of as unwanted or not useful which merely distracts the user from more useful material and/or wastes the user's time. The knowledge probability estimator is used to estimate the probability that the user already knows about various documents. Factors which may be used in generating the knowledge probability are popularity of the item, user's experience in the subject, user's occupation, the amount of time a user has been on the Internet, the overall salience of an item, the amount of time an item has been accessible by the public, or on the server, demographic information about the user. The results are displayed so that the user can review them (Abstract, column 7, lines 59-67, column 8, column 9, lines 1-19, 51-67, column 10, column 16, lines 35-42).

(*Id.* at 3-4.) The Examiner then found that one of skill in the art would have been motivated to compute the probability of the user's interest in a document, given the information that was already stored in Breese's user representation. (*Id.* at 4.)

49. In their September 14, 2004 Reply, Applicants again disputed that Breese renders the claimed invention invalid, arguing that the claims required "generalization" <u>beyond</u> the documents and interactions stored within the user representation:

> Generalization **<u>predicts beyond</u>** items in the past and even beyond the user itself; it **<u>estimates probability</u>** of something to happen in the future. It is exactly this generalization that is claimed in claims 1 and 32 by:
>
> (1) using the monitored actions to **<u>estimate parameters of a learning machine</u>**, and
>
> (2) using the learning machine **<u>to estimate the probability</u>** that a document is of interest to a user.

(September 14, 2004 Reply at 5.)

50. On November 17, 2004, the Examiner maintained its rejection of all claims as invalid in light of Breese. The Examiner again found that the "transparently monitoring,"

"updating user-specific files," and "analyzing documents" limitations were present in Breese. (November 17, 2004 Office Action at 2-3.) In addition, the Examiner found that the "providing personalized services" limitation was present, and implicitly stated that "estimating probabilities" was also present (*Id.* at 3.) While the Examiner agreed that the "learning machine" and "User Model" limitations were not present in Breese, the Examiner pointed out that Breese does disclose "mak[ing] future predictions and estimations on other information the user would find interesting. These predictions and estimations are based on a user's profile, which include information about previous searches/user actions, user's knowledge of information, gender, age." (*Id.*) The Examiner thus concluded that it would be obvious to one of skill in the art to refer to the "user profile" as a learning machine or user model, "since the same functionalities of analyzing the information the user interacts with and profiling the user is achieved." (*Id.*)

51. In their December 28, 2004 Reply, the Applicants amended both independent claims to require that the "estimating a probability" step operate only on unseen documents:

> **e) estimating a probability $P(u|d)$ that ~~the~~ an unseen document d is of interest to the user u, wherein the probability $P(u|d)$ is estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model; and**

(December 28, 2004 Reply at 2.) The Amendment was pursuant to phone interviews that occurred on December 13 and 20, wherein the PTO apparently agreed that Breese was not prior art to the claims and that amending the claims to specify "unseen documents" would render the claims patentable. (*Id.* at 20.) The Applicants again argued that "A personal of average skill in the art clearly understands that the teaching of Breese are merely memorization and not learning. Breese does not teach and not even address the problem of generality and predictability beyond a memory model and can therefore not render the present claims obvious." (*Id.* at 23.)

52. On July 8, 2005, the Examiner rejected most claims (including all asserted claims of the '040 Patent) in light of U.S. Patent No. 5,991,735 to Gerace.[7] Gerace discloses a content delivery system that tracks user interactions with that content and stores them in User Action History Objects. (Gerace at 2:11-13; 6:58 – 7:22.) The system uses the monitored interactions to compute the user's categories of interest, which are stored in a User Interface Object. (*Id.* at 5:12-14; 6:22-31.) These two User Objects (and others) are then used to present targeted content and advertisements to the user. (5:19-25.) The Examiner accordingly found that all six limitations of the independent claims were present in the Gerace reference. (July 8, 2005 Office Action, pp. 3-4.)

53. In their August 8, 2005 Reply, the Applicants traversed the rejection of claims based on Gerace, referencing an interview that occurred on August 3, 2005. Applicants again argued that the prior art taught "memorization" rather than "generalization," noting that Gerace was unable to process unseen documents, as required by the Applicants' earlier amendment:

> **If the AD or document belongs to a category X that is not listed or not part of the set of existing users, then *Gerace's* system has to present this Ad or unseen document to a random set of users until sufficient statistics about the users that like this has emerged. In other words, it is not taught nor is it suggested how the first set of users or the first user are/is presented with an unseen document or an unseen Ad. *Gerace* has no answer to that problem!**

(August 8, 2005 Reply at 4.) All claims were subsequently allowed.

[7] Of note, the Examiner who authored the previous Office Actions was Barbara N. Burgess. The Examiner who authored this rejection and who authored subsequent communications with the Applicants was Bharat N. Barot.

2. The '276 Patent

54. The application that would become the '276 Patent was filed on January 8, 2008—one week before the predecessor patent (U.S. Patent No. 7,320,031) issued, which in turn was filed five days before the '040 Patent issued. In the December 24, 2008 Office Action, the Examiner rejected claims 30-36 but allowed claims 1-29. In their June 19, 2009 Reply, the Applicants canceled rejected claims 30-36, and the remaining claims were issued.

C. The Relevant Claims

55. The relevant claims of the patents-in-suit are reproduced below:

1. The '040 Patent

1. A computer-implemented method for providing automatic, personalized information services to a user u, the method comprising:

> a) transparently monitoring user interactions with data while the user is engaged in normal use of a computer;
>
> b) updating user-specific data files, wherein the user-specific data files comprise the monitored user interactions with the data and a set of documents associated with the user;
>
> c) estimating parameters of a learning machine, wherein the parameters define a User Model specific to the user and wherein the parameters are estimated in part from the user-specific data files;
>
> d) analyzing a document d to identify properties of the document;
>
> e) estimating a probability P(u|d) that an unseen document d is of interest to the user u, wherein the probability P(u|d) is estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model; and
>
> f) using the estimated probability to provide automatic, personalized information services to the user.

11. The method of claim **1** further comprising estimating a posterior probability P(u|d,q) that the document d is of interest to the user u, given a query q submitted by the user.

22. The method of claim **1** wherein the monitored user interactions include a sequence of interaction times.

32. A program storage device accessible by a central computer, tangibly embodying a program of instructions executable by the central computer to perform method steps for providing automatic, personalized information services to a user u, the method steps comprising:

> a) transparently monitoring user interactions with data while the user is engaged in normal use of a client computer in communication with the central computer;
>
> b) updating user-specific data files, wherein the user-specific data files comprise the monitored user interactions with the data and a set of documents associated with the user;
>
> c) estimating parameters of a learning machine, wherein the parameters define a User Model specific to the user and wherein the parameters are estimated in part from the user-specific data files;
>
> d) analyzing a document d to identify properties of the document;
>
> e) estimating a probability P(u|d) that an unseen document d is of interest to the user u, wherein the probability P(u|d) is estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model; and
>
> f) using the estimated probability to provide automatic, personalized information services to the user.

34. The program storage device of claim **32** wherein analyzing the document d provides for the analysis of documents having multiple distinct media types.

2. <u>The '276 Patent</u>

1. A computer-implemented method for providing personalized information services to a user, the method comprising:[8]

> a) transparently monitoring user interactions with data while the user is engaged in normal use of a browser program running on the computer;
>
> b) analyzing the monitored data to determine documents of interest to the user;

[8] The letter designations are added for clarity.

c) estimating parameters of a user-specific learning machine based at least in part on the documents of interest to the user;

d) receiving a search query from the user;

e) retrieving a plurality of documents based on the search query;

f) for each retrieved document of said plurality of retrieved documents: [i] identifying properties of the retrieved document, and [ii] applying the identified properties of the retrieved document to the user-specific learning machine to estimate a probability that the retrieved document is of interest to the user; and

g) using the estimated probabilities for the respective plurality of retrieved documents to present at least a portion of the retrieved documents to the user.

3. The method of claim **1**, wherein transparently monitoring user interactions with data comprises monitoring user interactions with data during multiple different modes of user interaction with network data.

5. The method of claim **1**, further comprising analyzing the monitored data to determine documents not of interest to the user, and wherein estimating parameters of a user-specific learning machine further comprises estimating parameters of a user-specific learning machine based at least in part on the documents not of interest to the user.

6. The method of claim **1**, wherein monitoring user interactions with data for a document comprises monitoring at least one type of data selected from the group consisting of information about the document, whether the user viewed the document, information about the user's interaction with the document, context information, the user's degree of interest in the document, time spent by the user viewing the document, whether the user followed at least one link contained in the document, and a number of links in the document followed by the user.

7. The method of claim **1**, wherein said plurality of retrieved documents correspond to a respective plurality of products.

21. The method of claim **1**, wherein using the estimated probabilities for the respective plurality of retrieved documents to present at least a portion of the retrieved documents to the user comprises presenting to the user at least said portion of the retrieved documents based on the estimated probability that the retrieved document is of interest to the user and the relevance of the retrieved document to the search query.

22. The method of claim **1**, wherein identifying properties of the retrieved document comprises identifying properties selected from the properties consisting of a topic associated with the retrieved document, at least one product feature extracted from the retrieved document, an author of the retrieved document, an

age of the retrieved document, a list of documents linked to the retrieved document, a number of users who have accessed the retrieved document, and a number of users who have saved the retrieved document in a favorite document list.

D. The Court's Claim Constructions

56. On January 25, 2012, the Court issued a *Markman* Order construing several terms in the asserted claims. (*See* Docket Entry 348 ("Markman Order")). The Court's constructions are as follows:

1. "user" and "user u" mean "a person operating a computer or the associated representation of the user."

2. "user-specific data files" means "the monitored user interactions with data and a set of documents associated with the user."

3. "monitored user interactions with the data" means "the collected information about the user's interactions with data."

4. "parameters" means "values or weights."

5. "estimating parameters of a learning machine" means "estimating values or weights of the variables of a learning machine."

6. "learning machine" means a "mathematical function and/or model used to make a prediction, that attempts to improve its predictive ability over time by altering the values/weights given to its variables, depending on a variety of knowledge sources, including monitored user interactions with data and a set of documents associated with the user."

7. "User Model specific to the user" means "an implementation of a learning machine updated in part by data specific to the user."

8. "user-specific learning machine" means "a learning machine [as construed] specific to the user."

9. "document" means "an electronic file including text or any type of media."

10. "estimating" means "approximating or roughly calculating."

11. "probability" means "numerical degree of belief or likelihood."

12. "unseen document" means "document not previously seen by the user."

13. "estimating a probability P(u/d) that an unseen document d is of interest to the user u" means "approximating or roughly calculating a numerical degree of belief or likelihood that an unseen document d is of interest to the user u given the information that is known about the unseen document."

14. "estimating a posterior probability P(u/d,q) that a document d is of interest to the user u given a query q submitted by the user" means "approximating or roughly calculating a numerical degree of belief or likelihood that a document d is of interest to the user u given the information that is known about the document, and given a query q."

15. "present" and "presenting" mean "to provide or make available."

16. "documents of interest to the user" means "documents [i.e., electronic files (including text or any type of media)] for which the user has a positive response."

17. "documents not of interest to the user" means "documents [i.e., electronic files (including text or any type of media)] for which the user has a negative response or has ignored."

18. "user interest information derived from the User Model" means "interests or other information inferred from the User Model."

19. "set" means "group or collection."

20. "set of documents associated with the user" means "group or collection of documents associated with the user."

21. “automatic” means “without human intervention.”

22. “central computer” means “computer on the server side of a client-server relationship.”

E. Characteristics of the Asserted Claims

57. As indicated above, all three independent claims require the same six basic elements:

a. **transparently monitoring user interactions** ['040 Patent, claims 1(a) and 32(a); '276 Patent, claim 1(a)[9]]

b. **updating user-specific data files** ['040 Patent, claims 1(b) and 32(b); '276 Patent, claim 1(b)]

c. **estimating parameters of a learning machine that define a User Model specific to the user** ['040 Patent, claims 1(c) and 32(c); '276 Patent, claim 1(b)]

d. **analyzing a document to identify its properties** ['040 Patent, claims 1(d) and 32(d); '276 Patent, claim 1(f[i])]

e. **estimating the probability that the user is interested in the document by applying the document properties to the parameters of the User Model** ['040 Patent, claims 1(e) and 32(e); '276 Patent, claim 1(f[ii])]

f. **providing personalized services based on the estimated probability** ['040 Patent, claims 1(f) and 32(f); '276 Patent, claim 1(g)]

Those elements are discussed more fully in Section IV.A, above.

58. In addition, independent claim 1 of the '276 Patent further requires that the personalized service involve receiving a search query (claim 1[d]), retrieving documents based on the search query (claim 1[e]), and using the computed probability to present some of those retrieved documents to the user (claim 1[g]). Claim 1 of the ‘276 Patent also requires transparently monitoring the user’s interactions with a browser, rather than with a computer in general as with the claims of the ‘040 Patent (claim 1[a]). Where applicable, I will point out

[9] The claims of the '276 Patent further specify that the monitored user interactions be with a browser.

where the prior art discloses receiving and responding to a search query and monitoring browser activity.

59. The following chart compares the elements of the three relevant independent claims:

Claim 1 of '040 Patent	Claim 32 of '040 Patent	Claim 1 of '276 Patent
1. A computer-implemented method for ***providing automatic, personalized information services to a user*** u, the method comprising:	32. A program storage device accessible by a central computer, tangibly embodying a program of instructions executable by the central computer to perform method steps for ***providing automatic, personalized information services to a user*** u, the method steps comprising:	1. A computer-implemented method for ***providing personalized information services to a user***, the method comprising:
a) ***transparently monitoring user interactions*** with data while the user is engaged in ***normal use of a computer***;	a) ***transparently monitoring user interactions*** with data while the user is engaged in ***normal use of a client computer*** in communication with the central computer;	***transparently monitoring user interactions*** with data while the user is engaged in normal use of ***a <u>browser program</u> running on the computer***;
b) ***updating user-specific data files***, wherein the user-specific data files ***comprise*** the ***<u>monitored user interactions with the data</u>*** and ***a set of documents associated with the user***;	b) ***updating user-specific data files***, wherein the user-specific data files ***comprise*** the ***<u>monitored user interactions with the data</u>*** and ***a set of documents associated with the user***;	analyzing the monitored data to determine ***documents of interest to the user***;
c) ***estimating parameters of a learning machine***, wherein the parameters define a User Model specific to the user and ***wherein the parameters are estimated in part from the user-specific data files***;	c) ***estimating parameters of a learning machine***, wherein the parameters define a User Model specific to the user and ***wherein the parameters are estimated in part from the user-specific data files***;	***estimating parameters of a user-specific learning machine*** based at least in part on the ***documents of interest to the user***;
		<u>receiving a search query</u> from the user;

		retrieving a plurality of documents based on the search query;
d) analyzing a document d to ***identify properties of the document***;	d) analyzing a document d to ***identify properties of the document***;	for each retrieved document of said plurality of retrieved documents: ***identifying properties of the retrieved document***, and
e) ***estimating a probability*** P(u|d) ***that an unseen document d is of interest to the user*** u, wherein the probability P(u|d) is estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model; and	e) ***estimating a probability*** P(u|d) ***that an unseen document d is of interest to the user*** u, wherein the probability P(u|d) is estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model; and	applying the identified properties of the retrieved document to the user-specific learning machine to ***estimate a probability that the retrieved document is of interest to the user***; and
f) using the estimated probability to ***provide automatic, personalized information services*** to the user.	f) using the estimated probability to ***provide automatic, personalized information services*** to the user.	using the estimated probabilities for the respective plurality of retrieved documents to ***present at least a portion of the retrieved documents to the user***.

1. Dependent Claims

60. Where applicable, I will also point out where the prior art contains one or more of the following dependent claim steps:

(a) Estimating a posterior probability P(u|d,q) that the document d is of interest to the user u, given a query q submitted by the user ('040 Patent, Claim 11)

61. Claim 11 requires computing the probability the user will find the document interesting given the query. As the patent explains, "P(u|q,d) is the posterior probability of the event that a document d is of interest to a user u having an information need q." (28:1-3.) Accordingly, claim 11 of the '040 Patent has similar requirements as claim 1 of the '276 Patent, which also requires processing a query. *See* above.

(b) <u>Monitored user interactions include a sequence of interaction times ('040 Patent, Claim 22)</u>

62. Claim 22 requires that the monitored interactions of claim 1[a] include a sequence of interaction times. The specification details several aspects of the user interaction that can be monitored, one of which is "how long the user spent viewing the document." (22:39-40.) This information is gathered while the user browses, presumably as a threshold to determine whether the browsed document should be added to the user documents: "In network browsing, the user browses among linked documents, and each document is added to the buffer, <u>along with its interaction time</u>." (23:29-31.) Interaction times are also used when evaluating expert advice: "In finding expert advice mode, the user's interest in expert advice is recorded; interest may be measured by <u>the interaction time with an email</u> from an expert, a user's direct rating of the quality of information received, or other suitable measure." (23:41-46.)

(c) <u>Analysis of documents having multiple distinct media types ('040 Patent, Claim 34)</u>

63. The Court has construed "document" as "an electronic file including text or any type of media." (Markman Order at 2). The specification indicates that different media types may be analyzed in different ways:

> In step 84, documents 80 are parsed and separated into text, images and other non-text media 88, and formatting. Further processing is applied to the text, such as stemming and tokenization to obtain a set of words and phrases 86, and information extraction… Additional processing is applied to images and other non-text media 88. For example, pattern recognition software determines the content of images, and audio or speech recognition software determines the content of audio.

(17:55 – 18:6.) I have reviewed PUM's infringement contentions and understand that PUM interprets claim 34 of the '040 Patent to require that the document only contain multiple media types, not that those multiple media types actually be analyzed. (Plaintiff's April 2012 Infringement Contentions, Tab B at 11 (asserting that Search Ads meets the limitation because

"Google Search Ads may be displayed as text, images, or video.")) I understand that a patent claim must be interpreted consistently for purposes of infringement and invalidity. Thus, I will adopt PUM's interpretation of this element for purposes of this report and will discuss prior art systems that process documents which can contain more than one media type, such as systems that process web pages.

(d) Monitoring user interactions with data during multiple different modes of user interaction with network data ('276 Patent, Claim 3)

64. The specification lists several modes of user interaction that can be monitored, including "network searching, network navigation, network browsing, email reading, email writing, document writing, viewing pushed information, finding expert advice, product information searching, and product purchasing." (21:66 – 22:4.) These interactions are "transparently monitored while the user is engaged in normal use of his or her computer" in order to update "the set of user documents and the parameters of each user representation in the User Model." (21:63-66; 22:4-7.) Accordingly, claim 3 of the '276 Patent requires that more than one mode of interaction be "transparently monitored."

(e) Determining documents not of interest to the user, and updating User Model based on those documents ('276 Patent, Claim 5)

65. In addition to determining which documents are of interest to the user and updating the User Model accordingly, the specification also describes determining which documents are not of interest to the user and updating the User Model accordingly:

> Through his or her actions, the user creates positive and negative patterns... ***Negative examples are documents that are not of interest to the user***, and include search results that are ignored although appear at the top of the search result, deleted bookmarks, and ignored pushed news or email. Conceptually, positive and negative examples can be viewed as additions to and subtractions from the user data and resources.
>
> ...

> Negative examples provide words, sites, and topics that can be used in several ways. ***The measure of any item obtained from the negative example may be reduced in the user distribution***. For example, if the negative example is from a particular site that is in the user site distribution, then the probability or mutual information of that site is decreased. ***Alternatively, a list of informative negative items may be stored. The negative items are obtained from negative examples and are used to reduce the score of a document containing negative items***.

(22:15-26; 23:10-19.) Accordingly, claim 5 of the '276 Patent asserts that the User Model may store or update information about documents the user did not find interesting, not just information about documents the user found interesting.

(f) Monitoring requires gathering a specific type of information ('276 Patent, Claim 6)

66. Claim 6 of the '276 Patent requires that at least one type of data from the following group be gathered:

(a) information about the document,

(b) whether the user viewed the document,

(c) information about the user's interaction with the document,

(d) context information,

(e) the user's degree of interest in the document,

(f) time spent by the user viewing the document,

(g) whether the user followed at least one link contained in the document, and

(h) a number of links in the document followed by the user.

Note that "time spent by the user viewing the document" is similar to the "interaction times" required by Claim 22 of the '040 Patent.

(g) The retrieved documents correspond to products ('276 Patent, Claim 7)

67. Claim 7 of the '276 Patent asserts that the retrieved documents may correspond to products. I have reviewed PUM's infringement contentions and understand that it interprets

claim 7 of the '276 Patent to assert that the document may mention a product, not that there be any type of special handing for products. (Plaintiff's April 2012 Infringement Contentions, Tab B at 21 (asserting that advertisements—which may or may not correspond to products—meet this limitation).) Again, I understand that a patent claim must be interpreted consistently for purposes of infringement and invalidity. Thus, I will adopt PUM's interpretation of this element for purposes of this report and will discuss prior art systems that can process documents which happen to mention products, such as systems that process web pages or advertisements.

(h) Presentation is based on both probability of interest to user and responsiveness to search query ('276 Patent, Claim 21)

68. While Claim 1 of the '276 Patent requires using the search query to retrieve the initial set of candidate documents, it does not specify using that query when determining whether to present a document from the set of candidate documents to the user. Claim 21 adds this limitation. Claim 21 of the '276 Patent is therefore similar to Claim 11 of the '040 Patent, which requires computing a posterior probability P(u|d,q) that the document d is of interest to the user u, given a query q submitted by the user. (*See* above; *see also* 28:1-16.)

(i) Identifying properties requires identifying a specific type of information ('276 Patent, Claim 22)

69. Much as Claim 6 of the '276 Patent required that certain information be considered when monitoring user interactions, Claim 22 requires that certain information be identified from candidate documents retrieved in response to a search query:

(a) a topic associated with the retrieved document,

(b) at least one product feature extracted from the retrieved document,

(c) an author of the retrieved document,

(d) an age of the retrieved document,

(e) a list of documents linked to the retrieved document,

(f) a number of users who have accessed the retrieved document, and

(g) a number of users who have saved the retrieved document in a favorite document list.

V. THE SCOPE AND CONTENT OF THE PRIOR ART

A. The Prior Art Generally

70. The field of machine learning began in the early 1980's, with roots in mathematical and engineering fields such as statistics, information theory, signal processing, approximation theory, pattern recognition and control theory. The general idea of machine learning at its outset was that computers should be able to learn from their experience, so that it would not be necessary for a programmer to build in all of the possible actions of the computer in advance. The computer should be able to adapt to its particular surroundings, including the network to which it is attached, the users with whom it interacts, the data that it is provided with initially and the data that arises during its operation.

71. The main conceptual problem that machine learning has aimed to attack from its inception is that of "generalization." An algorithm that can generalize is one that can give a reasonable response when presented with objects or situations that it has not seen before. This differs from classical computer systems such as databases or information retrieval systems, where the goal is efficient storage and retrieval of given data. For example, we might consider a database containing background information on people together with some score that measures their credit worthiness. A classical database would store such data in a way such that it would quickly retrieve the credit score of a given individual in the database. But if no credit score were provided for that individual the database system would not be able to return an answer. A machine learning algorithm, on the other hand, might aim to predict the credit score of the individual from the credit scores of other similar individuals.

72. Machine learning algorithms can also be applied to problems involving the activities of a single individual. An example is spam filtering, where a machine learning system is presented with a set of documents and a corresponding set of ratings of "spam" or "not-spam" by a single individual. The goal of such a system is to make rating predictions for new documents not yet seen by the individual. Examples of such systems include Sahami et al. (1998) and Cohen (1996).

73. By the late 1980's there existed many algorithms for prediction problems of this kind. Examples include decision trees, neural networks, nearest neighbor methods, naive Bayes models, locally linear regression and discriminant analysis. All of these algorithms had their roots in the fields of statistics and pattern recognition, but were explored in great depth during the 1980's by machine learning researchers. By the mid-1990's these methods had been joined by a host of other methods, including the support vector machine, boosting, Bayesian networks, wavelets and Gaussian processes. Textbook treatments of machine learning methods that were available in the 1990's include Bishop (1996), Cherkassky and Mulier (1998), and Mitchell (1997).

74. Machine learning algorithms for prediction are generally based on a computation that brings together two kinds of information. The first is a representation of an object in the given problem domain. Returning to the description of individuals and the prediction of credit score, we might represent individuals using a long list of features such as hair color, height, number of years of education, number of visits to England, etc. The second is a list of numerical values, or "parameters" as the term was construed by the Court. Combining a parameter value with the corresponding feature value contributes toward the prediction of the credit score of that individual.

75. Thus machine learning algorithms for prediction combine "features" and "parameters." Features are often known in advance; they are given as input to a machine learning system. Parameters are not known in advance; they are internal to the machine learning system and are adjusted based on "training data." For prediction problems, the training data consists in a set of objects for which the rating is known. Thus the training data for the credit rating problem would consist in a set of individuals, the list of feature values for each individual and the known credit rating for each individual. Based on the training data, a learning algorithm adjusts the parameters so that the learning system makes good predictions on the training set. Put another way, the learning algorithm tries to "learn" what factors in the training data tended to correspond with high credit scores and which did not. Thus, if it is true in the training data that tall people have high credit scores, then the learning algorithm will increase the parameter value corresponding to height.

76. It is important to appreciate that different training sets generally yield different parameters and thus different predictions. For example, in the spam filtering problem, if the training data are the emails of Sue (where for each email she has labeled the email as "spam" or "not-spam"), then the parameters and the subsequent predictions will be different from the parameters and predictions obtained if the training data had come from Bob. This allows a spam filter to be personalized to Sue's tastes simply by being trained on Sue's emails; similarly, a spam filter trained on Bob's emails will be personalized to Bob's tastes.

77. Alternatively, it is also possible to lump together training data and make predictions that are appropriate for a group of people. For example, we can combine the training data for all individuals living in Portland, and obtain a Portland-specific spam filter.

78. The key point is that machine learning algorithms directly yield a notion of "personalization" by their very nature. Personalized predictions are obtained simply by

restricting the training data to a single person. Group-specific predictions are obtained by restricting the training data to a specific group. A generic learning machine is able to predictions at any level of specificity, from entirely personal to entirely general, simply by varying the source of the training data.

79. These issues were well appreciated in the machine learning field by the mid-1990's. Moreover, it was also well understood how to apply these ideas to problems in the categorization of text documents and multi-media documents. As we have discussed, machine learning systems are based on featural representations of objects in a domain, and in the case of documents, the field that studies such representations is called *information retrieval* (IR). By the mid-1990's, machine learning researchers had begun to make use of representations provided by IR researchers, specifically the standard "TF-IDF" representation (Salton & McGill, 1983), in which features are obtained by counting the occurrences of words within a document and across a corpus. Given such a representation, machine learning researchers built systems that combined TF-IDF representations (and other similar representations for multi-media content) with parameter vectors and training data, obtained machine learning systems that could make predictions of how a given individual might rate a given document. See, e.g., Yang (1999) for a summary of some of the results in this line of research.

80. Prediction is only one of the problems that are addressed in the field of machine learning. Two other notable examples are "feature selection" and "clustering." Algorithms to solve both of these problems were known by the middle of the 1980's and have continued to be developed in the ensuing decades.

81. The "feature selection" problem extends the machine learning paradigm as we have discussed it thus far to ask the learning algorithm to find useful ways to describe entities (e.g., objects, people, situations). Consider that we can find many hundreds of thousands of ways

to describe a person (hair color, height, number of years of education, number of visits to England, etc.), but we generally want to store and manipulate a subset of these features. We wish to find those features that are the most useful. The key is the definition of "useful" and many different definitions have been proposed and studied. One definition involves variability: if a feature does not vary across the population then it's presumed to not be useful; if instead a feature is highly variable it's presumed to be useful. Another definition ties feature selection to a prediction problem: features that are highly predictive of a response to be predicted (e.g., "spam" or "not-spam") are retained. Note that this algorithm does not actually make the prediction of the rating; it merely identifies which features are likely to be helpful in making that prediction. By the mid-1990's, a number of algorithms for feature selection had been proposed, many of which had their roots in information theory. Examples include the mutual information score, the Kullback-Leibler divergence and sufficient dimension reduction. Other examples include principal component analysis and canonical correlation.

82. The "clustering" problem involves using a learning algorithm to partition or segment a set of entities (e.g., objects, people, situations) into groups. This generally involves identifying some notion of "similarity" that assesses the extent to which two entities are related to each other. For example, if the features describing a person are binary ("lives in California" vs. "doesn't live in California"; "loves cats" vs. "doesn't love cats"; etc), then a similarity measure can simply count the number of features where two people agree. By the mid-1990's a wide variety of clustering algorithms had been proposed and studied, including hierarchical clustering, K-means clustering and spectral clustering.

83. In all of these machine learning problems and their algorithmic solutions there is a notion of "generalization" present. We have already alluded to this in the case of prediction problems, where an algorithm that can generalize is one that makes a prediction for an entity it

has not seen before. For feature selection, algorithms that can generalize are ones that can select features that are likely to be useful for data that has not yet been seen. For clustering, algorithms that can generalize are ones that can form groups such that future entities are likely to be close to entities already observed.

84. By the mid-1990's there had been many papers written on these topics. These papers explored aspects of the theory, practice and implementation of machine learning. Software systems using machine learning were developed and applications in technology and science were described.

85. It is important to appreciate that "generalization" isn't magic. To give a reasonable response when presented with objects that have not yet been seen, those objects must be similar in some sense to objects that have already been seen. Returning to the description of individuals and the prediction of credit score, where we represent individuals using a long list of features such as hair color, height, number of years of education, number of visits to England, etc., then we would hope to find that some of these features are predictive, either individually or in combination, of the credit score. Suppose that, based on the training data, we find that tall people have high credit scores. A new person, Mary Smith, arrives and, assuming that she is tall, we are able to predict that she will have a high credit score. As we have discussed, this is generally done via a model that contains a list of numerical values, or "parameters," one for each feature (or combination of features). For example, if tall people have high credit scores, then the parameter corresponding to height will be a large positive number, which, when combined with other parameters, will tend to increase the prediction of a high credit score.

86. If instead of using features to represent individuals, we had used social security numbers, then we would likely not be able to give a reasonable prediction of credit score for a new individual. Individuals with similar social security numbers presumably do not tend to have

similar credit scores. In general it is essential in machine learning to choose features that capture similarities that are useful for the prediction at hand. Prediction and feature selection go hand in hand. This was clearly appreciated by the mid-1990's.

87. It is also essential in machine learning to have sufficient data on which generalization can repose. If there is not sufficient data on a particular object or individual of interest, then it is common to cluster objects or individuals together so that there is sufficient data at the level of clusters. Thus prediction and clustering go hand in hand. This was also clearly appreciated by the mid-1990's.

88. Finally, I note that the notion of "personalization" has been part of the fabric of machine learning and statistics since the outset, long before the advent of the World Wide Web. In particular, by the mid-1990's there had been many developments in the area of modeling human choice behavior based on observed data, with applications in fields such as econometrics and psychology. There was also a great deal of work in the subfields of biometrics and risk analysis where the goal is to model the survival times and other responses associated with individuals. All of these developments provided an intellectual platform on which researchers approaching search and information retrieval problems in the 1990's were able to build. That is, with the advent of the World Wide Web, it was natural for researchers to make use of prior art in machine learning and statistics to develop personalization strategies for users of the World Wide Web, and indeed, as we will discuss in the following section, there were many researchers who did exactly that.

B. <u>Exemplary Prior Art References</u>

1. <u>U.S. Patent No. 6,182,068 to Culliss</u>

89. U.S. Patent No. 6,182,068 to Culliss, entitled "Personalized Search Methods," was filed on March 1, 1999 and issued on January 20, 2001. I understand that Culliss is

accordingly a prior art patent with respect to the asserted patents, which claim priority to a December 1999 provisional application.[10] As detailed in the specification, Culliss functions similarly to a traditional search engine in that it accepts a search query from a user and displays squibs of articles ranked by their comparison scores. (Culliss 2:39-42.) Those scores can be altered by additional factors that are specific to individual users, such as whether they were displayed to a user, whether they were selected by a user, how much time the user spent with the article, etc. (*Id.* at 2:43-46.)

90. Culliss describes transparently monitoring a user's actions to derive a user profile (*e.g.,* Culliss at 3:46-56) and using that profile to predict whether a web page is interesting or not to that particular user. (*e.g., id.* at 10:53-67.) The user profile generation phase consists of three stages: the observation phase, wherein the user is monitored and his actions on the documents and the documents themselves are recorded (*id.* at 5:49-59); the feature extraction phase, wherein features are extracted from the actions to form training examples (*id.* at 3:46-56); and the generation of a user profile by a machine learning algorithm (*id.* at 3:57-65). The user profiles are in turn used to classify new web pages as interesting or not, and search results are accordingly narrowed. (*id.* at 10:53-67.) This provides a personalized web experience to the user. For example, users who are identified with an interest in "shoes" will receive different results than users identified with an interest in "sports."

2. **Experience with Learning Agents which Manage Internet-Based Information (Edwards)**

91. Edwards, which was published in 1996, describes a machine learning system that learns a user's profile and uses that profile to predict whether a web page is interesting to that particular user. I understand that Edwards is accordingly a prior art publication with respect to

[10] I reserve the right to supplement my report should PUM's expert seek to opine on the conception or reduction to practice of the asserted patents.

the asserted patents, which claim priority to a December 1999 provisional application. User profile generation phase consists of three stages. First, the user is monitored and the "observations collected through the modified Web browser consist of HTML documents visited by the user and actions performed on these documents." (Edwards at 35.) Next, "[f]eatures are extracted from these observations, and used to create a training instance. (*Id.* at 33.) Finally, "[t]he training instances are then used to induce the user profile." (*Id.*)

92. The user profiles are used to compute confidence rates that classify a new web page as interesting to that user or not, with "more interesting" links highlighted. For each incoming document, "[f]eatures are extracted from each document, and the user profile employed to generate a classification (with an associated confidence rating). The confidence rating is used by the Prediction Stage to determine whether a prediction should be made." (Edwards at 33.) "The rating given to each link is used to order the search, with the highest scoring links being explored first. Pages are analyzed in a similar manner, using the page profile. Those classified as interesting are given a rating, and the highest rated pages are presented to the user." (*Id.* at 35.)

3. <u>Towards Practical Interface Agents which Manage Internet-Based Information (Green)</u>

93. Green, published in 1995, is based on the same underlying system as Edwards. I understand that Green is accordingly a prior art publication with respect to the asserted patents, which claim priority to a December 1999 provisional application. Here, several algorithms are compared and their performances analyzed for accuracy, training and prediction times, CPU and storage requirements. Green further focuses on user interface issues such as intrusions and disruptions. More specifically, Green considers the relative amount of feedback required from a user balanced against the notion of interfering as little as possible with the user's experience.

4. **WebWatcher and related articles**

94. WebWatcher is a system that suggests appropriate hyperlinks by following users' browsing actions and by analyzing Web page properties. The system is described in three papers: *WebWatcher: A Tour Guide for the World Wide Web*, by Joachims, Freitag, and Mitchell (1997) ("Joachims"); *WebWatcher: Machine Learning and Hypertext*, by Joachims, Freitag, Mitchell and Armstrong (1995) ("Joachims2"); and *Web Watcher: A Learning Apprentice for the World Wide Web*, by Armstrong, Freitag, Joachims, and Mitchell (1995) ("Armstrong"). I understand that the papers are accordingly prior art publications with respect to the asserted patents, which claim priority to a December 1999 provisional application. WebWatcher was "in operation from August, 1995 to February, 1997." (Joachims at 2.) I further understand that WebWatcher itself is likewise a prior use with respect to the asserted patents, as it was in use more than a year before December 1999.

95. Users can begin by entering a phrase that describes their current interest, such as "intelligent agents." (*Id.*) WebWatcher henceforth accompanies users as they navigate web pages. "Each time the user selects a hyperlink, WebWatcher accompanies the user to the next page, and logs this hyperlink selection as a training example for learning to improve future advice." (*Id.*) That advice comes in the form of highlighting selected hyperlinks by inserting "eyeball" icons around the link, as shown in the Figure below:

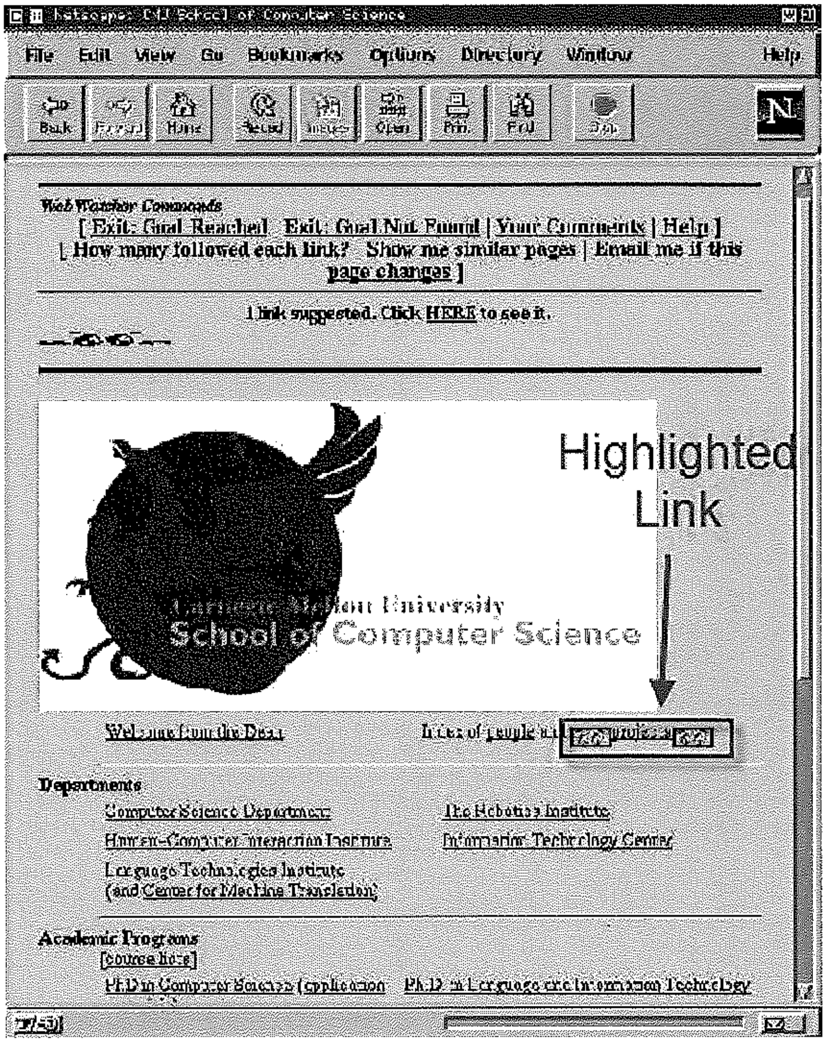

Figure 9: Joachims at 2 (notations in red added)

96. WebWatcher's predictions are based on an estimated "link quality," which takes into account the user's query, the current Web page, and target Web page. (Joachims at 3). More specifically,

$$LinkQuality : Page \times Interest \times Link \rightarrow [0, 1]$$

where "[t]he value of *LinkQuality* is interpreted as the probability that a user will select *Link* given the current *Page* and *Interest*." (*Id.*) "Interests and hyperlink descriptions are represented by very high-dimensional feature vectors," where the "elements of a vector are calculated using the TFIDF heuristic." (*Id.*) "[V]ector representation similarity is calculated as the cosine between vectors." (*Id.*) WebWatcher also analyzes and compares the title text of the corresponding web pages. (*Id.* at 4.)

97. Links that WebWatcher predicts will have "high quality" are highlighted. While the system learns to specialize to a specific Web domain, the authors also present the concept of a *personalized* WebWatcher; *i.e.*, a system that learns to specialize to a particular user. (Joachims at 6.) This concept was subsequently developed in detail by Mladenic as described below.

5. <u>Personal WebWatcher and Mladenic articles</u>

98. The Personal WebWatcher system is described in *Personal WebWatcher: design and implementation*, by Dunja Mladenic, Technical Report IJS-DP-7472, Department of Intelligent Systems, J. Sefan Institute, Slovenia (1996) ("Mladenic"). The same system is also described in *Machine learning for better Web browsing*, Proceedings of the Eight Electrotechnical and Computer Sc. Conference ERK'99, Ljubljana, Slovenia: IEEE section (1999) ("Mladenic2"). I understand that at least Mladenic is a prior art publication with respect to the asserted patents, which claim priority to a December 1999 provisional application. Furthermore, Personal WebWatcher was in public use by at least the date Mladenic was published.[11] I further understand that PWW is a prior use with respect to the asserted patents, as it was in use more than a year before December 1999.

99. As suggested by the name, Personal WebWatcher was based on the WebWatcher system described above. Rather than aggregating information for all users that visit a particular web site, Personal WebWatcher follows the activities of a particular user across all web sites.

[11] Note that the Mladenic paper as published contained Personal WebWatcher's source code. (Mladenic at 2.)

That is, "[u]nlike WebWatcher, Personal WebWatcher (PWW) is structured to specialize for a particular user, modeling his/her interests." (Mladenic at 3).

100. Like WebWatcher, Personal WebWatcher observes the user's activity by acting as a proxy or intermediary between the user and the Web. (Mladenic at 3 ("[Personal WebWatcher] 'watches over the user's shoulder' the similar way [sic] WebWatcher does, but avoids involving the user in its learning process"); *see also* Mladenic at 7-8, Figure 2.)

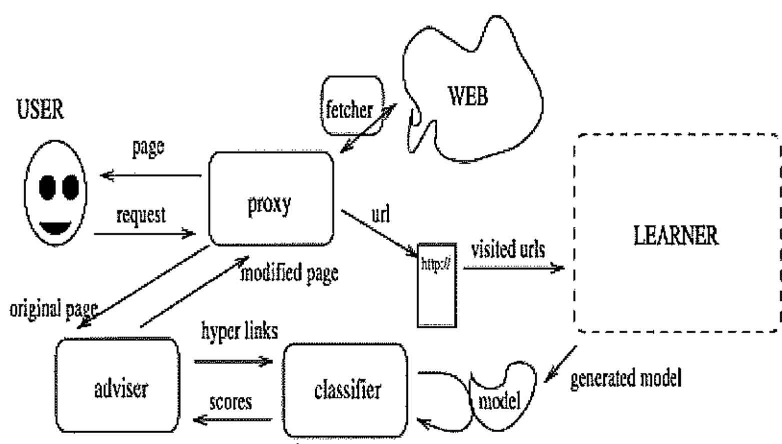

Figure 2: Structure of Personal WebWatcher. The learning part is described separately

Figure 10: Mladenic at 8, Fig. 2

Accordingly, users need not provide explicit feedback to make use of Personal WebWatcher—the system learns behavior simply by observing the user's actions.

101. Mladenic describes the operation of Personal WebWatcher thusly:

> Proxy waits in an infinite loop for a page request from the browser. On request, it fetches the requested document and; [sic] if it is an HTML-document adds advice and; [sic] forwards the document to the user. To add advice proxy forwards the page to adviser, that extracts hyperlinks from document [sic] and calls external code for classification that users generated user-model. A limited number of hyperlinks that are scored above some threshold are recommended to the user, indicating their scores using graphical symbols placed around each advised hyperlink.

(Mladenic at 7-8).