content of the pages and the user's preferences." (Wasfi at 60.) "By making content-based filtering, we can deal with pages unseen by others." (*Id.*)

227. Furthermore, because ProfBuilder relies on textual and conceptual analysis of the document in question, it is able to estimate the probability of unseen documents as well as seen documents. At least at some point, ProfBuilder likely analyzed a document that the current user had not previously seen. Under Plaintiff's own infringement allegations, the *likelihood* that an unseen document was presented is sufficient to meet this limitation. (Plaintiff's April 2012 Infringement Contentions, Tab A, p. 18: "Due to the extremely large number candidate search results in Google's indexes, and due to the frequent changes or updates made to the candidate search results, it is likely that candidate search results have not been previously seen by the user.") (emphasis added). Moreover, a content-based recommendation system that does not exclude unseen documents has the inherent capability to recommend unseen documents.

(f) Providing automatic, personalized information services to the user.

228. Wasfi/ProfBuilder provides personalized information services, specifically personalized links based on the user's preferences. "The result frame displays a list of the recommended pages represented by their respective titles.... ProfBuilder highlights each recommendation to show its relevance and access frequency (given the user's current path) by putting 'ball' and 'man' icons, respectively, in front of the title.... Titles in 'bold' font indicate unread pages, while titles in 'normal' font indicate pages have been read by the user." (Wasfi at 61.)

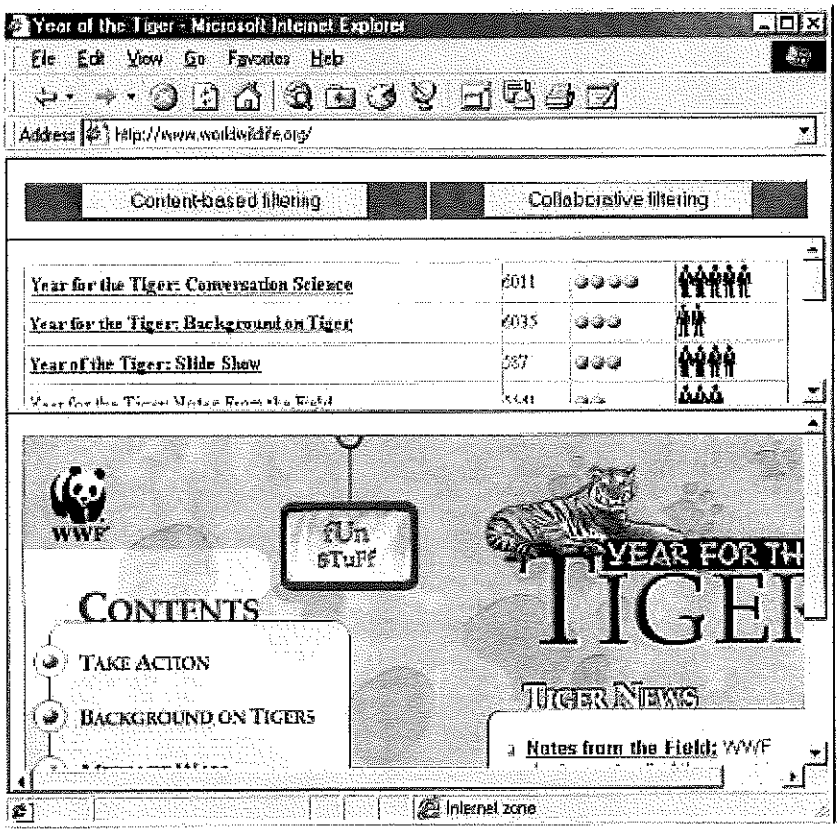


Figure 2 ProfBuilder's interface.

Figure 22: Wasfi at 62, Fig. 2

3. Wasfi and ProfBuilder Anticipate Claim 22 of the '040 Patent

229. Claim 22 requires "The method of claim 1 wherein the monitored user interactions include a sequence of interaction times." Wasfi discloses that ProfBuilder monitors interaction times: "a timeout mechanism is used to delete user's session information after a predetermined amount of idle time. So that, a connection after the specified period having the same IP is identified as a new user." (Wasfi at 60.)

4. Wasfi and ProfBuilder Anticipates Claim 32 of the '040 Patent

230. Claim 32 of the '040 patent requires "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," followed by the same method steps required in Claim 1 of the '040 patent. As detailed

above, Wasfi/ProfBuilder discloses all method steps required in claim 1 and thus it discloses all method steps required in claim 32.

231. Wasfi further discloses that "ProfBuilder inhabits a Web site," i.e., a central computer. (Wasfi at 58, 60.) Further, ProfBuilder is a "program of instructions" as it is written using the Java programming language. (*Id.* at 60.) Accordingly, Wasfi/ProfBuilder meets all limitations of claim 32 of the '040 Patent.

5. Wasfi and ProfBuilder Anticipate Claim 34 of the '040 Patent

232. Claim 34 requires "The program storage device of claim 32 wherein analyzing the document d provides for the analysis of documents having multiple distinct media types." As discussed in Section VI.A.5 above, analyzing web pages meets this limitation, as web pages contain both text and images. Further, ProfBuilder parses webpages that were written in HTML, which can contain multiple media types. (Wasfi at 61.) By processing HTML documents, Wasfi/ProfBuilder "provides for the analysis of documents having multiple distinct media types."

E. The Culliss Patent Anticipates Claims 1, 11, 22, 32, and 34 of the '040 Patent and Claims 1, 3, 6, 7, 21, and 22 of the '276 Patent under PUM's Infringement Theories.[18]

1. Background on Culliss

233. Culliss describes a search engine wherein the results of the search are personalized based on the characteristics of a user:

> [T]he invention can accept a search query from a user and a search engine will identify matched articles and display squibs of the matched articles in accordance with their comparison scores. Articles can have their key term scores or key term total scores altered according to whether they were displayed to a user, whether they were selected by a user, how much time the user spent with the article, etc.

[18] E.g., under PUM's theory that any score is the "probability" required by the claims. See below.

(Culliss at 2:39-46.) Articles can be "documents, files, databases, text collections, audio clips, video clips and samples of any other type of information." (*Id.* at 2:20-24.) Furthermore, the scores can be represented as "key term score, key term total score, key term probability score, comparison score, or other ranking score." (*Id.* at 2:48-51.) Culliss employs both content-based and collaborative filtering in presenting search results to a user:

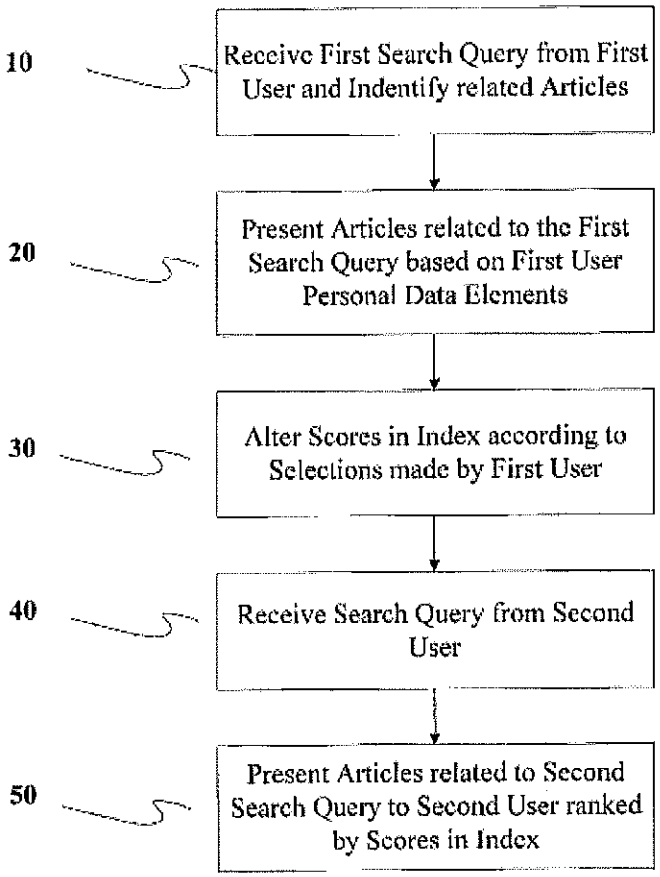


Figure 23: Culliss, Fig. 1

2. Culliss Anticipates Claim 1 of the '040 Patent under PUM's Infringement Theories.

234. Culliss discloses a computer-implemented method for providing automatic, personalized services to a user. "[T]he invention can accept a search query from a user and a search engine will identify matched articles and display squibs of the matched articles....

Articles can have their key term scores or key term total scores altered according to whether they were displayed to a user, whether they were selected by a user, how much time the user spent with the article, etc." (Culliss at 2:39-46.)

(a) Transparently monitoring user interactions with data while the user is engaged in normal use of a computer

235. Culliss transparently monitors user interactions during normal use of a computer: "Users can explicitly specify their own personal data, ***or it can be inferred from a history of their search requests or article viewing habits***. In this respect, certain key words or terms, such as those relating to sports (i.e. "football" and "soccer"), can be detected within search requests and used to classify the user as someone interested in sports. Also, certain known articles or URLs can be detected in a users searching or browsing habits, such as those relating to CNNfn (www.cnnfn.com) or Quote.com (www.quote.com), and also used to classify the user as someone interested in finance." (Culliss at 3:46-56.) Since Culliss identifies user characteristics based on the user's normal searching and browsing history, it meets this limitation.

(b) Updating user-specific data-files comprising the monitored user interactions and a set of documents associated with a user.

236. Culliss discloses updating user-specific data files: "***A cumulative score*** can be kept with regard to these occurrences of ***certain classified key terms, queries or visited URLs*** to quantify how strongly someone is associated with a particular item of personal data. The score can be normalized over time, frequency or other activity such as the number of searches performed, the amount of time spent online, the amount of time spent browsing on a particular subject, the number of URLs or articles selected for a particular subject, or otherwise." (3:57-65; *see generally* 3:66 – 5:43 (tracking example queries and URLs, and storing that information for the user)). Since the scores are ***cumulative***, Culliss tracks and updates user-specific data files that correspond to the user's queries and visited web sites.

237. Culliss further gathers various personal data from the user, including "data about past actions of the user, such as reading habits, viewing habits, searching habits, previous articles displayed or selected, previous search requests entered, previous or current site visits, previous key terms utilized within previous search requests, and time or date of any previous activity." (Culliss at 3:29-35.)

238. Culliss also updates "a set of documents associated with a user" under PUM's reading of the claims. As detailed in its infringement allegations, PUM interprets this claim element to require "updating ***information*** about the documents that the user has clicked on" rather than updating the document itself. (Plaintiff's April 2012 Infringement Contentions, Tab A, p. 6.) Culliss similarly updates information about documents, specifically the URL of the document. (Culliss at 3:46 – 4:54.) Accordingly, Culliss meets both aspects of the "user-specific data files" under PUM's reading of the claims.

(c) Estimating parameters of a learning machine

239. Culliss estimates parameters of a learning machine based on the user-specific data files. As detailed above, Culliss tracks the search queries and URLs visited by the user. The keywords and URLs are in turn correlated with certain interest categories. For example, "certain key words or terms, such as those relating to sports (i.e. 'football' and 'soccer'), can be detected within search requests and used to classify the user as someone interested in sports." (Culliss at 3:48-51.) Similarly, "certain known articles or URLs can be detected in a users searching or browsing habits, such as those relating to CNNfn (www.cnnfn.com) or Quote.com (www.quote.com), and also used to classify the user as someone interested in finance." (*Id.* at 3: 52-56.)

240. "[W]henever there is a match (whole or partial) between a search request or URL and an item of personal data, a record for the user can be updated to give a +1 for that item of

personal data. A cumulative score can be developed for the user for each item of personal data, called a personal data item score." (Culliss at 4:61-66.) As Culliss explains, "personal data" consists of categories like "Sports Interest" or "Finance Interest." (*See generally id.* at 4:1-53.) "When the personal data item score of the user reaches a certain threshold, then the item of personal data can be said to be associated with the user. Additionally or alternatively, the ***strength of the association can be determined by the cumulative personal data item score.***" (4:67 to 5:4.) Thus viewing several URLs that correspond to "Sport Interest" would cause the user's association with "Sports Interest" to increase. Those associations are "parameters of a learning machine based on the user-specific data files," *i.e.* the URLs visited. Formally, the learning machine presented by Culliss is a statistical method known as a loglinear model.

(d) Analyzing a document

241. Culliss discloses analyzing incoming documents to identify properties of the document:

> [T]he present invention maintains an index of key words, terms, data or identifiers in English or other languages, computer code, or encryption which are collectively referred to as key terms and represented herein by the generic labels "Alpha," "Beta," "Gamma," "Delta," "Epsilon," etc.
>
> ***The articles can each be associated with one or more of these key terms*** by any conceivable method of association now known or later developed. ***A key term score is associated with each article for each of the key terms.*** Optionally, a key term total score can also be associated with the article.

(Culliss at 2:26-37.) Accordingly, the properties of a document—*i.e.*, the key terms like "Shoes" or "Nike" *(id.* at 10-20-34)—are identified by analyzing the document.

(e) Estimating a probability that an unseen document is of interest to the user

242. Culliss estimates a probability that the document is of interest to the user under PUM's interpretations of the claim language. The Court has construed "probability" to mean "numerical degree of belief or likelihood." (Order at 2.) PUM has interpreted the term and the

Court's construction to allow for ***any*** numerical score or ranking of a document. (Plaintiff's April 2012 Infringement Contentions, Tab A, pp. 17-19.) Accordingly, under PUM's infringement arguments, Culliss need only generate match scores for documents based on the user profile in order to meet this claim element.

243. Culliss discloses generating match scores for each document based on the user profile. More specifically, Culliss discloses that "[w]hen a first user enters a search query, the personal data can be considered part of the request and stored within or added to the index, individually or in groupings with other items of data such as key terms, categories, or ratings." (5:18-21.) That personal data in turn influences the rank scores computed for the search results. *See, e.g.,* 6:37-63 (showing different relevancy scores for different articles retrieved for "shoes" dependent on whether the user is a man or a woman), 8:45-50 ("[f]or example, the query or key term grouping such as 'pump shoes' may have different relevancy scores depending on whether the previous searchers were women or men, whereas the rankings may not be different for a personal data characteristic such as profession.")

244. In essence, Culliss appends additional personal information to a query, such as whether the user is a man or woman, whether and how much the user likes sports, etc. That larger query—containing both the user's inputted keywords and the user's information—is used to retrieve relevant articles from the database. The relevancy scores for an article for a keyword query vary based on the user's demographic data, the same way they vary based on the inputted keywords. By adjusting the relevancy score based on the user's personal information, Culliss meets the "estimating a probability" limitation under PUM's interpretation of the claim.

245. Furthermore, because Culliss relies on textual and conceptual analysis of the document in question, it is able to estimate the probability of unseen documents as well as seen documents. Under Plaintiff's own infringement allegations, the *likelihood* of an unseen

document being presented is sufficient to meet this limitation. (Plaintiff's April 2012 Infringement Contentions, Tab A, p. 18: "Due to the extremely large number candidate search results in Google's indexes, and due to the frequent changes or updates made to the candidate search results, it is likely that candidate search results have not been previously seen by the user.") (emphasis added). Moreover, a content-based recommendation system that does not exclude unseen documents has the inherent capability to recommend unseen documents.

(f) Providing automatic, personalized information services to the user.

246. Culliss uses the relevance score to provide personalized information services to the user: "[T]he invention can accept a search query from a user and a search engine will identify matched articles and display squibs of the matched articles in accordance with their comparison scores. Articles can have their key term scores or key term total scores altered according to whether they were displayed to a user, whether they were selected by a user, how much time the user spent with the article, etc." (Culliss 2:39-46.) *See also id.* at 10:53 – 11:10 (describing personalized results for "shoe").

3. Culliss Anticipates Claim 11 of the '040 Patent under PUM's Infringement Theories.

247. Claim 11 requires "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." Culliss discloses a personalization system that operates on top of a search engines: "the invention can ***accept a search query from a user*** and a search engine will identify matched articles and display squibs of the matched articles in accordance with their comparison scores. Articles can ***have their key term scores or key term total scores altered according to whether they were displayed to a user***, whether they were selected by a user, how much time the user spent with the article, etc." (Culliss at 2:39-46.) The altered score would

constitute "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." Further note that the "score" in Culliss can be a "key term ***probability*** score." (2:46-51.)

4. Culliss Anticipates Claim 22 of the '040 Patent under PUM's Infringement Theories.

248. Claim 22 requires "The method of claim 1 wherein the monitored user interactions include a sequence of interaction times." Culliss discloses monitoring interaction times: "Personal activity data includes data about past actions of the user, such as reading habits, viewing habits, searching habits, previous articles displayed or selected, previous search requests entered, previous or current site visits, previous key terms utilized within previous search requests, and ***time or date of any previous activity***." (Culliss at 3:30-35.) These times and dates comprise "a sequence of interaction times," which can be used to determine how much time the user spent viewing an article. In fact, Curtiss explicitly notes that after being retrieved in response to a search query, "[a]rticles can have their key term scores or key term total scores altered according to 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.)

5. Culliss Anticipates Claim 32 of the '040 Patent under PUM's Infringement Theories.

249. Claim 32 of the '040 patent requires "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," followed by the same method steps required in Claim 1 of the '040 patent. As detailed above, Culliss discloses all method steps required in claim 1 and thus it discloses all method steps required in claim 32. Culliss further discloses "a program of instructions executable by the central computer" to perform those method steps. *See, e.g.,* Culliss at 2:39-42 (disclosing a

search engine, which is a program of instructions executable by a central computer). Accordingly, Culliss meets all limitations of claim 32 of the '040 Patent.

6. Culliss Anticipates Claim 34 of the '040 Patent under PUM's Infringement Theories.

250. Claim 34 requires "The program storage device of claim 32 wherein analyzing the document d provides for the analysis of documents having multiple distinct media types." Culliss discloses operating on the Internet, which contains "hundreds of millions of documents, files, databases, text collections, audio clips, video clips and samples of any other type of information." (Culliss at 2:19-22.) Further, it is known to one of skill in the art that documents on the World Wide Web are generally written in HTML, which can contain multiple media types. By processing HTML documents, Culliss "provides for the analysis of documents having multiple distinct media types." (*See, e.g.,* 9:14-17.)

7. Culliss Anticipates Claim 1 of the '276 Patent under PUM's Infringement Theories.

251. Claim 1 of the '276 patent is substantially similar to claim 1 of the '040, as described in Section IV.E above. It contains three additional limitations: a) requiring that the user be normally using a browser in particular rather than a computer in general, b) requiring that the method receive a search query from the user, and c) requiring that the documents analyzed be documents retrieved in response to the search query.

(a) Normal use of a browser program

252. Culliss discloses monitoring the user during normal use of a browser program. Specifically, Culliss monitors both the user's search requests and the user's article viewing habits. (3:46-48.) For example, "certain known articles or URLs can be detected in a users searching or browsing habits, such as those relating to CNNfn (www.cnnfn.com) or Quote.com

(www.quote.com), and also used to classify the user as someone interested in finance." (3:52-56.)

(b) Receiving a search query from the user

253. Culliss discloses receiving a search query from the user: "***the invention can accept a search query from a user*** and a search engine will identify matched articles and display squibs of the matched articles in accordance with their comparison scores." (2:39-42.)

(c) Retrieving a plurality of documents based on the search query

254. Culliss also discloses retrieving documents based on the query: "the invention can accept a search query from a user and ***a search engine will identify matched articles and display squibs of the matched articles in accordance with their comparison scores***." (2:39-42.) Culliss further explains that "***documents***, files, databases, text collections, audio clips, video clips and samples of any other type of information" are "collectively referred to as articles." (2:20-23.) Moreover, the Court has construed "document" as "an electronic file including text or any type of media (Order at 2), so Culliss' "matched articles" easily qualify as "documents" under the Court's construction.

8. Culliss Anticipates Claim 3 of the '276 Patent under PUM's Infringement Theories.

255. Claim 3 requires "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." Culliss discloses transparently monitoring both the user's searching and browsing activities: "Users can explicitly specify their own personal data, or it can be inferred from a history of their search requests or article viewing habits." (Culliss at 3:46-48.) Of note, claim 4 of the '276 patent depends from claim 3, and further requires that the "different modes of user interaction comprise a plurality of modes selected from the group consisting of a ***network searching mode***, a network navigation mode, and a ***network***

browsing mode." Accordingly, both "searching" and "browsing" are relevant "modes of user interaction" with regard to claim 3.

9. Culliss Anticipates Claim 6 of the '276 Patent under PUM's Infringement Theories.

256. Claim 6 requires "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." Culliss discloses at least information about the user's interaction with the document, specifically whether the user viewed the document: "Users can explicitly specify their own personal data, or it can be inferred from a history of their search requests or ***article viewing habits***." (Culliss at 3:46-48.)

10. Culliss Anticipates Claim 7 of the '276 Patent under PUM's Infringement Theories.

257. Claim 7 requires "The method of claim 1, wherein said plurality of retrieved documents correspond to a respective plurality of products." As detailed in section VI.A.9 above, PUM interprets this limitation as requiring that the documents ***may*** correspond to products. As any document on the internet may correspond to a product, and as Culliss does not make any attempt to distinguish pages that correspond to products from pages that do not correspond to products, any given set of documents analyzed by Culliss may also correspond to products. Accordingly, Culliss discloses this claim element under PUM's own infringement theories. (*See, e.g.*, 9:55-59 (processing a query for "High Heels").)

11. Culliss Anticipates Claim 21 of the '276 Patent under PUM's Infringement Theories.

258. Claim 21 requires "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." Culliss discloses presenting document based both on the query and on the computation of whether the document is of interest to the user: "the invention can accept a search query from a user and a search engine will ***identify matched articles and display squibs of the matched articles in accordance with their comparison scores***. Articles can have their ***key term scores or key term total scores altered according to whether they were displayed to a user***, whether they were selected by a user, how much time the user spent with the article, etc." (Culliss at 2:39-46.)

12. Culliss Anticipates Claim 22 of the '276 Patent under PUM's Infringement Theories.

259. Claim 22 requires "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." Culliss discloses extracting and saving keywords associated with the retrieved document: "The articles can each be associated with one or more of these key terms by any conceivable method of association now known or later developed. A key term score is associated

with each article for each of the key terms. Optionally, a key term total score can also be associated with the article." (Culliss at 2:33-37.)

F. The Refuah Patent Anticipates Claims 1, 11, 22, 32, and 34 of the '040 Patent and Claims 1, 3, 5, 6, 7, 21, and 22 of the '276 Patent under PUM's Infringement Theories.[19]

1. Background on Refuah

260. Refuah describes a persona-based and mood-based method of interacting with the Internet. The user's normal interactions are monitored to determine the user's persona and mood, which can be thought of as a temporary modification to the user's persona. (Refuah at 2:39-40.) The system then guides users to certain sites, affects searches, changes displays and layouts, etc. based on the user's profile. (*Id.* at 13:64 to 14:14.)

2. Refuah Anticipates Claim 1 of the '040 Patent under PUM's Infringement Theories.

261. Refuah discloses "a method of aiding information search and retrieval on an Internet. In a preferred embodiment of the invention, Internet searching is personalized to a particular user's profile. Alternatively or additionally, matching up of a supplier and a buyer, of a goods and/or a service, is facilitated, based on such personalization." (Refuah at 1:64 – 2:2.)

(a) Transparently monitoring user interactions with data while the user is engaged in normal use of a computer

262. Refuah discloses transparently monitoring the user's interactions with the Internet:

> In a preferred embodiment of the invention, the persona server tracks frequency, length and/or content of electronic communications with the friends, for example by e-mail, by chat group, by Internet telephone or by computer-dialer, to evaluate an instantaneous mood and/or to assess the relative effect of these friends.

(Refuah at 14:54-59.) Similarly, "a persona of a client may be automatically generated by tracking the way a client interacts with the Internet." (*Id.* at 19:20-22.) Based on the tracking,

[19] E.g., under PUM's theory that any score is the "probability" required by the claims. See below.

> [P]ersonality may be updated automatically. In a preferred embodiment of the invention, the mood is updated based on the one or more of the identification of sites visited by a user, the number of site visited, the dwell time at each site, the order in which sites are visited, the contents of the sites, services purchased, information downloaded, actions performed at the sites and/or a predefined or adaptive time-line based function.

(Refuah at 5:34-41.) "One aspect of this interaction, developed below, relates to an ability of automatically updating a mood based on actions of a user on the Internet." (*Id.* at 3:3-5.)

(b) Updating user-specific data-files comprising the monitored user interactions and a set of documents associated with a user.

263. Refuah discloses updating user-specific data files: "the mood is updated based on the one or more of the identification of sites visited by a user, the number of site visited, the dwell time at each site, the order in which sites are visited, the contents of the sites, services purchased, information downloaded, actions performed at the sites and/or a predefined or adaptive time-line based function." (Refuah at 5:35-41.) Refuah thus tracks the identity of the sites visited by the user, which the patents describe as comprising "monitored user interactions." *See, e.g.,* 22:27-33: "Information about each document that the user views is stored in a recently accessed buffer for subsequent analysis. The recently accessed buffer includes information about the document itself and information about the user's interaction with the document. One possible implementation of a buffer is illustrated in FIG. 14; however, any suitable data structure may be used."

264. Refuah also updates "a set of documents associated with a user" under PUM's reading of the claims. As detailed in its infringement allegations, PUM interprets this claim element to require "updating ***information*** about the documents that the user has clicked on" rather than updating the document itself. (Plaintiff's April 2012 Infringement Contentions, Tab A, p. 6.) Refuah similarly updates information about documents, specifically identification of

the site, presumably through the URL of that site. (Refuah at 5:35-41.) Accordingly, Refuah meets both aspects of the "user-specific data files" under PUM's reading of the claims.

(c) Estimating parameters of a learning machine

265. Refuah estimates parameters of a learning machine based on the user-specific data files:

> In one preferred embodiment of the invention, ***a persona is defined as a set of parameters with values associated with each parameter***. Additionally or alternatively, the parameters may be organized, for example by subject and/or in a hierarchical manner. In a preferred embodiment of the invention, the persona is organized in an object oriented manner. In a preferred embodiment of the invention, not all persona have the same parameters. In a preferred embodiment of the invention, two types of parameters are used, local and global. Local parameters affect only a small part of the interaction with the Internet. For example "***subject of interest=baseball***" does not affect browsing of business sites, except perhaps advertisements. However, "***image download tolerance time=3 sec***" affects the browsing of any site having images. Also "***Color scheme=garish***" will affect the search results of diverse searches. In contrast to such site-general parameters, a persona may also include site specific parameters, for example, "CNN subscriber number=123456", which affect substantially only interaction with the CNN web site.

(Refuah at 6:29-48.) Thus, the parameters of the user's profile are adjusted based on the monitored user's interactions. E.g., if a user does not wait very long for images to download, then the "image download tolerance time" variable may be smaller than average. Similarly, users who spend a lot of time on baseball sites will have their "subject of interest" modified to include "baseball." *See also* Refuah 6:49-64 (listing addition parameters like "subject of interest=chess," "reject=pornography," "reject if pornography level >3," and "baseball=5, basketball=3.")

266. The mood can act as an additional set of parameters layered on top of the persona: "In a preferred embodiment of the invention, a mood is defined as parameters with values that affect a persona. Such values may be, inter alia, single values, ranges of allowed values, scripts,

continuous values and/or discrete values. A mood may replace certain parameter values, affect their value and/or affect their relative weighting." (Refuah at 7:33-38.)

(d) Analyzing a document

267. Refuah discloses analyzing a content of a site:

> Preferably, analyzing a content, comprises ***determining at least one trait of said site***. Alternatively or additionally, analyzing a content comprises determining an ambiance of said site. Alternatively or additionally, analyzing comprises ***analyzing lexicographical characteristics of said site***. Alternatively or additionally, ***analyzing comprises analyzing graphical characteristics of said site***.

(Refuah at 9:50-56.) Refuah further discloses aspects of these analyses:

> In a preferred embodiment of the invention, an atmosphere of a site may be automatically evaluated by ***analyzing the content of a site***, in addition to or instead of utilizing a client's reaction to the site or statistics of accessing the site. Various characteristics of a site may be automatically determined. Each of these characteristics and/or combinations thereof may be used to estimate values for traits and/or atmosphere. The characteristics preferably include one or more of:
>
> (a) word length;
>
> (b) whether certain words and/or phrases used by or associated with the site belong to certain groups, such as "academic words", "swear words", "adult words", "new-age words", "sports words", "baseball words";
>
> (c) sentence complexity;
>
> (d) density of displayed text;
>
> (e) ratio between images and text;
>
> (f) size of text;
>
> (g) distribution of colors in image sand in background;
>
> (h) number of links; number of links visited, date of last visit, by the client, by the persona, by the mood and/or by other moods, personas and/or clients;
>
> (i) size of site;
>
> (j) key-words presented by the site; and/or
>
> (k) number of images; and/or

(l) number and/or type of multimedia files.

(Refuah at 21:6-30.) Accordingly, Refuah discloses analyzing a site to determine properties of the site, such as the lexicographical and graphical characteristics of a site. A site qualifies as a "document" under the Court's construction, as the Court construed "document" as "an electronic file including text or any type of media." (Order at 2).

(e) Estimating a probability that an unseen document is of interest to the user

268. Refuah estimates a probability that the document is of interest to the user under PUM's interpretations of the claim language. The Court has construed "probability" to mean "numerical degree of belief or likelihood." (Order at 2.) PUM has interpreted the term and the Court's construction to allow for ***any*** numerical score or ranking of a document. (Plaintiff's April 2012 Infringement Contentions, Tab A, pp. 17-19.) Accordingly, under PUM's infringement arguments, Refuah need only generate match scores for documents based on the user profile in order to meet this claim element.

269. Refuah discloses generating match scores for each document based on the user profile. More specifically, Refuah discloses that "a site obtains information on a persona and/or a mood of the accessing user and then tailors services and/or advertisements based on the mood or persona. In a preferred embodiment of the invention, when a user enters a book-seller's web site, ***even if the user has never been at the book-seller***, he may be offered books which match his persona and/or mood." (Refuah at 3:56 – 4:4.) Matching a persona to an ad or a book necessarily generates a "probability" under PUM's interpretation of the term. Further, since the user is presented with books even if he's never been to the book-seller, those books may be "unseen" by the user.

270. Regarding search results in another embodiment, "[v]arious combinations of thresholding, grading and sorting may be applied on search results, by comparing them to a

persona." (Refuah at 17:34-36.) "In the key-words technique, a search index includes a classification and/or key-words which match parameters such as those described above for a persona." (*Id.* at 17:39-41.) Similarly, "In the evaluation technique, a site is evaluated for suitability and/or for qualities which are preferred and/or match a particular persona. Example include: number of images in the site, expected download time and/or number of links from the site." (*Id.* at 17:44-47.) Furthermore, "the presentation of search results may also be parameters of the persona.... [o]ne or more parameters of a persona may define matching requirements, for example exactness of match and allowed error.") (*Id.* at 17:49-65.) By adjusting the relevancy score based on the user's personal information, Refuah meets the "estimating a probability" limitation under PUM's interpretation of the claim.

271. Furthermore, because Refuah relies on textual and conceptual analysis of the document in question, it is able to estimate the probability of unseen documents as well as seen documents. Under Plaintiff's own infringement allegations, the *likelihood* of an unseen document being presented is sufficient to meet this limitation. (Plaintiff's April 2012 Infringement Contentions, Tab A, p. 18: "Due to the extremely large number candidate search results in Google's indexes, and due to the frequent changes or updates made to the candidate search results, it is likely that candidate search results have not been previously seen by the user.") (emphasis added). Moreover, a content-based recommendation system that does not exclude unseen documents has the inherent capability to recommend unseen documents.

(f) Providing automatic, personalized information services to the user.

272. Refuah provides personalized information services to the user. As the specification recites:

> In a preferred embodiment of the invention, a persona and/or a mood may be used to have one or more of the following effects on the interaction between client 10 and Internet 16:

(a) preferentially guide client 10 to certain sites;

(b) affect the way searches are performed for information and/or web sites;

(c) affect the way a particular web site responds to a client's request;

(d) affect the display of information;

(e) affect the format and/or layout of a site on the client's terminal;

(f) affect the interpretation of a client's actions and/or data entry;

(g) target promotions and/or advertisements to a client; and/or

(h) protect a client from unwanted influences on Internet 16.

(Refuah at 13:64 to 14:15.) In other words, "one or more aspects of browsing and/or using the Internet may be affected by personality." (2:66 – 3:1.)

3. Refuah Anticipates Claim 11 of the '040 Patent under PUM's Infringement Theories.

273. Claim 11 requires "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." Refuah discloses that "a persona is used to personalize information retrieval.... including ***search engines***, name servers, intelligent agents, yellow pages, white pages, and searching inside a WWW site, such as searching for articles on Microsoft products inside the Microsoft WWW site." (Refuah at 17:21-43.) Refuah further states that "the presentation of search results may also be parameters of the persona.... [o]ne or more parameters of a persona may define matching requirements, for example exactness of match and allowed error." (*Id.* at 17:49-65.)

4. Refuah Anticipates Claim 22 of the '040 Patent under PUM's Infringement Theories.

274. Claim 22 requires "The method of claim 1 wherein the monitored user interactions include a sequence of interaction times." Refuah discloses monitoring interaction times:

> [P]ersonality may be updated automatically. In a preferred embodiment of the invention, the mood is updated based on the one or more of the identification of sites visited by a user, the number of site visited, ***the dwell time at each site***, the order in which sites are visited, the contents of the sites, services purchased, information downloaded, actions performed at the sites and/or a predefined or adaptive time-line based function.

(Refuah at 5:34-41.)

5. Refuah Anticipates Claim 32 of the '040 Patent under PUM's Infringement Theories.

275. Claim 32 of the '040 patent requires "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," followed by the same method steps required in Claim 1 of the '040 patent. As detailed above, Refuah discloses all method steps required in claim 1 and thus it discloses all method steps required in claim 32. Refuah further discloses "a program of instructions executable by the central computer" to perform those method steps. *See, e.g.,* Refuah at 4:5-32 ("[T]he personalization information may be stored by a persona-service. Preferably, a user enters some type of identification, such as a code number, so that the service identifies the user. In some preferred embodiments of the invention, the persona are stored at a central location.... The personalization information may be stored locally or may be acquired from a remote location, such as a persona server. Additionally or alternatively, the persona may be embodied by a proxy server, through which the Internet communication of a user must pass."). Accordingly, Refuah meets all limitations of claim 32 of the '040 Patent.

6. Refuah Anticipates Claim 34 of the '040 Patent under PUM's Infringement Theories.

276. Claim 34 requires "The program storage device of claim 32 wherein analyzing the document d provides for the analysis of documents having multiple distinct media types." As discussed in Section VI.A.5 above, analyzing web pages meets this limitation, as web pages

contain both text and images. Even if claim 34 is read to require actually analyzing different media content, Refuah still meets the limitation, as it discloses analyzing both the text and images of a site:

> Preferably, analyzing a content, comprises ***determining at least one trait of said site***. Alternatively or additionally, analyzing a content comprises determining an ambiance of said site. Alternatively or additionally, analyzing comprises ***analyzing lexicographical characteristics of said site***. Alternatively or additionally, ***analyzing comprises analyzing graphical characteristics of said site***.

(Refuah at 9:50-56.) Accordingly, Refuah "provides for the analysis of documents having multiple distinct media types."

7. Refuah Anticipates Claim 1 of the '276 Patent under PUM's Infringement Theories.

277. Claim 1 of the '276 patent is substantially similar to claim 1 of the '040, as described in Section IV.E above. It contains three additional limitations: a) requiring that the user be normally using a browser in particular rather than a computer in general, b) requiring that the method receive a search query from the user, and c) requiring that the documents analyzed be documents retrieved in response to the search query.

(a) Normal use of a browser program

278. Refuah discloses monitoring the user during normal use of a browser program: "The automatic updating may be performed at the user site, for example ***by tracking the activity of an Internet browser***. Such tracking is preferably achieved using a standalone program which monitors the browser and/or TCP/IP connections. Alternatively or additionally, a dedicated TCP/IP stack and/or driver is used. Additionally or alternatively, ***the updating may be performed by a server, such as a proxy server***, through which a significant portion of a user's requests and/or traffic, pass" (Refuah at 5:57-65.)

(b) Receiving a search query from the user

279. Refuah discloses that "a persona is used to personalize information retrieval.... including ***search engines***, name servers, intelligent agents, yellow pages, white pages, and searching inside a WWW site, such as searching for articles on Microsoft products inside the Microsoft WWW site." (Refuah at 17:21-43.) Refuah further states that "the presentation of search results may also be parameters of the persona.... [o]ne or more parameters of a persona may define matching requirements, for example exactness of match and allowed error." (*Id.* at 17:49-65.)

(c) Retrieving a plurality of documents based on the search query

280. Refuah also discloses retrieving documents based on the query, specifically the documents from a search engine. *See* above.

8. Refuah Anticipates Claim 3 of the '276 Patent under PUM's Infringement Theories.

281. Claim 3 requires "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." Refuah discloses transparently monitoring different modes of user interactions, including sites visited, services purchased, and information downloaded:

> [P]ersonality may be updated automatically. In a preferred embodiment of the invention, the mood is updated based on the one or more of ***the identification of sites visited by a user***, the number of site visited, the dwell time at each site, the order in which sites are visited, the contents of the sites, ***services purchased***, ***information downloaded***, actions performed at the sites and/or a predefined or adaptive time-line based function.

(Refuah at 5:34-41.)

9. Refuah Anticipates Claim 5 of the '276 Patent under PUM's Infringement Theories.

282. Claim 5 requires "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." Refuah discloses analyzing documents that are not of interest to the user and adjusting the user profile accordingly: "A parameter may also be negative, for example, a blacklisting: 'reject=pornography' or 'reject if pornography level >3.'" (Refuah at 6:51-53.)

10. Refuah Anticipates Claim 6 of the '276 Patent under PUM's Infringement Theories.

283. Claim 6 requires "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." Refuah discloses at least information about whether a user visited a site and how much time he spent on that site:

> [P]ersonality may be updated automatically. In a preferred embodiment of the invention, the mood is updated based on the one or more of ***the identification of sites visited by a user***, the number of site visited, ***the dwell time at each site***, the order in which sites are visited, the contents of the sites, services purchased, information downloaded, actions performed at the sites and/or a predefined or adaptive time-line based function.

(Refuah at 5:34-41.)

11. Refuah Anticipates Claim 7 of the '276 Patent under PUM's Infringement Theories.

284. Claim 7 requires "The method of claim 1, wherein said plurality of retrieved documents correspond to a respective plurality of products." As detailed in section VI.A.9 above, PUM interprets this limitation as requiring that the documents ***may*** correspond to products. As any document on the internet may correspond to a product, and as Refuah does not make any attempt to distinguish pages that correspond to products from pages that do not correspond to products, any given set of documents analyzed by Refuah may also correspond to products. Furthermore, Refuah explicitly discloses personalizing advertisements, which PUM asserts meets this claim element:

> Another aspect of some preferred embodiments of the invention relates to ***personalizing advertisements responsive to a mood and/or a persona.*** This personalization of advertisements may be in addition to or alternatively to personalization responsive to a particular search and/or other actions performed by a user at a site. In a preferred embodiment of the invention, a site obtains information on a persona and/or a mood of the accessing user and then tailors services and/or advertisements based on the mood or persona. In a preferred embodiment of the invention, ***when a user enters a book-seller's web site, even if the user has never been at the book-seller, he may be offered books which match his persona and/or mood.***

(Refuah at 3:56 – 4:1; *see also id.* at 11:33-42.) Accordingly, Refuah discloses this claim element under PUM's own infringement theories.

12. Refuah Anticipates Claim 21 of the '276 Patent under PUM's Infringement Theories.

285. Claim 21 requires "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." Refuah discloses presenting document based both on the query and on the computation of whether the document is

of interest to the user: "the presentation of search results may also be parameters of the persona.... [o]ne or more parameters of a persona may define matching requirements, for example exactness of match and allowed error." (Refuah at 17:49-65.)

13. Refuah Anticipates Claim 22 of the '276 Patent under PUM's Infringement Theories.

286. Claim 22 requires "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." Refuah discloses that "several characteristics of a site may be defined, which may be used in filtering out and/or prioritizing such a site." (Refuah at 7:53 – 8:6.) Some of those characteristics are "number of links; number of links visited, date of last visit, by the client, by the persona, by the mood and/or by other moods, personas and/or clients" as well as "key-words presented by the site." (*Id.* at 21:9-30.) The "number of links" correspond to "a list of documents linked to the retrieved document"; the "date of last visit... by other moods, personas, and/or clients" correspond to "a number of users who accessed the retrieved document," and the "keywords presented by the site" is the "topic[s] associated with the retrieved document." Accordingly, Refuah discloses this limitation.

G. Joachims and WebWatcher Anticipate Claims 1, 11, 32 and 34 of the '040 Patent and Claims 1, 6, 7, 21, and 22 of the '276 Patent under PUM's Infringement Theories.[20]

1. Background on WebWatcher

287. As previously noted, WebWatcher was a system developed under the umbrella of Carnegie-Mellon University's Text Learning Group. Users can begin by entering a phrase that describes their current interest, such as "intelligent agents." (Joachims at 2). 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.

288. 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.)

289. 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

[20] E.g., under PUM's theory that any score is the "probability" required by the claims and that a large, shared database can be a "User Model specific to the user" or a "user-specific User Model." See below.

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 above.

2. Joachims and WebWatcher Anticipate Claim 1 of the '040 Patent under PUM's Infringement Theories.

290. WebWatcher discloses a computer implemented method for providing automatic, personalized services to a user. As Joachims explains: "This paper describes a simple but operational tour guide, called WebWatcher, which has given over 5000 tours to people browsing CMU's School of Computer Science Web pages. WebWatcher accompanies users from page to page, suggests appropriate hyperlinks, and learns from experience to improve its advice-giving skills." (Joachims at Abstract).

(a) Transparently monitoring user interactions with data while the user is engaged in normal use of a computer

291. As disclosed in Joachims, "WebWatcher acts as a learning apprentice [Mitchell et al., 1994], observing and learning from its users' actions." (*Id.* at 1). "In general, the user may click on any hyperlink, recommended or not. 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.* at 2). Accordingly, WebWatcher uses *implicit* user feedback (whether the user clicked on a hyperlink) rather than explicit feedback such as ratings, thereby meeting the requirement that the user be transparently monitored during normal use of the computer. (*See, e.g.*, 2:23-28.)

(b) Updating user-specific data-files comprising the monitored user interactions and a set of documents associated with a user.

292. As noted immediately above, WebWatcher logs the user's hyperlinks selections to improve future advice. This log of the user's hyperlink selections constitutes "monitored user interactions with data."

293. Joachims also updates "a set of documents associated with a user" under PUM's reading of the claims. As detailed in its infringement allegations, PUM interprets this claim element to require "updating ***information*** about the documents that the user has clicked on" rather than updating the document itself. (Plaintiff's April 2012 Infringement Contentions, Tab A, p. 6.) Joachims similarly updates information about documents, "log[ging] this hyperlink selection as a training example for learning to improve future advice. (Joachims at 2.) Accordingly, Culliss meets both aspects of the "user-specific data files" under PUM's reading of the claims.

(c) Estimating parameters of a learning machine

294. WebWatcher estimates 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. The "Learning in WebWatcher" section of Joachims discloses three methods by which WebWatcher estimates parameters of a learning machine. "The first approach uses previously given tours as a source of information to augment the internal representation of each selected hyperlink. The second approach is based on reinforcement learning. The idea is to find tours through the Web so that the amount of relevant information encountered over the trajectory is maximized. The third approach is the combined method that includes both of the first two approaches." (Joachims at 3).

295. The parameters of WebWatcher's learning machine also "define a User Model specific to the user" under PUM's interpretation of the claims and the Court's construction. Even though WebWatcher does not create a *separate* User Model for each user, PUM apparently interprets the Court's construction as allowing for a massive data structure containing information from multiple users to meet this claim element. (Plaintiff's April 2012 Infringement

Contentions, Tab A, pp. 17-19.) Therefore, Joachims/WebWatcher meets this element under PUM's own infringement allegations.

(d) Analyzing a document.

296. WebWatcher analyzes a document to identify properties of the document. Specifically, WebWatcher analyzes keywords in hyperlinks to determine the hyperlinks' properties. *See* Joachims at 3 ("For each hyperlink, the list of associated words (including the original underlined words) is used to calculate its similarity to the current user's interest.") Moreover, Mladenic states that "Some later versions of the WebWatcher system change slightly the way of constructing text for learning, e.g., adding words in the document retrieved behind hyperlink." (Mladenic at 4.) In addition, WebWatcher had the ability to add *new* links to the presented webpage: "WebWatcher highlights related hyperlinks on the current page and/or adds new hyperlinks to the current page." (*Id.* at 2; *see also* Joachims at 2.) Since the links did not previously exist, WebWatcher could not have relied on anchor text within the current webpage to make those recommendations, indicating that it must have analyzed the content of the suggested page.

297. Regarding WebWatcher's analysis of hyperlinks, while links to a document may seem to not be a property of the document itself, the asserted patents claim otherwise:

> The probability is estimated by analyzing properties of the document and applying them to the learning machine. ***Documents of multiple distinct media types of analyzed, and identified properties include***: the probability that the document is of interest to users who are interested in particular topics, a topic classifier probability distribution, a product model probability distribution, product feature values extracted from the document, the document author, the document age, ***a list of documents linked to the document***, the document language, number of users who have accessed the document, number of users who have saved the document in a favorite document list, and a list of users previously interested in the document.

('040 Patent, 4:35-47.) This requirement is further an element of unasserted claim 7 of the '040 patent, which depends from claim 1:

7. The method of claim 1 ***wherein the identified properties of the document d comprise a user u-independent property selected from the group consisting of***:

a) a probability P(t,d) that the document d is of interest to users interested in a topic t;
b) a topic classifier discrete probability distribution P(t|d);
c) a product model discrete probability distribution P(p|d);
d) product feature values extracted from the document d;
e) an author of the document d;
f) an age of the document d;
g) a list of documents linked to the document d;
h) a language of the document d;
i) a number of users who have accessed the document d;
j) a number of users who have saved the document d in a favorite document list; and
k) a list of users previously interested in the document d.

298. Accordingly, analyzing the links *to* the document during the application phase meets the limitation of "analyzing a document to identify properties of the document."

(e) Estimating a probability that an unseen document is of interest to the user

299. The Court has construed "probability" to mean "numerical degree of belief or likelihood." (Order at 2.) PUM has interpreted the term and the Court's construction to allow for ***any*** numerical score or ranking of a document. (Plaintiff's April 2012 Infringement Contentions, Tab A, pp. 17-19.) Accordingly, under PUM's infringement arguments, Joachims need only generate match scores for documents based on the user profile in order to meet this claim element.

300. WebWatcher generates match scores: "The metric used to compute similarity between a user's stated interest and a hyperlink description is based on a technique from information retrieval." (Joachims at 3). "Interests and hyperlink descriptions are represented by very high-dimensional feature vectors, each dimension representing a particular word in the English language. The elements (called word-weights) of a vector are calculated using the TFIDF heuristic [Salton, 1991]. Based on this vector representation similarity is calculated as the cosine between vectors." (*Id.*)

301. Moreover, Joachims explicitly states that WebWatcher can compute scores for "unseen" documents : "Because WebWatcher cannot expect that users will always stick to pages it has already seen, a core question in implementing this approach is how to learn a general approximation for each of the Q-functions [] that applies even to unseen states (pages) and actions (hyperlinks)." (Joachims at 4) (parentheticals in original). Further, a content-based recommendation system that does not exclude unseen documents has the inherent capability to recommend unseen documents.

302. Accordingly, WebWatcher estimates the probability that an unseen document is of interest to the user.

(f) Providing automatic, personalized information services to the user.

303. WebWatcher provides automatic, personalized information services to the user. (*See* Joachims at 2 ("We can tell that WebWatcher accompanies us from the additions it makes to the original page. A page augmented with WebWatcher's additions is shown in Figure 1)"; Fig. 1:

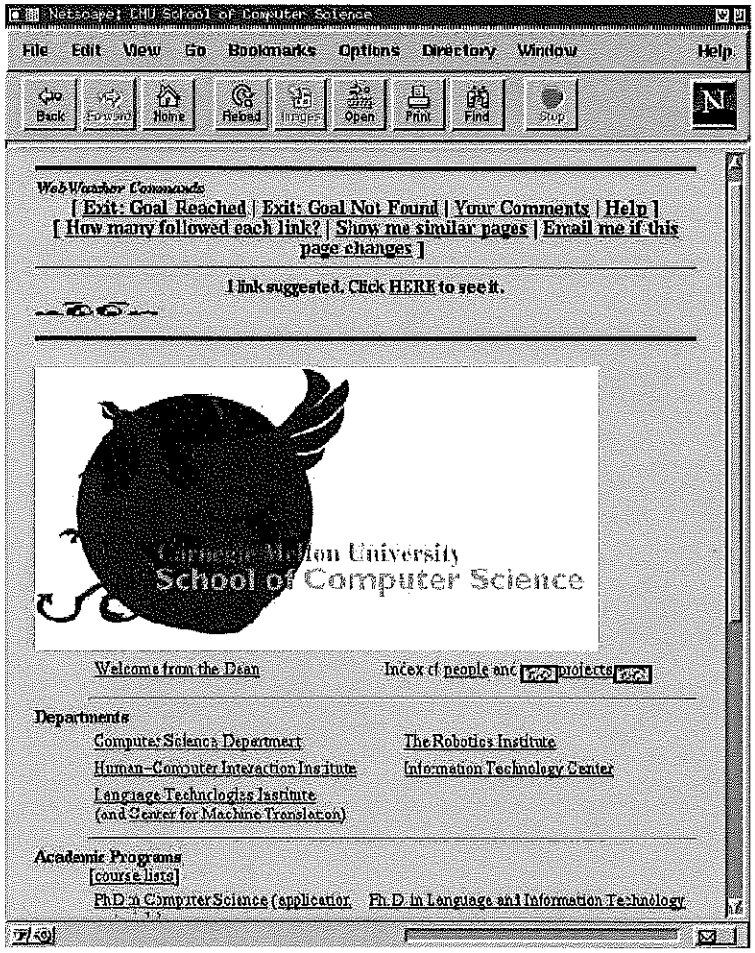


Figure 1: The Front-Door page with WebWatcher's additions.

Accordingly, Joachims/WebWatcher discloses this claim element.

3. Joachims and WebWatcher Anticipate Claim 11 of the '040 Patent under PUM's Infringement Theories.

304. Claim 11 requires "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." WebWatcher accepts a query from the user and uses that query to determine whether the document is of interest to the user: "Clicking on this hyperlink leads us to a page where we are asked for a short description of our current interest. In this example, we

enter the phrase 'intelligent agents' as our interest. WebWatcher now returns us to the initial page, prepared to guide our tour." (Joachims at 2; *see also* Mladenic at 2.)

4. Joachims and WebWatcher Anticipate Claim 32 of the '040 Patent under PUM's Infringement Theories.

305. Claim 32 of the '040 Patent requires "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," followed by the same method steps required in Claim 1 of the '040 patent. As detailed above, WebWatcher discloses all method steps required in claim 1 and thus it discloses all method steps required in claim 32. WebWatcher further includes "a program of instructions executable by the central computer" to perform those method steps. (*See, e.g.,* Joachims at 5 ("WebWatcher runs as a centralized server so that it can assist any Web user running any type of Web browser as well as combining training data from thousands of different users.")) Moreover, because WebWatcher is a computer program, it inherently embodies a "program of instructions." Accordingly, Joachims/WebWatcher meets all limitations of claim 32 of the '040 Patent.

5. Joachims and WebWatcher Anticipate Claim 34 of the '040 Patent under PUM's Infringement Theories.

306. Claim 34 requires "The program storage device of claim 32 wherein analyzing the document d provides for the analysis of documents having multiple distinct media types." As one would expect from its name, WebWatcher functions on Web documents, which are usually created in HTML. (*See* Joachims at 2 (describing inserting "eyeball icons" around selected hyperlinks).) HTML documents contain multiple media types, including text, images, and video. By processing HTML documents, WebWatcher "provides for the analysis of documents having multiple distinct media types." (*See, e.g.,* 9:14-17.)

6. Joachims and WebWatcher Anticipate Claim 1 of the '276 Patent under PUM's Infringement Theories.

307. Claim 1 of the '276 patent is substantially similar to claim 1 of the '040, as described in Section IV.E above. It contains three additional limitations: a) requiring that the monitoring occur during normal use of a browser rather than normal use of a computer, b) requiring that the method receive a search query from the user, and c) requiring that the documents analyzed be documents retrieved in response to the search query.

(a) Normal use of a browser program

308. WebWatcher monitors the user during normal use of a browser program, as WebWatcher acts upon users who are browsing the Web. (*See* Joachims at 1 ("This paper describes a simple but operational tour guide, called WebWatcher, which has given over 5000 tours to people browsing CMU's School of Computer Science Web pages."); 5 ("WebWatcher runs as a centralized server so that it can assist any Web user running any type of Web browser"). Accordingly, WebWatcher discloses this limitation.

(b) Receiving a search query from the user

309. WebWatcher receives a search query from the user. As detailed in Section VI.A.3, above, Joachims explains that WebWatcher accepts a query from the user and uses that query to determine whether the document is of interest to the user: "Clicking on this hyperlink leads us to a page where we are asked for a short description of our current interest. In this example, we enter the phrase 'intelligent agents' as our interest. WebWatcher now returns us to the initial page, prepared to guide our tour." (Joachims at 2; *see also* Mladenic at 2.)

(c) Retrieving a plurality of documents based on the search query

310. Joachims also explains that WebWatcher retrieves a plurality of documents based on the search query. As detailed above, the WebWatcher can also "add[] new hyperlinks to the current page" based on the few keywords entered by the user. "It also provides a user command

'Show me similar pages' in the command list, which causes WebWatcher to display a list of pages which are similar based on a metric derived from hypertext structure." (Joachims at 2.) Accordingly, those documents are retrieved and analyzed based on the search query entered by the user.

7. Joachims and WebWatcher Anticipate Claim 6 of the '276 Patent under PUM's Infringement Theories.

311. Claim 6 requires "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." Joachims discloses that WebWatcher monitors the user interactions for whether the user viewed the document: "In the first approach, learning is accomplished by annotating each hyperlink with the interests of the users who took this hyperlink on previous tours." (Joachims at 3.) Accordingly, WebWatcher contains this limitation.

8. Joachims and WebWatcher Anticipate Claim 7 of the '276 Patent under PUM's Infringement Theories.

312. Claim 7 requires "The method of claim 1, wherein said plurality of retrieved documents correspond to a respective plurality of products." As detailed in section VI.A.9 above, PUM interprets this limitation as requiring that the documents ***may*** correspond to products. As any document on the internet may correspond to a product, and as Joachims does not make any attempt to distinguish pages that correspond to products from pages that do not correspond to products, any given set of documents analyzed by Joachims may also correspond

to products. Accordingly, Joachims/WebWatcher discloses this claim element under PUM's infringement theories.

9. WebWatcher Anticipates Claim 21 of the '276 Patent under PUM's Infringement Theories.

313. Claim 21 requires "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." Mladenic explains that WebWatcher presents documents based both on the query and on the computation of whether a document is of interest to the user:

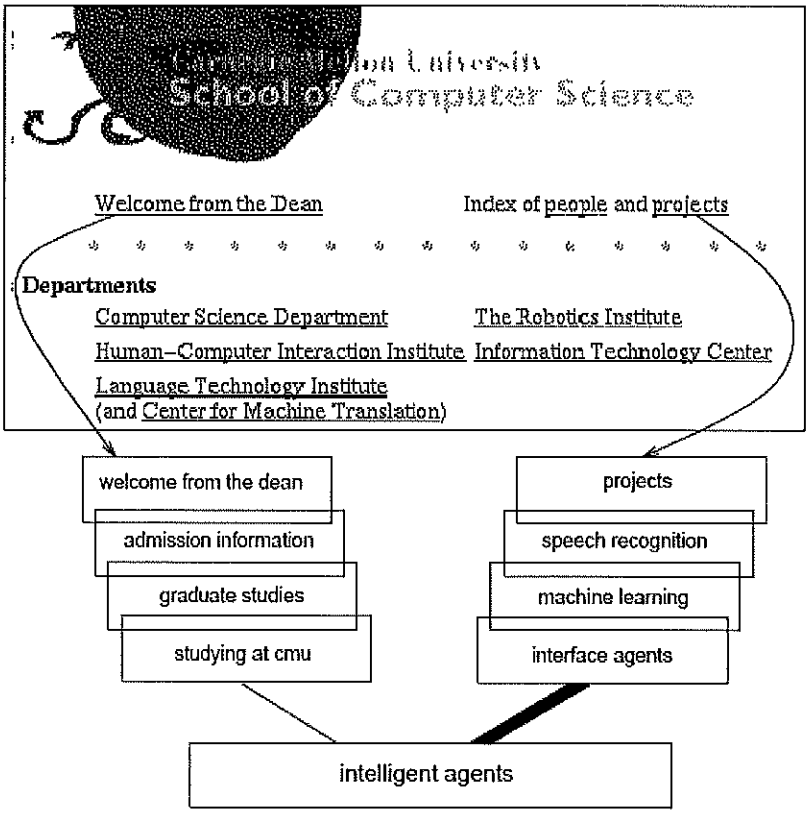
> WebWatcher can be described as an agent that assists users in locating information on the WWW. It learns by observing a user on her/his way through the WWW and suggests interesting hyperlinks ***whenever it is confident enough***. The idea is that ***the user provides a few keywords*** describing a search goal and WebWatcher highlights related hyperlinks on the current page ***and/or adds new hyperlinks to the current page***.

(Mladenic at 2.) Accordingly, WebWatcher suggests pages based both on the user's keywords (query) and on its computed confidence (probability) that the pages will be of interest to the user.

10. Joachims and WebWatcher Anticipate Claim 22 of the '276 Patent under PUM's Infringement Theories.

314. Claim 22 requires "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." As shown in Figure 3 of Joachims, WebWatcher analyzes a topic associated

with the document – the associated topics are the nested boxes displayed below the screenshot of the CMU School of Computer Science webpage:



VII. THE ASSERTED CLAIMS ARE OBVIOUS IN LIGHT OF THE PRIOR ART

315. The discussion above demonstrated that the asserted claims are anticipated by one or more of Mladenic, Autonomy, Montebello, Wasfi, Culliss, and Refuah. To the extent that any of those references do not disclose limitations in the asserted claims, this section demonstrates that those limitations consist only of obvious applications of art known to one of ordinary skill, and thus the claims are invalid for obviousness in light of each reference.

316. Exhibit 3 of this expert report is an element-by-element claim chart of each of the asserted claims in this case with references to the prior art, and is incorporated into the body of and is part of this report.

A. Standard for Obviousness

317. I understand that the Supreme Court in *KSR* expanded upon the framework for analyzing obviousness set forth in previous cases including *Graham v. John Deere.* It is my understanding that in *KSR*, the Supreme Court rejected the Federal Circuit's "rigid" application of the "teaching, suggestion, or motivation" test for obviousness in favor of an "expansive and flexible approach" using "common sense." I understand that in *KSR*, the Supreme Court specifically cautioned against granting patents that are nothing more than combinations of known elements driven by non-innovative factors such as market demands. The Court also provided guidance on how combination patents should be handled. The Supreme Court noted that "[g]ranting patent protection to advances that would occur in the ordinary course without real innovation retards progress and may, in the case of patents combining previously known elements, deprive prior inventions of their value or utility." The Supreme Court also stressed the need for "caution" before validating patents that are merely combinations of elements found in the prior art. In view of this caution, the Court explained that "[t]he combination of familiar elements according to known methods is likely to be obvious when it does no more than yield predictable results."

318. I further understand that the Supreme Court pointed to other factors which may show obviousness. For example, the Supreme Court observed, "[w]hen a work is available in one field of endeavor, design incentives and other market forces can prompt variations of it, either in the same field or a different one. If a person of ordinary skill in the art can implement a predictable variation," it is obvious. Similarly, the Court noted that "[i]f a technique had been used to improve one device, and a person of ordinary skill would recognize that it would improve similar devices in the same way, using the technique is obvious, unless its actual application is beyond his or her skill." Further, "[w]hen there is a design need or market

pressure to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill has good reason to pursue the known options within his or her technical knowledge." Finally, "[i]f a person of ordinary skill can implement a predictable variation of the prior art in the manner claimed, §103 likely bars its patentability."

319. I understand that in *KSR*, the Supreme Court also stated that the factors from *Graham v. John Deere* should be used in the obviousness analysis. These factors are:

(1) The scope and content of the prior art,

(2) Differences between the prior art and the claims asserted,

(3) The level of ordinary skill in the pertinent art.

(4) "Secondary considerations" of non-obviousness

B. The Asserted Patents Are a Combination of Prior Art Elements.

320. Each of the elements of the '040 and '276 Patents was present in the prior art.

321. Moreover, I note that PUM has not identified any element of the Asserted Patents that improved upon the prior art. PUM's Supplemental Response to Google's Interrogatory No. 4 states "the patents-in-suit improved upon the prior art at least because the patents-in-suit disclose a learning mechanism to personalize unseen content anywhere on the web, including, but not limited to, web content found through search." This, however, was disclosed in the anticipatory references identified above. Indeed, Montebello explicitly states that it describes "a system that reuses the information generated from search engines together with previously developed systems, and adapts it, by generating user profiles, to better meet the needs and interests of the users." (Montebello at 1.)

1. Transparent Monitoring

322. The notion of monitoring a user's interaction with data *transparently*—that is, without requiring explicit feedback from the user—pre-dates the alleged invention. For example,

Green explains that "[a] common method of developing a user profile is by observing and analysing user behaviour." (Green at 2.) *Wasfi* lists "indirect or transparent learning technique" as one of the means of obtaining information from a user. (Wasfi at 57.) Similarly, *Mladenic* states that it "watches over the user's shoulder… but avoids involving the user in its learning process." (Mladenic at 3.) Montebello likewise tracks a user's bookmarking activities. (Montebello at 3.) At least as early as 1992, persons of skill in the art were aware of the need for transparent or implicit monitoring. As Shoshana Loeb explained,

> Unlike the proactive users, the casual users are not likely to be willing to engage in lengthy interactions with the system in order to articulate current information needs and provide explicit feedback. Therefore, automating the personalized delivery of information to this class of users requires mechanisms that can cope with this fact. In particular, issues related to mechanisms for the creation of profiles for new users (e.g., by either using initial profiles based on stereotypes for users' groups or by building profiles directly from usage data) ***and to the detection of implicit feedback (e.g., skipped and revisited items)*** need further research.

(Loeb at 41). *See also* Ken Lang, Newsweeder:Learning to Filter Netnews (1995), at 1; Masahiro Morita and Yoichi Shinoda, Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval, at 3.2 and 4; Jon Orwant, Heterogeneous Learning in the Doppelganger User Modeling System; L. Ardissono, C. Barbero, A. Goy, G. Petrone, An Agent Architecture for Personalized Web Stores, at 185.

323. The asserted patents themselves concede that transparent monitoring was in the prior art, stating that "[t]he earliest collaborative filtering systems required explicit ratings by the users, ***but existing systems are implemented without the user's knowledge by observing user actions***. Ratings are inferred from, for example, the amount of time a user spends reading a document or whether a user purchases a particular product." (2:23-28.) The patents go on to describe the Mobasher system, which personalized through analyzing log files created during the user's normal interactions. (2:29-36.)

324. Of note, PUM has not identified transparent monitoring as a manner or technique in which the patents-in-suit improved upon the prior art. (PUM's Supplemental Response to Google's Interrogatory No. 4 ("the patents-in-suit improved upon the prior art at least because the patents-in-suit disclose a learning mechanism to personalize unseen content anywhere on the web, including, but not limited to, web content found through search")).

2. Updating User-Specific Files Comprising User Documents and Monitored User Interactions with said Documents

325. Storing files associated with the user as well as the user's interactions with those files was known in the art. For example, Mladenic "fetch[es] visited documents and documents one step behind the hyperlinks of visited documents and store[s] them as positive or negative examples of user interests, depending whether the user visited the document or not." (Mladenic at 8.) Edwards discloses that "[a]ctions performed by the user on a document (news article, Web page, etc.) are recorded together with the text of the document." (Edwards at 33.) Similarly, Payne states that "observations, consisting of articles and actions performed on them, are used to generate training examples, by passing them to the feature extraction module." (Payne at 6.)

326. As with transparent monitoring, the patents themselves concede that updating user-specific files and interactions with those files was in the prior art:

> Information filtering techniques focus on the analysis of item content and the development of a personal user interest profile. In the simplest case, ***a user is characterized by a set of documents, actions regarding previous documents***, and user-defined parameters, and new documents are characterized and compared with the user profile.

(1:32-37.) The patents go on to cite several patents that update user-specific files, concluding that "[o]ther techniques rate documents using the TFIDF (term frequency, inverse document frequency) measure. The user is represented as a vector of the most informative words in a set of user-associated documents." (1:37-57.)

327. PUM has not identified updating user-specific files as a manner or technique in which the patents-in-suit improved upon the prior art. (PUM's Supplemental Response to Google's Interrogatory No. 4 ("the patents-in-suit improved upon the prior art at least because the patents-in-suit disclose a learning mechanism to personalize unseen content anywhere on the web, including, but not limited to, web content found through search")).

3. Estimating Parameters of a Learning Machine

328. Creating a user-specific profile based on the user-specific files was also in the prior art. For example, Vannevar Bush's *As We May Think* in "The Atlantic" (July 1945) discloses a version of user modeling without even the use of a personal computer. User modeling was an established field prior to the invention of the patents-in-suit. There existed a journal entitled "User Modeling and User-Adapted Interaction" prior to the invention of the patents-in-suit, and in 1998, that journal distributed a special issue on machine learning for user modeling. (*See Preface to UMUAI Special Issue on Machine Learning for User Modeling*, "User Modeling and User-Adapted Interaction," vol. 8 (1998).) The preface states: "This special issue of User Modeling and User-Adapted Interaction brings together a collection of papers presenting a wide variety of machine learning techniques and their use in a diverse range of user modeling applications."

329. Payne discloses that "[t]he use of learning techniques to develop a *profile* of an individual user's preferences not only eliminates the need for programming rules, but also allows the agent to adapt to changes." (Payne 1 (emphasis in original).) Since personalization services "utilize[] personal data to further refine search results," that data needs to be stored in a manner permitting rapid access and determination of user interest. (Culliss 3:13-18.) The monitored interactions in the prior art often included a set of documents that were or could have been presented to the user as well as the user's interactions with those documents. (Mladenic 8, Wasfi

at 58.) In addition to storing the documents themselves, it makes sense to store the user's interactions with those documents. (Refuah, 5:34-50; Schuetze, 18:11-17.)

330. Nor is using a learning machine new. For instance, Mladenic discloses a LEARNER module that "transforms documents into examples" by "parsing each document, assigning an index to each word" and "calculating score (e.g. information gain) for each word." (Mladenic 9.) Similarly, Joachims discloses "three approaches to learning this target function [the LinkQuality] from experience. The first approach uses previously given tours as a source of information to augment the internal representation of each selected hyperlink. The second approach is based on reinforcement learning.... The third approach is the combined method that includes both of the first two approaches." (Joachims 3.)

331. Several texts discussing machine learning and/or learning algorithms pre-dated the claimed inventions. *See e.g.,* Vladimir Cherkassky and Filip Mulier, *Learning from Data, Concepts, Theory, and Methods*, IEEE Press (1998); John Hertz, Anders Krogh, Richard G. Palmer, *Introduction to The Theory of Neural Computation*, Addison-Wesley Publishing Company (1991); Richard O. Duda and Peter E. Hart, *Pattern Classification and Scene Analysis*, A Wiley-Interscience Publication (1973). Indeed, Cherkassky is expressly mentioned in the '040 patent.

332. Further, I note that named inventor Yochai Konig admitted that machine learning pre-dated the disclosed inventions. (Konig Depo., 20:17-21:10, 22:23-23:2.) In fact, Mr. Konig admitted that the use of user models in machine learning pre-dated the patents. (*Id.*, 23:19-24:16.) He similarly acknowledged that machine learning algorithms were previously used to provide personalization services. (*Id.*, 67:4-25) (acknowledging Amazon's feature). In addition, SRI did work in machine learning prior to 1999. (Bercow Dep. 75:24-776:8.) And, Konig's work at SRI involved building models using machine learning, estimating parameters,

calculating posterior probabilities using machine learning, and using Bayesian statistics for machine learning. (Franco Dep., 101:17-102:10.)

333. The asserted patents themselves concede that creating a user-specific user model was within the prior art. In the "Background Art" section, the patents disclose that "[t]he 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." (1:55-60.)

334. Again, PUM has not claimed that the creation of a user-specific profile was a novel aspect of the claimed invention. In response to Google's Interrogatory No. 4, PUM did not identify utilization of personal data or user modeling as manners or techniques in which the patents-in-suit improved upon the prior art. (PUM's Supplemental Response to Google's Interrogatory No. 4 ("the patents-in-suit improved upon the prior art at least because the patents-in-suit disclose a learning mechanism to personalize unseen content anywhere on the web, including, but not limited to, web content found through search")).

4. Analyzing Documents

335. Similarly, prior art was capable of analyzing and identifying features of a document. Even pre-computer libraries regularly analyzed documents to determine the author, title, and subject matter necessary to index them in card catalogs. Indeed, the patents state that analyzing the content of the document is what separates prior art information filtering techniques from prior art collaborative filtering methods. (1:32-34; 2:16-18.) Furthermore, it follows that *content-based* filtering would need some knowledge about the *content* of a document.

336. For example, Refuah discloses analyzing the word length, sentence complexity, ratio between images and text, number of links, number of images, and number of page accesses for a particular site, among other features. (Refuah 21:6-30.) Many known prior art references at

a minimum extracted the words from the document. (*See, e.g.,* Wasfi 61, Mladenic 4, Culliss 5:36-48.) As the patents themselves state, analyzing documents and rating its words through measures such as Term Frequency, Inverse Document Frequency was well-known in the art:

> Other techniques rate documents using the TFIDF (term frequency, inverse document frequency) measure. 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.

(1:53-60.) *See also* 11:12-20 (disclosing using TFIDF in a preferred embodiment of the claimed invention).

337. Again, in response to Google's Interrogatory No. 4, PUM did not identify analyzing documents as a manner or technique in which the patents-in-suit improved upon the prior art. (PUM's Supplemental Response to Google's Interrogatory No. 4 ("the patents-in-suit improved upon the prior art at least because the patents-in-suit disclose a learning mechanism to personalize unseen content anywhere on the web, including, but not limited to, web content found through search")).

5. Estimating a Probability

338. Determining the extent to which a user is interested in a document also predates the patents-in-suit. (*See e.g.,* Culliss, 2:43-51, 5:36-62.) Many user models contain or consist of a mathematical average of representations of the user's documents, as detailed above. New documents that are sufficiently "similar" to those previous documents would be favored over new documents which are not as "similar." Accordingly, prior art systems contained various ways to compare a vector representing the user's documents to a vector representing an incoming document. Known methods included nearest-neighbor algorithms, naïve Bayes classifiers, neural networks, support vector machines, IBPL1 and IBPL2 , CN2 and decision trees, etc.

339. For example, Mladenic "uses a Naïve (Simple) Bayesian classifier on frequency vectors to generate a model of user interests, that is used for advising hyperlinks." (Mladenic at 7.)[21] Autonomy similarly uses neural networks and Bayesian analysis "centered on calculating the probabilistic relationship between multiple variables and determining the extent to which one variable impacts another." (Autonomy WP at 4.) Similarly, Joachims discloses computing a LinkQuality between 0 and 1, wherein "[t]he value of LinkQuality is interpreted as the probability that a user will select Link given the current Page and Interest." (Joachims 3.) Finally, Wasfi discloses using the "importance" of a page to update the user profile, where importance is defined as a posterior probability: "If we consider the foregoing examples, we see that the importance of a page is inversely proportional to its probability following the page (or the sequence of pages) visited by the user." (Wasfi at 3).

340. Again, I note that several texts pre-dating the Asserted Patents discuss Bayesian statistics and computation of probabilities at length. *See e.g.,* Christopher D. Manning and Hinrich Schutze, *Foundations of Statistical Natural Language Processing*, The MIT Press (1999); Duda and Hart; Peter M. Lee, *Bayesian Statistics: An Introduction*, Oxford University Press (1989); Ricardo Baeza-Yates and Berthier Ribeiro-Neto, *Modern Information Retrieval*, ACM Press (1999); Keinosuke Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press (1990); Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, Wiley-Interscience (1991); Cherkassky and Mulier.

341. I also note that Mr. Konig admitted in deposition that Bayesian statistics preceded his inventions. (Konig Depo., 24:17-25:16, 26:4-9.) He further admitted that Bayesian models were previously used in information retrieval. (*Id.*, 25:17-20.) In addition, Konig's work at SRI

[21] Note that due to concerns about network speed, Mladenic compared the user profile to anchor text rather than the underlying document. *See* Section VI.A.2(d), *supra.*

involved building models using machine learning, estimating parameters, calculating posterior probabilities using machine learning, and using Bayesian statistics for machine learning. (Franco Dep., 101:17-102:10.)

342. Furthermore, PUM has taken the position that using ***any*** numerical score is sufficient to meet this claim element. (Plaintiff's April 2012 Infringement Contentions, Tab A, pp. 17-19.) Under this interpretation, virtually every prior content matching system would meet the claim limitation. For example, Culliss discloses computing "key term scores" whose values are altered based on the personal characteristics of the user. (Culliss at 2:43-51.) Edwards discloses computing a "confidence rating" by comparing features of the document to the user profile. (Edwards at 33.) Wasfi likewise discloses computing a similarity metric by taking the scalar product of the user vector and the document vector. (Wasfi at 60-61.)

343. The patents themselves agree that comparing profiles to documents was well known in the art:

> Other techniques rate documents using the TFIDF (term frequency, inverse document frequency) measure. 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.

(1:53-60.)

344. The claims also require that the "document" be "unseen," which the Court construed as "document not previously seen by the user." (Order at 2.) Again, virtually every prior art system would (at least occasionally) meet this limitation, as they are designed to personalize information services for a particular user regardless of whether the user had seen the document or not. (*See* above.) A prior art system that could *only* present documents that were seen by the user would have limited utility: there is less need to compute or estimate whether a

particular user would find a document interesting if the user already saw the document, and thus his reaction to that document can be recalled rather than estimated/computed.

345. Again, PUM's response to Google's Interrogatory No. 4 did not identify estimating a probability as a manner or technique in which the patents-in-suit improved upon the prior art. (PUM's Supplemental Response to Google's Interrogatory No. 4 ("the patents-in-suit improved upon the prior art at least because the patents-in-suit disclose a learning mechanism to personalize unseen content anywhere on the web, including, but not limited to, web content found through search")).

6. Providing Personalized Services

346. Providing personalized services was well-known in the art. The patents themselves list numerous prior art systems that personalized information for the user. (*See, e.g.*, 1:27 to 3:37.) Similarly, there can be little doubt that the prior art systems discussed above, such as Mladenic, Autonomy, Montebello, Wasfi, Culliss, and Refuah, provided personalized services for the user. In fact, providing personalized information to users dates back to at least Vannevar Bush's July 1945 *As We May Think*, in "The Atlantic."

347. As set forth in the discussion of Background Art in the shared specification of the Asserted Patents, information filtering and collaborative filtering techniques existed in the art, with both attempting to personalize information provided to users. The shared specification further notes that information filtering and collaborative filtering techniques had been combined prior to the invention of the Asserted Patents. Also existing in the 1990s, prior to the invention of the Asserted Patents, were focused crawling techniques of personalizing information for users at the collection phase, rather than the filtering phase.

348. Using ratings generated by comparisons as described above to personalize the service provided to the user existed in the art. Some prior art sources personalize the content

displayed to the user; some personalize search results; and some personalize both. It follows the common-sense understanding that a user is likely to be searching for the same types of documents which he previously located. Moreover, Mr. Konig admitted that the inventors did not invent personalized search. (Konig Depo. 10:6-21). In addition, SRI's FASTUS project provided Internet search support prior to July 1999. (Bercow Dep., 70:23-71:20.) And, another unnamed SRI project worked in personalization prior to 1999. (*Id.*, 77:7-24.)

349. Again, PUM's response to Google's Interrogatory No. 4 did not identify providing personalized services as a manner or technique in which the patents-in-suit improved upon the prior art. (PUM's Supplemental Response to Google's Interrogatory No. 4 ("the patents-in-suit improved upon the prior art at least because the patents-in-suit disclose a learning mechanism to personalize unseen content anywhere on the web, including, but not limited to, web content found through search")).

7. Responding to Users' Search Queries

350. Responding to users' search queries, as disclosed in claim 11 of the '040 Patent and claims 1 and 21 of the '276 patent, existed in the prior art.

351. For example, Culliss discloses responding to users' search queries:

> As described in my previous applications, the invention can accept a search query from a user and a search engine will identify matched articles and display squibs of the matched articles in accordance with their comparison scores. Articles can have their key term scores or key term total scores altered according to whether they were displayed to a user, whether they were selected by a user, how much time the user spent with the article, etc.

(Culliss at 2:39-46). Autonomy likewise responds to user queries:

> When you train an Agent you give it a few sentences of text describing the subject you want it to look for. The neural network is able to identify the key concepts in the text, and then use its knowledge of language to decide their relative importance.

(Autonomy UG at 4.) *See also* Autonomy WP at 3 (explaining that agents can be trained on user documents and/or user queries.) Montebello similarly employs queries to search engines, then personalizes the results: "Every query term is employed by the wrapper which will command the associated system to locate documents from its local index and return related results." (Montebello at 3.)

352. Claim 11 of the '040 Patent and claim 21 of the '276 Patent further require that the probability determination be based on the match between the query and the document, as well as the match between the user and the document. This element was present in the prior art. For example, Culliss states that

> As described in my previous applications, the invention can accept a search query from a user and a search engine will identify matched articles and display squibs of the matched articles in accordance with their comparison scores. Articles can have their key term scores or key term total scores altered according to whether they were displayed to a user, whether they were selected by a user, how much time the user spent with the article, etc.

(Culliss 2:39-46.) Refuah similarly discloses modifying the results of a search engine based on personalization. (Refuah at 1:63 – 2:2, 3:12-24.) Schuetze also discloses modifying search results based on personalization. (Schuetze at 22:31-48.)

8. Interaction Times

353. Monitoring user interactions through a sequence of interaction times, as disclosed in claim 22 of the '040 Patent, existed in the prior art. For example, Culliss discloses monitoring how much time a user spent with an article:

> Articles can have their key term scores or key term total scores altered according to whether they were displayed to a user, whether they were selected by a user, ***how much time the user spent with the article***, etc.

(Culliss at 2:39-46.) Since HTTP is a stateless protocol, one of skill in the art would understand that determining how much time the user spent with the article is derived from sequences in a user's interaction times, e.g. those present on server logs. Indeed, Culliss explicitly discloses

recording the "time or date of any previous activity." (*Id.* at 3:34-35.) Refuah similarly discloses monitoring user interactions times, which are used to determine whether the user is in a hurry and thus whether to steer him toward pages that load quickly:

> One aspect of this interaction, developed below, relates to an ability of automatically updating a mood based on actions of a user on the Internet. Thus, actions of a user affect can a style in which an Internet responds. In one example, a harried access to the Internet (not waiting for images to download, short dwell times) will result in the identification/definition of a rushed mood. Thereafter, search engines may steer the user away from sites which require long download times.

(Refuah 3:3-11; *see also id.* at 5:34-50 (tracking "the dwell time at each site" as well as identifying the user as "busy" based on interaction times).) Schuetze also discusses monitoring interaction times (Schuetze at 18:11-17 (tracking "time spent viewing documents").) Wasfi likewise monitors interactions times to distinguish users from the same IP address. (Wasfi at 60.)

9. Multiple Media Types

354. Analyzing documents having multiple distinct media types, as disclosed in claim 34 of the '040 Patent, existed in the prior art. For example, Schuetze discloses searches that are based on both an image and the text associated with the image:

> ***Set forth herein is an approach to document browsing and retrieval in which a user iteratively narrows a search using both the image and text associated with the image***, as well as other types of information related to the document, such as usage. Disparate types of information such as text, image features and usage are referred to as "modalities." Multi-modal clustering hence is the grouping of objects that have data from several modalities associated with them.
>
> The text surrounding or associated with an image often provides an indication of its context. ***The method proposed herein permits the use of multi-modal information, such as text and image features, for performing browsing and retrieval (of images, in the exemplary case described herein)***. This method is applicable more generally to other applications in which the elements (e.g., documents, phrases, or images) of ***a collection can be described by multiple characteristics, or features***.

(Schuetze at 4:12-35; *see also id.* at 5:48-58.) Refuah likewise analyzes both the text and images of a particular site. (Refuah at 9:50-56.)

355. Furthermore, claim 34 only requires that documents ***having*** "multiple distinct media types" be analyzed; there is no requirement that the multiple distinct media types themselves be analyzed. As documents on the World Wide Web are generally written in HTML—which allows for media to be embedded next to text—any system analyzing HTML or WWW documents would analyze documents having multiple distinct media types. See, *e.g.*, Mladenic at 3 (stating that Personal WebWatcher operates on HTML documents), Culliss at 2:19-24 (stating that the Internet articles may be "documents, files, databases, text collections, audio clips, video clips and samples of any other type of information," all of which may be analyzed by Culliss).

10. Using a Browser

356. Claim 1(a) of the '276 Patent is similar to claim 1(a) of the '040 Patent, save that it requires that the transparent monitoring be of the user's normal use of a "browser" rather than the more general "computer." Transparently monitoring browser use existed in the prior art. For example, *Mladenic* transparently observes the user's normal interactions with his browser through a proxy server:

> Personal WebWatcher consists of two main parts: a **proxy server** that interacts with the user via Web browser and a **learner** that provides the user-model to the server (see Figure 2). Communication between them is via disk; the **proxy** saves address of visited documents (URLs) and the **learner** uses them to generate model of user interests.

(Mladenic at 7; *see also id.* at 8, Fig. 2.) Refuah also discloses tracking the user's browser interactions: "The automatic updating may be performed at the user site, for example by tracking the activity of an Internet browser." (Refuah at 5:57-58.) In addition, Montebello also monitors the user's normal interactions with his web browser, specifically his "bookmarking" actions.

(Montebello at 3.) Wasfi similarly monitors the user's navigations on the Web. (Wasfi at Abstract.)

11. Monitoring Different Modes of User Interaction

357. Monitoring multiple different modes of user interactions with network data, as disclosed in claim 3 of the '276 Patent, existed in the prior art. For example, Culliss states that "Articles can have their key term scores or key term total scores altered according to whether they were displayed to a user, whether they were selected by a user, how much time the user spent with the article, etc." (Culliss at 2:43-46), and that "[p]ersonal activity data includes data about past actions of the user, such as reading habits, viewing habits, searching habits, previous articles displayed or selected, previous search requests entered, previous or current site visits, previous key terms utilized within previous search requests, and time or date of any previous activity." (*Id.* at 3:29-35.) Refuah similarly discloses monitoring a wide variety of user interactions with network data:

> In a preferred embodiment of the invention, the mood is updated based on the one or more of the identification of sites visited by a user, the number of site visited, the dwell time at each site, the order in which sites are visited, the contents of the sites, services purchased, information downloaded, actions performed at the sites and/or a predefined or adaptive time-line based function. Alternatively or additionally, a mood, for example a "rush" mood, may be identified by tracking whether a user waits until images are downloaded, whether a user waits for a complete site to download, whether a user follows links and how many links are followed, and/or rate of changing WWW pages and/or sites.

(Refuah at 5:34-50.) Schuetze likewise discloses monitoring "on-line purchasing habits, software usage, and time spent viewing documents." (Schuetze at 18:11-17.)

12. Documents Not of Interest

358. Determining and analyzing documents not of interest to the user, as disclosed in claim 5 of the '276 Patent, is also in the prior art. For example, Mladenic states that "Hyperlinks whose documents were visited by the user are considered to be positive examples, and all the

other to be negative examples of the user interests. The idea is that all hyperlinks were presented to the user and the user chose to visit some of them that meet her/his interests." (Mladenic at 8; *see generally id.* at 8-9.) Edwards similarly states that "Links explored by the user are used as training instances. Each instance is given a classification of either interesting or not interesting. If a link led to a page that the user saves as a bookmark, printed, or visited frequently then it is classified as interesting, otherwise as not interesting." (Edwards at 35-36.)

13. Monitoring Certain User Interactions

359. Claim 6 of the '276 Patent requires that the monitored user interactions be at least one of the following:

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

360. Monitoring at least one of these user interactions existed in the prior art. For example, Mladenic monitors whether or not the user viewed each document, which it in turn uses to distinguish documents of interest from documents not of interest. (Mladenic at 8.) Culliss similarly discloses monitoring whether the user viewed the document and how much time the user spent viewing the document. (Culliss at 2:43-46.) Schuetze likewise tracks how much time the user spent with the document. (Schuetze at 18:11-17.) Refuah similarly monitors a wide variety of user interactions. (Refuah at 5:34-50.)