it reported in its order.[2] This is precisely the analysis that must be undertaken by the USPTO in considering the meaning of the present claims for the purposes of applying the cited references in these reexamination proceedings. Hence, it follows that the USPTO should adopt the court's constructions as they already represent the broadest reasonable interpretation of these claims. See also, Declaration of Charles Nicholas, May 20, 2012 (hereinafter "Nicholas Dec.") at para. 23, explaining how the claims should be viewed from the standpoint of one of ordinary skill in the art.

The "User Model specific to the user" was construed as "an implementation of the learning machine updated in part by data specific to the user," and "user-specific data files" was construed as "the monitored user interactions with data and a set of documents associated with the user."[3]

The "parameters" of the learning machine were construed as "values or weights", and estimating those parameters was said to mean "estimating values or weights of the variables" of the learning machine, where "estimating" means "approximating or roughly calculating".[4]

The words "estimating a probability" were construed as "approximating or roughly calculating a numerical degree of belief or likelihood," and "estimating a probability P(u|d) that an unseen document d is of interest to the user u" was construed to mean "approximating or roughly calculating a numerical degree of belief or likelihood that an unseen document d is of interest to the user u given the information that is known about the unseen document."[5]

Thus, when given its broadest reasonable interpretation consistent with the specification, claim 1 can be understood as requiring, *inter alia*:

> approximating or roughly calculating values or weights of variables of a mathematical function and/or model used to make a prediction, that attempts to improve its predictive ability over time by altering the values/weights given in its variables, depending on a variety of knowledge sources, including monitored user

[2] *Cf. Multiform Desiccants, Inc. v. Medzam Ltd*., 133 F.3d 1473, 1477 (Fed. Cir. 1998) ("It is the person of ordinary skill in the field of the invention through whose eyes the claims are construed. . . The inventor's words that are used to describe the invention – the inventor's lexicography – must be understood and interpreted by the court as they would be understood and interpreted by a person in that field of technology.").

[3] *Markman Order, supra*, slip op. at 2.

[4] *Id.* at 1-2.

[5] *Id*. at 2.

Dockets.Justia.com

> interactions with data and a set of documents associated with the user,
>
> wherein the values or weights define an implementation of the mathematical function and/or model used to make a prediction, which implementation is updated in part by data specific to the user, and the values or weights are estimated in part from the monitored user interactions with data and a set of documents associated with the user;
>
> analyzing a document d to identify properties of the document; and
>
> approximating or roughly calculating a numerical degree of belief or likelihood that an unseen document d is of interest to the user u given the information that is known about the unseen document, wherein the numerical degree of belief or likelihood is approximated or roughly calculated by applying the identified properties of the document to the mathematical function and/or model used to make a prediction having the values or weights defined by the implementation.[6]

A number of these elements require increased attention.

First, consider the "numerical degree of belief or likelihood" that is calculated. This indicates that the present method is not one in which there are merely two possible outcomes of a decision, but rather one that can reveal a variation, gradation or range of interest on the part of a user. Nicholas Dec. at para. 24.

Second, consider the mathematical function and/or model used to make a prediction. Not just any model will suffice. It must be one that attempts to *improve* its predictive ability over time. This is done by altering the values/weights that define the implementation of the mathematical function/model in question, according to estimates of those values/weights derived from the monitored user interactions with data. Thus, the claim is concerned with a mathematical function/model that does more than merely memorize a user's past likes or dislikes. It is a mathematical function/model that generalizes or predicts an outcome.

The claim also requires that the mathematical function/model represent a user's interests independent of any particular query. This concept is succinctly explained in the specification, "The term P(u|d) represents the user interest in the document . . . and is calculated using the User Model." '040 Patent at col. 28, ll. 10-12, and see Nicholas Dec. at para. 25.

---

[6] Insofar as the limitations of claim 1 are common to claim 32, when they appear in claim 32 they should be understood in the same way as for claim 1. *Georgia-Pacific Corp. v. U.S. Gypsum Co.*, 195 F.3d 1322, 1331 (Fed. Cir. 1999) ("Unless the patent otherwise provides, a claim term cannot be given different meaning in the various claims of the same patent.").

Claim 1 then recites a method in which *a numerical degree of belief or likelihood that an unseen document is of interest to a user is approximated by applying identified properties of the document to a mathematical function/model that operates to make such a prediction and is defined by a collection of values/weights, where the values/weights are updated/altered, to allow the mathematical function/model to improve its predictive ability over time, according to monitored user interactions with data and a set of documents associated with the user*. Nicholas Dec. at para. 26.

With the above in mind, we turn now to the issues presented in the ACP. These issues will be addressed in the order presented therein.

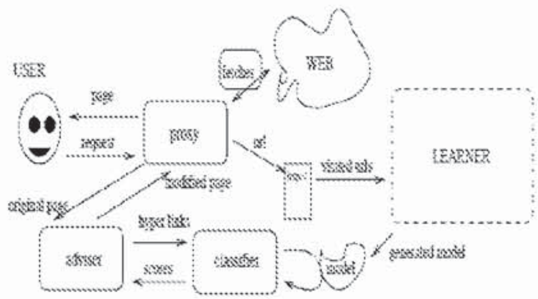
II. RESPONSE TO REJECTIONS BASED ON PRIMARY REFERENCE *MLADENIC*

A. *Mladenic* Teaches a Recommendation System Intended to Suggest Hyperlinks on Pages Requested by a User.

Mladenic's "Personal Web Watcher" ("PWW") is described as a personal "agent" that "looks over a user's shoulder" as he/she browses the Web, recommending hyperlinks it believes will be of interest. *Mladenic* at p. 3. Rather than analyzing subject documents to determine whether or not they would be of interest to a user, however, when a user is engaged in use of his/her computer PWW "actually predict[s] interestingness of [a] document based on the [extended representation of a] hyperlink pointing to it, and not on [the] document itself . . ." *Id*. at p. 10. The PWW's strategy for giving advice regarding hyperlinks is learned by revisiting pages from earlier web browsing sessions and incorporating the knowledge so gained within a model of user interests. *Id*. at p. 3. This takes place when the user (and the PWW advisor) is offline, typically at night. The updated model is then used the next time the user requests and views a web page, to provide advice concerning the links thereon. *Id*. and see Nicholas Dec. at para. 7.

In order to determine how best to make the hyperlink recommendations, *Mladenic* conducted experiments in which two different models of user interest were trialed. The first model, $User_{HL}$, was constructed using extended representations of hyperlinks (a representation that takes into account underlined words, words in a window around a hyperlink and words in all the headings above a hyperlink) that occurred on documents presented to the user (e.g., web pages the user actually visited) as training examples, and the model is intended to predict whether a new hyperlink (considered in its extended representation) is a positive or negative example of the user's interests. *Id*. The second model, $User_{DOC}$, was constructed using the documents behind (pointed to by) the hyperlinks on documents presented to the user as the training examples, thereby providing a model that predicts the interestingness of hyperlinks based on those documents. *Id*. at pp. 10-11. Importantly for the purposes of the present proceedings, both of these models produced and used by the PWW *predict the interestingness of hyperlinks and associated text and headings* on web pages that a user actually visits. Nicholas Dec. at para. 8.

As shown in *Mladenic's* Figure 2, reproduced below, the model(s) of user interests is employed when a user requests a web page.



In response to a request by the user, the web page is fetched and passed to an "adviser". The adviser extracts the extended representations of hyperlinks on the returned page and passes them to a classifier for comparison with the model of user interests.[7] The comparison results are returned to the adviser, which modifies the original web page to highlight suggested links when the page is finally presented to the user. *Id*. at pp.7-8, and see Fig. 3; Nicholas Dec. at para. 9.

The "learner" portion of the PWW is responsible for creating and updating the model of user interests and does so using the same information derived from an extended representation of a hyperlink that is used by the adviser when responding to a user request for a web page. *Id*. at p. 8. *Mladenic* describes two different ways of updating each model (the so-called "learning algorithm"): one using a k-Nearest Neighbor approach, and the other using a Naïve Bayesian

[7] Classification decisions (interesting/not interesting) are made based on weighted keyword vector representations of the hyperlinks. *Id*. at pp. 4-5.

Classifier. *Id.* at p. 11.[8] The k-Nearest Neighbor classifier is designed to determine Euclidean distances between a sample under test (in this case, a hyperlink or document behind a hyperlink being evaluated) and a specified training set (in this case, the models $User_{HL}$ and $User_{DOC}$). *Id.* at pp. 11-12. The Naïve Bayesian Classifier uses frequency vectors made up of keywords in the training data to generate the models of user interests such that the occurrence of certain words in the extended representation of the hyperlink described above indicates the probability (as estimated using Bayes Theorem) that the hyperlink belongs to one class or another. *Id.* at p. 7; Nicholas Dec. at para. 10.

B. Because *Mladenic* Does Not Teach Analyzing a Document to Identify Properties of the Document and Estimating a Probability that the Document is of Interest to a User by Applying the Identified Properties to a Learning Machine, *Mladenic* Cannot Anticipate Claims 1, 11, 32 and 34.

In the ACP, claims 1, 11, 32 and 34 were rejected under 35 U.S.C. §§ 102(a) and (b) as being anticipated by *Mladenic*. This was an error. In fact, *Mladenic* does not (indeed, cannot) anticipate any of the present claims.

"A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference." *Verdegaal Bros. v. Union Oil Co. of Californi*a, 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987) … See also MPEP § 2131.02. "The identical invention must be shown in as complete detail as is contained in the … claim." *Richardson v. Suzuki Motor Co.*, 868 F.2d 1226, 1236, 9 USPQ2d 1913, 1920 (Fed. Cir. 1989)…" Accordingly, "there must be no difference between the claimed invention and the reference disclosure, as viewed by a person of ordinary skill in the field of the invention." *Scripps Clinic & Research Found. v. Genentech, Inc.*, 927 F.2d 1565, 1576 (Fed. Cir. 1991). Here, this test is not met.

Mladenic is unequivocal in stating that the PWW "actually predict[s] interestingness of [a] document based on the hyperlink pointing to it, and not [on the] document itself." *Mladenic* at p. 10. The hyperlink (really its extended representation) on which the prediction is made is located on a current document that a user is viewing. *Id.* and see *Mladenic's* Fig. 3 showing

[8] Apparently, neither approach produced very good results – the k-Nearest Neighbor approach was said to yield no better than default accuracy if a negative class is predicted (that is, the approach yielded no better results than if one assumed that none of the links on a page would be of interest to a user), and the Naïve Bayesian approach provided even worse results. *Id.* at p. 12. Given these poor results, it is questionable whether anyone of ordinary skill in the art would ever adopt any of *Mladenic's* teachings when trying to solve the problem of determining hyperlinks (let alone actual unseen documents) of interest to a user. Nicholas Dec. at para. 12.

highlighted (recommended) links on a page currently being viewed in a browser. These links (in their extended representations) are classified using the model(s) ($User_{HL}$ and $User_{DOC}$) in order to determine whether or not they should be recommended. Both models are intended to predict interestingness of the presented link. *Id*. at pp. 10-11; Nicholas Dec. at para. 11.

Immediately then, one can recognize significant differences between *Mladenic's* PWW and the methods recited in the present claims. Claims 1 and 32 recite estimating a probability that an unseen document is of interest to a user by applying identified properties of the document, obtained from an analysis of the document, to a learning machine. *Mladenic*, on the other hand, describes analyzing a currently viewed web page to identify extended representations of hyperlinks on that web page, which extended representations of hyperlinks are used to determine whether the hyperlinks (not the currently viewed web page or any web page(s) pointed to by the hyperlink(s)) are of interest to the user. Nicholas Dec. at para. 11. Even assuming the models, $User_{HL}$ and $User_{DOC}$, can reasonably be considered implementations of "learning machines" (and this point is not conceded), the models do not (nor does the classifier that operates on the models) *approximate or roughly calculate a numerical degree of belief or likelihood that an unseen document is of interest to the user*, as required by the claims. Instead the classifier uses the models to "predict if a new hyperlink is a positive or negative example of the user interests." *Id*. at p. 10. The new hyperlink exists on a page the user is currently viewing and so neither that page nor the hyperlink thereon that is being evaluated is "unseen", and even if they were, it is the interestingness of the hyperlinks, not the document they are on nor the documents they point to, that is predicted. Nicholas Dec. at para. 12.

Mladenic then does not (and cannot[9]) teach the claimed feature of "estimating a probability P(u|d) that an unseen document d is of interest to the user u, wherein the probability

---

[9] Notice that the claim requires that "a" probability be estimated for "a" document. While "a" or "an" can mean one or more in patent parlance, where there is no indication that the inventors intended the indefinite article to have a meaning other than its normal singular meaning, courts construe it to be singular. *See, e.g., N. Am. Vaccine, Inc. v. Am. Cyanimid Co.*, 7 F.3d 1571, 1576-77 (Fed. Cir. 1993), holding that the fact that all references to a term in the specification were singular supported the argument that the term itself should be construed in the singular where the claim used the indefinite article "a".

In examining the '040 patent, one finds that all references to the estimate of probability of interestingness indicate that the inventors intended that a single numeric value be calculated. See, e.g., '040 patent at col. 28, ll. 5 et seq. in which P(u | d) is used in an expression to calculate P(u | q, d) and is said to represent *the* user interest - i.e., the single value for user interest. Hence, when construing the claims (even under its broadest *reasonable* interpretation), one must recognize that the estimate of probability is a single value.

P(u|d) is estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model". Accordingly, *Mladenic* does not anticipate claims 1 and 32, or any of their respective dependent claims, under 35 U.S.C. §§ 102(a) or (b).

C. *Mladenic's* Table 2 Does Not Demonstrate PWW Estimating a Probability that an Unseen Document is of Interest to a User.

At p. 9, in setting out the argument that *Mladenic's* PWW estimates the probability that an unseen document is of interest to a user as recited in the claims, the ACP identifies Table 2 at p. 12 (reproduced below). Presumably, although it is unstated, the ACP is identifying the column entitled "probability of interestingness" in support of this proposition. Such reliance is misplaced.

Mladenic's Table 2 reports "data characteristics" for certain Internet users[10] for the two user model formulations, $User_{HL}$ or $User_{DOC}$. *Mladenic* does not specifically address the "probability of interestingness" column (or any other items represented in the table for that matter), but what is clear is that the table is not reporting interestingness of an unseen document for a user as determined by the PWW. Indeed, hyperlinks are deemed interesting if and only if the user has already seen the documents to which they point. *Id*. at p. 8. Yet *Mladenic* expresses an interest in positive examples. *Id*. at p. 12; Nicholas Dec. at para. 13.

Mladenic's PWW, on the other hand, may provide conflicting values for the same "document". Consider that on any given web page, it is common to encounter multiple links to the same page or document. Because *Mladenic* "actually predict[s] interestingness of [a] document based on the hyperlink pointing to it, and not [on the] document itself", and because different hyperlinks that point to a common document will have different extended representations that are the basis upon which *Mladenic* predicts interestingness, it would not be uncommon to find situations where the PWW recommends some links and does not recommend other links, even though all of the links point to the same document. This is not providing "an" estimate of the probability of interestingness of the document, as claimed in the '040 Patent. It is, if anything, providing contradictory recommendations regarding such a document and cannot serve as a basis for finding the present claims anticipated by or obvious in view of *Mladenic*.

[10] The users were studied as part of a Carnegie Mellon University project called HomeNet, and not Mladenic's PWW project. *Mladenic* at p. 12.

UserId and data source	probability of interestingness	number of examples	data entropy
usr150101			
Doc	0.094	1 333	0.449
HL	0.104	2 528	0.480
usr150202			
Doc	0.107	3 415	0.492
HL	0.053	4 798	0.301
usr150211			
Doc	0.089	2 038	0.436
HL	0.044	2 221	0.259
usr150502			
Doc	0.100	1 272	0.468
HL	0.100	2 498	0.468

Table 2: Data characteristics for document (Doc) and hyperlink (HL) data for each of the four HomeNet users.

Consider, for example, that *Mladenic* has defined $User_{HL}$ and $User_{DOC}$ as yielding {pos, neg}. Any classification decision for a subject hyperlink (represented in its extended format) must be one of these two outputs, not a decimal measure. See *Mladenic* at pp. 10-11. So, if Table 2 does not report results provided by the PWW, what exactly does it report?

Based on Mladenic's own description at pp. 11-12, Table 2 is reporting nothing more than characteristics of the data that were used to train and evaluate the two user models.[11] Nicholas Dec. at para. 14. "Probability of interestingness" appears to be the probability of positive class values[12] in the data sets used to train and evaluate the HL or Doc models. That is, it is a report concerning the actual user behavior (clicked a link or not) represented in the data set used for training/evaluating these models.

Thus, the "probability of interestingness" relied upon by the ACP is not a result provided by the PWW and is instead simply a characteristic of data used to train/evaluate the PWW. Such a statistic says nothing about the ability of the PWW to estimate the interestingness of an unseen document to a user, and teaches one of ordinary skill in the art nothing with respect to determining the same. Nicholas Dec. at para 13. At best, it provides some insight into a set of training data that helps one understand the later classification and precision results reported in *Mladenic*, but this is an insufficient basis upon which to find the present claims unpatentable.

[11] Note, this is *not* the same as the characteristics of the results provided those models, which characteristics are reported in Mladenic's Figures 6-9, showing classification accuracies, and Figures 10-13, showing classification precision.

[12] Presumably, the probability was determined by considering the number of selected links in view of the total number of possible links which could have been selected in all of the available documents (reported in the "number of examples" column in *Mladenic's* Table 2).

D. Because *Mladenic* Does Not Teach Estimating a Posterior Probability P(u|d, q) that a Document, d, is of Interest to a User, u, Given a Query, q, *Mladenic* Cannot Anticipate Claim 11.

Unlike the WebWatcher, which "assist[ed] user[s] in locating information on the World Wide Web [by] taking keywords from the user, suggesting hyperlinks and receiving evaluation", *id*. at p. 1, and see *id*. at p. 2, the PWW does not involve the use of queries. Instead, the PWW observes HTML pages that are specifically requested by a user. *Id*. at p. 2, and see *id*. at p. 3 ("it doesn't ask the user for any keywords"). The addresses of these pages are stored, *id*. at p. 3, ll. 5-6, so that during the offline, learning phase, the PWW can revisit those pages in order to analyze them and generate/update the model of user interests. *Id*.

Mladenic, then, describes no query. Claim 11, on the other hand, specifically recites estimating a posterior probability that a document is of interest to a user, *given a query submitted by a user*. While the prior WebWatcher may have interacted with its users in the context of a query-response mechanism, reliance on such a statement in the context of an anticipation rejection is an error inasmuch as *Mladenic* specifically teaches a system that avoids such devices.

In setting out an anticipation rejection, an Examiner is not permitted to pick and choose various teachings from a cited reference:

> In an anticipation rejection, "it is not enough that the prior art reference . . . includes multiple, distinct teachings that [an ordinary] artisan might somehow combine to achieve the claimed invention."

Net MoneyIN, Inc. v. VeriSign, Inc., 545 F.3d 1359, 1371 (Fed. Cir. 2008). Rather, the reference must clearly and unequivocally disclose the claimed invention or direct those skilled in the art to the [invention] without any need for picking, choosing, and combining various disclosures not directly related to each other by the teachings of the cited reference. *In re Arkley*, 455 F.2d 586, 587 (CCPA 1972). While "[s]uch picking and choosing may be entirely proper in the making of a 103, obviousness rejection, . . . it has no place in the making of a 102, anticipation rejection." *Id. at* 587-88.[13]

[13] See also, *Lindemann Maschinenfabrik GMBH v. Am. Hoist & Derrick Co.*, 730 F.2d 1452, 1459 (Fed. Cir. 1984) ("The requirement that the prior art elements themselves be 'arranged as in the claim' means that claims cannot be 'treated . . . as mere catalogs of separate parts, in disregard of the part-to-part relationships set forth in the claims and that give the claims their meaning."), and *Net MoneyIN, supra,* 545 F.3d at 1371 ("[U]nless a reference discloses within the four corners of the document not only all of the limitations claimed but also all of the limitations arranged or combined in the same way as recited in the claim, it cannot be said to prove prior invention of the thing claimed and, thus, cannot anticipate under 35 U.S.C. § 102.").

Mladenic is quite clear that no query-response mechanism is employed by PWW ("it doesn't ask the user for any keywords"). *Id*. at p. 3. Consequently, *Mladenic* cannot anticipate claim 11.

E. Because *Culliss* does not Cure the Deficiencies of *Mladenic* with Respect to Estimating a Probability that the Document is of Interest to a User by Applying the Identified Properties to a Learning Machine, Claims 11, 22 and 34 Remain Patentable Over the Combination of *Mladenic* and *Culliss*.

Although *KSR*[14] may have loosened the required reasoning that may be employed for combining prior art references in an obviousness rejection, it remains the case that any such rejection under 35 U.S.C. § 103 must rest on a factual basis for finding each of the features of a rejected claim within the combined teachings of the cited references. See *In re Wilson*, 424 F.2d 1382 (CCPA 1970), MPEP 2143.03. If all of the elements of a claim are not so taught or suggested, no *prima facie* case of obviousness exists. *CFMT, Inc. v. Yieldup Int'l. Corp.*, 349 F.3d 1333 (Fed. Cir. 2003). Here, this test is not met in the combination of *Mladenic* and *Culliss*.

The *Culliss* reference is described in greater detail below. For the moment, however, it is important to recognize that *Culliss* does not disclose "estimating a probability P(u|d) that an unseen document d is of interest to the user u, wherein the probability P(u|d) is estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model" as recited in independent claims 1 and 32. In fact, *Culliss* cannot teach such a step because *Culliss* requires the existence of a query in order to produce any results. Nicholas Dec. at para. 20. Accordingly, these independent claims, and their respective dependent claims 11, 22 and 34, remain patentable over the combination of *Mladenic* and *Culliss*.

To better appreciate the above, consider that the claimed estimate of probability is the probability, P(u|d). This is specifically defined in the '040 Patent as "the user interest in the document *regardless of the current information need*". '040 Patent at col. 28, ll. 10-11

---

[14] *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 418 (2007).

(emphasis added). In other words, it is a probability determined without reference to a query (the query being indicative of an information need).[15]

In marked contrast, in the system described in *Culliss* only after a user enters a search query does the personal data that the system has stored about the user become combined with the query, and the revised query used to retrieve personalized search results. Nicholas Dec. at paras. 20-22. Absent the query, the *Culliss* system is unable to estimate any user interest in a document. See *Culliss* at col. 5:18-35.[16]

Combining the teachings of *Mladenic* and *Cullis* thus still does not yield the methods recited in independent claims 1 and 32. At a minimum, such a combination still would not yield the step of "estimating a probability P(u|d) that an unseen document d is of interest to the user u, wherein the probability P(u|d) is estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model", inasmuch as neither reference teaches or suggests such estimation. Accordingly, the claims are patentable over this combination of references.

Culliss' reliance on the existence of a query is also a reason why one of ordinary skill in the art would not combine its teachings with those of *Mladenic*. Nicholas Dec. at paras. 28-29. *Mladenic* chose to categorically reject approaches that required any user involvement. See *Mladenic* at p. 3 (the PWW "avoids involving the user in its learning process (it doesn't ask the user for any keywords or opinions about pages)"). This despite the fact that the author was well aware of the WebWatcher system that relied upon such actions in the form of search queries. *Id.* at p. 2. When evaluating the propriety of a rejection under 35 U.S.C. 103, one must be mindful to consider any prior art reference in its entirety, i.e., as a whole, including portions that would lead away from the claimed invention. *W.L. Gore & Associates, Inc. v. Garlock, Inc.*, 721 F.2d 1540, 220 USPQ 303 (Fed. Cir. 1983), *cert. denied*, 469 U.S. 851 (1984). Here, one of ordinary skill in the art would not be lead to incorporate teachings, such as those provided by *Culliss*,

---

[15] The Third Party Requester may allege that because the system discussed in *Culliss* works for every query, the system is somehow "query independent". This misses the point. It is not the subject of the query that matters, it is the existence of, indeed the need for, the query that is important when considering the *Culliss* reference in the context of the '040 Patent.

[16] As discussed in greater detail below, the previous-user relevant score (PRS) describd by *Culliss* is not an estimate of a probability either. Rather it is a mechanism to associate an article with one or more key terms. *Id.* at col. 3, ll. 31-37, 46-51. PRS may be a measure of the relationship between a query and articles based on the aggregated action of multiple users, but it is not an estimate of the probability that an unseen document d is of interest to the user u.

which require the exact thing which *Mladenic* is trying to avoid – user specification of keywords in the form of a search query. Accordingly, the very combination of references proposed by the ACP is not one which would be made and so the claims must be deemed patentable over these references.

F. Because *Refuah* does not Cure the Deficiencies of *Mladenic* with Respect to Estimating a Probability that the Document is of Interest to a User by Applying the Identified Properties to a Learning Machine, Claim 21 Remains Patentable Over the Combination of *Mladenic* and *Refuah*.

Refuah is discussed in greater detail below. For the present analysis, it is sufficient to recognize that *Refuah* does not estimate probabilities or interestingness, as recited in independent claim 1, from which claim 21 depends. Hence, even if the teachings of *Refuah* were combined with those of *Mladenic*, one would not arrive at the methods recited in claim 21.

Consider first that nowhere does *Refuah* discuss the concept of probabilities. Instead, "personalities" are used as a filter to manage information for a client. *Refuah* at col. 13:57-61; Nicholas Dec. at paras. 18-19. Stated differently, *Refuah* may be said to qualify a query based on a persona or mood, but this necessarily produces results that are query-dependent. It was demonstrated above that query-dependent results are incongruent with estimating the probability $P(u|d)$ as recited in the present claims and so this alone would be enough to find the present claims patentable over the combination of *Mladenic* and *Refuah*.

Even further, however, *Refuah* fails to teach estimating a probability that a retrieved or collected document is of interest to the user at all. By focusing on the phrase, "estimating a probability" in isolation, without considering the entire meaning of "*applying* the identified properties of the retrieved document *to the user-specific learning machine to estimate a probability* that the retrieved document is of interest to the user", the ACP misses the point. It is the application of identified properties of a retrieved document to the user-specific learning machine that is used to estimate a probability. There is no learning machine in *Refuah* in or to which to apply the identified properties of the retrieved document, and therefore *Refuah's* approach cannot be said to be that recited in the present claims.[17]

[17] *Refuah* describes using two factors, a persona and/or mood (collectively referred to as "personality"), to personalize web browsing on the Internet. *Refuah* at col. 13:55-59. This "personality" may be stored as one or more cookies on a user's computer, *id.* at col. 4., ll. 5-6, and is updated in response to one or more of the types and/or contents of sites that a client accesses. *Id.* at col. 22:6-8. Parameters in *Refuah* refer to "subject of interest = chess," "reject = pornography," or relative preference of subjects of interest, for example, "baseball=5," and "basketball=3." *Id.* at col. 6:49-55.

As discussed above, "applying the identified properties of the retrieved document to the user-specific learning machine to estimate a probability that the retrieved document is of interest to the user" has been construed as applying the identified properties of the retrieved document to a mathematical function (learning machine) specific to the user (user-specific learning machine) that attempts to improve its predictive ability over time by altering the values/weights given to its variable by approximating (estimating) a numerical degree of belief or likelihood (probability) that the retrieved document is of interest to the user. In *Refuah*, a site is evaluated in view of a particular persona, a snapshot view of a user's current interest (e.g., a query-defined interest) and not on either a mathematical function or model.[18] *Refuah* always makes evaluations in the context of the current information need as modified by the persona or mood. Thus, any combination of *Mladenic* and *Refuah* would require such a query-dependent process and would not suggest the subject matter recited in independent claim 1 or its dependent claim 21.

Further, *Refuah* is directed to evaluating a "website" or a "site" instead of "analyzing a document" to identify properties of the document or "estimating a probability" that an unseen document is of interest to the user. A website in *Refuah* is different from a document in the present claims. In fact, *Refuah* made this precise argument when describing his invention to the USPTO: "Herz describes identifying WWW [web] pages of interest based on their content. (Col. 67, lines 30-35). Hertz does not teach or suggest determining a trait of a site." See *Refuah* prosecution history, Amendment filed January 5, 2004, at p. 23. Even if sites and documents

---

In the concurrent litigation, the term "learning machine" was determined to mean a "*mathematical function* and/or model used to *make a prediction*, that attempts to *improve its predictive ability over time* by altering the values/weights given to its variables, depending on a variety of knowledge sources, including monitored user interactions with data and a set of documents associated with the user." Markman Order, pp. 2-3. In the Opinion accompanying the Markman Order, the court further explained that "the learning machine can reasonably be said to be a mathematical function." *Personalized User Model LLP v. Google Inc.*, Civ. No. 09-525-LPS, Claims Construction Opinion dated 25 Jan 2012, pp. 23-23, USDC D. Del. (hereinafter "Markman Opinion").

The use of cookies and subject categories in updating a persona and/or mood is not equivalent to estimating probabilities using a learning machine. The personality cookie described by *Refuah* is not and does not provide a "mathematical function and/or model used to make a prediction, that attempts to improve its predictive ability over time by altering the values/weights given to its variables, depending on a variety of knowledge sources, including monitored user interactions with data and a set of documents associated with the user."

[18] Indeed, *Refuah* describes a rudimentary approach in which a site is automatically evaluated by tracing the personas and/or moods of clients, *Refuah* at 20:31-34, and an even more manual approach where sites are evaluated by querying the operator thereof by way of email. *Id*. at col. 20, ll. 28-31.

were interchangeable, evaluation[19] of suitability as taught by *Refuah* is not equivalent to the estimation of probability as recited in the claims. The evaluation in the suitability as described by *Refuah* is not based on the content of a document (or a web page), but rather by a site-specific matching with a persona and/or mood. *Refuah* at 3:64 – 4:1, 6:23-26, and 17:44-48. Accordingly, there are numerous distinctions between the methods recited in the present claims and those taught by *Refuah* and, as a result, combining the teachings of *Mladenic* and *Refuah* would still not yield the presently claimed methods.

G. Because *Yang* does not Cure the Deficiencies of *Mladenic* with Respect to Estimating a Probability that the Document is of Interest to a User by Applying the Identified Properties to a Learning Machine, Claims 1, 11, 32 and 34 Remain Patentable Over the Combination of *Mladenic* and *Yang*.

The ACP cites *Yang* for certain teachings regarding a modified k-Nearest Neighbor algorithm to determine the relevance of an unseen document to a predetermined category. ACP at p. 11. According to the ACP, it would have been obvious for one of ordinary skill in the art to incorporate the teachings of *Yang* in this regard with the teachings of *Mladenic*, in part because *Mladenic* itself cites *Yang* as an example of a "learning algorithm" in the art. *Id*. at p. 12, and see *Mladenic* at p. 6. Utterly absent from the ACP, however, is any explanation as to what the resulting combination of these combined teachings would be. As such, the rejection itself is deficient as a matter of law. *In re Kahn*, 441 F.3d 977, 988, 78 USPQ2d 1329, 1336 (Fed. Cir. 2006) ("[R]ejections on obviousness cannot be sustained with mere conclusory statements; instead, there must be some articulated reasoning with some rational underpinning to support the legal conclusion of obviousness.").

When one actually considered what the resulting combination of *Mladenic* and *Yang* would be, the ACP's reliance on *Yang* for teaching the claimed estimating of a probability that a document is of interest to a user is seen to be wrong. Consider that the ACP appears to rely on *Yang's* formulation of a relevance score for a request:

$$rel(c_k \mid X) \approx \sum_{j=1}^{n} sim(X, D_j) \times P_r(C_k \mid D_j)$$

---

[19] The word "evaluation" has been defined as "the comparison of actual impacts against strategic plans. It looks at original objectives, at what was accomplished and how it was accomplished. It can be formative, that is taking place during the life of a project or organization, with the intention of improving the strategy or way of functioning of the project or organization. It can also be summative, drawing lessons from a completed project or an organization that is no longer functioning." http://en.wikipedia.org/wiki/Evaluation.

Yang at p. 16, eq. 3. But in the very next sentence *Yang* states: "*sim*(X, D_j) is a cosine value, *not* a probability. Consequently, *rel*(c_k|X) is *not* the probability of category c_k being related to request X." *Id*. at p. 16 (emphasis added). In other words, *Yang* itself refutes the ACP's contention that it teaches estimating probabilities and states explicitly that no such probabilities are estimated.[20]

Thus, *Yang* does not cure the underlying deficiencies of *Mladenic* concerning the claimed feature of estimating a probability P(u|d) that an unseen document d is of interest to the user u, wherein the probability P(u|d) is estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model, and the ACP's reliance on *Yang* is misplaced. The combination of *Mladenic* and *Yang* still would not include such teachings and so the claims are patentable over this combination of references.

III. RESPONSE TO REJECTIONS BASED ON PRIMARY REFERENCE WASFI

A. *Wasfi* Teaches a Recommendation System Intended to React to Perceived Changes in a User's Interests and to Make Recommendations for Web Pages at a Single Web Site Using Separate Content-Based and Collaborative Filters.

Wasfi describes ProfBuilder, a scheme for recommending web pages using separate content-based and collaborative filters, each of which creates its own recommendation list for pages of a single web site[21] that a user is currently visiting. See, e.g., *Wasfi* at Fig. 2, and pp. 61-62. The collaborative filter relies on navigation behaviors of past visitors to the web site to determine probabilities that a current user, given the current user's navigation history to that point, will visit a certain next page. *Id*. at p. 61. The content-based filter translates pages of the web site to a vector space representation used for a user profile and determines similarities

[20] *Yang* does indicate that a relevancy score computed using the cosine value described in eq. 3 may provide rankings similar to what probabilities might do, but this is not the point. Simply because two procedures may provide similar outcomes does not mean that one of ordinary skill in the art would be inclined to substitute one for the other. Quite the contrary; if anything, it suggests that persons of ordinary skill in the art may not seek to adjust the explicit teachings regarding the use of cosine values because there is no need to do so.

Indeed, from the literature cited in this reexamination proceeding, it appears that the accepted method for determining similarities is precisely the cosine method taught by *Yang*. *Mladenic* appears to adopt a similar approach (through use of a k-Nearest Neighbor approach) and so does *Wasfi* (see *Wasfi's* similarity calculation of vector scalar products at p. 61). If anything, this common reliance on a cosine value across different references suggests that it, rather than the use of probabilities, was the accepted practice of those of ordinary skill in the art at the time of the present invention and, as a result, the claimed methods are sufficiently different as to be non-obvious in view of these references.

[21] See *Wasfi* at p. 63, rt. col. ("Profbuilder assists a user by finding relevant information on only one Web site.").

between the pages and the profile. *Id.* Similarity is computed as the scalar product of weighted coefficients (specifically, keywords) of the page vector and the user profile vector. *Id.*; Nicholas Dec. at para. 14.

To ensure the similarity measures provided by the content filter accurately reflect a user's current interests, *Wasfi* proposes a "learning mechanism" for user profile reformulation. *Wasfi* at p. 58. The learning mechanism is based on adjusting a vector, Q_j, that represents the user according to:

$$Q_j = Q_j + t_{ij} * D_i$$

Id. at p. 58. In this expression, D_i represents a subject page (one which the user actually visits) in the same vector space as the user profile, and t_{ij} is a non-negative variable that represents the importance of the page to understanding changes in the user's interests. *Id.*; Nicholas Dec. at para. 14.

It appears that in order to ensure the user model is updated as rapidly as possible in response to changes in a user's interests, t_{ij} is defined in terms of what *Wasfi* styles the "entropy" of a page.[22] *Id.* at p. 59, left col. Because this entropy is a measure of unexpectedness, the more different a current page is from pages previously visited by the user, the better measure it is (according to *Wasfi*) of a change in the user's interest, hence, the greater effect it is given in updating the user model.[23] *Id.* at p. 58, rt. col. – p. 59, left col. The entropy of a given page of the subject web site is determined based upon the conditional probability of the user having arrived at that page. The probability distribution of pages is based on the collective actions of past users (i.e., the site navigation paths followed by those previous users) during their visits to the site. *Id.* at p. 59, left col. et seq.; Nicholas Dec. at para. 15.

---

[22] *Wasfi* defines entropy, H, of a probability, p, as $H(p) = -\log(p)$. *Id.* at p. 59, left col. This is a rather unusual formulation inasmuch as the classical (e.g., Shannon) definition of entropy of a discrete random variable, X, is $H(X) = -\sum_x p(x) \log p(x)$, where $X = \{x_1, x_2, \ldots, x_n\}$, and $p(x)$ is the probability of some x.

It is possible that *Wasfi* may have been trying to postulate the entropy of a page as the so-called *Wiener entropy*: $H_w(p) = -\log_2(p)$, where p is the probability of the event (here, presumably, a page) of interest. Nicholas Dec. at para. 15, fn. 1. If so, tij will not be bounded between 0 and 1, but instead will be bounded at infinity.

[23] In fact, because of this characteristic *Wasfi's* approach is subject to several shortcomings. Imagine, for example, a user arriving at a site for which no prior-user browsing history exists. *Wasfi's* methods are unable to provide any recommendations for our user because there is no basis for updating the user model as the user starts browsing the site. Indeed, in such circumstances, *Wasfi* is forced to assume that p is proportional to n_i/N, where n_i is the occurrence frequency of a page and N is the total number of page visits; thus, user model updates become nothing more than weighted page frequencies.