14. <u>Documents Correspond to Products</u>

361. Having retrieved documents that correspond to products, as disclosed in claim 7 of the '276 Patent, is also in the prior art. As an initial matter, PUM's infringement contentions as to AdWords do not require that <u>all</u> documents retrieved by AdWords correspond to products. For example, AdWords ads may correspond to charities, political campaigns, services, etc. (*See* Section VI.A.9, *supra.*) Rather, PUM contends that AdWords meets claim 7 because AdWords' retrieved documents <u>may</u> correspond to products.

362. Under PUM's interpretation of claim 7, virtually any prior art personalization reference also has documents that may correspond to products, as personalization references generally do not take steps to exclude product information from non-product information. *See, e.g.,* Culliss at 9:55- 10:13 (disclosing retrieving documents for "high heels"), Schuetze at 35:66 – 36:8 (disclosing retrieving documents that correspond to copiers), Refuah at 3:56 – 4:4 and 18:40-55 (disclosing retrieving advertisements, which may or may not correspond to products).

363. Even if the claim scope is limited exclusively to products, however, the limitation is still present in the prior art. For example, Refuah discloses "matching up of a supplier and a buyer, of a goods and/or a service" through its personalization techniques. (Refuah at 1:63 – 2:2.) Autonomy likewise describes services explicitly tailored to e-Commerce. (Autonomy WP at 10-11.) The patents themselves acknowledge that using personalization technologies for e-Commerce purposes was common in the art, pointing to Amazon.com's "Customers Who Bought" feature. (2:40-47.)

15. <u>Identifying Certain Properties in Documents</u>

364. Claim 22 of the '276 Patent requires that the identified properties of a document be at least one of the following:

a. a topic associated with the retrieved document,

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

c. an author of the retrieved document,

d. an age of the retrieved document,

e. a list of documents linked to the retrieved document,

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

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

365. Identifying at least one of these properties exists in the prior art. For example, Mladenic discloses identifying the headlines or topics associated with a retrieved document. (Mladenic at 4.) Schuetze discloses identifying a number of features, including the subject of the document and the links both to and from the document. (Schuetze at 6:58 – 7:15, 10:40-56.) Wasfi tracks and displays to the user the number of other users who have accessed the retrieved document. (Wasfi at 61.) Refuah discloses identifying a number of features, including the number of links on the document and the number of those links visited by users. (Refuah at 21:6-30.)

C. <u>The Combinations In the Asserted Patent Claims Are Predictable And Do Not Yield Any Unpredictable Results.</u>

366. The Supreme Court in *KSR* stated "[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 can implement a predictable variation, §103 likely bars its patentability." The Supreme Court also stated 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."

1. The Combinations In the Asserted Patents Are Predictable

367. Combining the elements of the asserted patents was predictable. The elements were available in combination and only with slight variations in the very same field of information retrieval. It is my opinion that this combination adds nothing to the nature and quality of each of the individual elements on its own, which I understand the Supreme Court has emphasized in KSR.

368. **Transparent monitoring**: As Loeb explains, there are two ways to receive feedback from users: directly—*i.e.*, requiring the user to enter additional information as to his preferences—or indirectly, *i.e.* deducing those preferences from the user's actions. (Loeb at 40-41.) There are also known benefits to using indirect or transparent monitoring to gather information. For example, Loeb observes that "casual users are not likely to be willing to engage in length interactions with the system in order to articulate current information needs and provide explicit feedback," and thus implicit means are needed to ensure their participation. (*Id.* at 41.) Similarly, Mladenic states that it gathers information from users transparently to "minimize users involvement in the learning process and enable learning without asking user for page rating." (Mladenic at 8-9.) Montebello observes that transparently monitoring is "more reliable than asking users to assign ratings, as it is less demanding on the user's time." (Montebello at 3.) Autonomy similarly notes that by not requiring manual intervention, it can dynamically track a user's needs and interests as they change over time. (Autonomy WP at 2.) Wasfi lists two types of direct learning as well as indirect learning, and gives a quick synopsis of the benefits of indirect learning. (Wasfi at 57.) Accordingly, it would have been obvious to one of ordinary skill to use transparent monitoring to obtain feedback from the user.

369. **Update user-specific data files**: Without direct feedback from a user, a system needs some manner of deciding whether the user's implicit data indicates interest in the

document or not. Mladenic employs a simple approach: documents viewed by the user were interesting to the user, and documents not viewed by the user were not interesting to that user. (Mladenic at 8.) Other systems are more complex, for example correlating interest with the amount of time the user spends reading the document. (Culliss at 2:43-46; Refuah at 5:34-50.) Regardless of the metric, it would have been obvious to one of skill in the art to store monitored interactions with the data so as to separate interesting documents from non-interesting documents.

370. For prior art personalization systems that rely on content analysis—such as the prior art systems presented in this report—the systems generally require access to the content of the corresponding preferred documents. As the systems need to access the content anyway when presenting the personalized services to the user, it would have been obvious to one of skill in the art to store the document content at that point rather than requiring a second, subsequent access to the content during the "analysis" phase. (*See, e.g.,* Mladenic at 8-9, Wasfi at 58, 60; Refuah at 5:34-50.)

371. **Estimating parameters of a learning machine based on the user-specific files**: One embodiment of a procedure for estimating parameters in a user model is to take a mathematical average of the user's documents. To use a simplistic example, suppose one were trying to determine how much ice cream a person wants to eat for dinner. A very simple means for estimating that preference would be to keep track of how much ice cream that person eats over 20 or 50 or 100 desserts, then average the amounts together to arrive at (say) half a cup of ice cream. That half-cup would represent the parameter that corresponds to the user's interest in ice cream: not as interested as someone who averages 1 cup, but more interested than someone who averages 1/4 a cup (or nothing). Doing this for many different kinds of foods yields a vector (a list of numbers) that defines a user profile.

372. The same "averaging" idea can be employed to derive a user profile from user documents. For instance, one can determine how important "Philadelphia Eagles" is to a document by simply counting the number of times that word appears.[22] (*See* 11:1-4.) One can do the same for every keyword in the document, i.e. count the number of times that "football" or "touchdown" or "flowers" appears in the document. These counts can be grouped into a keyword *vector*, which is simply all the scores for the keywords in a document. Figure 1 from Mladenic demonstrates such a vector using a short sample document, with "journal" appearing 3 times, "learning" not appearing at all, etc. (*See also* Wasfi at 58.)

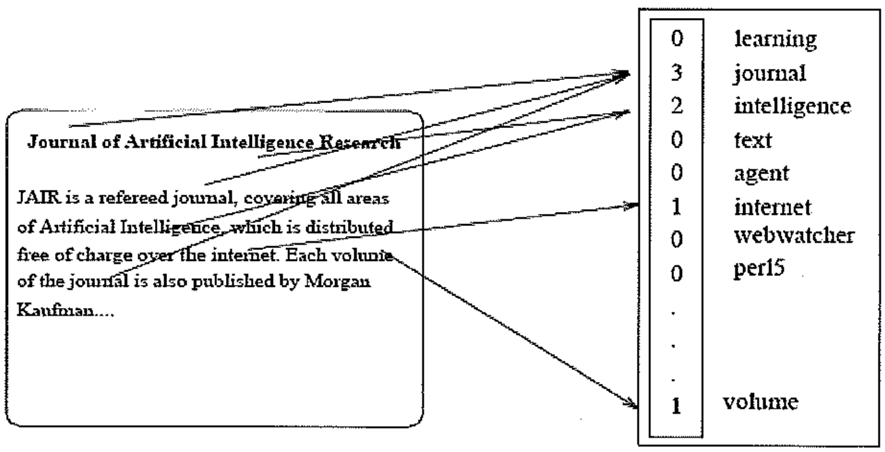

Figure 1: Bag-of-words representation using frequency vector.

Figure 24: Mladenic, Fig. 1

373. A user vector operates under the same "averaging" principle, using keywords instead of ice cream. By averaging the keyword vectors for a particular user, one determines an "average" document for that user that contains the keywords that the user tends to like. For

22 In practice, the "score" for a keyword is generally computed in a more complicated manner than its count. For example, Term Frequency / Inverse Document Frequency is a well-known metric that based the score not just on how often the word appears in the document, but on how infrequently the word appears in other documents. (11:12-20 (describing the prior art TFIDF metric); *see also* Mladenic 3.) The idea is that words that appear in lots of other documents (e.g., "person") are probably not that significant, whereas words that rarely appear in other documents (e.g., "Federal Court") probably are.

example a sports fan's documents would probably contain a lot of mentions of "football" or "baseball," and his user vector—that is, his "average" document—will similarly have high scores for "football" and "baseball."

374. Averaging keyword vectors from user documents was well-known in the art, and in fact was frequently practiced. See Section VII.B.3, *supra*. It would have been obvious to one of skill in the art to create a user model—that is, an "average" user document—from the user documents. (11:1 – 25:27; *see also* Fig. 4A.) Indeed, the patents themselves acknowledge that creating a user profile was known and readily apparent to those of skill in the 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. For example, U.S. Pat. No. 5,933,827, issued to Cole et al., discloses a system for identifying new web pages of interest to a user. ***The user is characterized simply by a set of categories,*** and new documents are categorized and compared with the user's profile. U.S. Pat. No. 5,999,975, issued to Kittaka et al., describes an online information providing scheme that ***characterizes users and documents by a set of attributes, which are compared and updated base on user selection of particular documents.*** U.S. Pat. No. 6,006,218, issued to Breese et al., discloses a method for retrieving information based on a user's knowledge, in which the probability that a user already knows of a document is calculated based on ***user-selected parameters or popularity of the document.*** U.S. Pat. No. 5,754,939, issued to Herz et al., discloses a method for identifying objects of interest to a user based on ***stored user profiles*** and target object profiles. 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.

(1:32-60.)

375. **Analyzing documents**: Prior art systems that rely on content analysis—like the systems cited in this report—unsurprisingly tend to examine the content of the document under consideration. As detailed above, content-matching systems usually create a user vector or user profile, which often takes the form of an "average" of the user's preferred documents. In order

to compare a new document with the "average" document, that new document needs to be in the same form as the "average" document. For instance, if the user profile consists of TFIDF scores for the keywords in the user's documents, then the new document should have its TFIDF scores computed as well. *See, e.g.,* Mladenic at 6 ("[a] new document is then represented as a vector in the same vector space as the generated model"), Wasfi at 58 ("[i]n the vector-space model, pages and queries (profiles) are both represented as vectors in some hyper-space.... [b]oth pages and profiles can then be represented as weight vectors.") Accordingly, it would obvious to one of skill in the art to analyze a prospective document in the same manner as the user documents, in order that the two may be compared.

376. **Estimating a probability**: Once the prospective document has been turned into a keyword vector, there needs to be some way of comparing it to the user profile—again, the "average" user document. Prospective documents that are "close" to the user profile will have similar words—e.g., "Philadelphia Eagles," "football"—to documents the user liked, and are thus likely to be interesting to the user. Prospective documents that are "far" from the user profile will not have similar words to other user-preferred documents, and are thus not likely to be interesting to the user.

377. There are various ways to compute the "scores" that correspond to the keywords in the user profile and in the prospective document. There are also various ways to compare the user profile and the document vector. Mladenic contains a concise summary of some of the known ways of representing documents and of comparing documents, including frequency representations, TFIDF representation, bigrams, mutual information, k-nearest neighbor, etc. (Mladenic at 3-6.) Some of those methods estimate probabilities, e.g. Naïve Bayes—the method ultimately employed in Mladenic. (*Id.* at 7; *see also* Autonomy WP at 4.)

378. To the extent PUM believes that "estimating a probability" can be met by any computation of a value, it would be obvious if not inherent to do so in content-matching system. As detailed above, both user profiles and prospective documents are often represented as vectors—a series of numbers that correspond to keywords within the documents. Computing or estimating some value with those numbers, whether by dot product, cosine computation, nearest-neighbor computation, etc. is the normal and expected means of comparing two vectors. Should "estimating a probability" be limited to actual probabilities, it would still be obvious to one of skill in the art to compute such estimates. A common way to do this is to transform the value obtained by dot product, cosine computation, via a "sigmoid function" that smoothly interpolates between zero and one. (See, e.g., Bishop, Mitchell and Cherkassky & Mulier).

379. The claims also require that the "document" be "unseen," which the Court construed as "document not previously seen by the user." (Order at 2.) 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. To the extent a prior art system can only estimate probabilities or compute scores for documents that were seen by the user, it would be obvious to one of skill in the art to extend that art to (at least occasionally) operate on documents not previously seen by the user. Indeed, one of the primary uses of the content-analysis systems described in this report is to analyze *content*, regardless of whether that content had been previously viewed. Further, there is little need to compute or estimate whether a particular user would find a document interesting if the user already saw the document; merely recalling his reaction to that document would suffice rather than implementing the computationally more expensive content analysis. Moreover, prior art references explicitly disclose the evaluation of unseen documents. (Mladenic at 8.)

380. Again, the patents themselves acknowledge that comparing an incoming document to a user profile was known and reading apparent to those of skill in the 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.*** For example, U.S. Pat. No. 5,933,827, issued to Cole et al., discloses a system for identifying new web pages of interest to a user. The user is characterized simply by a set of categories, and ***new documents are categorized and compared with the user's profile.*** U.S. Pat. No. 5,999,975, issued to Kittaka et al., describes an online information providing scheme that characterizes users and documents by a set of attributes, which are compared and updated base on user selection of particular documents. U.S. Pat. No. 6,006,218, issued to Breese et al., discloses a method for retrieving information based on a user's knowledge, in which the probability that a user already knows of a document is calculated based on user-selected parameters or popularity of the document. U.S. Pat. No. 5,754,939, issued to Herz et al., discloses a method for identifying objects of interest to a user based on stored user profiles and ***target object profiles.*** 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.***

(1:32-60.)

381. **Providing personalized services**: Using the computed probability to provide personalized services would have been obvious to one of skill in the art. Indeed, the whole point of comparing an incoming document to a user profile is to determine whether the user will find the incoming document interesting. (*See, e.g.,* Wasfi at 61: "The filtering process consists of translating pages to their vector space representation, finding pages that are similar to the profile, and selecting the top-scoring pages for presentation to the user"; Refuah at Abstract: "A method of a user interacting with an Internet, comprising: tracking interactions of the user with an Internet; analyzing said tracked interactions to determine at least one aspect of a user's interaction with the Internet; and modifying future interactions of said user with said Internet"; Schuetze at 1:29-33: "The invention relates to a recommendation system, and more particularly

to a system and method capable of providing document recommendations to a user based on various users' information browsing and retrieval histories.")

382. **Receiving query, retrieving documents based on query, probability of interest given query**: By 1999, search engines had become the preferred means of finding information on the Internet. Users could enter a few keywords on sites like Yahoo, Altavista, or Ask Jeeves, and receive a number of search results ranked according to how well they match the query. Given the popularity of search engines for finding information, it would have been obvious to one of skill in the art to design a personalization system to work with a search engine. Indeed, prior art systems like Montebello were explicitly designed to work with existing search engines. (Montebello at 1.)

383. Further, it would have been obvious to one of skill to rank search results according to both the output of the search engine and the predicted probabilities generated by a personalization engine. Absent the combination, the search results would either ignore the predicted probabilities—rendering the personalization worthless—or would ignore the match scores, which would result in *lower* quality results from the perspective of the user, since the search engine functionality would effectively be removed. Considering both the search engine matching and the probability of interest generated by the personalization system is nothing more than the combination of familiar elements according to known methods to yield predictable results.

384. **Interaction times**: As discussed above, transparent monitoring requires some means to distinguish documents the user considered "interesting" from documents the user considered "not interesting." One known means of making that distinction is to track how much time the user spends viewing a document. For example, if a user spends one second looking at a document, it is likely that the user was not interested in the content, and thus moved onto another

document. If a user spends a minute on the document, however, it is more likely the user considered the document interesting. Accordingly, tracking the length of time a user spends on a document would have been obvious to one of skill in the art, as it provides a means of distinguishing "interesting" documents from "not interesting" documents for training purposes.

385. **Analysis of documents with multiple media types**: Most web pages are written in HTML, a document formatting language that supports multiple media types. Even back in 1999, most web pages contained images, and some contained video or audio as well. It would have been obvious to one of skill in the art to analyze documents having multiple distinct media types, as such documents are the most common type of document on the Internet.

386. To the extent claim 34 of the '040 Patent is read to require analyzing the different media types, it would have been obvious to do so as well. Images on web sites are typically accompanied by text, either surrounding the image, as a caption for the image, or in the "ALT" tag for that image, which designates the text that appears if the image cannot be loaded. "The text surrounding or associated with an image often provides an indication of its context." (Schuetze at 4:26-27.) While image comparison software was not well developed in 1999, text analysis tools (e.g., keyword matching) were in widespread use. It would have been obvious to one of skill in the art to supplement image matching by matching the text relating to the images, as text analysis techniques were better developed and more accurate. (*See, e.g.,* Schuetze at 4:12-35.)

387. **Using a browser**: By 1999, web browsers were the preferred means of accessing content on the Internet. In fact, browsers were considered so important that Netscape and Microsoft fought a highly public "Browser War," with each seeking to become or remain the dominant web browser for the Internet. (*See, e.g.,* http://en.wikipedia.org/wiki/Browser_war.) It would have been obvious to one of ordinary skill to transparently monitor a user's use of a

browser, simply because that was the most common software used to examine documents on the Internet.

388. **Multiple modes of interaction**: As discussed above, transparent monitoring requires some means of deducing which documents the user considered interesting and which documents he did not find interesting. At the time of the alleged invention, there were multiple heuristics in place that offered such determinations, such as checking whether a user clicked on a link to the document, tracking how much time the user spent on the document, checking whether the user saved the document to his bookmarks, etc. It would have been obvious to one of skill in the art to combine multiple metrics in making that distinction. For example, saving a document to a bookmark list is a pretty strong signal of user interest, but it may not happen frequently enough to build a sufficient sample of user documents. One of skill in the art would know that he could gather the user's bookmarked documents, and supplement the sample with viewed documents if needed.

389. **Documents not of interest**: As discussed above, transparent monitoring requires some means of distinguishing documents of interest from documents not of interest. Some user models simply use the documents of interest and ignore the documents not of interest, under the logic that it is more important to know what the user likes than to know what the user does not like. However, it would be obvious to one of skill in the art to also adjust the user profile based on documents the user does not like, particularly if the "not of interest" documents are those the user never wants to see (e.g., pornography, racist literature, etc.) (*See, e.g.,* Refuah at 6:51-53.)

390. **Monitoring certain user interactions**: As discussed above, transparent monitoring requires some means of deducing which documents the user considered interesting and which documents he did not find interesting. At the time of the alleged invention, there were multiple heuristics in place that offered such determinations, including checking whether the user

viewed the document, checking how much time the user spent viewing the document, and checking whether the user followed links on the document. (*See* Section B.13.) It would have been obvious to one of skill in the art to employ any of those means to determine whether the document was of interest to the user. Doing so would constitute nothing more than the combination of familiar elements according to known methods to yield predictable results.

391. **Documents correspond to products**: To the extent claim 7 of the '276 Patent requires that the documents correspond to advertisements or products, it would have been obvious to one of skill in the art to do so. Targeted advertising was a well-known concept in 1999. Furthermore, some search engines—such as Goto.com—were entirely composed of advertisements. As it would have been obvious to one of skill in the art to combine a personalization system with search engines, so too would it have been obvious to combine a personalization system with any given search engine, including Goto.com. Doing so would constitute nothing more than the combination of familiar elements according to known methods to yield predictable results.

392. Furthermore, e-commerce was a growing field in the internet. As even the patents acknowledged, e-commerce sites employed various filtering techniques "primarily for increasing the number and size of customer purchases." (2:38-39.) It would have been obvious to one of skill in the art to add a personalization system to an e-commerce site, as that would achieve the predictable result of showing more interesting items to customers and thus increasing sales.

393. **Identifying certain properties of docs**: Claim 22 of the '276 patent requires that the document analysis identify certain properties of the document. As Mladenic explains, many document analysis algorithms only consider the words of the document. (Mladenic at 4.) However, it was known that extracting additional features from the HTML of the document—including the document's headlines or topics—may improve system performance. (*Id.*)

Accordingly, it would be obvious to one of skill in the art to extract additional information from the document, beyond its keywords, in order to improve the accuracy of the system. Doing so would constitute nothing more than the combination of familiar elements according to known methods to yield predictable results.

(a) **Mladenic and Personal Web Watcher**

394. As detailed in Section VI.A, Mladenic anticipates all asserted claims of the '040 Patent except claim 22, which requires the "the monitored user interactions include a sequence of interaction times." In logging user interactions and other events in a computing system, it has long been standard practice to include a record of times ("temporal tags") in the log. While this is implicit in Mladenic, it is made explicit in Mladenic's follow-up conference paper, *Machine Learning for Better Web Browsing*, where Figure 1 contains a timeline:

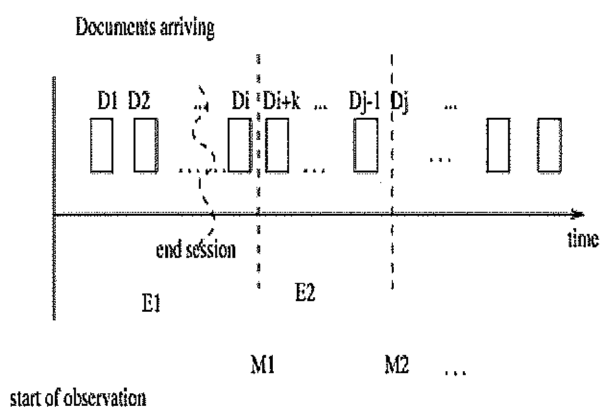

Figure 1: Illustration of the data collection process while the user is browsing the Web. Text documents D_1, D_2, ... are arriving over time as request by the user.

(Machine Learning for Better Web Browsing at 2, figure 1). Furthermore, tracking interaction times was a known means for determining user interest in a document. *See* Section VII.B.8, above. Accordingly, it would have been obvious to one of skill in the art to use interactions times to provide another measurement of the user's interest in a document.

395. Similarly, Mladenic anticipates all asserted claims of the '276 Patent except claim 3. Mladenic does not disclose monitoring multiple modes of user interaction; it only discloses monitoring whether the user clicked on or viewed a document. However, monitoring multiple modes of user interactions would be obvious to one of skill in the art. For example, Montebello discloses monitoring whether users saved documents to their "bookmarks" or "favorites" (Montebello at 3), while Culliss discloses monitoring the amount of time a user spent on the document. (Culliss at 2:43-46.) As discussed above, it would be been obvious to one of skill in the art to further refine the "documents of interest" determined by Mladenic to include additional indicia of interest. In fact, several prior art references explicitly included monitoring multiple modes of interaction for that reason. (Schuetze at 18:8-16.)

396. Personal WebWatcher discloses all the limitations of Mladenic except for receiving and processing a search query. While that limitation is disclosed in Mladenic itself through its discussion of the predecessor WebWatcher system, that element is not present in Personal WebWatcher itself. Nonetheless, it would be obvious to one of skill in the art to use Personal WebWatcher with a search engine, *e.g.* as disclosed in Montebello (Montebello at 1-2, explicitly combining Mladenic with search engines). Furthermore, the Mladenic reference itself discloses the needed functionality, so one of skill in the art would be motivated to modify Personal WebWatcher based on the teachings of the paper that describe Personal WebWatcher.

(b) <u>Autonomy</u>

397. As detailed in Section VI.B, Autonomy anticipates all asserted claims of the '040 Patent except claim 22, which requires the "the monitored user interactions include a sequence of interaction times." In logging user interactions and other events in a computing system, it has long been standard practice to include a record of times ("temporal tags") in the log.

Furthermore, tracking interaction times was a known means for determining user interest in a document. *See* Section VII.B.8, above. Accordingly, it would have been obvious to one of skill in the art to use interactions times to provide another measurement of the user's interest in a document.

398. Similarly, Autonomy anticipates all asserted claims of the '276 patent except claims 5 and 6. Claim 5 requires analyzing documents *not* of interest to the user in addition to documents that are of interest to the user. As disclosed in section VII.B.12, this is a known means of using additional data to better refine the user profile. Claim 6 requires one of various means of transparently monitoring user interactions. While the Autonomy references are silent as to the specifics of the monitoring, one of skill in the art would expect that they be one or more of the known methods described in claim 6, as disclosed in section VII.B.13. To the extent that Autonomy does not include one of those methods, it would have been obvious to one of skill in the art to, say, determine whether a user had viewed a document. Accordingly, both claims 5 and 6 of the '276 Patent would have been obvious in light of Autonomy.

(c) <u>Montebello and PEA</u>

399. As detailed in Section VI.C, Montebello/PEA anticipates all asserted claims of the '040 Patent except claim 22, which requires the "the monitored user interactions include a sequence of interaction times." In logging user interactions and other events in a computing system, it has long been standard practice to include a record of times ("temporal tags") in the log. Furthermore, tracking interaction times was a known means for determining user interest in a document. *See* Section VII.B.8, above. Accordingly, it would have been obvious to one of skill in the art to use interactions times to provide another measurement of the user's interest in a document.

400. Similarly, Montebello/PEA anticipates all asserted claims of the '276 Patent except claims 3 and 5. Montebello does not disclose monitoring multiple modes of user interaction; it only discloses monitoring whether the user bookmarked a document. However, monitoring multiple modes of user interactions would be obvious to one of skill in the art. For example, Mladenic discloses monitoring whether users viewed documents (Mladenic at 8), while Culliss discloses monitoring the amount of time a user spent on the document. (Culliss at 2:43-46.). As discussed above, it would be been obvious to one of skill in the art to further refine the "documents of interest" determined by Montebello/PEA to include additional indicia of interest. In fact, several prior art references explicitly included monitoring multiple modes of interaction for that reason. (Schuetze at 18:8-16.)

401. Claim 5 requires analyzing documents *not* of interest to the user in addition to documents that are of interest to the user. As disclosed in section VII.B.12, this is a known means of using additional data to better refine the user profile. Accordingly, both claims 3 and 5 of the '276 Patent would have been obvious in light of Montebello/PEA.

(d) <u>Wasfi and ProfBuilder</u>

402. As detailed in Section VI.D, Wasfi/ProfBuilder anticipates all asserted claims of the '040 Patent except claim 11, as Wasfi/ProfBuilder does not disclose accepting a search query. Since it lacks the search query element, Wasfi/ProfBuilder also does not anticipate any of the claims of the '276 Patent. However, all of those claims would have been obvious in light of Wasfi.

403. Claim 11 of the '040 patent requires estimating a posterior probability that the document is of interest to a user given a query. While Wasfi does not disclose accepting queries, one of skill it would have found it obvious to do so. In discussing means to build the user profile, Wasfi itself discloses "direct learning technique" where "[t]he user provides a set of

keywords to describe his/her interests." (Wasfi at 57.) In the same section, Wasfi discloses "indirect or transparent learning technique," where "[t]he system learns user preferences transparently without any extra effort from the user." (*Id.*) Wasfi further discloses mixing different means of obtaining user information "to get a compromise model between user effort and predictability." (*Id.*) While Wasfi does not explicitly disclose combining keywords with a user profile built on transparent monitoring, it would have been obvious to one of skill in the art to do so. Indeed, Wasfi explicitly discloses systems that use other combinations of user profile building, such as Antagonomy (which allows for both queries and explicit ratings) and SiteHelper and WebWatcher (which allow for both explicit ratings and transparent monitoring.) (*Id.* at 57-58.) Further and as discussed above, search engines were the preferred means to retrieve document content by 1999.

404. Claim 1 of the '276 patent has similar limitations as claim 1 of the '040, save that it also requires accepting and using search queries, as well as monitoring normal use of a browser. Wasfi discloses monitoring normal use of a browser: "Users navigate in the collection by using any navigation technique such as selecting hypertext links, specifying page addresses, or selecting pages from interest lists." (Wasfi at 58.) While Wasfi does not disclose using a search query, it would have been obvious to one of skill to add this feature to ProfBuilder, as described for claim 11 of the '040 Patent above.

405. Claim 3 of the '276 Patent 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." Wasfi monitors multiple user interactions with network data, including "selecting hypertext links, specifying page addresses, or selecting pages from interest lists." (Wasfi at 58.) Accordingly, claim 3 would have been obvious in light of Wasfi/ProfBuilder.

406. Claim 5 of the '276 Patent requires analyzing documents *not* of interest to the user in addition to documents that are of interest to the user. As disclosed in section VII.B.12, this is a known means of using additional data to better refine the user profile.

407. Claim 6 of the '276 Patent 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." Wasfi discloses at least information about whether a user visited the document: "To read the contents of the page, the user clicks on its title. Titles in 'bold' font indicate unread pages, while titles in 'normal' font indicate pages have been read by the user." (Wasfi at 61.) Accordingly, claim 6 would have been obvious in light of Wasfi/ProfBuilder.

408. Claim 7 of the '276 Patent 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 Wasfi/ProfBuilder 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 Wasfi may also correspond to products. Accordingly, Wasfi/ProfBuilder discloses this claim element under PUM's own infringement theories.

409. Claim 21 of the '276 Patent 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." While Wasfi does not disclose using a search query, it would have been obvious to one of skill to add this feature to ProfBuilder, as described for claim 11 of the '040 Patent above. Furthermore, the search query described in Wasfi would be used to augment the user's profile, since both it and the transparently monitored actions would modify the profile. Computed similarity scores would thus account for both the query and the original "profile" (based on transparently monitored interactions). Accordingly, claim 21 would have been obvious in light of Wasfi/ProfBuilder.

410. Claim 22 of the '276 Patent 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." Wasfi discloses at least identifying the number of users who have accessed the document: "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.... The number of men shows level of ***access frequency*** logarithmically: one-man pages are visited once, two-man pages visited two to three times, three-man pages are visited four to seven times, four-man pages are visited eight to fifteen times, and so on." (Wasfi at 61.) Wasfi further discloses identifying the topic of the retrieved document: "The vector representation is obtained by a text analysis of HTML pages. This is done by

extracting keywords from ***page titles. all level of headings***, and anchor hypertext." (*Id.*) Accordingly, claim 22 would have been obvious in light of Wasfi/ProfBuilder.

(e) Culliss

411. As detailed in Section VI.E, Culliss anticipates all asserted claims of the '040 Patent. Culliss also anticipates all asserted claims of the '276 Patent except claim 5. Claim 5 requires analyzing documents *not* of interest to the user in addition to documents that are of interest to the user. As disclosed in section VII.B.12, this is a known means of using additional data to better refine the user profile. Accordingly, claim 5 of the '276 Patent would have been obvious in light of Culliss.

(f) Refuah

412. As detailed in Section VI.F, Refuah anticipates all asserted claims of the '040 Patent. Refuah also anticipates all asserted claims of the '276 Patent.

(g) Joachims and WebWatcher

413. As detailed in Section VI.G, Joachims/WebWatcher anticipates all asserted claims of the '040 Patent except claim 22, which requires the "the monitored user interactions include a sequence of interaction times." In logging user interactions and other events in a computing system, it has long been standard practice to include a record of times ("temporal tags") in the log. Furthermore, tracking interaction times was a known means for determining user interest in a document. *See* Section VII.B.8, above. Accordingly, it would have been obvious to one of skill in the art to use interactions times to provide another measurement of the user's interest in a document.

414. Similarly, Joachims/WebWatcher anticipates all asserted claims of the '276 patent except claims 3 and 5. Joachims/WebWatcher does not disclose monitoring multiple modes of user interaction; it only discloses monitoring whether the user clicked on or viewed a document.

However, monitoring multiple modes of user interactions would be obvious to one of skill in the art. For example, Montebello discloses monitoring whether users saved documents to their "bookmarks" or "favorites" (Montebello at 3), while Culliss discloses monitoring the amount of time a user spent on the document. (Culliss at 2:43-46.) As discussed above, it would be been obvious to one of skill in the art to further refine the "documents of interest" determined by WebWatcher to include additional indicia of interest. In fact, several prior art references explicitly included monitoring multiple modes of interaction for that reason. (Schuetze at 18:8-16.)

415. Claim 5 requires analyzing documents *not* of interest to the user in addition to documents that are of interest to the user. As disclosed in section VII.B.12, this is a known means of using additional data to better refine the user profile. Accordingly, both claims 3 and 5 of the '276 Patent would have been obvious in light of Joachims and WebWatcher.

416. Joachims does not directly disclose claim 21 of the '276 Patent, which requires "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." However, 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. (Mladenic at 2.) It would have been obvious to one of skill in the art to combine the teachings of Joachims with the teachings of Mladenic, as both describe the same WebWatcher system.

2. <u>**The Combinations In the Asserted Patents Do Not Yield Unpredictable Results**</u>

417. As the Supreme Court observed in *KSR*, "[t]he combination of familiar elements according to known methods is likely to be obvious when it does no more than yield predictable results." In my opinion, there is nothing unpredictable that results from combining the elements

of the asserted patents. Each of the various elements was well known in the prior art, and their combination introduces nothing new. Indeed, many of the prior art references disclose goals and advantages very similar to those claimed in the asserted patents, as discussed above.

418. For example, Green discloses that "[t]he recent, rapid growth of the Internet has led to enormous amounts of on-line information. However, as the volume of this information has increased, so have the problems encountered by users in dealing with it." (Green at 1). Autonomy discloses that "[a]s the amount of unstructured text available to users explodes, companies are in even greater need of systems that can automate content management and distribution while reducing the level of manual effort required to get the right information to the right people at the right time." (Autonomy WP at 11). Likewise, Mladenic observes that "[w]ith the growing availability of information sources, especially non-homogenous, distributed sources like the World Wide Web, there is also a growing interest in tools that can help in making a good and quick selection of information we are interested in." (Mladenic at 1). Refuah aims to "provide a method of aiding information search and retrieval on the internet" where internet searching is "personalized to a particular user's profile" or involves "matching up of a supplier and a buyer, of a [sic] goods and/or a service." (Refuah 1:63-2:1).

D. One Skilled In The Art Would Have Been Motivated To Pursue The Claimed Combinations Through Market Forces And Trends

419. In *KSR*, the Supreme Court also observed, that "when 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 grasp. If this leads to the anticipated success, it is likely the product not of innovation but of ordinary skill and common sense. In that instance the fact that a combination was obvious to try might show that it was obvious under sec. 103."

420. Here, one of ordinary skill would have been motivated to pursue the claimed combinations, as market forces had already revealed the benefits of personalizing document retrieval. Successful systems like Autonomy occupied the market space, showing the benefits of personalization, and others were already researching the same concepts even earlier. Several of prior art references begin by reviewing the extensive work already done is the personalized information services space. (*See, e.g.,* Mladenic at 1-2, Edwards at 32, Green at 1-4, etc.). Moreover, the patent itself acknowledges other similar attempts at personalization, including Amazon recommendation services. ('040 Patent at 2:37-52). The market incentive for personalization is simple: by showing users documents that are more interesting to them, a web site can increase the amount of time the user spends on the site, which in turn increases web traffic, advertising revenue, and ecommerce revenue.

421. The application of inferential concepts, including statistical pattern recognition, mathematical modeling, stochastic modeling, machine learning, parameter estimation, and Bayesian posterior inference, to the problem of personalized information services was already an established trend by 1999. Konig himself admits that he applied these techniques at his previous employer, SRI, and other employees of SRI have testified that these same concepts/techniques were widely known and applied prior to the filing of the patents-in-suit.

422. From reviewing deposition transcripts, I understand that Mr. Konig applied statistical pattern matching (Konig Depo Tr. at 294:4-6), stochastic modeling (294:7-9, 305:16-25), neural networks (294:10-19), Bayesian learning algorithms (294:20-23), machine learning (305:1-15, 307:1-4) while at SRI. Moreover, Konig stated that several other concepts including pattern classification and posterior probabilities were described in a "classical book" on the subject of machine learning that he cited in a paper written while at SRI. (Konig Depo Tr. at 506:11-508:9). Other SRI employees confirm this: Mr. Sonmez testified that Mr. Konig was

involved in research at SRI that utilized Bayesian statistics and specifically posterior probabilities (Somnez Depo Tr. at 43:1-46:4); Mr. Stolcke testified that the techniques of models for individuals, neural networks, and Bayesian learning algorithms were all used by Konig at SRI in his voice recognition project before the conception of the patents-in-suit. (Stolcke Depo Tr. at 18:12-19:25).

423. The testimony of the other SRI employees corroborates the widespread use of these techniques. For example, Mr. Sonmez testified that statistical pattern recognition, mathematical modeling, stochastic modeling, machine learning, parameter estimation, and Bayesian posterior inference are all techniques that were generally known and practiced by experts in the field of machine learning. (Sonmez Depo Tr. at 100:22-102:10). Moreover, Mr. Stolcke testified that the machine learning, statistical pattern matching, and stochastic modeling techniques used by Mr. Konig at SRI are all general terms in the field and applicable to any of the research SRI was carrying out. (Stolcke Depo Tr. at 16:11-18:9). Mr. Konig had suggested future research possibilities to SRI involving document classification, unsupervised learning, and clustering (Somnez Depo Tr. at -47:4-52:13; see also Stolcke Ex. 4; see also Stolcke Depo Tr. at 28:4-32:3). Mr. Stolcke also confirmed that each of these techniques was widely known in the field prior to 1994. (Stolcke Depo Tr. at 74:8-77:1).

E. The PTO Has Rejected All Asserted Claims on Reexamination

1. The '040 Patent

424. I understand that, on or about May 27, 2011, the United States Patent & Trademark Office issued a non-final Office Action rejecting all asserted claims of the '040 patent on the bases set forth in Google's request for *inter partes* reexamination. (*See* 5/27/2011 Office Action, Application No. 95/001,569.)

425. With respect to the asserted claims, the PTO found that claims 1, 11, 32, and 34 are invalid under 35 U.S.C. §§ 102(a) and (b) as being anticipated by Mladenic, *Personal WebWatcher: design and implementation* (Chart C-5); claims 1, 22, and 32 are invalid under 35 U.S.C. §§ 102(a) and (b) as being anticipated by Wasfi, Wasfi, *Collecting User Access Patterns for building User Profiles and Collaborative Filtering*, Proceedings of the 4th International Conference on Intelligent User interfaces. (January 1999) (Chart C-10); claims 1, 11, 22, 32, and 34 are invalid under 35 U.S.C. § 102(e) as being anticipated by Refuah '032 (Chart C-8); and claims 1, 11, 22, 32, and 34 are invalid under 35 U.S.C. § 102(e) as being anticipated by Cullis '068 (Chart C-1).

426. The Examiner further found that each of the asserted claims is invalid under 35 U.S.C. § 103(a) as obvious over Mladenic, Wasfi, Refuah, and Culliss. (*See* 5/27/2011 Office Action, Application No. 95/001,569.) The Examiner noted that it would be obvious to combine the elements of these references for the reasons set forth in Google's request for *inter partes* reexamination. (*Id.*)

427. For example, the Examiner noted that claim 11 is obvious over Mladenic in view of Culliss because personalized document recommendations disclosed in Mladenic are drawn from the Internet as a whole. But, it would be just as feasible (and obvious) to use Personal WebWatcher to analyze a pool of search results generated in response to a user's query. (*Id.*, at 13.)

428. The Examiner further noted that combining Mladenic and Culliss to modify Personal WebWatcher to include an analysis of user interaction times in judging which documents would be of most interest to a user (claim 22), would be "the mere application of a known technique to a known system ready for improvement and would yield a predictable result." (*Id.*, at 14.)

429. The Examiner similarly found that it would have been obvious to combine Culliss with Mladenic or Wasfi such that those systems could be used to analyze web pages of multiple media types (claim 34), because by the time the '040 patent was filed, "it was well known that web pages could contain the multiple types of media disclosed by Culliss." (*Id.*, at 21.)

430. The Examiner further found that combining elements of Refuah and Mladenic "would merely have involved fusing two known pieces of prior art, each retaining its ordinary and established function," and this combination would have been "obvious to try" and within the grasp of a person of ordinary skill in the art. (*Id.*, at 27.) The same is true of combining Culliss and Mladenic. (*Id.*, at 33-34.)

2. The '276 Patent

431. I understand that, on or about February 10, 2012, the United States Patent & Trademark Office issued an Action Closing Prosecution (ACP) rejecting all asserted claims of the '276 patent.

432. With respect to the asserted claims of the '276 patent, the Examiner found that claims 1, 3, 5, 6, 21, and 22 were obvious in light of Wasfi in view of Mladenic. In particular, the Examiner agreed that Wasfi discloses a user-specific learning machine, as if discloses "content-based filtering that is performed by comparing the vector-space representations of pages with the vector-space representations of user interests." (ACP at 25.) The Examiner further agreed that "calculating the similarity between a given document an a user profile… is an indication of the probability that the given document is of interest to the user." (*Id.* at 26.) The Examiner further found that claim 7 was obvious in light of Wasfi in view of Mladenic and Refuah.

433. The Examiner also found that all asserted claims of the '276 patent were anticipated by Refuah. In particular, the Examiner found that "personas and moods in Refuah

need not be manually selected by a user, but rather may be updated based on a user's activities," and thus that Refuah discloses estimating parameters of a user-specific learning machine. (ACP at 29.) The Examiner also found that "in the 'evaluation technique' described by Refuah (Refuah col. 17, lns. 44-46), evaluating a site for suitability and/or qualities which are preferred by a particular persona involves estimating the probability that the site is of interest to the user." (*Id.* at 31.) The Examiner similarly found that "Refuah's use of thresholding and grading, and determining strong matches, is further evidence of estimating a probability, or likelihood, that a document is of interest to the user." (*Id.* at 32.)

434. In addition, the Examiner found that asserted claims 1, 3, 6, 7, 21, and 22 were anticipated by Culliss. In particular, the Examiner found that Culliss discloses transparent monitoring because a user's personal data "can be inferred from a history of their search requests or article viewing habits." (ACP at 35.) The Examiner also found that "the calculations in Culliss are consistent with estimations as disclosed by the '276 patent" and that "[u]sing mathematical formulas to calculate a score does not preclude those calculations from being considered estimates," and thus Culliss teaches estimating parameters of a user-specific learning machine. (*Id.* at 37.) The Examiner also found that Culliss "determines how relevant a given document is to the searching user based on the document's relevancy score," that "the relevancy scores, even if directly computed, serve to estimate a probability, or likelihood, that the document is relevant to the user" and thus "Culliss teaches estimating a probability that the retrieved/collected document is of interest to the user." (*Id.* at 39.) The Examiner also found that claim 5 was rendered obvious by Culliss in view of Mladenic, and specifically found that "both Mladenic and Culliss are in the field or providing personalized information recommendations to Internet users" and thus in the same field of art. (*Id.* at 41.)

435. The Examiner also found that asserted claims 1, 3, 6, 7, 21, and 22 were anticipated by Montebello. More specifically, the Examiner found that "extracting features of a document, comparing them against a user's profile, and suggesting documents to a user that fit the user's interest"—as Montebello does—meets the limitation of "applying identified properties of retrieved/collected documents to a user-specific learning machine to estimate the probability that the retrieved document is of interest to the user." (ACP at 44.) The Examiner also found that claim 5 was rendered obvious by Montebello in view of Mladenic. (*Id.* at 45.)

436. Of note, the Examiner found that claims 1, 5, 6, 21, and 22 were ***not*** anticipated by Mladenic. The Examiner agreed that Mladenic teaches the "transparently monitoring" limitation because it "learns by observing user actions rather than requiring explicit feedback from the user." (ACP at 17.) In the same vein, the Examiner rejected the patent owner's argument that highlighting links in a document violated transparently monitoring, as that "relates to how Mladenic *presents* interesting hyperlinks to the user, not how Mladenic *monitors* user behavior to learn user interests," and since the '276 Patent describes highlighting links as a preferred embodiment. (*Id.*) The Examiner also found that "the scored word map of Mladenic constitutes a user-specific learning machine that is consistent with the description of a user-specific learning machine provided by the disclosure of the '276 patent." (*Id.* at 21.)

437. However, the Examiner found that Mladenic does not disclose "receiving a search query from the user" or "retrieving a plurality of documents based on the search query" because "the Personal WebWatcher disclosed by Mladenic does not receive a search query from the user." (ACP at 14.) While the Examiner agrees that the prior WebWatcher system also disclosed in Mladenic receives a search query, the Examiner states that "[i]n the context of a 35 USC 102 rejection, the Requester may not pick and choose certain features of both the WebWatcher system and the Personal WebWatcher system to reject the claims." (*Id.* at 15.) To

the extent the Examiner's reasoning is found valid, Mladenic would still render "receiving a search query from the user" obvious in light of Mladenic itself. It would be obvious to one of skill in the art to combine the "receiving a query" aspect of WebWatcher with the Personal WebWatcher personalization system, as both are described in the same paper and they are both based off the same body of research. In fact, I note that the Examiner agreed that combining the "receiving a query" aspect of WebWatcher with the prior art Barrett and Asnicar references would have been obvious to one of skill in the art. (*Id.* at 51, 63.) If it is obvious to combine two systems described in two different papers, then it is certainly obvious to perform the same type of combination with two systems described in the *same* paper, particularly when those two systems were developed at the same university from the same set of research.

438. The Examiner also found that Mladenic did not teach "applying the identified properties of the retrieved document" limitation, since Personal WebWatcher analyzes the anchor text that points to a prospective document rather than contents of the document itself. (ACP at 18-19.) As discussed in section VI.A.2(d), I disagree with the Examiner's conclusion, as both the specification and the claims indicate that the patents consider links to a document to be a property of that document. I further note that the WebWatcher system discloses analyzing the text of the prospective document itself. However, to the extent the Examiner's reasoning is found valid, analyzing the document itself would be obvious to one of skill in the art. As with the "receiving a search query" term, the element is present in the WebWatcher system described within the *same* paper. Furthermore, Mladenic states that it is "predicting interestingness of document based on the hyperlink pointing to it" because "retrieving documents behind the requested hyperlink is usually time consuming," but explicitly chooses to retrieve the entire document when it "can afford using more time." (Mladenic at 10.) While time constraints as to retrieving and analyzing documents may have been prohibitive in 1996—when Mladenic was

written—they would not have been so in 1999. Between 1996 and 1999 usage of the Internet exploded, and both network speeds and processor speeds rose dramatically to keep pace with online growth. Thus, by 1999 it would have been obvious to analyze the entire document rather than just the anchor text, in light of the reasoning presented in Mladenic itself.

F. The *Graham* Factors Demonstrate That the '040 and '276 Patent Claims Which Merely Combine Known Elements Are Obvious

439. I understand that the Supreme Court in *KSR* instructed that the factors in *Graham v. John Deere Co. of Kansas City*, 383 U.S. 1 (1966), for applying the statutory language of 35 U.S.C. § 103 are as follows:

> Under § 103, the scope and content of the prior art are to be determined; differences between the prior art and the claims at issue are to be ascertained; and the level of ordinary skill in the pertinent art resolved. Against this background the obviousness or nonobviousness of the subject matter is determined.

Graham also set forth a broad inquiry and invited courts, where appropriate, to look at any secondary considerations that would prove instructive:

> Such secondary considerations as commercial success, long felt but unresolved needs, failure of others, etc., might be utilized to give light to the circumstances surrounding the origin of the subject matter sought to be patented.

1. The Scope and Content of the Prior Art

440. The first *Graham* factor, "the scope and content of the prior art," shows the asserted patents to be obvious. As detailed throughout this report, each element of the patent at issue existed in the prior art. See Section VII.B.

441. In particular, the idea of identifying documents that the user previously deemed interesting, creating a user profile out of those documents, comparing a new document to the existing user profile, and providing personalized services based on how well the new document matched the user profile—*e.g.*, the "average" document preferred by that user—was well known to persons of skill in the art. *See* discussions regarding Mladenic, Wasfi, Culliss, Refuah,

Schuetze, Montebello, and Autonomy above; *see generally* Section V.A. The patents themselves acknowledge that all of this is in the prior art. (1:32-60.)

2. Differences Between the Prior Art and the Claims at Issue

442. As to the second factor, the "differences between the prior art and the claims asserted," each element of the asserted patents existed before and each claim of the patents is anticipated as detailed above. To the extent there is any difference at all between the prior art and the claims, however, as detailed herein in Section VI.A.10 and elsewhere, it would be obvious to one of ordinary skill to add any missing elements of the asserted claims to each prior art reference described above. See Section VII.C.1.

443. In particular, machine learning and learning machines based on parameter estimation and Bayesian posterior inference were well-known in the prior art. Their combination use to create personalized information systems was not novel. For example, Autonomy discusses Bayesian analysis and Shannon's information theory as part of its "intellectual foundations." (Autonomy WP at 4). Mladenic similarly discloses using a Naïve Bayesian classifier on frequency vectors. (Mladenic at 7.) Likewise, it was well understood that a principal benefit of using machine learning methods is that they provide "generalization" to entities not seen in their training data; in particular they make it possible to predict a user's interest in "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 $Q_w(s, a)$ that applies even to unseen states (pages) and actions (hyperlinks)." (Joachims at 4.) Wasfi similarly states that "[b]y making content-based filtering, we can deal with pages unseen by others." (Wasfi at 60.)

444. The various limitations added by the dependent claims are options that were readily used in personalization systems with expected effects. As discussed more thoroughly in Section VII.C.1, using personalization with a search engine would have been obvious given that search engines were the preeminent means of locating information on the internet in 1999. Similarly, tracking interactions times is obvious in most HTTP servers given the time stamps that

appear on server logs, and could be readily used to determine how much time a user spent viewing a document. Analyzing documents having multiple media types is virtually inherent on the Internet given that almost all web sites contain images and text, and using the text surrounding an image would be obvious as it provides additional information about that image. There were multiple means of inferring user interest in documents from different user actions, and it would have been obvious to combine two or more of those means so as to get a more robust sample of user-preferred documents. Supplementing the user model with documents *not* of interest to the user, or with additional properties of the documents beyond their words (e.g., author, time written, etc.) would also be obvious, as it would provide more data for the user model. Finally, using personalization with web pages that represent products would be obvious, as it would increase the potential revenue to the web site.

3. <u>Level of Ordinary Skill in the Pertinent Art</u>

445. The third *Graham* factor is the level of ordinary skill in the pertinent art. As the Supreme Court recognized in *KSR*, "[a] person of ordinary skill is also a person of ordinary creativity, not an automaton."

446. The asserted patents apply non-novel mathematical and statistical techniques to the problem of determining the relevance of documents to a user. One skilled in the art would be familiar with the underlying math and statistics, and would immediately see the possibility of applying them to the problem of the patents, as evidenced by the numerous prior art systems using the same techniques towards the same end.

447. In my opinion, an individual holding a BS degree in computer science or its equivalent with 2-3 experience in the field of information retrieval would be aware of the scope and content of the prior art. As discussed above, the patents-in-suit admit that its Bayesian statistics, information theory, TFIDF measures and more would have been known to one versed in the art.

4. The Secondary Considerations Set Forth in *Graham* Do Not Alter the Conclusion of Obviousness

448. I understand that secondary considerations that could prove instructive on the issue of obviousness include commercial success, long felt but unresolved needs, and failure of others, as well as additional considerations discussed in this report. In this case, it is my opinion that there are no secondary considerations that overcome the obviousness determination.

(a) Commercial Success

449. I understand that PUM has identified "Google's commercial success" and "acquisition of Kaltix Corp. and Outride, Inc." as evidence of non-obviousness in its interrogatory response. (PUM Response and Supplemental Responses to Google's Interrogatory No. 3). However, PUM has failed to show or even allege any nexus between the allegedly infringing aspects of the accused products (or the products of the acquired companies) and the alleged invention.

450. I reserve my right to supplement my report should PUM present alleged evidence of a nexus between the patented claims and the commercial success of Google.

451. PUM also cites "the interest of others in purchasing the patented technology from Utopy." However, it is my understanding that Utopy's licensing program was a failure. (Konig Dec. 2, 2010 Dep. at 36:4-7).

(b) Long-Felt But Unresolved Need

452. Likewise, PUM identifies a long-felt but unresolved need as a secondary consideration supporting the non-obviousness of its alleged invention. (PUM Response and Supplemental Response to Google's Interrogatory No. 3). As evidence of long-felt need, PUM cites "Google's success" and "Google's attempts to obtain patent protection for the personalization of web search." (PUM Response to Google's Interrogatory No. 3). However, PUM has failed to show or even allege any nexus between the allegedly infringing aspects of the

accused products and the alleged invention. Furthermore, Google's attempts to obtain patent protection on *specific elements* of search engine technology does not indicate that the elements described in the asserted patent were of any value. Under PUM's logic, Google's attempts to obtain patent protection for "software" would show that a patent that covered "software" was valuable.

453. I reserve my right to supplement my report should PUM present alleged evidence of a nexus between the patented claims and the commercial success or patent activities of Google.

(c) Failure of Others

454. PUM identifies the failure of others as a secondary consideration supporting non-obviousness, including "Google's attempts to obtain patent protection for the personalization of web search" and "statements emphasizing the advantages and importance of personalization." (PUM Response and Supplemental Response to Google's Interrogatory No. 3) Yet PUM has failed to explain how this evidence shows that others were unable to solve the problem of personalized information services.

455. PUM also opaquely cites "statements in Google's U.S. patent applications, including, but not limited to, Application No. 10/676,711, regarding the disadvantages of the prior art as evidence of the failure of others." Again, however, PUM has failed to demonstrate or even allege that these patent applications sought to solve the same problem or make use of the same techniques as the alleged invention.

456. In fact, an assessment of the prior art shows that others had not failed - numerous prior art systems including Mladenic's Personal WebWatcher, Autonomy's Agentware, Wasfi's ProfBuilder, and others, succeeded in providing users with personalized information services generated from learning machines. See above, discussing prior art systems and market trends.

VIII. CONCLUSIONS

457. None of the Asserted Claims is valid.

458. All the Asserted Claims are anticipated.

459. All the Asserted Claims are obvious.

Executed on April 11, 2012, in Berkeley, California.

Michael I. Jordan

EXHIBIT 1

Exhibit 1
Materials Considered

All documents cited in the report.

U.S. Patent No. 6,981,040.
U.S. Patent No. 7,320,031.
U.S. Patent No. 7,685,276.
File History of U.S. Patent No. 6,981,040.
File History of U.S. Patent No. 7,685,276.
May 27, 2011 Office Action, Re-examination of '040 Patent
February 10, 2012 Action Closing Prosecution, Re-examination of the '276 Patent

January 25, 2012 Order.
January 25, 2012 Opinion.

Plaintiff's December 2011 Infringement Contentions.
Plaintiff's April 2012 Infringement Contentions.

December 2, 2010 Deposition of Yochai Konig.
April 14, 2011 Deposition of Douglas Bercow.
April 14, 2011 Deposition of Horatio Franco.
July 6, 2011 Deposition of Andreas Stolcke.
August 15, 2011 Deposition of Mustafa Sonmez.

U.S. Patent No. 6,182,068
U.S. Patent No. 6,567,797
U.S. Patent No. 7,631,032

Armstrong, R. et al (1995), *Web Watcher: A Learning Apprentice for the World Wide Web*, 1995 AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments.

"Autonomy Releases the First Intelligent Internet Software to Deliver Personalized Information of Demand," March 10, 1997 Press Release, Autonomy Inc.

"Autonomy Agentware User Guide," Copyright 1996 Autonomy Systems Limited.

Autonomy Technology Whitepaper (1999).

Bharat, K. et al. (1997), *Personalized, Interactive News on the Web*, May 5, 1997.

Edwards, P. et al. (1996), *Experience with Learning Agents which Manage Internet-Based Information*, AAAI Technical Report SS-96-05.

Green, C. et al. (1995) *Towards Practical Interface Agents which Manage Internet-Based Information,* Intelligent Agents Workshop, BCS Specialist Group on Expert Systems and Representation & Reasoning Special Interest Group.

Hofferer, M. et al. (1994), *An Evolutionary Approach to Intelligent Information Filtering*, Proc of the 2nd Singapore International Conference on Intelligent Systems.

Joachims, T. et al (1997), *WebWatcher: A Tour Guide for the World Wide Web*, In Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence.

Joachims, T. et al. (1995), *WebWatcher: Machine Learning and Hypertext*, Lernen. University of Dortmund.

Kamba, T. et al. (1995), *The Krakatoa Chronicle - An Interactive, Personalized, Newspaper on the Web*, In Proceedings of the Fourth International World Wide Web Conference.

Lieberman, H. (1995), *Letizia: An Agent That Assists Web Browsing*, International Joint Conference on Artificial Intelligence.

Lieberman, H. (1997), *Autonomous Interface Agents*, Proceedings of the SIGCHI conference on Human factors in computing systems.

Mladenic, D. (1996), *Personal WebWatcher: design and implementation*, Technical Report IJS-DP-7472, Department of Intelligent Systems, J. Sefan Institute, Slovenia.

Mladenic, D. (1996), *Machine learning for better Web browsing*, Proceedings of the Eight Electrotechnical and Computer Sc. Conference ERK'99, Ljubljana, Slovenia: IEEE section (1999)

Montebello, M. et al. (1998), *A Personal Evolvable Advisor for WWW Knowledge-Based Systems*, Proceedings of Workshop on Reuse of Web information at the Seventh International World Wide Web Conference.

Payne, T. et al. (1996), *Experience with Rule Induction and k-Nearest Neighbor Methods for Interface Agents that Learn*, IEEE Transactions on Knowledge and Data Engineering, Vol. 9, No. 2, March-April 1997.

Payne, T. and Edwards, P. (1995), *Learning Mechanisms for Information Filtering Agents*, BCS/DTI Intelligent Agents Workshop, November 1995.

Tan, A, and Teo, C. (1998), *Learning User Profiles for Personalized Information Dissemination*, Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on In Neural Networks Proceedings, 1998.

Wasfi, A. (1999), *Collecting User Access Patterns for Building User Profiles and Collaborative Filtering*, IUI '99 Proceedings of the 4th international conference on Intelligent user interfaces.

Ardissono, C. et al. (1999) *An Agent Architecture for Personalized Web Stores*, In Proc. 3rd Int. Conf. on Autonomous Agents

Baeza-Yates, R. & Ribeiro-Neto, Berthier (1999) *Modern Information Retrieval*, ACM Press.

Bishop, C. (1996). *Neural Networks for Pattern Recognition,* Oxford University Press.

Bush, V. (1945). *As We May Think*, The Atlantic (July 1945).

Cherkassky, V. & Mulier, F. (1998). *Learning from Data: Concepts, Theory, and Methods,* Wiley-Interscience.

Cohen, W. W. (1996). "Learning rules that classify email." In *Proceedings of the 1996 AAAI Spring Symposium on Machine Learning in Information Access*.

Duda, R. & Hart, P. (1973) *Pattern Classification and Scene Analysis*, A Wiley-Interscience Publication.

Cover, T. & Thomas, J. (1991) *Elements of Information Theory*, Wiley-Interscience.

Fukunaga, K. (1990) *Introduction to Statistical Pattern Recognition*, Academic Press.

Hertz, J. et al. (1991). *Introduction to The Theory of Neural Computation*, Addison-Wesley Publishing Company.

Lang, K. (1995). *Newsweeder:Learning to Filter Netnews*, In Proceedings of the 12th International Machine Learning Conference.

Lee, P. (1989). *Bayesian Statistics: An Introduction*, Oxford University Press (1989).

Loeb, S. (1992). *Architecturing Personalized Delivery of Multimedia Information*, Communications of the ACM, Dece. 1992, Vol. 35, No. 12.

Manning, C. and Schutze, H. (1999) *Foundations of Statistical Natural Language Processing*, The MIT Press.

Mitchell, T. (1997). *Machine Learning*, McGraw-Hill.

Morita, M. & Shinoda, Y. (1994) *Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval*, Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval.

Orwant, J. (1995) *Heterogeneous Learning in the Doppelganger User Modeling System*, Interaction, Vol. 4, pp. 107-130.

Preface to UMUAI Special Issue on Machine Learning for User Modeling, "User Modeling and User-Adapted Interaction," vol. 8 (1998).

Sahami, M. Dumais, S., Heckerman, D. and Horvitz, E. (1998). "A Bayesian approach to filtering junk e-mail". *AAAI'98 Workshop on Learning for Text Categorization.*

Salton, G. and M. J. McGill (1983). *Introduction to Modern Information Retrieval*. McGraw-Hill.

Yang, Y. (1999). An evaluation of statistical approaches to text categorization, Journal of *Information Retrieval, 1,* 67-88.

EXHIBIT 2

Vitae

MICHAEL I. JORDAN

Department of Electrical Engineering and Computer Science
Department of Statistics
University of California
Berkeley, CA 94720-1776
Phone: (510) 642–3806
Email: jordan@cs.berkeley.edu

EDUCATION

PhD in Cognitive Science, 1985
University of California, San Diego.

MS in Mathematics (Statistics), 1980
Arizona State University.

BS *magna cum laude* in Psychology, 1978
Louisiana State University.

RESEARCH AND TEACHING EXPERIENCE

Professor – Department of Electrical Engineering and Computer Science, Department of Statistics, University of California, Berkeley, 1998 – present.

Professor – Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 1997 – 1998.

Associate professor with tenure – Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 1994 – 1997.

Associate professor – Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 1992 – 1994.

Assistant professor – Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 1988 – 1992.

Postdoctoral researcher – Department of Computer and Information Science, University of Massachusetts, Amherst, 1986 – 1988.

RESEARCH INTERESTS

Statistical machine learning

Bayesian nonparametric statistics

Graphical models

Variational inference methods

Computational biology, statistical genetics

Human motor control, speech production, cognitive modeling

Distributed statistical inference

Spectral methods

Convex optimization

Adaptive signal processing

HONORS

Elected Member, International Statistical Institute (ISI), 2012.

Fellow, Society for Industrial and Applied Mathematics (SIAM), 2012.

Member, American Academy of Arts and Sciences (AAAS), 2011.

Member, National Academy of Sciences (NAS), 2010.

Member, National Academy of Engineering (NAE), 2010.

Fellow, Association for Computing Machinery (ACM), 2010.

Fellow, Cognitive Science Society (CSS), 2010.

ACM/AAAI Allen Newell Award, 2009.

Honorary Professor of Hebei University, China, 2009.

SIAM Activity Group on Optimization Prize, 2008.

Miller Research Professorship, University of California, Berkeley, 2008.

Fellow, American Statistical Association (ASA), 2007.

Fellow, American Association for the Advancement of Science (AAAS), 2006.

IEEE Neural Networks Pioneer Award, 2006.

Pehong Chen Distinguished Professorship, University of California, 2006.

Diane S. McEntyre Award for Excellence in Teaching, 2006.

Fellow, Institute of Mathematical Statistics (IMS), 2005.

Fellow, Institute of Electrical and Electronics Engineers (IEEE), 2005.

Fellow, American Association for Artificial Intelligence (AAAI), 2002.

MIT Class of 1947 Career Development Award, 1992 – 1995.

NSF Presidential Young Investigator Award, 1991 – 1996.

NAMED LECTURES

Vincent Meyer Colloquium, Israel Institute of Technology, 2012.

Constance van Eeden Colloquium, University of British Columbia, 2012.

Neyman Lecture, Institute of Mathematical Statistics, 2011.

Ernst Ising Lecture, Brown University, 2011.

Dertouzos Lecture, Massachusetts Institute of Technology, 2011.

George A. Bekey Lecture, University of Southern California, 2011.

Thomas E. Noonan Lecture, Georgia Institute of Technology, 2011.

R. L. Anderson Lecture, University of Kentucky, 2011.

S. James Press Endowed Lecture, University of California, Riverside, 2010.

Posner Lecture, Neural Information Processing Systems Annual Conference, 2010.

Morris H. DeGroot Memorial Lecture, Carnegie Mellon University, 2009.

Pao-Lu Hsu Lecture, Beijing University, 2009.

Institute Medallion Lecturer, Institute of Mathematical Statistics, 2004.

Paul Rockwood Memorial Lecture, Institute for Neural Computation, 1996.

BEST PAPER AWARDS

Best student paper award (with P. Wang, K. Laskey and C. Domeniconi), SIAM International Conference on Data Mining (SDM), 2011.

Best student paper award (with J. Duchi and L. Mackey), International Conference on Machine Learning (ICML), 2010.

Best student paper award (with P. Liang), International Conference on Machine Learning (ICML), 2008.

IEEE Signal Processing Society young author award (with X. Nguyen and M. Wainwright), 2007.

Best student paper award (with P. Flaherty and A. Arkin), Neural Information Processing Systems (NIPS), 2005.

Best paper award (with X. Nguyen and M. Wainwright), International Conference on Machine Learning (ICML), 2004.

Best paper award honorable mention (with F. Bach and G. Lanckriet), International Conference on Machine Learning (ICML), 2004.

Best student paper award (with D. Blei, T. Griffiths and J. Tenenbaum), Neural Information Processing Systems (NIPS), 2003.

Best paper award nominee (with B. Sinopoli, M. Franceschetti, L. Schenato, K. Poolla, and S. Sastry), 42nd IEEE Conference on Decision and Control (CDC), 2003.

Best student paper award runner-up (with E. Xing and S. Russell), Uncertainty in Artificial Intelligence (UAI), 2003.

Best student paper award (with T. Jaakkola), Uncertainty in Artificial Intelligence Conference (UAI), 1996.

Best paper award (with R. Jacobs), American Control Conference (ACC), 1991.

EDITORIAL BOARDS

Foundations and Trends in Machine Learning (Editor-in-Chief, 2007-)

Bayesian Analysis (Editor, 2006-2011)

Stochastic Analysis and Applications (Honorary Editorial Board, 2010-)

Information and Inference (Associate Editor, 2011-)

IEEE Signal Processing Magazine (Editorial Board, 2010-)

IEEE Signal Processing Magazine (Guest Editor, Special Issue on Graphical Models, 2010)

Journal of the American Statistical Association (Associate Editor, 1998-2001)

Journal of Machine Learning Research (Action Editor, 2000-2009)

Neural Computation (Associate Editor, 1989-)

Statistical Analysis and Data Mining (Associate Editor, 2006-2009)

Machine Learning (Action Editor, 1993-1999)

Journal of Artificial Intelligence Research (Editorial Board, 1998-2001)

International Journal of Machine Learning and Cybernetics (Advisory Board, 2010-)

Cognition (Editorial Board, 1992-1998)

International Journal of Neural Systems (Editorial Advisory Board, 2002-2010)

Neural Networks (Editorial Board, 1994-2008)

Neurocomputing (Editorial Board, 1994-2003)

Neural Processing Letters (Editorial Board, 1994-2007)

OTHER PROFESSIONAL ACTIVITIES

President, International Society for Bayesian Analysis (ISBA), 2010-2011

ACM Turing Award Committee, 2011-2014

IMS Committee on Special Lectures, 2011-2014

Membership Committee, American Academy of Arts and Sciences (AAAS), 2011-

Series Editor, Springer-Verlag Series on Statistics and Information Sciences

Series Editor, MIT Press Series on Adaptive Computation and Machine Learning

Executive Committee, International Society for Bayesian Analysis (ISBA), 2009-2012

Prize Committee, International Society for Bayesian Analysis (ISBA), 2009-2010

Advisory Board, Bayesian Analysis (Journal of the International Society for Bayesian Analysis)

Scientific Advisory Board, Institute of Mathematical Statistics, Tokyo, Japan, 2008-

External Advisory Board, Statistics and Operational Research Doctoral Training Centre, Lancaster University, 2010-

Founding Board Member of the International Machine Learning Society (IMLS), 2001-2009

Member of the Neural Information Processing Systems (NIPS) Foundation Board, 1998-

Session Organizer, IMS Annual Meeting, 2010

Chair, MIT Press Editorial Advisory Board, 1994-1998

Advisory Council for the International Association for the Study of Attention and Performance, 1994-2002

Program Chair, NIPS (Neural Information Processing Systems Conference), 1996

General Chair, NIPS (Neural Information Processing Systems Conference), 1997

Advisory Editor, MIT Encyclopedia of the Cognitive Sciences

Director – NATO ASI Summer School on Learning in Graphical Models, Erice, Italy, September, 1996

GRADUATE AND POSTDOCTORAL SUPERVISION

Graduate Student Supervision

Eric Loeb, 1989–1995; Zoubin Ghahramani, 1990–1995; John Houde, 1990–1997; Wey Fun, 1991–1995; Philip Sabes, 1991–1996; Tommi Jaakkola, 1992–1997; Emanuel Todorov, 1992–1998; Marina Meila, 1992–1999; Andrew Ng, 1997–2003; David Blei, 1999–2004; Alice Zheng, 1999–2005; Eric Xing, 2000–2004; Jon McAuliffe, 2000–2005; Francis Bach, 2000–2005; Gert Lanckriet, 2000–2005; Brian Vogel, 2001–2005; Patrick Flaherty, 2001–2007; XuanLong Nguyen, 2001–2007; Barbara Engelhardt, 2001–2007; Romain Thibaux, 2003–2008; Simon Lacoste-Julien, 2003–2009; Guillaume Obozinski, 2003–2009; Sarah Moussa, 2003–2005; Ben Blum, 2004–2008; Alex Simma, 2004–2010; Peter Bodik, 2004–2010; Junming Yin, 2005–2010; Alexandre Bouchard-Cote, 2005–2010; Sriram Sankararaman, 2005–2010; Percy Liang, 2005–2011; Chris Hundt, 2006–2008; Alex Shyr, 2006–2011; Kurt Miller, 2006–2011; Daniel Ting, 2006–2011; Ariel Kleiner, 2006–; Fabian Wauthier, 2007–; Lester Mackey, 2007–; John Duchi, 2008–; Tamara Broderick, 2009–; Teodor Moldovan, 2009–; Andre Wibisono, 2010–

Postdoctoral Supervision

Robert Jacobs, 1990–1992; Marios Mantakas, 1990–1991; Yoshua Bengio, 1991–1992; Lei Xu, 1992–1993; David Cohn, 1992–1995; Daniel Wolpert, 1992–1995;

Satinder Singh, 1993–1995; Lawrence Saul, 1994–1996; Thomas Hofmann, 1997–1999; Yair Weiss, 1998–2001; Chiranjib Bhattacharyya, 2000–2002; Sekhar Tatikonda, 2000–2002; Michal Rosen-Zvi, 2002–2003; Martin Wainwright, 2002–2004; Yee-Whye Teh, 2003–2005; Matthias Seeger, 2003–2005; Ben Taskar, 2004–2006; Fei Sha, 2006–2007; Zhihua Zhang, 2006–2008; Erik Sudderth, 2006–2009; Gad Kimmel, 2006–2008; Charles Sutton, 2007–2009; Emily Fox, 2010–2011; Justin Ma, 2010–; Ameet Talwalkar, 2010–; Purnamrita Sarkar, 2010–; John Paisley, 2011–; Jennifer Tom, 2011–; Venkat Chandrasekaran, 2011–

JOURNAL ARTICLES

Broderick, T., Jordan, M. I., & Pitman, J. (to appear). Beta processes, stick-breaking, and power laws. *Bayesian Analysis.*

Bouchard, A., Sankararaman, S., & Jordan, M. I. (2012). Phylogenetic inference via sequential Monte Carlo. *Systematic Biology*, doi:10.1093/sysbio/syr131.

Obozinski, G., Wainwright, M. & Jordan, M. I. (2011). Support union recovery in high-dimensional multivariate regression. *Annals of Statistics*, *39*, 1-47.

Fox, E. B., Sudderth, E., Jordan, M. I., & Willsky, A. S. (2011). A sticky HDP-HMM with application to speaker diarization. *Annals of Applied Statistics*, *5*, 1020-1056.

Engelhardt, B., Jordan, M. I., Srouji, J., & Brenner, S. (2011). Genome-scale phylogenetic function annotation of large and diverse protein families. *Genome Research*, doi/10.1101/gr.104687.109.

Sutton, C. A. & Jordan, M. I. (2011). Bayesian inference for queueing networks and modeling of Internet services. *Annals of Applied Statistics*, *5*, 254-282.

Fox, E. B., Sudderth, E., Jordan, M. I., & Willsky, A. S. (2011). Bayesian nonparametric inference of switching dynamic linear models. *IEEE Transactions on Signal Processing*, *59*, 1569-1585.

Wauthier, F., Jordan, M. I., & Jojic, N. (2011). Nonparametric combinatorial sequence models. *Journal of Computational Biology*, *18*, 1649-1660.

Carin, L., Baraniuk, R. G., Cevher, V., Dunson, D., Jordan, M. I., Sapiro, G., & Wakin, M. B. (2011). Learning low-dimensional signal models. *IEEE Signal Processing Magazine*, *28*, 39-51.

Zhang, Z., Dai, G., & Jordan, M. I. (2011). Bayesian generalized kernel mixed models. *Journal of Machine Learning Research*, *12*, 111139.

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

Blum, B., Jordan, M. I., & Baker, D. (2010). Feature space resampling for protein conformational search. *Proteins: Structure, Function, and Bioinformatics*, *78*, 1583-1593.

Nguyen, X., Wainwright, M., & Jordan, M. I. (2010). Estimating divergence functionals and the likelihood ratio by convex risk minimization. *IEEE Transactions on Information Theory*, *56*, 5847-5861.

Ting, D., Wang, G., Shapovalov, M., Mitra, R., Jordan, M. I., & Dunbrack, R. (2010). Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. *PLoS Computational Biology*, *6*, e1000763.

Sankararaman, S., Sha, F., Kirsch, J., Jordan, M. I., & Sjolander, K. (2010). Active site prediction using evolutionary and structural information. *Bioinformatics*, *26*, 617-624.

Obozinski, G., Taskar, B. & Jordan, M. I. (2010). Joint covariate selection and joint subspace selection for multiple classification problems. *Statistics and Computing*, *20*, 231-252.

Fox, E. B., Sudderth, E., Jordan, M. I., & Willsky, A. S. (2010). Bayesian nonparametric methods for learning Markov switching processes. *IEEE Signal Processing Magazine*, *27*, 43-54.

Ding, C., Li, T., & Jordan, M. I. (2010). Convex and semi-nonnegative matrix factorizations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*, 45-55.

Zhang, Z., Dai, G., Xu, C., & Jordan, M. I., (2010). Regularized discriminant analysis, ridge regression and beyond. *Journal of Machine Learning Research*, *11*, 2141-2170.

Sankararaman, S., Obozinski, G., Jordan, M. I., & Halperin, E. (2009). Genomic privacy and the limits of individual detection in a pool. *Nature Genetics*, *41*, 965-967.

Nguyen, X., Wainwright, M., & Jordan, M. I. (2009). On surrogate loss functions and f-divergences. *Annals of Statistics*, *37*, 876-904.

Fukumizu, K., Bach, F. R., & Jordan, M. I. (2009). Kernel dimension reduction in regression. *Annals of Statistics*, *37*, 1871-1905.

Yin, J., Jordan, M. I., & Song, Y. (2009). Joint estimation of gene conversion rates and mean conversion tract lengths from population SNP data. *Bioinformatics*, *25*, i231-i239.

Sankararaman, S., Kimmel, G., Halperin, E., & Jordan, M. I. (2008). On the inference of ancestries in admixed populations. *Genome Research, 18*, 668-675.

Wainwright, M. & Jordan, M. I. (2008). Graphical models, exponential families and variational inference. *Foundations and Trends in Machine Learning, 1*, 1-305.

Kimmel, G., Karp, R., Jordan, M. I., & Halperin, E. (2008). Association mapping and significance estimation via the coalescent. *American Journal of Human Genetics, 83*, 675-683.

Zhang, Z., & Jordan, M. I. (2008). Multiway spectral clustering: A margin-based perspective. *Statistical Science, 23*, 383-403.

Flaherty, P., Radhakrishnan, M. A., Dinh, T., Jordan, M. I. & Arkin, A. P. (2008). A dual receptor cross-talk model of G protein-coupled signal transduction. *PLoS Computational Biology, 4*, e1000185.

Nguyen, X., Wainwright, M., & Jordan, M. I. (2008). On optimal quantization rules for some sequential decision problems. *IEEE Transactions on Information Theory, 54*, 3285-3295.

Obozinski, G., Grant, C. E., Lanckriet, G. R. G., Jordan, M. I., & Noble, W. S. (2008). Consistent probabilistic outputs for protein function prediction. *Genome Biology, 9*, S7.

Pena-Castillo, L., Tasan, M., Myers, C., Lee, H., Joshi, T., Zhang, C., Guan, Y., Leone, M., Paganini, A., Kim, W., Krumpelman, C., Tian, W., Obozinski, G., Qi, Y., Mostafavi, S., Lin, G., Berriz, G., Gibbons, F., Lanckriet, G., Qiu, J., Grant, C., Barutcuoglu, Z., Hill, D., Warde-Farely, D., Grouios, C., Ray, D., Blake, J., Deng, M., Jordan, M., Noble, W., Morris, Q., Klein-Seetharaman, J., Bar-Joseph, Z., Chen, T., Sun, F., Troyanskaya, O., Marcotte, E., Xu, D., Hughes, T. & Roth, F. (2008). Quantitative gene function assignment from genomic datasets in M. musculus. *Genome Biology, 9*, S2.

D'Aspremont, A., El Ghaoui, L., Jordan, M. I., & Lanckriet, G. R. G. (2007). A direct formulation for sparse PCA using semidefinite programming. *SIAM Review, 49*, 434-448.

Kimmel, G., Jordan, M. I., Halperin, E., Shamir, R., & Karp, R. (2007). A randomization test for controlling population stratification in whole-genome association studies. *American Journal of Human Genetics, 81*, 895-905.

Xing, E. P., Jordan, M. I., & Sharan, R. (2007). Bayesian haplotype inference via the Dirichlet process. *Journal of Computational Biology, 14*, 267-284.

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

Bartlett, P., Jordan, M. I., & McAuliffe, J. D. (2006). Convexity, classification and risk bounds. *Journal of the American Statistical Association*, *101*, 138-156.

Bach, F. R., & Jordan, M. I. (2006). Learning spectral clustering, with application to speech separation. *Journal of Machine Learning Research*, *7*, 1963-2001.

Wainwright, M. & Jordan, M. I. (2006). Log-determinant relaxation for approximate inference in discrete Markov random fields. *IEEE Transactions on Signal Processing*, *54*, 2099-2109.

McAuliffe, J. D., Blei, D. M., & Jordan, M. I. (2006). Nonparametric empirical Bayes for the Dirichlet process mixture model. *Statistics and Computing*, *16*, 5-14.

Taskar, B., Lacoste-Julien, S., & Jordan, M. I. (2006). Structured prediction, dual extragradient and Bregman projections. *Journal of Machine Learning Research*, *7*, 1627-1653.

Blei, D. M., Franks, K. Jordan, M. I. & Mian, S. (2006). Mining the Caenorhabditis Genetic Center bibliography for genes related to life span. *BMC Bioinformatics* *7*, 250-269.

McAuliffe, J. D., Jordan, M. I. & Pachter, L. (2005). Subtree power analysis and species selection for comparative genomics. *Proceedings of the National Academy of Sciences*, *102*, 7900-7905.

Engelhardt, B., Jordan, M. I., Muratore, K., & Brenner, S. (2005). Protein function prediction via Bayesian phylogenomics. *PLoS Computational Biology*, *1*, e45.

Blei, D. M., & Jordan, M. I. (2005). Variational inference for Dirichlet process mixtures. *Bayesian Analysis*, *1*, 121-144.

Gyaneshwar, P., Paliy, O., McAuliffe, J., Popham, D. L., Jordan, M. I., & Kustu, S. (2005). Lessons from *Escherichia coli* genes similarly regulated in response to nitrogen and sulfur limitation. *Proceedings of the National Academy of Sciences*, *102*, 3453-3458.

Nguyen, X., Wainwright, M., & Jordan, M. I. (2005). Nonparametric decentralized detection using kernel methods. *IEEE Transactions on Signal Processing*, *53*, 4053-4066.

Lee, W., St. Onge, R. P., Proctor, M., Flaherty, P., Jordan, M. I., Arkin, A. P., Davis, R. W., Nislow, C., & Giaever, G. (2005). Genome-wide requirements for resistance to functionally distinct DNA-damaging agents. *PLoS Genetics*, *1*, 235-246.

Gyaneshwar, P., Paliy, O., McAuliffe, J., Jones, A., Jordan, M. I., & Kustu, S. (2005). Sulfur and nitrogen limitation in *Escherichia coli* K12: specific homeostatic responses. *Journal of Bacteriology*, *187*, 1074-1090.

Nguyen, X., Jordan, M. I., & Sinopoli, B. (2005). A kernel-based learning approach to ad hoc sensor network localization. *ACM Transactions on Sensor Networks, 1*, 134-152.

Flaherty, P., Giaever, G., Kumm, J., Jordan, M. I., & Arkin, A. P. (2005). A latent variable model for chemogenomic profiling. *Bioinformatics, 21*, 3286-3293.

Jordan, M. I. (2004). Graphical models. *Statistical Science, 19*, 140-155.

Giaever, G., Flaherty, P., Kumm, J., Proctor, M., Jaramillo, D. F., Chu, A. M., Jordan, M. I., Arkin, A. P. and Davis, R. W. (2004). Chemogenomic profiling: Identifying the functional interactions of small molecules in yeast. *Proceedings of the National Academy of Sciences, 3*, 793-798.

McAuliffe, J. D., Pachter, L., & Jordan, M. I. (2004). Multiple-sequence functional annotation and the generalized hidden Markov phylogeny. *Bioinformatics, 20*, 1850-1860.

Lanckriet, G. R. G., De Bie, T., Cristianini, N., Jordan, M. I., & Noble, W. S. (2004). A statistical framework for genomic data fusion. *Bioinformatics, 20*, 1-10.

Bach, F. R., & Jordan, M. I. (2004). Learning graphical models for stationary time series. *IEEE Transactions on Signal Processing, 52*, 2189-2199.

Fukumizu, K., Bach, F. R., & Jordan, M. I. (2004). Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. *Journal of Machine Learning Research, 5*, 73-99.

Sinopoli, B., Schenato, L., Franceschetti, M., Poolla, K., Jordan, M. I., & Sastry, S. (2004). Kalman filtering with intermittent observations. *IEEE Transactions on Automatic Control, 49*, 1453-1464.

Xing, E. P., Wu, W., Jordan, M. I., & Karp, R. M. (2004). LOGOS: A modular Bayesian model for *de novo* motif detection. *Journal of Bioinformatics and Computational Biology, 2*, 127-154.

Lanckriet, G. R. G., Cristianini, N., Bartlett, P., El Ghaoui, L., & Jordan, M. I. (2004). Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research, 5*, 27-72.

Bhattacharyya, C., Grate, L. R., Jordan, M. I., El Ghaoui, L., & Mian, I. S. (2004). Robust sparse hyperplane classifiers: application to uncertain molecular profiling data. *Journal of Computational Biology, 11*, 1073-1089.

Corbin, R. W., Paliy, O., Yang, F., McAuliffe, J., Shabnowitz, J, Platt, M., Lyons, Jr., C. E., Root, K., Jordan, M. I., Kustu, S., Soupene, G., & Hunt, D. F. (2003). Toward a protein profile of *Escherichia coli*: comparison to its transcription profile. *Proceedings of the National Academy of Sciences, 100*, 9232-9237.

Bach, F., & Jordan, M. I. (2003). Beyond independent components: Trees and clusters. *Journal of Machine Learning Research, 4*, 1205-1233.

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

Grate, L. R., Bhattacharyya, C., Jordan, M. I., & Mian, I. S. (2003). Integrated analysis of transcript profiling and protein sequence data. *Mechanisms of Ageing and Development, 124*, 109-114.

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

Bhattacharyya, C., Grate, L. R., Rizki, A., Radisky, D., Molina, F. J., Jordan, M. I., Bissell, M. J. & Mian, I. S. (2003). Simultaneous classification and relevant feature identification in high-dimensional spaces: Application to molecular profiling data. *Signal Processing, 83*, 729-743.

Andrieu, C., De Freitas, J., Doucet, A., & Jordan, M. I. (2003). An introduction to MCMC for machine learning. *Machine Learning, 50*, 5-43.

Todorov, E., & Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience, 5*, 1226-1235.

Bach, F. R., & Jordan, M. I. (2002). Kernel independent component analysis. *Journal of Machine Learning Research, 3*, 1–48.

Houde, J., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research, 45.*

Lanckriet, G. R. G., El Ghaoui, L., Bhattacharyya, C., & Jordan, M. I. (2002). A robust minimax approach to classification. *Journal of Machine Learning Research, 3*, 555-582.

Jaakkola, T., & Jordan, M. I. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing, 10*, 25–37.

Ma, J., Xu, L., & Jordan, M. I. (2000). Asymptotic properties of the convergence rate of the EM algorithm for Gaussian mixtures. *Neural Computation, 12*, 2881–2908.

Saul, L. K., & Jordan, M. I. (2000). Attractor dynamics in feedforward neural networks. *Neural Computation, 12*, 1313–1335.

Meila, M., & Jordan, M. I. (2000). Learning with mixtures of trees. *Journal of Machine Learning Research, 1*, 1–48.

Saul, L. K., & Jordan, M. I. (1999). Mixed memory Markov models: Decomposing complex stochastic processes as mixture of simpler ones. *Machine Learning, 37*, 75–87.

Desmurget, M., Prablanc, C., Jordan, M. I., & Jeannerod, M. (1999). Are reaching movements planned to be straight and invariant in the extrinsic space? *Quarterly Journal of Experimental Psychology, 52*, 981–1020.

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

Jaakkola, T., & Jordan, M. I. (1999). Variational probabilistic inference and the QMR-DT network. *Journal of Artificial Intelligence Research, 10*, 291–322.

Houde, J., & Jordan, M. I. (1998). Adaptation in speech production. *Science, 279*, 1213–1216.

Sabes, P. N., Jordan, M. I., & Wolpert, D. M. (1998). The role of inertial sensitivity in motor planning. *Journal of Neuroscience, 18*, 5948–5959.

Todorov, E., & Jordan, M. I. (1998). Smoothness maximization along a predefined path accurately predicts the speed profiles of complex arm movements. *Journal of Neurophysiology, 80*, 696–714.

Smyth, P., Heckerman, D., & Jordan, M. I. (1997). Probabilistic independence networks for hidden Markov probability models. *Neural Computation, 9*, 227–270.

Desmurget, M., Jordan, M. I., Prablanc, C. & Jeannerod, M. (1997). Constrained and unconstrained movements involve different control strategies. *Journal of Neurophysiology, 77*, 1644–1650.

Sabes, P. N., & Jordan, M. I. (1997). Obstacle avoidance and a perturbation sensitivity model for motor planning. *Journal of Neuroscience, 17*, 7119–7128.

Ghahramani, Z., & Jordan, M. I. (1997). Factorial Hidden Markov models. *Machine Learning, 29*, 245–273.

Desmurget, M., Rossetti, Y., Jordan, M. I., Meckler, C. & Prablanc, C. (1997). Viewing the hand prior to movement improves accuracy of pointing performed toward the unseen contralateral hand. *Experimental Brain Research, 115*, 180–186.

Xu, L., & Jordan, M. I. (1996). On convergence properties of the EM algorithm for Gaussian mixtures. *Neural Computation, 8*, 129–151.

Saul, L. K., Jaakkola, T., & Jordan, M. I. (1996). Mean field theory for sigmoid belief networks. *Journal of Artificial Intelligence Research, 4*, 61–76.

Alpaydin, E., & Jordan, M. I. (1996). Local linear perceptrons for classification. *IEEE Transactions on Neural Networks*, *7*, 788–792.

Cohn, D., Ghahramani, Z., & Jordan, M. I. (1996). Active learning with statistical models. *Journal of Artificial Intelligence Research*, *4*, 129–145.

Jordan, M. I., & Bishop, C. (1996). Neural networks. *Computing Surveys*, *28*, 73–75.

Ghahramani, Z., Wolpert, D., & Jordan, M. I. (1996). Generalization to local remappings of the visuomotor coordinate transformation. *Journal of Neuroscience*, *16*, 7085-7096.

Jordan, M. I. (1995). The organization of action sequences: Evidence from a relearning task. *Journal of Motor Behavior*, *27*, 179–192.

Wolpert, D., Ghahramani, Z., & Jordan, M. I. (1995). Are arm trajectories planned in kinematic or dynamic coordinates? An adaptation study. *Experimental Brain Research*, *103*, 460–470.

Wolpert, D., Ghahramani, Z., & Jordan, M. I. (1995). An internal forward model for sensorimotor integration. *Science*, *269*, 1880–1882.

Houde, J. & Jordan, M. I. (1995). Adaptation in speech production to transformed auditory feedback. *Journal of the Acoustical Society of America*, *97*, 3243.

Jordan, M. I., & Xu, L. (1995). Convergence results for the EM approach to mixtures-of-experts architectures. *Neural Networks*, *8*, 1409–1431.

Houde, J. & Jordan, M. I. (1995). Patterns of generalization in speech sensorimotor adaptation. *Journal of the Acoustical Society of America*, *100*, 2663.

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

Saul, L. K., & Jordan, M. I. (1994). Learning in Boltzmann trees. *Neural Computation*, *6*, 1173–1183.

Jaakkola, T., Jordan, M. I., & Singh, S. P. (1994). On the convergence of stochastic iterative dynamic programming algorithms. *Neural Computation*, *6*, 1183–1190.

Wolpert, D., Ghahramani, Z., & Jordan, M. I. (1994). Perceptual distortion contributes to the curvature of human reaching movements. *Experimental Brain Research*, *98*, 153–156.

Jordan, M. I., Flash, T., & Arnon, Y. (1994). A model of the learning of arm trajectories from spatial targets. *Journal of Cognitive Neuroscience*, *6*, 359–376.

Perkell, J. S., Matthies, M. L., Svirsky, M. A., & Jordan, M. I. (1993). Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot motor equivalence study. *Journal of the Acoustical Society of America, 93*, 2948–2961.

Jacobs, R. A. & Jordan, M. I. (1993). Learning piecewise control strategies in a modular neural network architecture. *IEEE Transactions on Systems, Man, and Cybernetics, 23*, 337–345.

Hirayama, M., Kawato, M., & Jordan, M. I. (1993). The cascade neural network model and a speed-accuracy tradeoff of arm movement. *Journal of Motor Behavior, 25*, 162–175.

Jordan, M. I. (1992). Constrained supervised learning. *Journal of Mathematical Psychology, 36*, 396–425.

Jordan, M. I., & Rumelhart, D. E. (1992). Forward models: Supervised learning with a distal teacher. *Cognitive Science, 16*, 307–354.

Jacobs, R. A., & Jordan, M. I. (1992). Computational consequences of a bias towards short connections. *Journal of Cognitive Neuroscience, 4*, 331–344.

Jacobs, R. A., Jordan, M. I., Nowlan, S., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation, 3*, 1–12.

Mazzoni, P., Andersen, R., & Jordan, M. I. (1991). A more biologically plausible learning network model for neural networks. *Proceedings of the National Academy of Sciences, 88*, 4433–4437.

Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991). Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks. *Cognitive Science, 15*, 219–250.

Mazzoni, P., Andersen, R., & Jordan, M. I. (1991). A more biologically plausible learning rule than backpropagation applied to a network model of cortical area 7a. *Cerebral Cortex, 1*, 293–307.

Bailly, G., Jordan, M. I., Mantakas, M., Schwartz, J-L., Bach, M., & Olesen, O. (1990). Simulation of vocalic gestures using an articulatory model driven by a sequential neural network. *Journal of the Acoustical Society of America, 87:S105.*

Jordan, M. I. (1990). A non-empiricist perspective on learning in layered networks. *Behavioral and Brain Sciences, 13*, 497–498.

INVITED JOURNAL ARTICLES