IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

PERSONALIZED USER MODEL, L.L.P., Plaintiff, v. GOOGLE, INC., Defendant.	C.A. No. 09-525 (LPS)
GOOGLE, INC. Counterclaimant, v. PERSONALIZED USER MODEL, L.L.P. and YOCHAI KONIG, Counterclaim-Defendants.	**PUBLIC VERSION**

VOLUME 1 OF 3 - DECLARATION OF JAIME G. CARBONELL IN SUPPORT OF PLAINTIFF PERSONALIZED USER MODEL, L.L.P.'S ANSWERING BRIEF IN OPPOSITION TO GOOGLE'S MOTION FOR SUMMARY JUDGMENT ON INVALIDITY

OF COUNSEL:

Marc S. Friedman
SNR Denton US LLP
1221 Avenue of the Americas
New York, NY 10020-1089
(212) 768-6700

Mark C. Nelson
SNR Denton US LLP
2000 McKinney Avenue, Ste. 1900
Dallas, TX 75201
(214) 259-0901

Jennifer D. Bennett
SNR Denton US LLP
1530 Page Mill Road, Ste. 200
Palo Alto, CA 94304-1125
(650) 798-0300

MORRIS, NICHOLS, ARSHT & TUNNELL LLP
Karen Jacobs Louden (#2881)
Jeremy A. Tigan (#5239)
1201 N. Market Street
P.O. Box 1347
Wilmington, DE 19899-1347
(302) 658-9200
klouden@mnat.com

Attorneys for Personalized User Model, L.L.P.

Original Filing Date: January 14, 2013

Redacted Filing Date: January 23, 2013

IN THE UNITED STATES DISTRICT COURT
FOR THE DISTRICT OF DELAWARE

PERSONALIZED USER MODEL, L.L.P.,	)	
	)	
Plaintiff,	)	
	)	
v.	)	
	)	
GOOGLE, INC.,	)	
	)	C.A. No. 09-525 (LPS)
Defendant.	)	
	)	**PUBLIC VERSION**
	)	
GOOGLE, INC.	)	
	)	
Counterclaimant,	)	
	)	
v.	)	
	)	
PERSONALIZED USER MODEL, L.L.P. and YOCHAI KONIG,	)	
	)	
Counterclaim-Defendants.	)	

DECLARATION OF JAIME G. CARBONELL IN SUPPORT OF PLAINTIFF PERSONALIZED USER MODEL, L.L.P.'S ANSWERING BRIEF IN OPPOSITION TO GOOGLE'S MOTION FOR SUMMARY JUDGMENT ON INVALIDITY

I, Jaime G. Carbonell, declare:

1. I submit this Declaration in Support of Plaintiff's Personalized User Model, L.L.P's ("PUM") Answering Brief in Opposition to Google's Motion for Summary Judgment on Invalidity.

2. I have been retained by counsel for PUM to respond to the Report of Defendant's Expert Michael I Jordan, Ph.D., Concerning Invalidity of Claims 1, 11, 22, 32 and 34 of U.S. Patent No. 6,981,040 and Claim 1, 3, 5, 6, 7, 21 and 22 of U.S. Patent No. 7,685,276, and to respond where appropriate.

3. Attached hereto as Exhibit A is a true and correct copy of my Rebuttal Expert Witness Report signed November 28, 2012, and selected exhibits. I hereby incorporate by reference and adopt all statements and opinions reflected in my Expert Witness Report as if set forth fully herein.

I declare under penalty of perjury under the laws of the United States of America that the foregoing is true and correct.

Executed this 12th day of January 2013, at Pittsburgh, Pennsylvania.

Jaime G. Carbonell

CERTIFICATE OF SERVICE

I hereby certify that on January 14, 2013, I caused the foregoing to be electronically filed with the Clerk of the Court using CM/ECF which will send electronic notification of such filing to all registered participants.

Additionally, I hereby certify that true and correct copies of the foregoing were caused to be served on January 14, 2013, upon the following individuals in the manner indicated:

BY E-MAIL

Richard L. Horwitz
David E. Moore
POTTER ANDERSON & CORROON LLP
1313 N. Market St., 6th Floor
Wilmington, DE 19801

BY E-MAIL

Brian C. Cannon
QUINN EMANUEL URQUHART
& SULLIVAN, LLP
555 Twin Dolphin Dr., 5th Floor
Redwood Shores, CA 94065

Charles K. Verhoeven
David A. Perlson
Antonio R. Sistos
Andrea Pallios Roberts
Joshua Lee Sohn
QUINN EMANUEL URQUHART
& SULLIVAN, LLP
50 California Street, 22nd Floor
San Francisco, CA 94111

/s/ Jeremy A. Tigan
Jeremy A. Tigan (#5239)

EXHIBIT A

REDACTED

IN ITS

ENTIRETY

Exhibit 1

Personal WebWatcher: design and implementation

Dunja Mladenić
Dpt. for Intelligent Systems, J.Stefan Institute,
Jamova 39, 11000 Ljubljana, Slovenia
Phone: (+38)(61) 1773 272, Fax: (+38)(61) 1258-158
E-mail: Dunja.Mladenic@ijs.si
http://www-ai.ijs.si/DunjaMladenic/home.html
School of Computer Science, Carnegie Mellon University
Pittsburgh, PA, USA

1 Introduction

With the growing availability of information sources, especially non-homogeneous, 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. Recent work that arises at the intersection of Information Retrieval and Machine Learning offers some novel solutions to this problem, as well as work in Intelligent Agents. For example, Armstrong et al. [2] developed WebWatcher, a system that assists user in locating information on the World Wide Web taking keywords from the user, suggesting hyperlinks and receiving evaluation. Balabanovic et al. [3] developed "a system which learns to browse the Internet on behalf of a user". It searches the World Wide Web taking bounded amount of time, selects the best pages and receives an evaluation from the user. The evaluation is used to update the search and selection heuristics. Pazzani et al. [26] collect ratings of the explored Web pages from the user and learn a user profile from them. Pages are separated according to their topic and a separate profile is learned for each topic. Mitchell et al. [25] proposed a system connected to the user's electronic calendar, that generates sets of rules capturing the user's scheduling preferences and some other information about individual attendees of meetings. It uses these rules to provide advice to the user for new, unscheduled meetings. Lang [21] developed a system for electronic news filtering that uses text-learning to generate models of user interests. Krulwich and Burkey [20] proposed "The ContactFinder agent" that reads and responds to bulletin board messages; assists users by referring them to other people who can help them and; categorizes messages and extracts their topic areas. Maes [24] described "interface agents" that learn from the user as well as from other agents. As examples of such agents they developed agents for: electronic mail handling, meeting scheduling, electronic news filtering and entertainment recommendation. Some of these agents use the context of documents and adopt Information Retrieval approaches (eg. news filtering). Others rely on correlation between different users — performing "social

learning" (eg. entertainment recommendation). Holte and Drummond [15], Drummond et al. [9] designed a system that assists browsing of software libraries, taking keywords from the user and using a rule-based system with forward chaining inference, assuming that the library consists of one type of item and the user goal is a single item. Etizioni and Weld [11] offer an integrated interface to the Internet combining UNIX shell and the World Wide Web to interact with Internet resources. Their agent accepts high-level user goals and dynamically synthesizes the appropriate sequence of Internet commands to satisfy those goals. Hammond et al. [**?**] and Burke et al. [6] developed a system that uses a "natural language question-based interface to access distributed text information sources" and helps the user to find answers to her/his question in a databases such as FAQ files.

One of the available information sources is the World Wide Web and it is currently growing quickly, attracting many users with different interests. Since interaction with the World Wide Web (WWW) is by means of computer, one can use computers to "observe" and record user actions and use this information to help users find their way in the WWW.

Personal WebWatcher is a system that observes users of the WWW and suggests pages they might be interested in. It learns user interests from the pages requested by the user. The learned model of user interests is then used to suggest hyperlinks on new HTML-pages requested by and presented to the user via Web browser that enables connection to "proxy" eg. NETSCAPE. Section 2 gives an overview of Personal WebWatcher, its functionality and some directions for further development. Section 3 describes some of the research problems related to Personal WebWatcher's learning. The structure of the system and its implementations are described in Section 4 and Perl code is given in the Appendix. Results of the first experiments are given in Section 5.

2 "Personalizing" WebWatcher

Personal WebWatcher is mainly inspired by WebWatcher: "a Learning Apprentice for the World Wide Web" [2], [16] and some other work related to learning apprentice and learning from text [17], [19], [21], [25]. The idea of a learning apprentice is to automatically customize to individual users, using each user interaction as a training example.

WebWatcher can be described as an agent that assists users in locating information on the WWW. It learns by observing a user on her/his way through the WWW and suggests interesting hyperlinks whenever it is confident enough. The idea is that the user provides a few keywords describing a search goal and WebWatcher highlights related hyperlinks on the current page and/or adds new hyperlinks to the current page. It can also suggest pages related to the current page using information stored in the structure of hypertext without considering the text itself [16], or send an e-mail message to the user whenever specified pages change. The same WebWatcher version is designed to serve all users, collecting information and sharing it between users. For example, if someone recognizes a page as being related to the keywords she/he typed in the system at the beginning of the search, this can be useful for any user searching for similar

information.

Unlike WebWatcher, Personal WebWatcher (PWW) is structured to specialize for a particular user, modeling her/his interests. It "watches over the user's shoulder" the similar way WebWatcher does, but it avoids involving the user in its learning process (it doesn't ask the user for any keywords or opinions about pages). It solely records the addresses of pages requested by the user and highlights hyperlinks that it believes will be of interest. In the learning phase (typically during the night), requested pages are analyzed and a model of user interests is generated/updated. This model is used to give advice for hyperlinks on retrieved HTML-pages requested by and presented to the user via Web browser.

Since each user has her/his own copy of the system - her/his own agent, these agents can communicate and exchange information on a base of similarity between their users, often referred to as collaborative or social learning [22], [31]. There are also some other types of "Personal agents" the user could use, for example, an agent for Calendar Apprentice [25], and these agents can exchange information they have about the same user in different fields/activities. we plan to investigate these in the future work on Personal WebWatcher.

3 PWW "Behind the stage"

There are many research question behind Personal WebWatcher, that wait to be answered. Here, we consider some of them that are related to learning. There are also many others, for example, how to design communication between user and agent and between different agents, how to provide privacy to the user, to which extent agent should involve user to its learning/exploration process, ...

If we concentrate on learning, we first want to know what kind of problem are we faced with, how to represent it to some learning algorithm and which kind of algorithm to use. Currently we restricted our work to text documents (plain text and HTML) so we are faced with the problem of text-learning having mainly short to medium size documents with varying vocabulary. This section gives different approaches related to (1) document representation, (2) feature selection and (3) learning used on text-learning problem. Table 1 summarizes them over some related papers in order to give an idea about current trends.

3.1 Document representation

The frequently used document representation in Information Retrieval and text-learning is the so called $TFIDF$-vector representation. It is a *bag-of-words representation*: all words from the document are taken and no ordering of words or any structure of text is used (see Figure 1). Since most of our documents are in HTML format, there is a well defined underlying structure that could be used. There is also additional information in plain text, for example structure of sentences, position of words or neighboring words. The question is how much can we gain considering additional information in learning (and what information to consider) and what is the price we have to pay for it? There is currently no well studied comparison or directions that we are aware of. There is

Paper reference	Doc. Representation	Feature Selection	Learning
Apté et al. [1]	bag-of-words (frequency)	stop list+ frequency weight	Decision Rules
Armstrong et al. [2]	bag-of-words (Boolean)	mutual info.	TFIDF, Winnow, WordStat
Balabanovic et al. [3]	bag-of-words (frequency)	stop list+stemming+ keep 10 best words	TFIDF
Bartell et al. [4]	bag-of-words (frequency)	latent semantic indexing using SVD	—
Berry et al. [5] Foltz and Dumais [12]	bag-of-words(frequency)	latent semantic indexing using SVD	TFIDF
Cohen [8]	bag-of-words (Boolean)	infrequent words pruned	Decision Rules ILP
Joachims [17]	bag-of-words (frequency)	in/frequent words+ mutual info.	TFIDF, PrTFIDF, Naive Bayes
Lewis et al. [23]	bag-of-words (Boolean)	log likelihood ratio	logistic regression combined with Naive Bayes
Maes [24]	bag-of-words+ header info.	mail/news header info.+ selecting keywords	Memory-Based reasoning
Pazzani et al. [26]	bag-of-words (Boolean)	stop list+ mutual info.	TFIDF, Naive Bayes, Nearest Neighbor, Decision Trees
Sorensen and Mc Elligott [33], [10]	n-gram graph (only bigrams)	weighting graph edges	connectionist combined with Genetic Algorithms
Yang [35]	bag-of-words (Boolean, frequency, TFIDF)	stop list	adapted k-Nearest Neighbor

Table 1: Document representation, feature selection and learning algorithms used in some related work.

some evidence in Information Retrieval research, that for long documents, considering information additional to bag-of-words is not worth efforts. But our documents are mostly HTML-documents on the WWW and they are not especially long!

In the process of using text to learn how to give advice for a hyperlink, different approaches can be used to decide which part of text to use and how to represent it. Personal WebWatcher uses an approach similar to that of WebWatcher. In WebWatcher the bag-of-words representation is used, where considered text consists of underlined words, words in the sentence containing the hyperlink, words in all the headings above the hyperlink and words given as keywords by the user [2]. Some later versions of the WebWatcher system change slightly the way of constructing text for learning, eg. adding words in the document retrieved behind hyperlink. Many current systems that learn on text use the bag-of-words representation using either Boolean features indicating if specific word occurred in document (eg. [2], [8], [23], [26], [35]) or frequency of word in a given document (eg. [1], [3] [4], [5], [17], [35]). There is also some work that uses additional information such as word position [8] or word tuples called n-grams [33].

We decided to use the bag-of-words representation using frequency of word and observe success of given advice (whether user selected the advised hyperlink). In case of poor system performance, some additional information from HTML-structure could be added, for example, frequency of word in headlines of a given document. We would

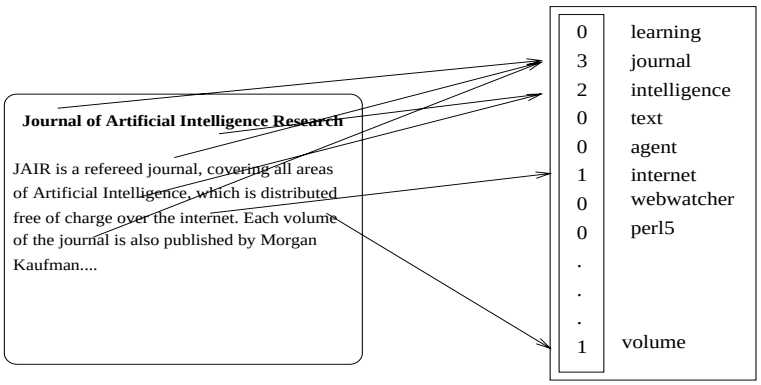


Figure 1: Bag-of-words representation using frequency vector.

like to spend time on extracting and using this additional information only when highly probable that we will gain a decent improvement in system performance. This is currently under research.

3.2 Feature selection

If we decide to ignore all the additional information and use the bag-of-words approach, we still end up with several tens of thousands of different words that occur in our documents. Not only is using all these words time-consuming but also many of them are not really important for our learning task. One of the approaches to reduce number of different words is to use "stop-list" containing common English words (eg. a, the, with) or pruning the infrequent and/or very frequent words ([8], [17]). There is also the possibility of word stemming. Many approaches introduce some sort of word weighting and select only the best words (eg. [1], [2], [3], [17], [23], [26]) or reduce dimensionality using latent semantic indexing with singular value decomposition (eg. [4], [5], [12]). Some of the Machine Learning techniques for feature selection could also be used (eg. [7], [32], [18]) but most of them take too long in situations with several tens of thousands of features.

In the current implementation of PWW we weight words using mutual information between word occurance and class value [34], the same way as used in [2] or [17]. The research topic is the number of best words to consider and we observe its influence on classification accuracy and precision of the best suggested hyperlink (see Section 5). Mutual information assigns higher weight to the words that make better distinction between interesting and uninteresting documents. One of the practical problems during classification is that a new document often contains very few or even non of the selected words. Since we are more interested in positive class (interesting documents) and we want to have words that are frequent, it might be better to include in the weighting formula the probability of a word occurring in the positive class or frequency of the word ([23]) $TF(w)\log(\frac{P(w|c)}{P(w|\bar{c})}$, where w is a selected word, c is the positive class and $TF(w)$ is the frequency of word w.

We plan to make additional experiments using the proposed combined weighting as

well as using some other weighting methods. For example, combining a stop list with weighting words by their frequency and keeping the most frequent words [35] or using word weighting used in the odds ratio method [27]. Odds ratio is the method of document ranking according to their relevance for a given problem (eg. being interesting for user). Ranking of documents is defined as $ranking(d,c) = \log \frac{P(c|d)}{P(\bar{c}|d)} = \log \frac{P(c)P(d|c)}{P(\bar{c})P(d|\bar{c})} = \log \frac{P(c)\Pi_{w \in d} P(w|c)}{P(\bar{c})\Pi_{w \in d} P(w|\bar{c})} = \sum_{w \in d} weight(w) + const.$ Word weight that it defines can be used for feature selection, weighting all words and selecting words with the highest weight. This word weight is defined as

$$weight(w) = \log \frac{p(w)(1 - \bar{p}(w))}{(1 - p(w))\bar{p}(w)}$$

where $p(w) = \frac{TF(w,c)+const.}{\#docs(c)+1}$, where $docs(c)$ is number of documents in class c and $\bar{p}(w)$ is the same as $p(w)$ except that c is substituted with $\bar{c}$. Shaw [28] proposed special handling of singularities in the above formulas for $p(w)$ and $\bar{p}(w)$, namely, $p(w) = \frac{1}{\#docs^2}$ when $TF(w,c) = 0$ and $p(w) = 1 - \frac{1}{\#docs^2}$ when $TF(w,c) = \#docs(c)$

3.3 Learning algorithm

One of the well-established techniques for text in Information Retrieval is to represent each document as a $TFIDF$-vector in the space of words that appeared in training documents [30], sum all interesting document vectors and use the resulting vector as a model for classification (based on [29] relevance feedback method). Each component of a document vector $d^{(i)} = TF(w_i, d)IDF(w_i)$ is calculated as the product of TF (Term Frequency — number of times word w_i occurred in document) and $IDF = log \frac{D}{DF(w_i)}$ (Inverse document Frequency), where D is the number of documents and document frequency $DF(w_i)$ is the number of documents word w_i occurred in at least once. The exact formulas used in different approaches may slightly vary (some factors are added, normalization performed [30]) but the idea remains the same. A new document is then represented as a vector in the same vector space as the generated model and the distance between them is measured (usually defined as a cosine of angle between vectors) in order to classify the document. This technique has already been used in Machine Learning experiments on World Wide Web data (eg. [2], [3], [5], [17], [26]).

There are also some other techniques for model generation that have been used in text-learning. Armstrong et al. [2] used a statistical approach they called WordStat that assumes mutual independence of words and defines probability of class c as $P(c) = 1 - \Pi_w(1 - P(c/w))$. Pazzani et al. [26] used a Naive (Simple) Bayesian classifier on Boolean vectors, that assumes independence of words and defines probability of class c for given document doc that contains words w as proportional to $P(c)\Pi_w P(c/w)$. They also used Nearest Neighbor and symbolic learning using Decision Trees. A variant of k-Nearest Neighbor was also used by Yang [35], where relevance of class c given document doc is defined as $rel(c/doc) = \sum_{i=1}^{k} similarity(doc, D_i)P(c/D_i)$ and $similarity$ is measured by cosine between vectors and $P(c/D_i) = \frac{\#D_i Inc}{\#D_i Intrainingdata}$ (same document may occur several times being classified in different categories). Joachims [17] introduced Probabilistic $TFIDF$ that takes into account document representation Θ and de-

fines probability of class c for given document doc that contains words w as $P(c/doc) = \sum_w \frac{P(c)P(w/c)}{\sum_i P(c_i)P(w/c_i)} P(w/doc, \Theta)$ where $P(w/c) = \frac{TF(w,c)}{\sum_i TF(w_i,c)}$ and $P(w/doc, \Theta) = \frac{TF(w,doc)}{\sum_i TF(w_i,doc)}$. He also used Naive (Simple) Bayesian classifier on frequency vectors, the same as we used in PWW. It assumes independence of words and defines probability of class c for given document doc that contains words w as

$$P(c/doc) = \frac{P(c)\Pi_w P(w/c)^{TF(w,doc)}}{\sum_i P(c_i)\Pi_w P(w/c_i)^{TF(w,doc)}}$$

where $P(w/c) = \frac{1+TF(w,c)}{\#words+\sum_i TF(w_i,c)}$; $TF(w,c)$ denotes frequency of word w in documents of class c and $TF(w,doc)$ denotes frequency of word w in document doc.

Apté et al. [1] used Decision Rules and observed that in case of different topics being categories, it is better to select features for each given topic (using stop-list and frequency weighting) than for all topics at once, even if the set of features is additionally reduced for each topic using entropy-based measure to weight features. Cohen [8] used Decision Rules and the Inductive Logic Programming systems FOIL and FLIPPER. Lewis and Gale [23] used a combination of a Naive Bayesian classifier and logistic regression defining probability of class c for given document doc that contains words w as $P(c/doc) = \frac{exp(a+b\sum_w log\frac{P(w/c)}{P(w/\bar{c})})}{1+exp(a+b\sum_w log\frac{P(w/c)}{P(w/\bar{c})})}$.

Maes [24] used Memory-Based reasoning, McElligot and Sorensen [10], [33] used a connectionist approach combined with Genetic Algorithms.

We decided to test different learning algorithms on PWW data (see Section 5), since it is not clear which algorithm is the most appropriate. The current version of PWW uses a Naive (Simple) Bayesian classifier on frequency vectors to generate a model of user interests, that is used for advising hyperlinks.

4 Structure of Personal WebWatcher

Personal WebWatcher consists of two main parts: a **proxy server** that interacts with the user via Web browser and a **learner** that provides the user-model to the server (see Figure 2). Communication between them is via disk; the **proxy** saves addresses of visited documents (URLs) and the **learner** uses them to generate model of user interests. The whole system is implemented in approximately 2500 lines of Perl code and 1500 lines of C^{++} code.

The **proxy server** consists of three main parts, each implemented as a Perl script: **proxy** (additionally calls external **fetcher** code to fetch the page), **adviser** and **classifier** (calls external C^{++} code for classification). **Proxy** waits in an infinite loop for a page request from browser. On request, it fetches the requested document and; if it is an HMTL-document adds advice and; forwards the document to the user. To add advice **proxy** forwards the page to **adviser**, that extracts hyperlinks from document and calls external code for **classification** that uses generated user-model. A limited number of hyperlinks that are scored above some threshold are recommended to the

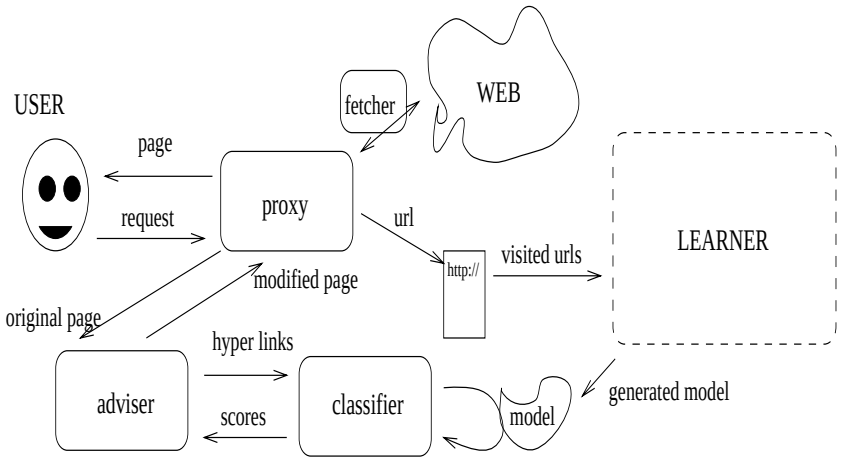


Figure 2: Structure of Personal WebWatcher. The learning part is described separately .

user, indicating their scores using graphical symbols placed around each advised hyperlink. For example, in Figure 3 three hyperlinks are suggested by PWW: "Machine Learning Information Services" and two project members (Dayne Freitag, Thorsten Joachims). There is a banner at the top of the page showing that PWW is "watching over the user's shoulder".

4.1 Structure of the learning module

Learner works in two versions: learning a new model from scratch (LEARNER) or updating an existing model (UPDATER) as shown in Figure 4. The difference is that the first one has to define the domain (words to be used) and starts learning with an empty model, while the second one has already defined which words to use in representing documents as frequency vectors and has an existing model to modify. Both versions fetch visited documents and documents one step behind the hyperlinks of visited documents and store them as positive or negative examples of user interests, depending whether the user visited the document or not (getDoc in Figure 4). Hyperlinks from visited HTML-documents are extracted and stored in an extended hyperlink format, the same that is used by **adviser**. Each hyperlink is represented in extended format, taking into account underlined words, words in a window around a hyperlink and words in all the headings above the hyperlink. Using hyperlinks represented only with underlined words is often a bad idea, eg. click here...

Hyperlinks whose documents were visited by the user are considered to be positive examples, and all the other to be negative examples of the user interests. The idea is that all hyperlinks were presented to the user and the user chose to visit some of them that meet her/his interests. This simplification is introduced to minimize users

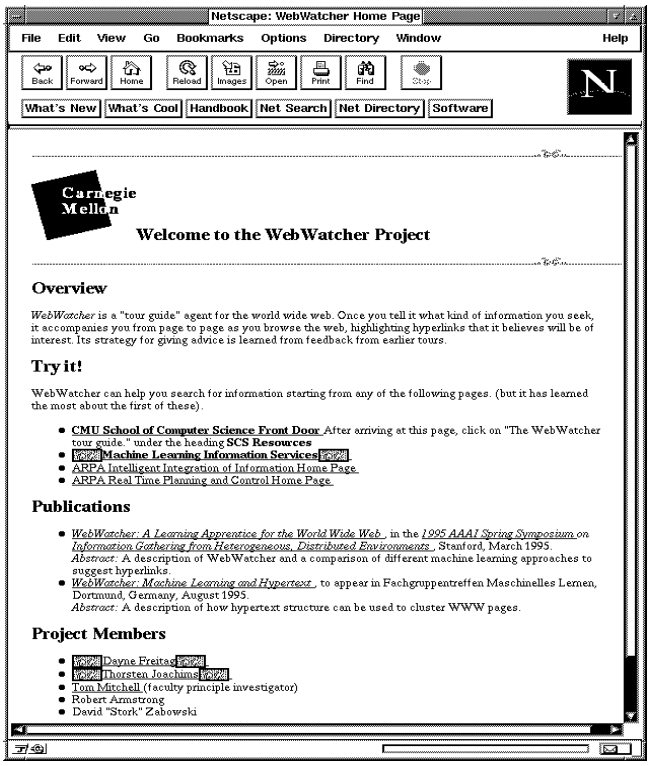


Figure 3: Example of HTML-page presented to the user by PWW.

involvement in the learning process and enable learning without asking user for page rating. We plan to make it optional in some later versions of the system.

LEARNER transforms documents into examples in two phases: (1) (docs2exs and docs2addexs in Figure4) parsing each document, assigning an index to each word and representing it in three files as a line of word indices containing: all words, only headline words, only underlined words. (2) (exs2vec in Figure4) calculating score (eg. information gain) for each word, selecting some top words and represent documents as bag-of-words keeping frequency for each of the top words.

UPDATER uses given top words (domain definition) and represent documents as bag-of-words, the same way LEARNER does (docs2ddexs in Figure 4).

During the model generation(GenModel in Figure 4), the system can ask for additional information about some words (stating which kind of information - functions and on which words - basic attributes). The kind of information it could ask for is specified as so called background knowledge eg. feature saying how many times a word occurred in document headlines (genAttr in Figure 4). This is currently under development.

Learning part consists basically of eight Perl scripts that call external C^{++} code for model generation/updating. Two scripts integrate parts of LEARNER (UPDATER) and the other six that are represented with rectangles in Figure 4 (getDocs, docs2exs, docs2addexs, exs2vec, genAttr, docs2ddexs). Two additional rectangles in Figure 4 GenModel, UpdateModel represent C^{++} code.

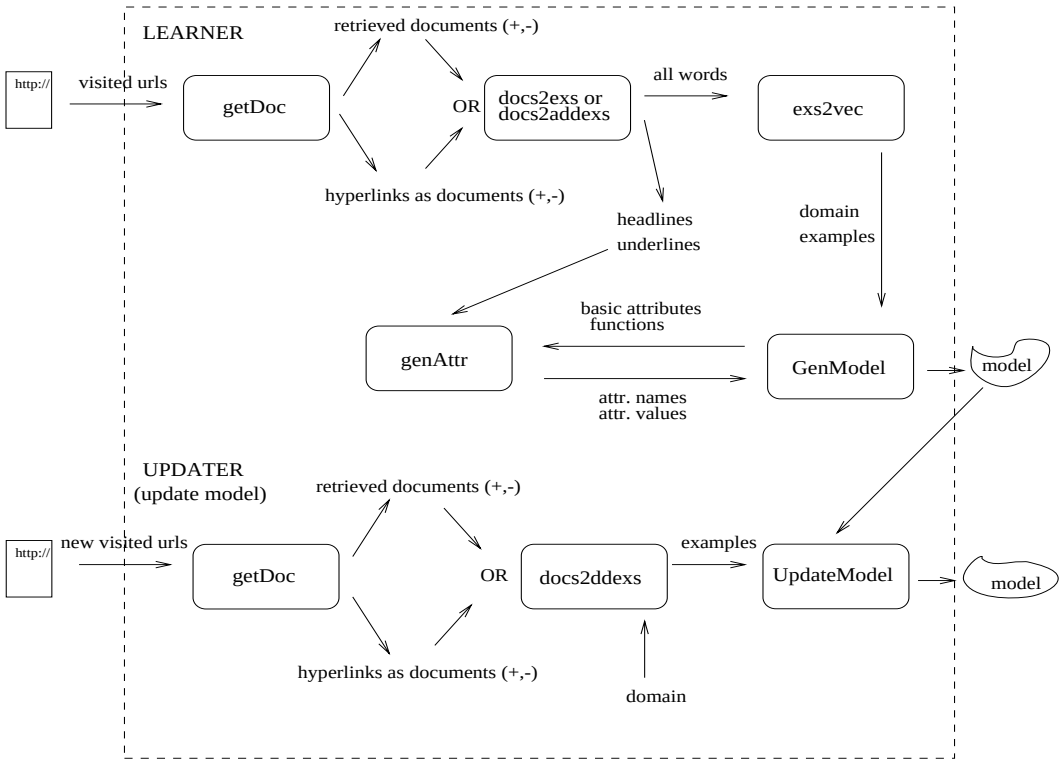


Figure 4: Structure of PWW Learner.

4.2 Model of user interests

The model of user interests is designed to predict if some document is positive or negative example of user interests. It is used to advice hyperlinks on the HTML-document requested by the user. Since the prediction should be performed while user is waiting for a HTML-document, we are actually predicting interestingness of document based on the hyperlink pointing to it, and not document itself (retrieving documents behind the requested hyperlinks is usually time consuming)

The model of user interests is generated "off-line", usually during the night and thus its generation is not so critical in time as its usage for prediction. One of the simplest idea for learning is to use hyperlinks that occurred on the documents presented to the user as training examples and learn to predict if a new hyperlink is positive or negative example of the user interests:

$$User_{HL} : HyperLink \rightarrow \{pos, neg\}$$

What we use is an extended representation of hyperlink (see Section 4.1), that tries to capture information related to the document behind a hyperlink. But during the learning phase we can afford using more time than when adding advice, so why not retrieving documents behind hyperlinks, instead of using the extended hyperlink representation? In that case, we can learn the model of user interests directly from documents whose

interestingness we are trying to predict:

$$User_{DOC} : Document \rightarrow \{pos, neg\}$$

In this case, we end up with using the model generated from documents to predict interestingness of hyperlinks. Since our hyperlinks are represented as short HTML-documents (including headlines and some portions of text) it is not so unusual, but we could also learn a model that predicts document content based on a given hyperlink:

$$Document_{HL} : HyperLink \rightarrow Document$$

and then predict interestingness of so predicted document content using the above described model $User_{DOC}$. Our first experiments are in learning the first two models.

5 First experimental results

In order to select a good document representation, feature selection method and learning algorithm we decide to test different possibilities and compare them. Since PWW has to recommend interesting hyperlinks to the user, we are interested in measuring the precision of our system on the most highly recommended hyperlink. Precision is frequently used as a metric in Information Retrieval and it is defined as the percent of positive suggestions among all suggestions made. In our case, it is either zero (in case the best suggestion is a negative example and shouldn't be suggested to the user) or one (if we made a correct suggestion). We also measured traditional Machine Learning quality estimate classification accuracy, defined as percent of correctly classified examples (over all classes). Both quality estimates are calculated using 10-fold cross-validation technique (see Figure 5) using the same example splits for all tested algorithms.

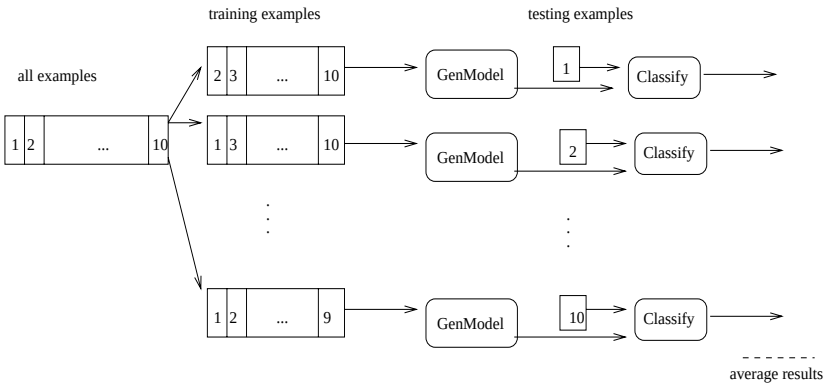


Figure 5: Illustration of 10-fold cross-validation experiments.

In first experiments we observe how vector length (number of features selected) influences model quality for two learning algorithms: Naive Bayesian classifier on frequency vector as used by Joachims [17] (see Section 3.3) and k-Nearest Neighbor approach on

frequency vectors using Euclidean distance between examples and summing class probabilities predicted by k-neighbors. We tested both algorithms on data for documents behind hyperlinks $User_{DOC}$ and data for hyperlinks $User_{HL}$ (see Section 4.2). Documents are currently represented using the bag-of-words approach (see Section 3.1) and feature selection is performed using mutual information approach (see Section 3.2). Our experiments are performed on data collected for four users participating in the HomeNet project [14] with the data characteristics given in Table 2.

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

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

In all experiments k-Nearest Neighbor achieved slightly higher classification accuracy than the Naive Bayesian classifier (see Figures 6, 7, 8, 9), but the difference is significant only in one out of six experiments (see Figure 8). Adding more than approximately 50 features doesn't appear to help for classification accuracy, but it also doesn't hurt. High classification accuracy achieved for all four users by k-Nearest Neighbor algorithm is in fact default accuracy if negative class is predicted for all documents. So what we are really interested in is making good predictions for positive documents, and that is why we decide to measure precision of the best suggested hyperlink.

Precision varies much more with vector size than classification accuracy (see Figures 10, 11, 12, 13), it actually drops in seven out of eight experiments. It seems that k-Nearest Neighbor is more stable in precision than Naive Bayesian classifier, achieving higher precision for longer vectors and about the same for up to 100 features, with exception for one user (see Figure 12), where the precision of Naive Bayesian classifier is for most vector sizes 0 for the document model and much better on short vectors than k-Nearest Neighbor for the hyperlink model.

In order to draw some conclusion about vector size and quality of algorithms, we need to perform more experiments on different users. These first experiments show that increasing vector size probably isn't as beneficial as one could expect and it even could hurt precision of the best suggestion. There is no evidence that algorithms differ substantially in classification accuracy, although k-Nearest Neighbor seems to be

more promising both in accuracy and precision. If further experiments confirm the hypothesis that long vectors are not advisable, a closer look at the short vectors (eg. see Figure 14) should give an idea about number of features that work well for both accuracy and precision.

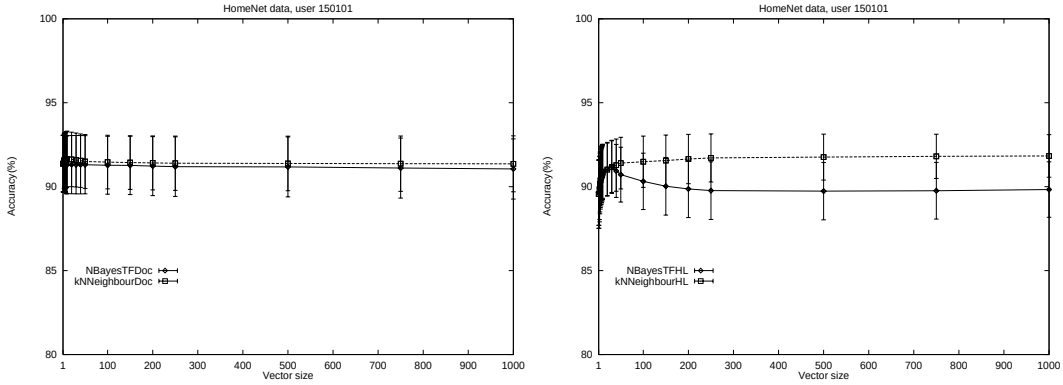


Figure 6: Influence of vector size to classification accuracy of model based on documents (left) and hyperlinks (right) for the HomeNet project user with id: 150101. Notice that classification accuracy scale starts at 80% accuracy. Error bars show standard deviation since accuracy is calculated as average of 10 results.

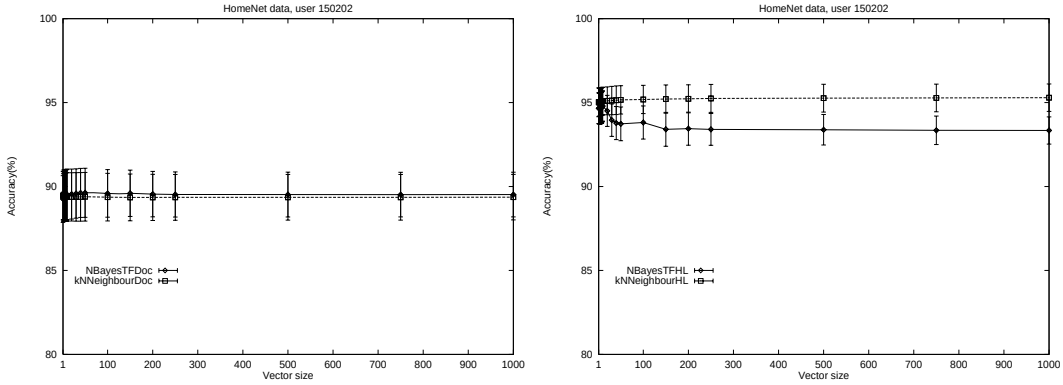


Figure 7: Influence of vector size to classification accuracy of model based on documents (left) and hyperlinks (right) for the HomeNet project user with id: 150202. Notice that classification accuracy scale starts at 80% accuracy. Error bars show standard deviation since accuracy is calculated as average of 10 results.

References

[1] Apté, C., Damerau, F., Weiss, S.M., Toward Language Independent Automated Learning of Text Categorization Models, *Proc. of the 7th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, Dubline, 1994.

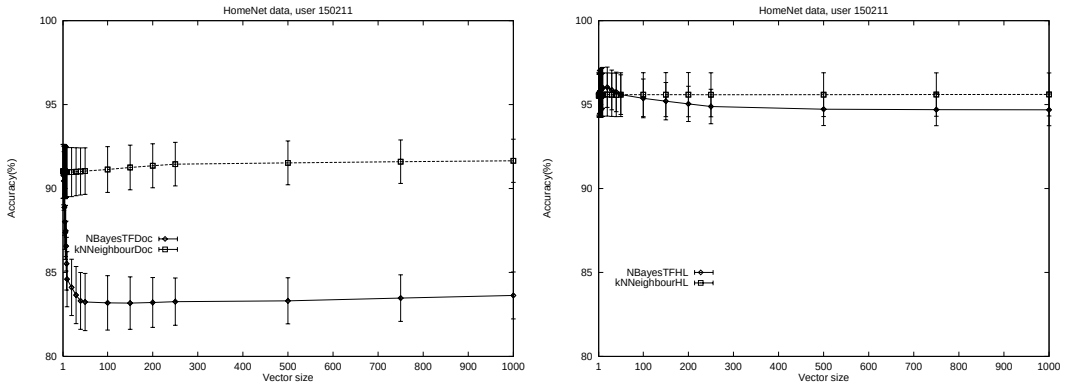


Figure 8: Influence of vector size to classification accuracy of model based on documents (left) and hyperlinks (right) for the HomeNet project user with id: 150211. Notice that classification accuracy scale starts at 80% accuracy. Error bars show standard deviation since accuracy is calculated as average of 10 results.

[2] Armstrong, R., Freitag, D., Joachims, T., Mitchell, T., WebWatcher: A Learning Apprentice for the World Wide Web, *AAAI 1995 Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, Stanford, March 1995. URL: http://www.cs.cmu.edu/afs/cs/project/theo-6/web-agent/www/webagents-plus.ps.Z

[3] Balabanović, M., Shoham, Y., Learning Information Retrieval Agents: Experiments with Automated Web Browsing, *AAAI 1995 Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments*, Stanford, March 1995. URL: http://robotics.stanford.edu/people/marko/papers/lira.ps

[4] Bartell, B.T., Cottrell, G.W., Belew, R.K., Latent Semantic Indexing is an Optimal Special Case of Multidimensional Scaling, *Proceedings of the ACM SIG Information Retrieval*, Copenhagen, 1992.

[5] Berry, M.W., Dumais, S.T., OBrein, G.W., Using linear algebra for intelligent information retrieval. *SIAM Review*, Vol. 37, No. 4., pp. 573–595, December 1995.

[6] Burke, R., Hammond, K., Kozlovsky, J., Knowledge-based Information Retrieval for Semi-Structured Text, *Working Notes from AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*, pp. 19–24, American Association for Artificial Intelligence, 1995.

[7] Caruana, R., Freitag, D., Greedy Attribute Selection, *Proc. of the 11th International Conference on Machine Learning ICML94*, pp. 28—26, 1994.

[8] Cohen, W.W., Learning to Classify English Text with ILP Methods, *Workshop on Inductive Logic Programming*, Leuven, September 1995.

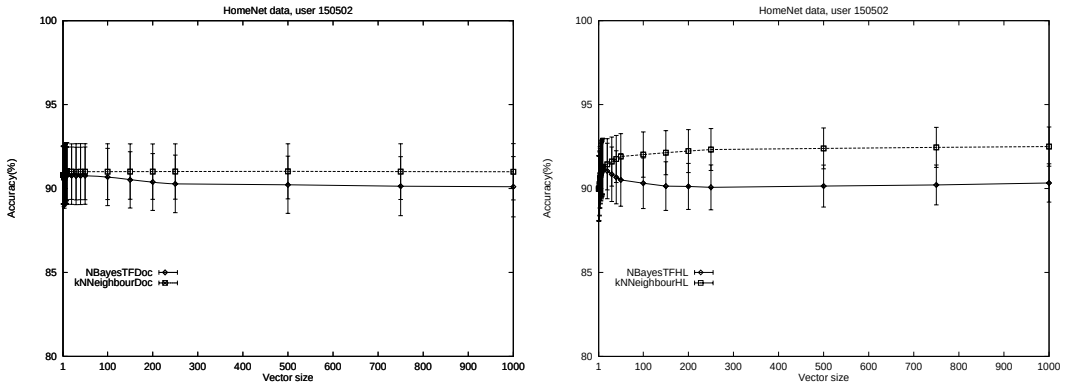


Figure 9: Influence of vector size to classification accuracy of model based on documents (left) and hyperlinks (right) for the HomeNet project user with id: 150502. Notice that classification accuracy scale starts at 80% accuracy. Error bars show standard deviation since accuracy is calculated as average of 10 results.

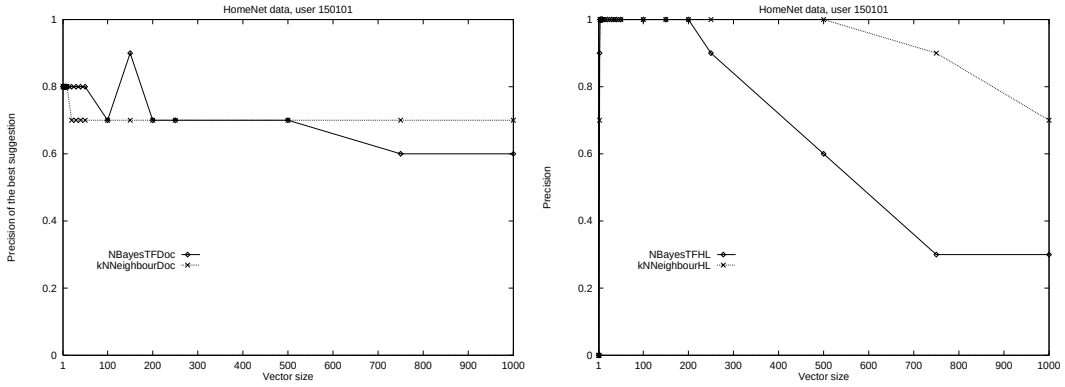


Figure 10: Influence of vector size to precision of model based on documents (left) and hyperlinks (right) for the HomeNet project user with id: 150101.

[9] Drummond,C., Ionescu, D., Holte, R., A Learning Agent that Assists the Browsing of Software Libraries, *Technical Report TR-95-12*, Computer Science Dept., University of Ottawa, 1995.

[10] Mc Elligott, M., Sorensen, H., An emergent approach to information filtering, *Abakus. U.C.C. Computer Science Journal*, Vol 1, No. 4, December 1993.

[11] Etizioni, O., Weld, D., A Softbot-Based Interface to the Internet, *Communications of the ACM* Vol. 37, No. 7, pp.72–79, July 1994.

[12] Foltz, P. W. and Dumais, S. T., Personalized information delivery: An analysis of information filtering methods. *Communications of the ACM*, 35(12), pp.51—60, 1992.

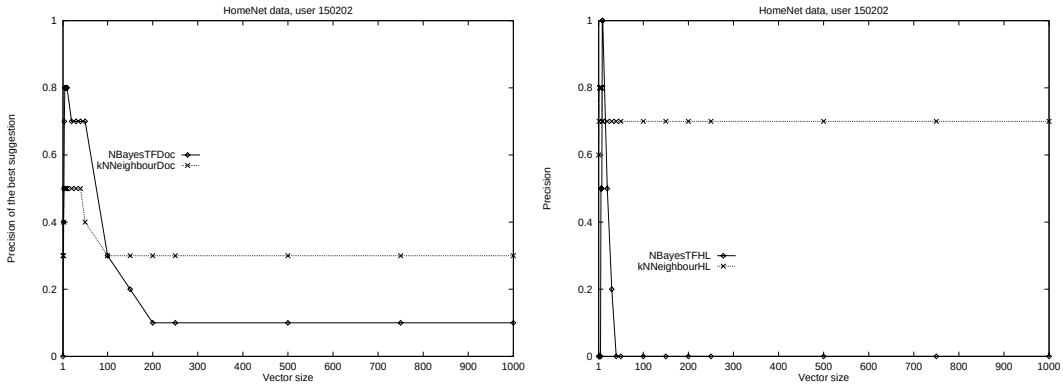


Figure 11: Influence of vector size to precision of model based on documents (left) and hyperlinks (right) for the HomeNet project user with id: 150202.

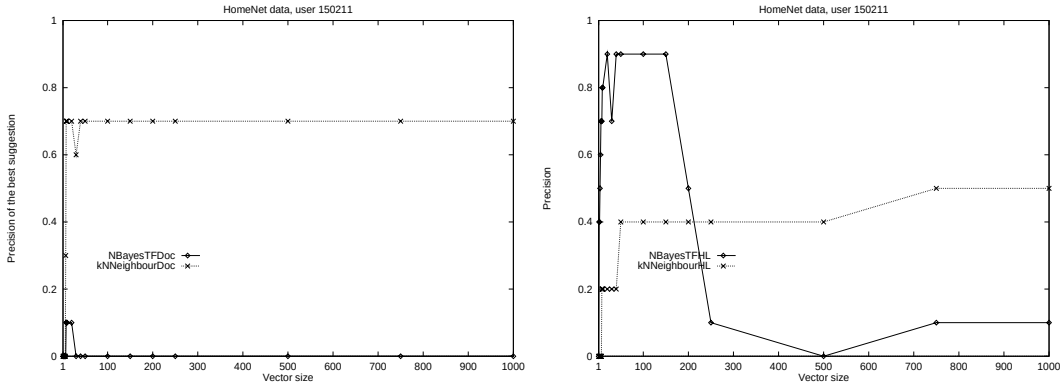


Figure 12: Influence of vector size to precision of model based on documents (left) and hyperlinks (right) for the HomeNet project user with id: 150211.

[13] Hammond, K., Burke, R., Schmitt, K., A Case-Based Approach to Knowledge Navigation, *AAAI Workshop on Indexing and Reuse in Multimedia Systems*, pp. 46–57, American Association for Artificial Intelligence, 1994.

[14] HomeNet: a research project at Carnegie Mellon University, Pittsburgh, PA, USA, 1996, URL:http://homenet.andrew.cmu.edu/progress/

[15] Holte, R.C., Drummond, C., A Learning Apprentice For Browsing, *AAAI Spring Symposium on Software Agents*, 1994.

[16] Joachims, T., Mitchell, T., Freitag, D., Armstrong, R., WebWatcher: Machine Learning and Hypertext, *Fachgruppentreffen Maschinelles Lernen*, Dortmund, August 1995.

[17] Joachims, T., A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization, *Technical report, CMU-CS-96-118*, School of Computer Science, Carnegie Mellon University, March 1996.

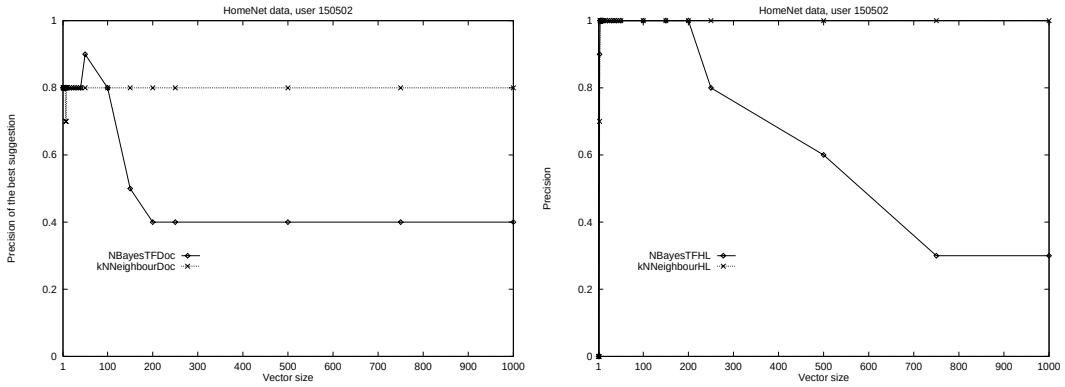


Figure 13: Influence of vector size to precision of model based on documents (left) and hyperlinks (right) for the HomeNet project user with id: 150502.

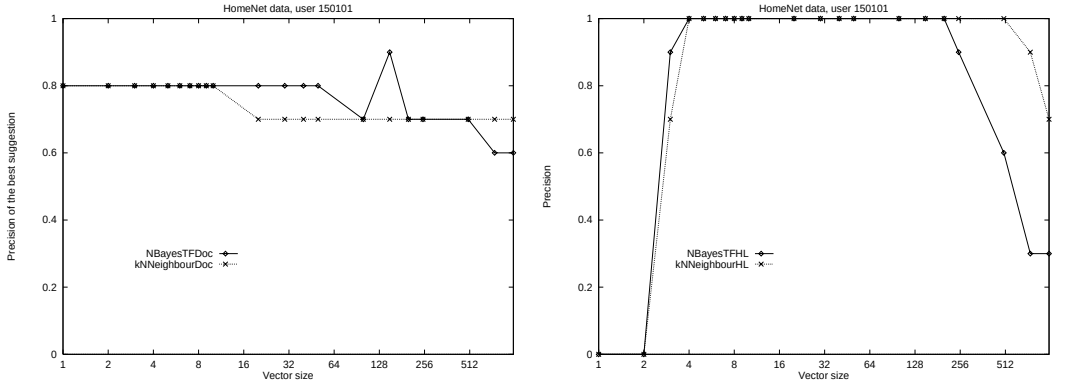


Figure 14: Influence of vector size to precision of model based on documents (left) and hyperlinks (right) for the HomeNet project user with id: 150101. Notice that log-scale on x-aix.

[18] John, G.H., Kohavi, R., Pfleger, K., Irrelevant Features and the Subset Selection Problem, *Proc. of the 11th International Conference on Machine Learning ICML94*, pp. 121—129, 1994.

[19] de Kroon, H.C.M, Mitchell, T., Kerckhoffs, E.J.H., Improving Learning Accuracy in Information Filtering, *ICML-96 Workshop Machine learning meets human computer interaction*, 1996. URL: http://www.ics.forth.gr/ mous-taki/ICML96_HCI_ML/kroon.ps

[20] Krulwich, B., Burkey, C., The ContactFinder agent: Answering buletin board questions with referrals, *Proceedings of the Thirteenth National Conference on Artificial Intelligence AAAI 96*, pp.10—15, 1996.

[21] Lang, K., News Weeder: Learning to Filter Netnews, *Proc. of the 12th International Conference on Machine Learning ICML95*, 1995.

[22] Lashkari, Y., The WebHund Personalized Document Filtering System, 1995.

[23] Lewis, D.,D., Gale, W., A., A Sequential Algorithm for Training Text Classifiers, *Proc. of the 7th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, Dubline, 1994.

[24] Maes, P., Agents that Reduce Work and Information Overload, *Communications of the ACM* Vol. 37, No. 7, pp.30–40, July 1994.

[25] Mitchell, T., Caruana, R., Freitag, D., McDermott, J., Zabowski, D., Experience with a Learning Personal Assistant, *Communications of the ACM* Vol. 37, No. 7, pp.81–91, July 1994. URL: http://www.cs.cmu.edu/afs/cs/user/mitchell/ftp/cacm.ps.Z

[26] Pazzani, M., Muramatsu, J., Billsus, D., Syskill & Webert: Identifying interesting web sites, *AAAI Spring Symposium on Machine Learning in Information Access*, Stanford, March 1996 and *Proceedings of the Thirteenth National Conference on Artificial Intelligence AAAI 96*, pp.54—61, 1996.

[27] van Rijsbergen, C.J,. Harper, D.J., Porter, M.F., The selection of good search terms, *Information Processing & Management*, 17, pp.77—91, 1981.

[28] Shaw Jr, W.M., Term-relevance computations and perfect retrieval performance, *Information Processing & Management*, 31(4), pp.491—498, 1995.

[29] Rocchio, J., Relevance Feedback in Information Retrieval, in *The SMART Retrieval System: Experiments in Automatic Document Processing*, Chapter 14, pp.313—323, Prentice-Hall Inc., 1971.

[30] Salton, G., Buckley, C., Term Weighting Approaches in Automatic Text Retrieval, *Technical report, COR-87-881*, Department of Computer Science, Cornell University, November 1987.

[31] Shardanand, U., Maes, P., Social Information Filtering: Algorithms for Automating "Word of Mouth", *CHI'95 Mosaic of Creativity*, pp. 210—217, 1995.

[32] Skalak, D.B., Prototype and Feature Selection by Sampling and Random Mutation Hill Climbing Algorithms, *Proc. of the 11th International Conference on Machine Learning ICML94*, pp. 293—301, 1994.

[33] Sorensen, H., McElligott, M., PSUN: A Profiling System for Usenet News, *CIKM'95 Intelligent Information Agents Workshop*, Baltimore, December 1995.

[34] Quinlan, J.R., Constructing Decision Tree in *C4.5: Programs for Machine Learning*, pp.17–26, Morgan Kaufman Publishers, 1993.

[35] Yang, Y., Expert Network: Effective and Efficient Learning form Human Decisions in Text Categorization and Retrieval, *Proc. of the 7th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieva*l, Dubline, 1994.

Exhibit 4

A Personal Evolvable Advisor for WWW Knowledge-Based Systems

M.Montebello, W.A.Gray, S.Hurley
Computer Science Department
University of Wales, Cardiff.
email: (m.montebello,w.a.gray,s.hurley)@cs.cf.ac.uk

Abstract

The immense size of the distributed WWW knowledge-base and the dramatic rapid increase in the volume of data on the Internet, requires techniques and tools that reduce users' information overload and improve the effectiveness of online information access. Despite the potential benefits of existing indexing, retrieving and searching techniques in assisting users in the browsing process, little has been done to ensure that the information presented is of a high recall and precision standard. In this position paper we present a system that reuses the information generated from search engines together with previously developed systems, and adapts it, by generating user profiles, to better meet the needs and interests of the users by improving recall and precision measures.

Introduction - Background and Motivations

In recent years there has been a well-publicized explosion of information available on the Internet, and a corresponding increase in usage. This is particularly true of the World-Wide Web (WWW) [Berners-Lee et al., 1994] and its associated browsers which allow relative easy access to the information available, and thus make it accessible to a wider audience. The WWW is a major knowledge dissemination system that makes the world's staggering wealth of knowledge and experience, stored on server machines scattered across the internet, accessible to the on-line world.

When people access the web, they are either searching for specific information, or they are simply browsing, looking for something new or interesting (often referred to as *surfing*). The WWW's sheer scale and its exponential growth renders the task of simply finding information, tucked away in some Web site, laborious, tedious, long-winded and time consuming. The fact that a user's time is valuable and that relevant information might not be accessed, imposes serious restrictions on the efficient use of the WWW and the benefits that users can expect from their interaction.

It is well documented that traditional search engines provide services which are far from satisfactory [DeBra and Post, 1994,Spetka, 1994,Srinivasan et al., 1996]. Users are faced with the problem of these search engines being too generalised and not focused enough to their real and specific needs. This triggered further research to develop more sophisticated techniques and agent like systems that make use of the user profile to personalise the service they provide and add value to the information they presented [Pazzani et al., 1996,Green and Edwards, 1996,Mladenic, 1996,Pazzani et al., 1996].

The Personal Evolvable Advisor (PEA), presented in this position paper, is a system we have developed to reuse information generated by search engines and utilise previously developed retrieval systems. Conceptually, the PEA is similar to a meta-search engine, but with the major difference that it employs user profiling to specifically target documents for individual users. In this way duplication and redundancy of information is significantly reduced, while the real needs and interests of the users are fully addressed in a more focussed retrieval.

PEA - Current Implementation

Our goal with PEA is to achieve a high recall and high precision performance score on the information presented to the user. Recall measures how efficient the system is at retrieving the relevant documents from the WWW, while precision measures the relevance of the retrieved set of documents to the user requirements. In order to obtain a high recall execution we make use of the hits returned by a number of traditional search engines together with the output from retrieval systems that have been previously developed. The reason for doing this is twofold. Firstly, we could have developed our own search engine and argued that it utilises the ultimate retrieval techniques and produced results similar to other systems. However, by making use of what other systems generate, we ensure that we obtain all the information that all of them would retrieve at the same time, and not have the problem of developing an ultimate system. Secondly, there are numerous WWW crawlers available, bombarding servers and clogging networks. By using them, we simply use other systems' knowledge-bases, rather than duplicating it, and move up to the next level of the information ``food chain'' [Selberg and Etzioni, 1995], in this way our recall is as good as can be achieved with any current system. On the other hand in order to add value to the retrieved results and maximise the precision and efficiency with which the system achieves high recall scores, we generate user profiles to predict and suggest the most suitable information for specific users. Through various

interactions the system will be able to optimize the targeting and predicting of what users are interested in, thereby improving the precision factor of the retrieved information.

The processes required by an information retrieval and filtering system include several tasks that PEA decomposes into a number of simpler tasks. Figure 1 shows the major components of the system: the WWW and the external systems at the bottom level, the underlying application software on the next level up, and the GUI at the top.

The WWW is one of the components over which we have no control. It requires no local development, but its heterogeneous, unstructured and uncensored nature causes developers

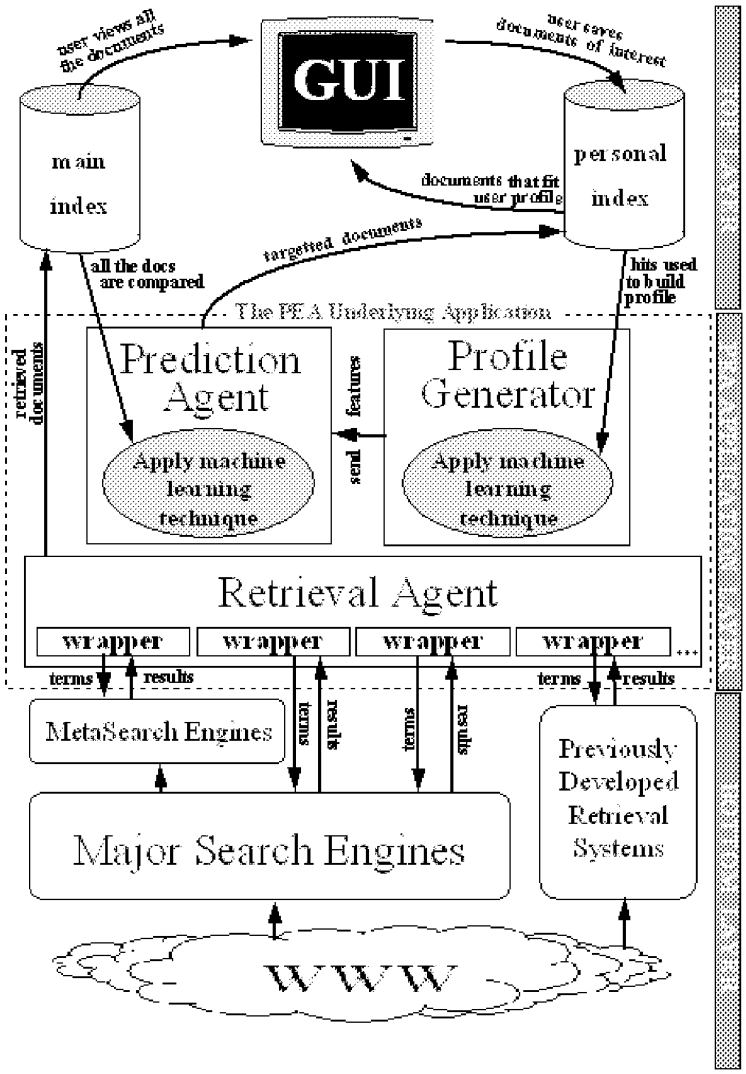


Figure 1: PEA Architecture

to face awkward coding situations in order to be able to cater for all the kinds of data found on the WWW. The WWW can be assumed to be a very large heterogeneous distributed digital information database. In order to optimize the management and exploit the potential of the WWW's vast knowledge-base we require to search and retrieve efficiently and effectively specific information for users.

The external systems utilised include some of the major search engines and also some other retrieving systems that have been developed by other research groups[Eichmann and Wu, 1996,Mladenic, 1996,Selberg and Etzioni, 1995]. They use the WWW as their source of input and we use their output as the input for PEA. All the external systems are considered to be black boxes and action is taken upon the information they output. Wrappers are used to manage the appropriate and proper handshaking between the diverse search engines and the other retrieving systems and the application layer.

Query terms are used to locate documents and retrieve results from the external systems. These results need substantial re-formatting as they usually include completely useless information like advertisements, local links and site specific information.

The underlying application layer has the difficult task of performing all the work required, transparently from the user. It makes use of the information retrieved from the external systems and attempts to improve on the recall/precision metrics mentioned earlier. By using state-of-the-art external systems we attempt to achieve a high recall rate, while using a personalised profile with specific interests for each user, we attempt to also achieve a high precision rate.

The three main components of the PEA underlying application layer (retrieval agent, profile generator and prediction agent - Figure 1) perform the necessary work to satisfy our initial motivations.

The Retrieval Agent

The retrieval agent, is responsible for aggregating all the hits returned by the external systems. It collates the results, by removing duplicates and ensuring integrity, and stores the formatted and pre-ranked results as a single list in a local database, known as the *main index*. The Java programming language was employed to develop this part of the application due to its ease in performing TCP/IP connections to allow retrieval of documents and their processing. This agent interacts with the external systems via appropriate wrappers. Every query term is employed by the wrapper which will command the associated system to locate documents from its local index and return related results. These results are basically a series of document addresses (URLs - Universal Resource Locator) which are listed within an HTML page that the external system returns. A scan through the WWW page will quickly identify the URL links and list them. Some of the links are useless to the user, so the retrieval agent initially removes adverts, duplicates, and site specific links. It then analyses the vetted URLs and accesses the document on-line. This will identify whether the link is still accessible, has moved or been removed completely. If the document is valid, then an initial paragraph from the document is extracted and saved locally in the main database index together with the reference search term, its reference within the index, the URL, and the document title. All these details will be available to the user through the GUI, and also to the prediction agent to identify if the particular document is relevant to a particular user or not.

The Profile Generator

The task that the profile generator sets out to achieve is to analyse each users' personal index and generate a profile. If users have different interests stored in their personal index, then a separate profile is required and generated for each interest. No novel machine learning technique has been developed for the profile generator. It uses specific techniques previously employed by other similar systems [Edwards et al., 1995,Green and Edwards, 1996,Payne and Edwards, 1997]. The difference is that users are able to select which technique they would like to use to generate their profile, and predict other relevant documents in future interactions. This profile generation utilizes the *term frequency/inverse document frequency* machine learning technique[Salton and McGill, 1983], but other machine learning techniques are being implemented. Profile generating systems like MAGI and UNA were relatively easy because specific and fixed fields were provided in the data they were extracting information from. Documents like email and USENET news articles have inherently static field holders embedded in them, e.g. ``to'', ``from'', ``date'', and ``subject''. These are typical examples of anchored features a developer can rely on when designing the filtering procedures. On the other hand, when considering how to perform the same task on WWW documents (normally HTML), no fixed fields are provided. Even though HTML version 3 introduced the META tag, which allows authors to specify indexing information; it is unreliable as authors can fail to use it. A developer cannot assume that HTML document authors abide by standard conventional fields within documents e.g. ``<HTML>'', ``<TITLE>'', and ``<BODY>'', due to the weak typing nature of HTML. Despite this, there are many systems that filter HTML documents e.g. WebHunter [Lashkari, 1995], LIRA [Balabanovic and Shoham, 1995], Letizia [Lieberman, 1995], WebWatcher [Armstrong et al., 1995][Joachims et al., 1997], SULLA [Eichmann and Wu, 1996], Personal WebWatcher [Mladenic, 1996], and others described in [Etzioni and Weld, 1994][Holte and Drummond, 1994] and [Perkowitz and Etzioni, 1995].

We assume that normally, when searching or even browsing, a user bookmarks a page of interest and proceeds with the activity he/she was performing. Taking this activity into perspective, all that is required is to take into consideration what the user bookmarks, and utilise this information to generate the profile. While this method may have problems of over identification, it is more reliable than asking users to assign ratings, as it is less demanding on the user's time. Another problem that many of these HTML filtering systems ignore is that machine learning techniques have a slow learning curve and require a sufficient number of examples before they can make accurate predictions. As a result a profile generator encounters problems when dealing with completely new situations. Generally this is true for all such systems and as [Maes and Kozierok, 1993] rightly argue, the user and the profile agent will gradually build up a trust relationship over time. Issues regarding how many profiles to generate for a user - one specific profile per user, a general profile for a group of users, different profiles for different users or different profiles for the same users - have been tackled differently. Some profile generators develop the `specific user profile', especially those systems which have been produced to cater for specific items like emails or newsgroups, while others specialise in a `specific topic profile', like WebFind [Monge and Elkan, 1995], MetaCrawler [Selberg and Etzioni, 1995], PAINT [Oostendorp et al., 1994], and CURRY [Krishnamurthy and Tsangaris, 1996] which recommend documents to users with the same interests or needs. Other systems, like Syskill and Webert [Pazzani et al., 1996], learn a separate profile for each topic of each user. They argue that many users have multiple interests and it will be possible to learn a more accurate profile for each topic separately since the factors that make one topic interesting are unlikely to make another interesting. We take this argument one step further, and argue that what one user finds

interesting in a specific topic, differs from what another user describes as interesting about the same topic. Therefore, different profiles need to be generated for every different interest a user has if the predicted results are to be focused accurately.

The Prediction Agent

The user interest profile generated by the profile generator will be used by the prediction agent in combination with the extracted features from documents in order to predict and suggest new interesting documents to a user. Documents that have been retrieved and stored within the main index by the retrieval agent will have their features extracted and compared to the profile of each individual user generated by the profile generator. This is performed on every item a user has shown interest in, and if any of the documents from the main index happen to fit the user's interests or needs, then they will be eventually suggested to the user the next time the user logs in (Figure 2). Each suggestion, if considered interesting, may be explicitly added to the personal database by the user, or deleted completely. The user might even prefer that he/she is notified, via email, that documents of interest have been located. The machine learning techniques employed to generate the user profile is also applied to extract features from documents. In this way the targetted documents reflect, and are consistent, with the specific user profile generated.

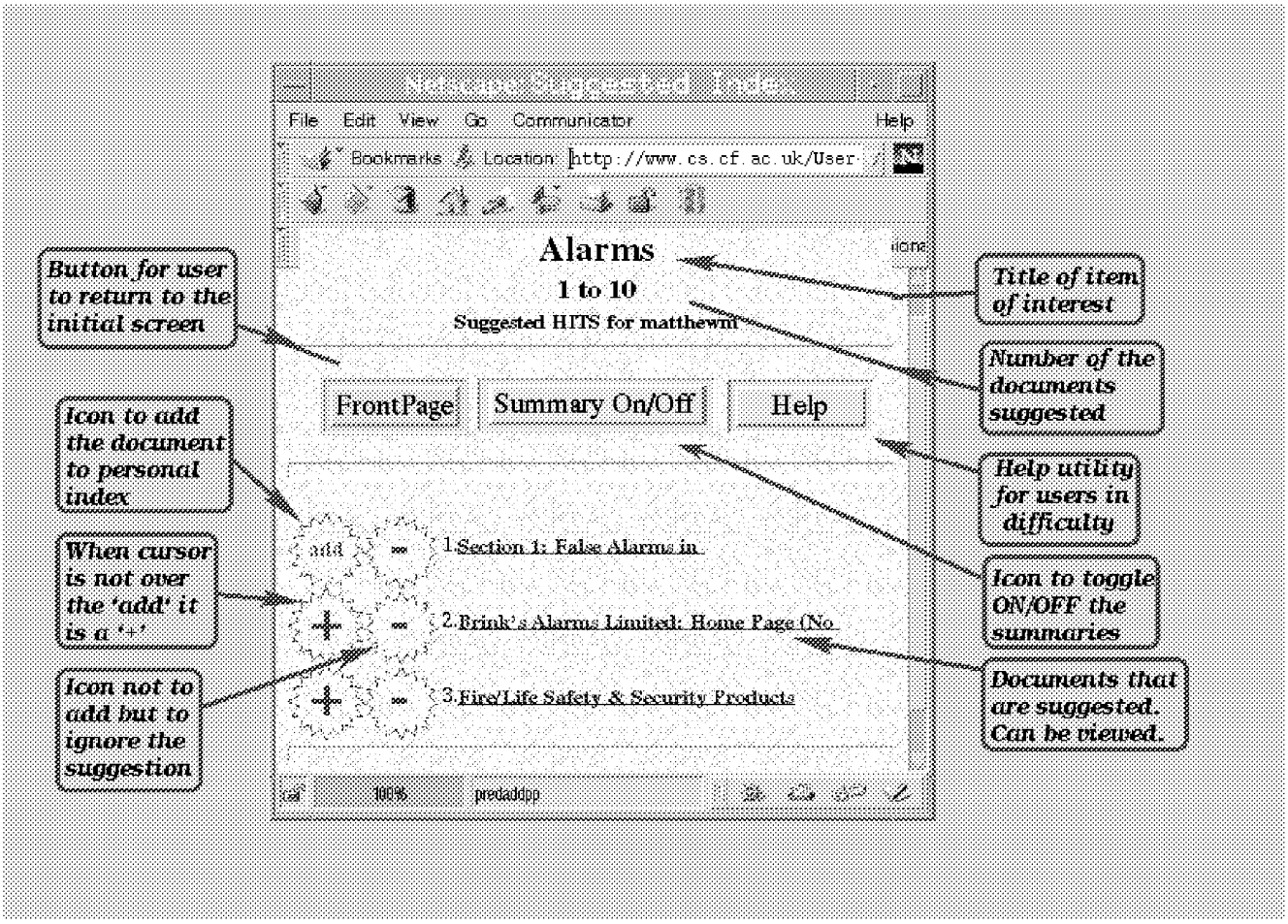


Figure 2: Suggested Documents

Evolvability

One final point to make is about the evolvability of PEA. The choice of external systems coupled with the appropriate wrapper within the retrieval agent, and the machine learning technique employed in the application layer can be selected from a list of systems and techniques incorporated in the system. PEA allows this list to be amended and hence other systems and techniques that might be developed in the future can be easily incorporated. The use of available systems means the system evolves as they evolve, relatively easily. The only amendments to PEA being if a new wrapper is required.

PEA GUI - An Example

PEA requires an administrator to manage the general needs and demands of a specific interest group of users. Search terms tailored to any type of interest group can be initialised by the administrator and furthermore users will be able to suggest any other terms to add to the main search list. Documents relevant to the specific area of interest are retrieved and stored by the underlying application within the main index, and when a user logs-in he/she is able to benefit from the systems' high recall fidelity. Having analysed the documents, individual users can bookmark and highlight specific items as interesting and appealing. These will be saved inside their personal database index. At this stage the underlying application plays another important role in attaining precise targeting of documents to individual users by generating a profile from the personal database index and predicting other documents from within the main index. Users can decide to add the suggested documents to their personal database index or remove them completely. As new and suggested documents are entered in the personal database index the user profile becomes more focused and finely tuned, as a result of which higher precision results will be achieved.

Related Work

Several research systems and commercial off-the-shelf agents have been developed which are similar to our work, but the closest systems are the so called metasearch engines. [Selberg and Etzioni, 1995] fed search terms to six major search engines and made use of the outcome within the MetaCrawler system. A number of front-end metasearch engines have also been introduced on the market, among them Surfbot, WebCompass, WebFerret, and WebSeeker. These all have very similar capabilities to the MetaCrawler plus additional features such as monitoring specific documents, verifying links and providing relevance ranking. All these related systems are only as reliable as the search engines that they depend upon. This means that their recall score might be very high because a search is done on many of the most popular search engines on the WWW with a single command, potentially retrieving all possible indexed documents. On the other hand, their low precision factor will require the users to check through the documents returned by the metasearchers to identify which ones are of interest. In our system, this task is performed by the profile/prediction components within the underlying application, which combines the optimization of both recall and precision.

Concluding Comments

In this position paper we have presented a system, PEA, that adds value to the information traditional search engines and other metasearch engines generate from the WWW. We argue that by reusing the information output from several retrieving/indexing systems we ensure a high recall score, while generating a specific user profile to predict and target other documents to specific users, we also ensure a high precision score. Users are able to select their own profile generator/prediction agent from a number of alternatives, reflecting different machine learning techniques employed. New techniques can be integrated into this evolvable system by the system administrator, who can also easily maintain the system's resources and update the search terms specific to a user group. In the future we will be investigating the integration of other machine learning techniques that have been developed and employed by other systems. This will help us to evaluate which technique is best suited to cater for the needs of different users. Evaluation of the recall/precision scores is also required to ensure that value is added to the normal services provided by the search engines and the meta-search engines. This will be done by analysing the feedback given from a group of users who are presently making use of the system and who will eventually assess the extent to which the information presented is of high recall/precision quality.

References

Armstrong et al., 1995
Armstrong, R., Freitag, D., Joachims, T., and Mitchell, T. (1995).
WebWatcher: A Leraning Apprentice for the World Wide Web.
AAAI Spring Symposeum on Information Gathering.

Balabanovic and Shoham, 1995
Balabanovic, M. and Shoham, Y. (1995).
Learning Information Retrieval Agents: Experiments with Automated Web Browsing.
AAAI Spring Symposeum on Information Gathering.

Berners-Lee et al., 1994
Berners-Lee, T., Caillian, R., Luotonen, A., Nielsen, H. F., and Secret, A. (1994).
The World-Wide Web.
Communications of the ACM, 37(8):76-82.

DeBra and Post, 1994
DeBra, P. M. E. and Post, R. D. J. (1994).
Searching for arbitrary information in the WWW: the fish-search for mosaic.
In *Proceedings of the 2nd. international world wide web conference*.

Edwards et al., 1995
Edwards, P., Bayer, D., Green, C. L., and Payne, T. R. (1995).
Experience with Learning Agents which Manage Internet-Based Information.
In *Proceedings of ML95 Workshop on Agents that Learn from Other Agents*.

Eichmann and Wu, 1996
Eichmann, D. and Wu, J. (1996).
Sulla - a user agent for the web.
In *5th. International World Wide Web Conference*.

Etzioni and Weld, 1994
Etzioni, O. and Weld, D. S. (1994).
A Softbot-Based Interface to the Internet.
Communications of the ACM, 37(7):72-79.

Green and Edwards, 1996
Green, C. L. and Edwards, P. (1996).
Using Machine Learning to enhance software tools for internet information management.
In Franz, A. and Kitamo, H., editors, *AAAI-96, Workshop on Internet-Based Information Systems*, pages 48-55. AAAI Press.

Holte and Drummond, 1994
Holte, R. and Drummond, C. (1994).
A Learning Apprentice for Browsing.
AAAI Spring Symposeum on Software Agents.

Joachims et al., 1997
Joachims, T., Mitchell, T., and Freitag, D. (1997).
WebWatcher: A Tour Guide for the World Wide Web.
IJCAI97.

Krishnamurthy and Tsangaris, 1996
Krishnamurthy, B. and Tsangaris, M. (1996).
Curry: A customizable url recommendation repository.
In *5th. International World Wide Web Conference*.

Lashkari, 1995
Lashkari, Y. (1995).
Feature Guided Automated Collaborative Filtering.
Master's thesis, MIT, department of Media Arts and Sciences.

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

Maes and Kozierok, 1993
Maes, P. and Kozierok, R. (1993).
Learning Interface Agents.
In *Proceedings of the 11th National Conference on Artificial Intelligence*, pages 450-465.

Mladenic, 1996
Mladenic, D. (1996).
Personal WebWatcher: Implementation and Design.
Technical Report IJS-DP-7472, J Stefan Institute, Llubljana, Slovenia.

Monge and Elkan, 1995
Monge, A. E. and Elkan, C. P. (1995).
Integrating external information sources to guide Worldwide Web Information Retrieval.
Technical Report CS96-474, University of California, San Diego.

Oostendorp et al., 1994
Oostendorp, K. A., Punch, W. F., and Wiggins, R. W. (1994).
A tool for individualizing the Web.
In *Proceedings of the 2nd. WWW conference '94: Mosaic and the Web*.

Payne and Edwards, 1997
Payne, T. R. and Edwards, P. (1997).
Interface Agents that Learn: An Investigation of Learning Issues in a Mail Agent Interface.
Applied Artificial Intelligence, 11(1):1-32.

Pazzani et al., 1996
Pazzani, M., Muramatsu, J., and Billsus, D. (1996).
Syskill and Webert: Identifying Interesting Web Sites.
AAAI Conference.

Perkowitz and Etzioni, 1995
Perkowitz, M. and Etzioni, O. (1995).
Category Translation: Learning to understand Information on the Internet.
In *International Joint Conference on Artificial Intelligence*.

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

Selberg and Etzioni, 1995
Selberg, E. and Etzioni, O. (1995).
Multi-service search and comparison using the meta-crawler.
The Web Revolution. Proceedings of the 4th. international world wide web conference.

Spetka, 1994
Spetka, S. (1994).
The TkWWW Robot: Beyond browsing.
In *Proceedings of the 2nd. WWW conference '94: Mosaic and the Web*.

Srinivasan et al., 1996
Srinivasan, P., Ruiz, M. E., and Lam, W. (1996).
An investigation of indexing on the www.
ASIS '96 Annual Meeting of American Society for Information Science., 33:79-83.

Footnotes

...Systems
Research funded by the Radiocommunications Agency, UK.

...Surfbot
http://www.surflogic.com/

...WebCompass
http://www.quarterdeck.com/qdeck/products/webcompass/

...WebFerret
http://www.webferret.com/netferret/webferret.htm

...WebSeeker
http://www.ffg.com/seeker/

M Montebello
3/5/1998

Exhibit 5

Collecting User Access Patterns for Building User Profiles and Collaborative Filtering

Ahmad M. Ahmad Wasfi
School of Computer Sciences
University of Science, Malaysia
Penang 11800, Malaysia
ahmadm@cs.usm.my

ABSTRACT
The paper proposes a new learning mechanism to extract user preferences transparently for a World Wide Web recommender system. The general idea is that we use the entropy of the page being accessed to determine its interestingness based on its occurrence probability following a sequence of pages accessed by the user. The probability distribution of the pages is obtained by collecting the access patterns of users navigating on the Web. A finite context-model is used to represent the usage information. Based on our proposed model, we have developed an autonomous agent, named ProfBuilder, that works as an online recommender system for a Web site. ProfBuilder uses the usage information as a base for content-based and collaborative filtering.

Keywords
Autonomous agent, Classical information theory, Finite context-model, Content-based filtering, Collaborative filtering.

INTRODUCTION
The World Wide Web hypertext system is a very large distributed digital information space. Some estimates suggested that the Web included about 150 million pages and this number doubled every four months [7]. As more information becomes available, it becomes increasingly difficult to search for information without specialized aides.

Agent-based recommenders have been proposed as a solution to this problem. During a browsing session, these computer systems work collaboratively with a user without the need of an explicit initiation. It has a static or dynamic collection of pages to be recommended. It assists the user by recommending pages that match his/her needs.

Since the system involves repeated interaction with the user and this may extend over a long period of time, the user's interests cannot be assumed to stay constant. The change in interests could be anything from a slight shift in relative priorities to completely transforming into other domains. Thus, the recommender system must be able to specialize to the current interests of the user and adapt as they change over time. A variety of learning mechanisms have been employed within recommender systems. They mostly revolve around the following three basic techniques [14].

- Direct learning technique:
  The user provides a set of keywords to describe his/her interests (e.g. SIFT [22]). These keywords may include objective terms (such as author name and date of publication) or content keywords to reflect the information manifested in the desired documents. One advantage of this technique is its predictability. The user can usually guess why such information has been delivered. The problem with this technique is that it requires a lot of effort on the part of the user. The user should continually update the keywords to reflect his/her new interests. The user may not also formulate effective keywords to describe his/her interests.
- Partially direct learning technique:
  The system learns user preferences by eliciting explicitly some kind of user feedback (e.g. SysKill & Webert [13], InfoFinder [10], Ringo [18], and Fab [3]). When the user moves from one page to another, the user rates how interesting the current page is. It needs less user effort than the first technique. However, the user still has a greater mental load compared when he/she browses normally.
- Indirect or transparent learning technique:
  The system learns user preferences transparently without any extra effort from the user. However, current methods are not adequately practical. For example, Letizia [11] infers user preferences from observing user-browsing behavior such as saving a reference to a page. This approach is intuitively reasonable to indicate an interest to the page. During a browsing session, however, a user may refer to many pages in his/her bookmark just for future examination and not because of their interestingness.

To address the problems of the basic techniques, some systems use a mixture of them to get a compromise model between user effort and predictability. For example, Anatagonomy [14] learns user preferences by joining both

IUI 99 Redondo Beach CA USA

the second and third techniques. SiteHelper [19] and WebWatcher [2] use the first technique to specify the initial user area of interests and use the second technique to get feedback from the user to refine their recommendations. However, the problems of the basic techniques are still inherent in the recommender systems.

With the drawbacks of the learning techniques in mind, we propose a new mechanism that learns user interests and adapts automatically to their changes without user intervention. It relies on using the probability distribution of the pages to be accessed and tools derived from classical information theory. In order to illustrate how the learning mechanism plays out in a real application, we present an agent-based recommender, called ProfBuilder, that inhabits a Web site and is assigned the goal of being online responsive to the information needs of the site's users.

The following section provides a detail of our new proposed learning mechanism. We then present an overview of the architecture of ProfBuilder. Finally we outline some related work, followed by some possible ideas to address the limitations of ProfBuilder.

THE LEARNING MECHANISM

Background

A profile is a description of user interests. To deliver information a user wants to see, we should search for pages that are similar to his/her profile. An appropriate representation for profiles and pages is based on vector-space representation, commonly used in information retrieval (IR) literature [15].

In the vector-space model, pages and queries (profiles) are both represented as vectors in some hyper-space. The model assumes that there is an available keyword set **k**, where each element k_i is a keyword. Both pages and profiles can then be represented as weight vectors of the form

$$D_i = < d_i >$$

and

$$Q_i = < q_i >$$

where d_i and q_i represent the weight of k_i in vector D_i or Q_i, respectively. d_i (or q_i) is set equal to 0 when k_i is absent from the vector.

For this representation, the method for profile reformulation in response to the changes of user's interest is based on vector adjustment. Since profiles and pages are both vectors, the profile should move closer to the vectors representing pages which are relevant and away from the vector representing pages which are non-relevant. The implicit assumption of this is that pages resemble each other are represented by reasonably similar vectors.

Consider that page s_i is the current page of user u_j. Let us assume that variable t_{ij}, which is a nonnegative number between zero and one, indicates the relevance or importance of page s_i to user u_j. A reformulation of vector Q_j representing the user profile is obtained by taking Q_j and adding the vector elements D_i representing page s_i after it is changed in proportion to t_{ij},

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

i.e. the weight of each word in D_i is modified proportional to t_{ij}. The resulting effect is that, for those words already present in the profile, the word-weights are modified in proportion to $t_{ij} * d_i$. Words which are not in the profile are added to it.

It remains to find an effective way for inferring t_{ij}.

The Method

Before introducing the algorithm, let us first get our bearings by considering a few examples. Suppose we have a hypertext collection of 1000 pages (e.g. a Web site). Users navigate in the collection by using any navigation technique such as selecting hypertext links, specifying page addresses, or selecting pages from interest lists. Suppose we have collected the access patterns of a large number of users. Consider that *A* and *B* are two pages in the collection. If the conditional probability of visiting *A* given *B* is very high (*i.e.* it is very probable that when a user visits *A*, the user will jump next to *B*), one would consider that there is a strong interdependency or relationship between *A* and *B*. One form of relationship is that there is a high resemblance between *A* and *B* in content. In effect this means that the user is still interested in the same domain of *A*. Hence visiting *B* does not give much new information about the user's interests as the content of *B* is redundant to *A*'s. To guarantee that the agent can adapt quickly to the changes of user interests, the importance of *B* should be low.

Another form of relationship is that *B* has a great informative value from *A*. This is because *B* may contain important hyper information (hyperlinks pointing to other pages), an important content, or may be both. In this case, *B* may not be considered to reflect the actual user's interests. The user may visit *B* just because of its importance at some time. As an example, consider a user reads the top-headline story "Iraq Standoff as Diplomatic Efforts Continue, Sabers Rattle" from the main page of CNN site. The user may read the story just because of its significance as it is the top-headline story and is not interested to get more information about Iraq. Since *B* gives a little information about user interests because a user will likely visit *B* regardless of his/her interests, the importance of *B* should be low.

If, on the other hand, the probability of visiting *B* following *A* is very low, one may consider that there is no relationship between *A* and *B* (such as in content). Thus, if a user chooses to jump from *A* to *B*, *B* gives much information about his/her interest. For example, suppose a user reads a non-headline story about space from the main page of CNN site. Since it is reasonable to assume that the probabilities of visiting non-headline stories are low in comparison to the headline ones, one would expect that the user is interested in space stories. In other cases, *B* may be

considered to be a threshold point of losing interest in some domains and gaining interest in others. Thus, the importance of B should be high.

If we consider the foregoing examples, we see that the importance of a page is inversely proportional to its probability following the page (or the sequence of pages) visited by the user.

The importance of classical information theory is embodied in the idea that the value of information content H of messages sent from message sources to message receivers can be measured quantitatively. Because of the redundancy that occurs due to the dependency between successive messages, the value of a message is assumed to be inversely proportional to the probability with which the message could have been predicted by the receiver before the message arrived [17].

H of a message is then defined as a decreasing function $H(p)$, also known as the entropy, of the probability p of that message. Because the information content of two messages should be an additive function of their individual content values, that is, $H(p_1p_2) = H(p_1) + H(p_2)$, and the value of a message received with the probability one should be zero, the information content of a message can be defined as (for formal prove see[1]),

$$H(p) = -\log(p)$$

If the browsing process is considered as a transference of information from a collection of pages such as a Web site (message source) to a user (message receiver), it is possible to use tools derived from information theory to quantify the information value of a page (message) as it relates to each user.

The importance or interestingness t_{ij} of page s_i to user u_j is then assumed to be its entropy $H(pr)$ based on its conditional probability pr of being accessed following the sequence of pages accessed by user u_j,

$$t_{ij} = H(pr) = -\log(pr)$$

Representation for Web Navigation

The probability distribution of the pages to be accessed is based on collecting the visiting patterns of many users. The usage history of a collection of pages, such as a Web site, was represented previously as a directed graph of pages [21] [9]. This means that the occurrence of a page depends only upon the previous page. A more general type representation is that the occurrence of a page may depend upon a context consisting of a finite number m of preceding pages. Such a representation is called finite-context modeling which is commonly used in statistical modeling of an information source (e.g. for text compression [8]). The advantage of using the context model is that it better reflects the actual distribution of the pages. The model is specified by giving the finite set **s** consisting of q pages (where each element s_i is a page) and the set of conditional probabilities:

$$pr(s_i \mid s_{j1}, s_{j2}, \ldots, s_{jm}) \text{ for } i = 1, 2, \ldots, q; jp = 1, 2, \ldots, q$$

where pr is the conditional probability of visiting page s_i, given the sequence of pages $s_i \mid s_{j1}, s_{j2}, \ldots, s_{jm}$.

A context model may be an *order-m fixed model* based on a fixed number m of previous pages in its probability determination or may be an *order-m mixed model* based on contexts of several lengths with a maximum-length of m. In an order-m fixed model, the probability (or, more accurately, frequency) of a given page is known only if we know the m preceding pages. For instance, if $m = 0$ then no context is used and the probability of the current page is the probability of its occurrence in the collection. If $m = 1$ then the previous page is used to determine the probability of the current page. If $m = 2$ then the previous two pages are used, and so on. At any one time, therefore, we shall call the m preceding pages the *state* of the order-m model at that time. Since there are q pages, an order-m model will have q^m possible states. A handy way to illustrate the behavior of a context model is through the use of a *state diagram*. In a state diagram we represent each of the q^m possible states of the model by a circle, and the possible transitions from state to state by arrows.

As an example, suppose we have a very simple Web site consisting of two pages A and B. That is, **s** = {A, B} and q = 2. When $m = 2$ then the probability distribution of page s_i (i = 1 or 2) is determined by the previous two pages s_j and s_k (j, k = 1 or 2). Let us assume that the conditional distribution of s_i given s_j and s_k is as follows:

$pr(s_i = A \mid s_j = A, s_k = A) = 1/4$

$pr(s_i = A \mid s_j = A, s_k = B) = 1/5$

$pr(s_i = A \mid s_j = B, s_k = A) = 1/7$

$pr(s_i = A \mid s_j = B, s_k = B) = 2/3$

$pr(s_i = B \mid s_j = \alpha, s_k = \beta) = 1 - pr(s_i = A \mid s_j = \alpha, s_k = \beta)$

where (α, β = A or B).

Because q is equal to 2 and we have assumed an order-*2* fixed context model, we have four states – *AA*, *AB*, *BA*, *and BB*. The state diagram of this model is shown in Figure 1.

The possible states of the site are indicated by the four circles. The possible state transitions are indicated by arrows from state to state, with the probability of a transition shown by a number associated with each arrow. For example, if we are in state *AA* we can go to either *AA* or *AB* but not to state *BA* or *BB*. The probability of remaining in state *AA* is shown as 1/4, and the probability of going to state *AB* is shown as 3/4.

On the other hand, in an order-m mixed model, if $m = 2$ then we use the previous two pages, one predecessor if two-pages context fails to determine the probability of a page, and the probability of the page occurrence if both two-pages and one-page contexts fail. A mixed model may be either fully or partially mixed. The model is fully mixed if it contains all the fixed sub-models whose orders equal

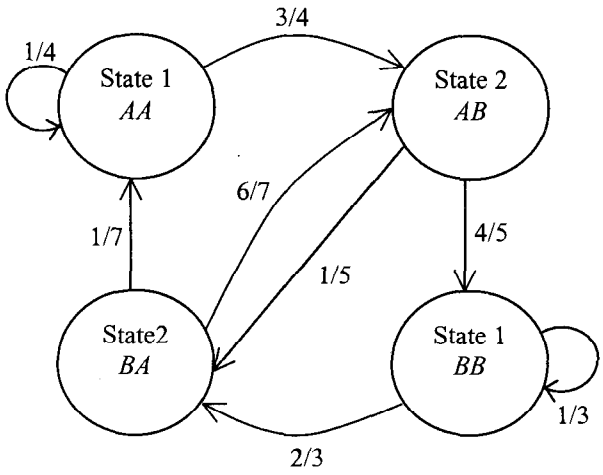


Figure 1. State diagram of an order-2 fixed model with $q = 2$.

or less than the order of the model. That is, a fully mixed model with $m = 3$ bases its determination on sub-models of orders 3, 2, 1, and 0. The model is partially mixed if it uses some, but not all, of the sub-models.

The Order of a Context Model

Determining the order m of a model is critical to reflect the actual probability distribution of a collection of pages. Let us consider the behavior of the model as m equals to zero and when m gets larger. When $m = 0$, the probability of a page is just the probability of its occurrence in the collection, which is obviously not enough to show the dependency among the pages. As m increases, it reflects better the actual dependency among the pages. But when m gets very large, the dependence of a page on the previous m pages becomes very weak. Thus, an order of few pages is very reasonable to determine the dependency among the pages.

Calculating the Probability

The general mechanism to calculate the probability pr of page s_i in a fully order-m mixed model is dictated by the m most recent pages of the user path, if page s_i has occurred in this particular path before by past users. In this case, only the order-m probability distribution is used. Otherwise, sub-models of lower orders are consulted. If the order-0 sub-model is consulted (*i.e.* the page s_i has never occurred in the context of any higher sub-model before), pr is assumed to be proportional to,

$$n_i/N$$

where n_i is the occurrence frequency of page s_i, and N is the total number of times of visiting all the pages. This will guarantee to supply a probability for any page in collection **s**.

ARCHITECTURE

This section discusses ProfBuilder (acronym for Profile Builder), a transparent, adaptive, autonomous agent which works as a recommender system. ProfBuilder's learning mechanism is implemented based on the algorithm described in the preceding section.

ProfBuilder inhabits a Web site and is assigned the goal of finding relevant local pages for the site's users. The advantage of this architecture is that ProfBuilder does not need to search the Web to collect the pages to be recommended, as they are the site's pages. Thus, ProfBuilder has the benefit of operating without using bandwidth from the Internet except a trivial amount when it delivers its recommendations to the users. Moreover, such a system is also transparent to the users requiring to external installation.

ProfBuilder is autonomous as it can take actions relating to page filtering on the user's behalf. It is adaptive as it learns the preferences of the user and adapts as they change over time. It is transparent as it extracts the preferences without user intervention.

ProfBuilder keeps track of each individual user and provides that person online assistance. The assistance includes two lists of recommendations based on two different filtering paradigms: content-based and collaborative. ProfBuilder updates the lists each time the user changes his/her current page. Content-based filtering is based on the correlation between the content of the pages and the user's preferences. The collaborative filtering is based on a comparison between the user's path of navigation and the access patterns of past users. Combining the two paradigms may eliminate the shortcomings in each approach. By making collaborative filtering, we can deal with any kind of content and explore new domains to find something interesting to the user. By making content-based filtering, we can deal with pages unseen by others [3].

To overcome the problem of stateless connection in HTTP, ProfBuilder follows users through tracking their IP addresses. To track user presence, a timeout mechanism is used to delete user's session information after a predetermined amount of idle time. So that, a connection after the specified period having the same IP is identified as a new user. This method is fairly easy to implement. The problem with this way is that many users connect to the Internet though proxy servers. Consequently, the IP of a proxy server may represent two or more people who are accessing the same Web site simultaneously in their browsing sessions, causing an obvious conflict. However, the reality is that many large sites use this method and have not had any clashes [20].

Fast performance is a key requirement as ProfBuilder is in the middle of every Web transaction. ProfBuilder was built in a highly multi-threaded fashion using Java language so that no information that ought to be delivered to the browser gets stuck somewhere in the agent system.

The architecture of ProfBuilder can be broken down into three modules.

- The graphical user interface module is responsible for displaying ProfBuilder's interface.
- The learning module is responsible for maintaining the mapping between the actual interests of the user and the user profiles.
- The filtering module is responsible for the content-based and collaborative filtering.

The Graphical User Interface

The graphical user interface of ProfBuilder is a separate resizable HTML frame at the top of the current page. Figure 2 illustrates ProfBuilder's interface[1]. The result frame displays a list of the recommended pages represented by their respective titles. The title is also associated with the page size which may be useful to distinguish among pages having the same title, as well as allows users to estimate the time and space it will take to retrieve the page.

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 balls shows levels of relevance: one-ball pages are poor, two-ball pages are neutral, three-ball pages are good, and so on. 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.

To read the content 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.

In addition to the list, the frame also shows two buttons for choosing the type of filtering. In the content-based filtering, the selected pages are sorted in decreasing order of their interestingness, while in the collaborative filtering are sorted in decreasing order of their access frequencies.

The Learning Module

The learning module handles the task of mapping user interests to the profile and maintaining the correlation between the two. It is implemented on the basis of our proposed learning mechanism.

The Web site usage is represented as a full order-1 mixed model. A mixed model is desirable and essentially unavoidable to show the actual probability distribution of the pages. The order has been chosen because of its space and time effectiveness as well as to the reasons mentioned in the previous section.

The frequency of occurrence of each page in the order-0 sub-model is initialized based on its visibility (roughly it is the number of pages pointing to the page [4]). The visibility is a sign of popularity [12] and frequency of access. For example, we expect that a page with a visibility of 10 will be accessed more than a page with a visibility of 1. The visibility data is obtained by querying Infoseek search engine.

The context model is built progressively as users jump from one page to another using any navigation technique. The general mechanism of building is to update the frequency of the occurrence of the current page in order-0 sub-model, and update its frequency of occurrence in order-1 sub-models based on the user's previous page.

The Information Filtering Module

Content-based Filtering

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.

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 hypertexts. This narrow analysis leads to retrieval of fewer pages, but most of the retrieved materials are likely to be helpful to the user; as it is reasonable to assume that the author of a Web page used these words to give the main aspect of the page. Stop words [5] are filtered out and word stemming [6] is then performed to improve IR performance. The keywords are weighted based on the well-test algorithm TDIDF [16]. The weight of a keyword in one page is the product of its keyword frequency and the inverse of its document frequency. The weight of the keyword k_j is given by,

$$w_{ij} = tf_{ij} * idf_j$$

where tf_{ij} is the number of occurrences of k_j in page s_i, and idf_j is the inverse document frequency of k_j in the Web site.

The similarity metric between the vector D_i representing page s_i and the vector Q_j representing the interests of user u_j is calculated by taking a scalar product of the two vector,

$$Similarity(D_i, Q_j) = \sum_k w_{ik} * w_{jk}$$

Collaborative Filtering

The filtering process is based on the following hypothesis: making available the work of a large number of past users can be useful to find out relationships or interdependency between pages. Thus, it is reasonable to advise one user of what was done by others (the previous section discussed relationship types). For instance, when there is a high probability to visit page B given page A, it may indicate that B has an important content. This reduces the chances of missing something particularly significant.

The module finds pages from the user's current path which is in this application only his/her current page, and selecting the top-frequency pages for presentation to the user.

[1] For evaluation purposes; we have modified ProfBuilder to a proxy-based architecture, as there was no practical site physically available.

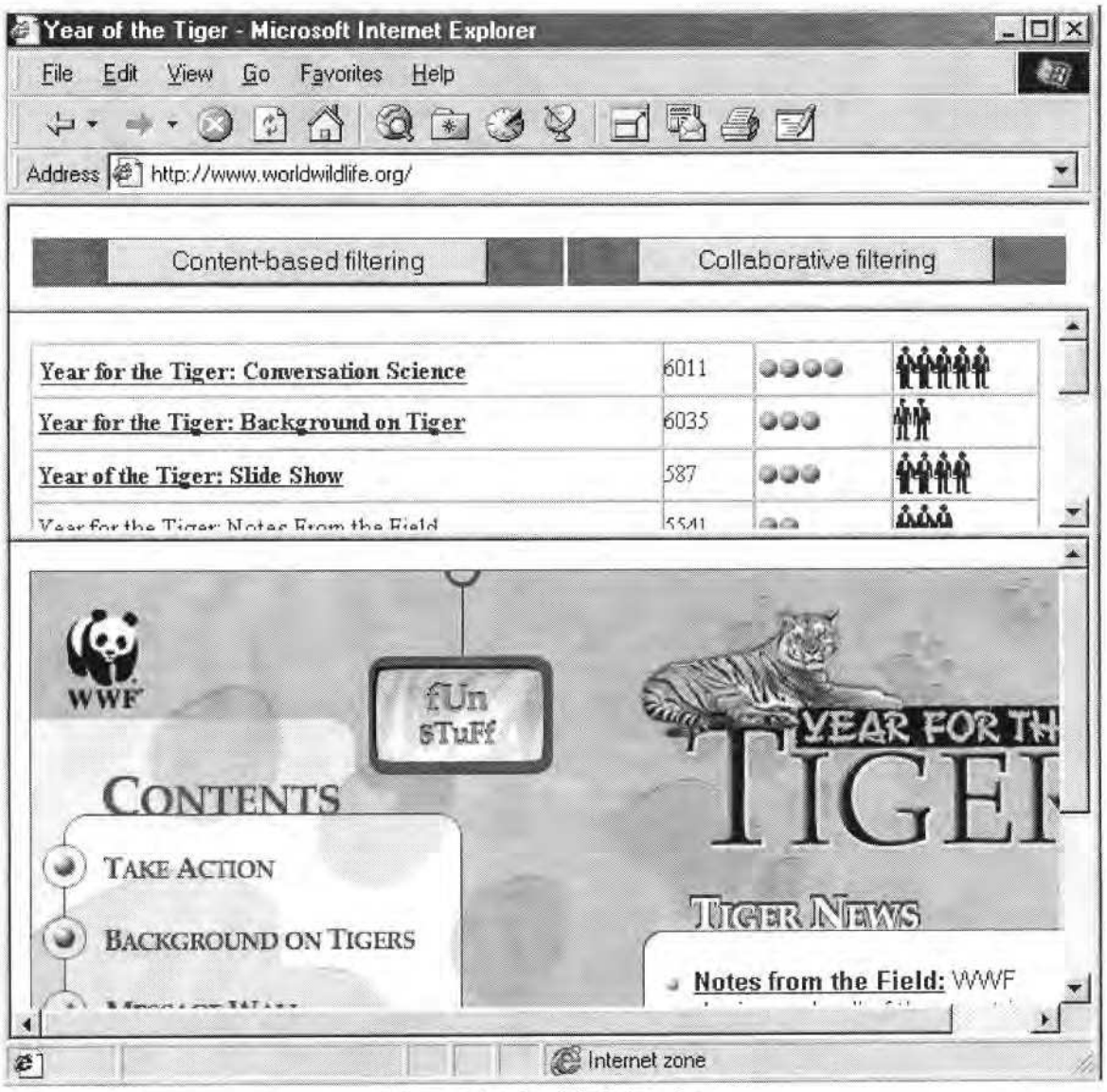


Figure 2 ProfBuilder's interface.

RELATED RESEARCH

This section compares ProfBuilder with other recommender systems. Among these systems, four are selected for comparison: Letizia [11], SiteHelper [19], Ringo [18], and Fab [3]. The systems were selected because they cover the best features of the other recommender systems.

Letizia

Letizia [11] is an agent that assists a user browsing the Web. Letizia uses the idle time spent by the user reading the current document to explore the neighborhood looking for pages that are related to the user's interest. Similar to ProfBuilder, the goal of Letizia agent is to autonomously provide assistance to the user without his/her intervention while the user is browsing. However, Letizia learns the areas that are of interest to a user by recording the user's browsing behavior. Letizia uses only content-based filtering, while ProfBuilder uses both content-based and collaborative filtering. Furthermore, Letizia requires considerable bandwidth to operate resulting in network overload and bandwidth shortages. On the other side, ProfBuilder inhabits a Web site and operate locally in assisting external users. Finally, Letizia requires a Macintosh and Netscape browser to operate, which severely limits its extent. In contrast, ProfBuilder is a platform independent.

SiteHelper

SiteHelper [19] is an agent that acts as a housekeeper for a Web site. It helps a user to find relevant information at the site. This is similar to ProfBuilder design. However, the learning mechanism is not transparent. SiteHelper prompts the user for a set of keywords and asks for an explicit feedback of rating the keywords. In contrast, the learning mechanism used by ProfBuilder works transparently without user intervention. In addition, SiteHelper's

recommendations are based only on content-based filtering.

Ringo

Ringo [18] is a system, which creates personalized recommendations for music albums and artists. It can be accessed through electronic mail or the Web. Users describe their interests to the system by explicitly rating some music. A user profile is a record of the user's interests (positive as well as negative) in specific items. Ringo makes recommendations based on comparisons among user profiles (collaborative filtering). Its processing time takes an hour, while ProfBuilder operates concurrently with the user in his/her browsing session.

Fab

Fab [3] is a system that helps Web users to discover new and interesting sites. Fab combines both content-based and collaborative filtering systems. The system maintains user profiles based on content analysis, and directly compares these profiles to determine similar users for collaborative recommendation. The system delivers a number of pages that it thinks would interest the users. Users evaluate the pages and provide explicit feedback to the system. Fab uses the Web for collecting pages for recommendation. This is an advantage over ProfBuilder's approach, since it is restricted to the site's pages. However, Fab uses considerable bandwidth for collecting these pages. Moreover, the system does not operate concurrently with users during their browsing session. Users get the recommendations by explicitly accessing their account in Fab's database through the Web.

FUTURE WORK

The limitations of ProfBuilder and possible solutions are as follows:

First, the main limitation of our system is that there is no real assessment of ProfBuilder. We plan to do serious evaluation studies in the near future. For example, we plan to examine different combinations of sub-models with various orders.

Second, one problem of its learning mechanism is that it needs a large number of users, in order to have enough data for reflecting the interdependency between pages. We intend to create simulated or virtual users who represent a particular taste to show the content relationships among the pages. In the CNN site for example, we can create a virtual 'Iraq' user, who visits only pages containing the word 'Iraq'. In the case of an order-1 fixed model, for instance, all the pages will be connected to form a complete graph. The frequency of a transition between two pages is assigned proportional to the product of their weights of the word 'Iraq'.

Third, since the number of states of an order-m fixed context model increases exponentially with m, it is even difficult to have a model with $m = 2$ as the space required to store all the context information is prohibitive. We intend to solve this problem as others have done (e.g. [8]) by using self-organizing lists and hashing techniques to provide a means of representing approximated context models of any order in a reasonable amount of memory.

Finally, ProfBuilder assists a user by finding relevant information on only one Web site. We intent to solve the problem by maintaining user profiles across different Web sites that use ProfBuilder. So that, when a user jumps to another site, the user's profile will also be transferred to the new site whose ProfBuilder will search for pages similar to the profile. Thus, the user can find relevant recommendations in the first page accessed in the new site.

CONCLUSION

In this paper, we have proposed a new learning mechanism to learn user preferences from the retrieved pages. It is based on their probabilities, which are obtained from collecting the visiting patterns of past users. We assumed that the importance of a page is its entropy based on its probability of visiting following a sequence of pages visited by the user.

We have also introduced ProfBuilder, an agent-based recommender for a Web site. ProfBuilder uses the site usage information to learn user interests and as a base for collaborative filtering. ProfBuilder helps the user to find relevant pages on the site by providing both content-based and collaborative filtering. ProfBuilder is independent of the browser as it assists the users by inserting its recommendations in the requested pages.

The efficiency of the learning mechanism and the usefulness of recommendations given by ProfBuilder have not been formally evaluated so far. We believe that ProfBuilder performs well for sites that consist of 1000 HTML URLs and above as it is infeasible to read the entire content of the sites.

REFERENCES

1. Aczel, J., and Daroczy, Z. 1975. *On Measures of Information and their Characterizations*. Academic Press.
2. Armstrong, R.; Freitag, D.; Joachims, T.; and Mitchell, T. 1995. WebWatcher: A Learning Apprentice for the World Wide Web. *In Proceedings of the AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments, Stanford, CA.*
3. Balabanovic, M., and Shoham, Y. 1997. Fab: Content-Based, Collaborative Recommendation. *Communications of the ACM* 40(3):66-72.
4. Bray, T. 1996. Measuring the Web. *In Proceedings of the Fifth International World Wide Web Conference, Paris.*
5. Fox, C. 1990. A Stop List for General Text. *SIGIR Forum* 24(1-2): 19-35.
6. Frakes, W. B. 1984. *Term Conflation for Information Retrieval.* New York: Cambridge university Press.

7. Gudivada, V. N.; Raghavan, V. V.; Grosky, W. I.; and Kasanagottu, R., 1997. Information Retrieval on the World Wide Web. *IEEE Internet Computing* 1(5):58-68.

8. Hirschberg, D. S., and Lelewer, D. A. 1992. *Context Modeling for Text Compression, Image and Text compression.* Kluwer academic Press, Norwell, MA.

9. Jaczynski M., Trousse, B. 1997. Broadway: A World Wide Web Browsing Advisor Reusing Past Navigations from a Group of Users.

   http://www.inria.fr/aid/papers/97/ukcbr/html/ukcbr.html

10. Krulwich, B., and Burkey, C. 1997. The InfoFinder Agent: Learning User Interests through Heuristic Phrase Extraction. *IEEE Intelligent System* 12(5):22-27.

11. Lieberman, H. 1995. Letizia: An Agent that Assists Web Browsing. *In Proceedings of International Joint Conference on Artificial Intelligence, Montreal.*

12. Marchiori, M. 1997. The Quest for Correct Information on the Web: Hyper Search Engines. *In Proceedings of the Sixth International World Wide Web Conference, Santa Clara.*

13. Pazzani, M.; Muramatsu, J.; and Billsus, D. 1996. Syskill & Webert: Identifying Interesting Web Sites. *Workshop on Machine Learning in Information Access, AAAI Spring Symposium Series, Stanford, CA.*

14. Sakagami H., and Kamba, T. 1997. Learning Personal on Online Newspaper Articles from User Behaviors. *In Proceedings of the sixth International World Wide Web Conference, Santa Clara.*

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

16. Salton, G., and Yang, C. S. 1973. On the Specification of Term Values in Automatic Indexing. *Journal of Documentation* 29(4):351-372.

17. Shannon, C. E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27(3):623-656.

18. Shardanand, U., and Maes, P. 1995. Social Information Filtering: Algorithms for automating "Word of Mouth". *In Proceedings of the Human Factors in Computing Systems Conference -- CHI'95, Denver.*

19. Siaw, D., and Ngu, W. 1997. SiteHelper: A Localized Agent that Helps Incremental Exploration of the World Wide Web. *In Proceedings of the Sixth International World Wide Web Conference, Santa Clara.*

20. Thomas, B. 1997. Recipe for E-commerce. *IEEE Internet Computing* 1(6):72-74.

21. Wexelblat, A., and Maes, P. 1997. Footprints: Visualizing Histories for Web Browsing. *In Proceedings of RIAO'97: Computer-Assisted Information Retrieval on the Internet, Montreal.*

22. Yan, T. W., and Garcia-Molina, H. 1995. SIFT – A Tool for Wide Area Information Dissemination. *USENIX Technical Conference.*

Exhibit 8

US007631032B1

(12) **United States Patent**
Refuah et al.

(10) **Patent No.: US 7,631,032 B1**
(45) **Date of Patent: Dec. 8, 2009**

(54) **PERSONALIZED INTERNET INTERACTION BY ADAPTING A PAGE FORMAT TO A USER RECORD**

(75) Inventors: **Aviv Refuah**, Tel-Aviv (IL); **Zeev Refuah**, Tel-Aviv (IL); **Eli Nhaissi**, Old Westbury, NY (US); **Maier Fenster**, Petach Tikva (IL)

(73) Assignee: **Net-Express, Ltd.**, Petach Tikva (IL)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/601,385**

(22) PCT Filed: **Jan. 28, 1999**

(86) PCT No.: **PCT/IL99/00056**

§ 371 (c)(1),
(2), (4) Date: **Jul. 28, 2000**

(87) PCT Pub. No.: **WO99/39281**

PCT Pub. Date: **Aug. 5, 1999**

(30) **Foreign Application Priority Data**

Jan. 30, 1998 (IL) 123129
Jul. 20, 1998 (IL) 125432

(51) **Int. Cl.**
G06F 15/16 (2006.01)

(52) **U.S. Cl.** **709/201**; 709/218; 709/228; 380/23

(58) **Field of Classification Search** 709/206, 709/217, 218, 220, 224, 227, 228, 203, 229, 709/232, 201, 223; 705/116, 34, 1, 10, 14; 370/251; 713/201, 202; 707/10
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

2,769,024 A 10/1956 Del Riccio et al.
4,870,579 A 9/1989 Hey
4,996,642 A 2/1991 Hey

(Continued)

FOREIGN PATENT DOCUMENTS

EP 0 387 226 9/1990

(Continued)

OTHER PUBLICATIONS

Falk, A. et al.; "PAWS: An Agent for WWW-Retrieval and Filtering"; Proceedings of the International Conference on The Practical Application of Intelligent Agents and Multi-Agent Technology; Apr. 22, 1996; pp. 169-179; XP002037595.

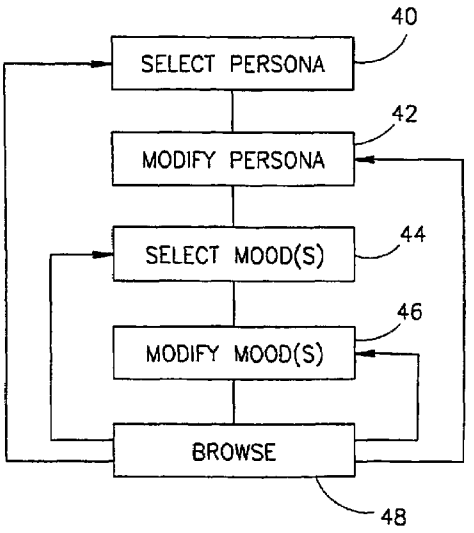
(Continued)

Primary Examiner—Khanh Q Dinh
(74) *Attorney, Agent, or Firm*—Ostrager Chong Flaherty & Broitman PC

(57) **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, responsive to said determined aspect, wherein said modified interactions comprise site-content related interactions with a plurality of unrelated sites. Preferably, the aspect is adapted in real-time to reflect changes in the tracked interactions.

34 Claims, 2 Drawing Sheets



U.S. PATENT DOCUMENTS

5,105,268	A		4/1992	Yamanouchi et al.
5,446,891	A		8/1995	Kaplan et al. 709/216
5,577,109	A		11/1996	Stimson et al.
5,644,724	A		7/1997	Cretzler
5,659,732	A		8/1997	Kirsch
5,671,342	A		9/1997	Millier et al.
5,721,827	A		2/1998	Logan et al.
5,721,902	A		2/1998	Schultz
5,722,067	A		2/1998	Fougnies et al.
5,727,950	A		3/1998	Cook et al.
5,732,219	A		3/1998	Blumer et al.
5,737,734	A		4/1998	Schultz
5,742,768	A		4/1998	Gennaro et al.
5,745,556	A		4/1998	Ronen
5,749,075	A		5/1998	Toader et al.
5,761,280	A		6/1998	Noonen et al.
5,761,499	A		6/1998	Sonderegger
5,764,906	A		6/1998	Edelstein et al. 705/32
5,768,521	A		6/1998	Dedrick
5,768,578	A		6/1998	Kirk et al.
5,769,906	A		6/1998	Kremer et al.
5,774,664	A		6/1998	Hidary et al.
5,794,210	A		8/1998	Goldhaber et al.
5,796,832	A		8/1998	Kawan
5,799,063	A		8/1998	Krane
5,806,043	A		9/1998	Toader
5,812,776	A		9/1998	Gifford
5,815,665	A		9/1998	Teper et al.
5,819,092	A		10/1998	Ferguson et al.
5,848,396	A	*	12/1998	Gerace 705/10
5,852,812	A		12/1998	Reeder
5,862,325	A		1/1999	Reed et al.
5,878,219	A		3/1999	Vance, Jr. et al.
5,884,262	A		3/1999	Wise et al.
5,892,919	A		4/1999	Nielsen
5,898,830	A		4/1999	Wesinger, Jr. et al.
5,898,835	A		4/1999	Truong
5,902,252	A		5/1999	Hohlfeld et al.
5,902,353	A		5/1999	Reber et al.
5,903,729	A		5/1999	Reber et al.
5,907,680	A		5/1999	Nielsen
5,911,145	A	*	6/1999	Arora et al. 715/514
5,915,093	A		6/1999	Berlin et al.
5,920,848	A		7/1999	Schutzer et al.
5,958,008	A	*	9/1999	Pogrebisky et al. 709/223
5,968,125	A		10/1999	Garrick et al.
5,970,477	A		10/1999	Roden
5,974,453	A		10/1999	Andersen et al.
5,987,440	A	*	11/1999	O'Neil et al. 705/44
5,987,506	A		11/1999	Carter et al.
5,991,735	A	*	11/1999	Gerace 705/10
6,009,459	A		12/1999	Belfiore et al.
6,021,496	A	*	2/2000	Dutcher et al. 713/202
6,023,698	A		2/2000	Lavey, Jr. et al.
6,029,195	A	*	2/2000	Herz 725/116
6,035,334	A	*	3/2000	Martin et al. 709/228
6,061,738	A		5/2000	Osaku et al.
6,070,157	A		5/2000	Jacobson et al.
6,097,441	A	*	8/2000	Allport 348/552
6,098,065	A		8/2000	Skillen et al.
6,101,482	A	*	8/2000	DiAngelo et al. 705/26
6,151,624	A		11/2000	Teare et al.
6,199,054	B1		3/2001	Khan et al.
6,243,816	B1	*	6/2001	Fang et al. 726/5
6,275,824	B1	*	8/2001	O'Flaherty et al. 707/9
6,292,551	B1		9/2001	Entman et al.
6,310,873	B1		10/2001	Rainis et al.
6,332,158	B1		12/2001	Risley et al.
6,356,936	B1	*	3/2002	Donoho et al. 709/206
6,397,219	B2		5/2002	Mills
6,446,260	B1		9/2002	Wilde et al.
6,446,261	B1	*	9/2002	Rosser 725/34
6,453,335	B1	*	9/2002	Kaufmann 709/203
6,480,850	B1	*	11/2002	Veldhuisen 707/9
6,507,872	B1		1/2003	Geshwind
6,546,002	B1	*	4/2003	Kim 370/351
6,546,399	B1		4/2003	Reed et al.
6,556,217	B1	*	4/2003	Makipaa et al. 345/667
6,571,279	B1	*	5/2003	Herz et al. 709/217
6,574,629	B1	*	6/2003	Cooke et al. 707/10
6,640,284	B1	*	10/2003	Shaw et al. 711/129
6,668,177	B2	*	12/2003	Salmimaa et al. 455/566
6,671,818	B1	*	12/2003	Mikurak 714/4
6,735,701	B1		5/2004	Jacobson
6,799,063	B2		9/2004	Carson
6,853,993	B2		2/2005	Ortega et al.
7,240,022	B1	*	7/2007	Bistriceanu et al. 705/14
7,289,971	B1	*	10/2007	O'Neil et al. 705/44
7,334,013	B1	*	2/2008	Calinov et al. 709/201
2001/0011222	A1	*	8/2001	Mclauchlin et al. 705/1
2002/0184534	A1		12/2002	Rangan et al.
2005/0203835	A1		9/2005	Nhaissi et al.

FOREIGN PATENT DOCUMENTS

EP	0 643 541	3/1995
EP	0 749 081	12/1996
EP	1051683	11/2000
GB	2312975	11/1997
JP	HEI09-305518	11/1997
JP	HEI09-319767	12/1997
JP	HEI10-021259	1/1998
WO	WO 96/23265	8/1996
WO	9642041	12/1996
WO	WO 97/07656	3/1997
WO	9719564	5/1997
WO	WO 97/26612	7/1997
WO	WO 97/31490	8/1997
WO	WO 97/41673	11/1997
WO	WO 98/04088	1/1998
WO	9826381	6/1998
WO	WO 98/53581	11/1998
WO	WO 99/39275	8/1999
WO	WO 99/39280	8/1999
WO	WO 99/39281	8/1999
WO	WO 00/05684	2/2000
WO	WO 01/06393	1/2001

OTHER PUBLICATIONS

Kamba, T. et al.; "Anatagonomy: A Personalized Newspaper on The World Wide Web"; International Journal of Human-Computer Studies; vol. 46; Jan. 1, 1997, pp. 789-803; XP 002086827.

Davison, A.; "A Graphical Internet Chat"; Web Techniques; Jan. 1998; Miller Freeman; USA; vol. 3; No. 1; pp. 54-57; XP002117568.

Duerst, M.; "Internationalization of Domain Names;" Jul. 1998.

Hahn, M.; "Uniform Resource Locators;" Dec. 1, 1995; EDPACS; vol. 3, No. 6; pp. 8-13, XP000566203.

Takada, T.; "Multilingual Information Exchange Through the World-Wide Web;" Nov. 1, 1994; Computer Networks and ISDN Systems; vol. 27, No. 2; pp. 235-241; XP004037994.

Cummings, R, "A URL Alternative",Sep. 12, 1997, keyword.com press release.

Auric Web Systems "The Intercard System. Prepaid Internet Access Cards", Auric Web Systems, 17 p., 1998.

Davies et al. "Networked Information Management", BT Technology Journal, 15(2): 194-208, 1998.

Estabrook "Sams' Teach Yourself Microsoft Internet Explorer 4 in 24 Hours", Sams.Net Publishing, Hour 6 & 9: 83-85, 131-132, 1997.

Lee "Microsoft Internet Explorer PowerToys", The Naked PC Newsletter (TNPC), 3 p., 1998.

Wong "Web Client Programming With Perl", O'Reilly Online Catalog, 5 p., 1997. Retrieved From the Internet: URL: http://www.orielly.com/openbook/webclient/ch01.html.

Moran, Joseph, "Using Windows 98's Address Bar to its Potential".

Soria et al. "Cytosolic Calcium Oscillations and Insulin Release in Pancreatic Islets of Langerhans", Diabetes & Metabolism, 24: 37-40, 1998.
W3C, Line Mode Browser Commands, Dec. 9, 1996.
Schloss, R. J., "Novel Business Uses of Independently Created Hyperlinks in the World Wide Web: Basic Mechanism and Examples," Proceedings of the Annual Hawaii International Conference on System Sciences, 1996, pp. 137-146.
Canadian Patent Office Action dated Sep. 10, 2008, in Canadian patent application No. 2,319,750 to Net-Express Ltd., entitled "WWW Addressing."
European Patent Office Examination Report Dated Oct. 9, 2007, in European Patent Application No. 99 901 873.2-2221 to Net-Express Ltd., entitled "WWW Addressing".
European Patent Office Examination Report Dated Mar. 12, 2004, in European Patent Application No. 99 901 873.2-2221 to Net-Express Ltd., entitled "WWW Addressing".
International Search Report dated Nov. 18, 1999, in International Patent Application No. PCT/IL99/00055 entitled "WWW Addressing".
International Preliminary Examination Report dated May 11, 2000, in International Patent Application No. PCT/IL99/00055 entitled "WWW Addressing".
Written Opinion dated Jan. 5, 2000, in International Patent Application No. PCT/IL99/00055 entitled "WWW Addressing".
Invitation to Pay Additional Fees dated Jul. 2, 1999, in International Patent Application No. PCT/IL99/00055 entitled "WWW Addressing".
International Preliminary Examination Report dated May 22, 2000, in International Patent Application No. PCT/IL99/00056, entitled "Personalized Internet Interaction".
Written Opinion dated Dec. 2, 1999, in International Patent Application No. PCT/IL99/00056, entitled "Personalized Internet Interaction".
Invitation to Pay Additional Fees dated Jul. 2, 1999 in International Patent Application No. PCT/IL99/00056, entitled "Personalized Internet Interaction".
Examination Report dated Nov. 28, 2007, in European Patent Application No. 99 933 101.0-1225, entitled "Internet Billing".
Supplementary European Search Report dated Jan. 23, 2007, in European Patent Application No. 99 933 101.0-1225, entitled "Internet Billing".
Corrected International Search Report dated Apr. 26, 2000, in International Patent Application No. PCT/IL99/00399, entitled "Internet Billing".
International Search Report dated Feb. 8, 2000, in International Patent Application No. PCT/IL99/00399, entitled "Internet Billing".
International Preliminary Examination Report dated Jun. 26, 2001, in International Patent Application No. PCT/IL99/00399, entitled "Internet Billing".
Written Opinion dated Jul. 27, 2000, in International Patent Application No. PCT/IL99/00399, entitled "Internet Billing".
International Search Report dated Dec. 22, 1999, in International Patent Application No. PCT/IL99/00433, entitled "User Interface Method".
"What Are Networds", Internet Archive, Oct. 15, 1997, http://www.web.archive.org/web/19971015221524/www.netword.com/n/help/what.html.
"Google Beta", Dec. 2, 1998, http://web.archive.org/web/19981202230410/http://www.google.com/.
Japanese Foreign Patent Office Action for Japanese Patent Application No. 2000-529668 dated Jul. 7, 2008.
Kikui et al, "Multiple Language Function in Navigation of Internet Information", Symp Application of Natural Lang Process, Information Processing Soc, Oct. 19, 1995, vol. 95, No. 6, pp. 97-105.
Sakamoto et al, "URL Number (Hot Code) Converting Server on Internet", NTT Technical J, Oct. 1, 1997, vol. 9, No. 10, pp. 92-94.
Yergeau, P. et al.; "Internationalization of URLs;" Sep. 9, 1996; pp. 1-3; Retrieved from Internet: <http://www.alis-com:8085/~yergeau/url-00.html>.
Lee "Microsoft Internet Explorer PowerToys", The Naked PC Newsletter (TNPC), 3 P., 1998.
Brin et al., "The Anatomy of a Large-Scale Hypertextual Web Search Engine," Computer Science Department, Stanford University, 1998, 20 pages.
Armstrong et al "WebWatcher: A Learning Apprentice for the World Wide Web," School of Computer Science, Carnegie Mellon University, 1995, 7 pages.
"Surfin' the World Wide Web with Japanese," by K. Kondo and C. Hemphill, Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2.
J. Kornblum: "N.Y. Times tries pop-up ads" CNET News. COM,—Jan. 16, 1998 (Jan. 16, 1998) XP002297192, Retrieved from the Internet: URL:http://news.cnet.com/2100-1023-207188.html, retrieved on Oct. 6, 2009.
W3C: "HTML 3.2 Reference Specification"—1997 XP002968017, retrieved from the Internet:URL:http://www.w3.org/TR/REC-html32, retrieved on Oct. 6, 2009.
"Changing url of a web site—meta refresh" Apr. 2, 1997 (Apr. 2, 1997), XP007909426, retrieved from the Internet:URL: http://lists.webjunction.org/wjlists/web4lib/1997-April/017651.html, retrieved on Oct. 6, 2009.
Inoue, Still not too late for Internet [8] Using Internet Explorer 3.0 (Part 2) (Inoue, Imakara-demo osokunai Internet [dai 8 kai] Internet Explorer 3. 0 wo tukau (2)), pp. 292-295, Nikkei Personal Computer, No. 281, Jan. 27, 1997, Nikkei BP (Dispatch No. 601380).

* cited by examiner

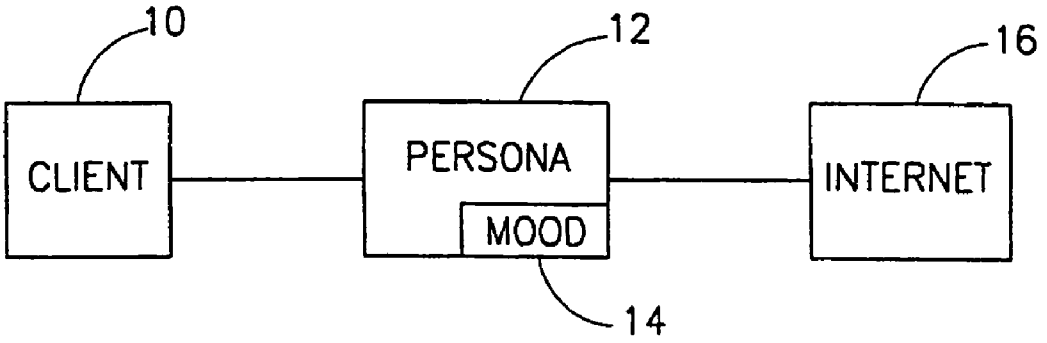

10
12
16
CLIENT
PERSONA
MOOD
INTERNET
14

FIG.1

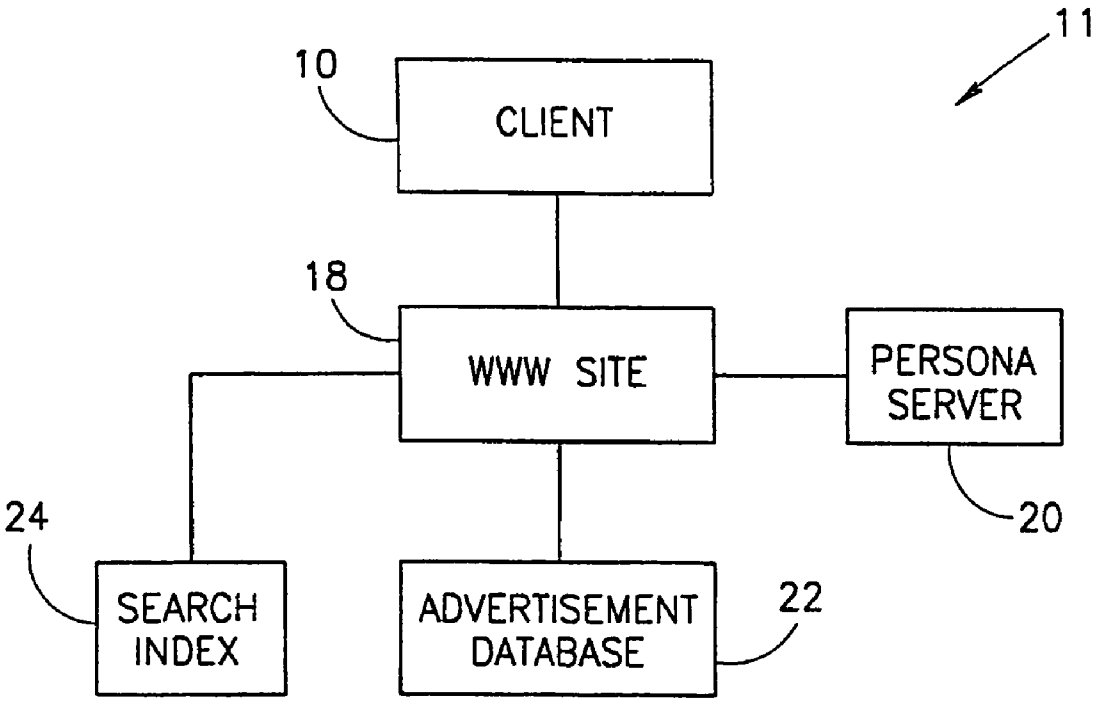

11
10
CLIENT
18
WWW SITE
PERSONA SERVER
20
24
SEARCH INDEX
ADVERTISEMENT DATABASE
22

FIG.2

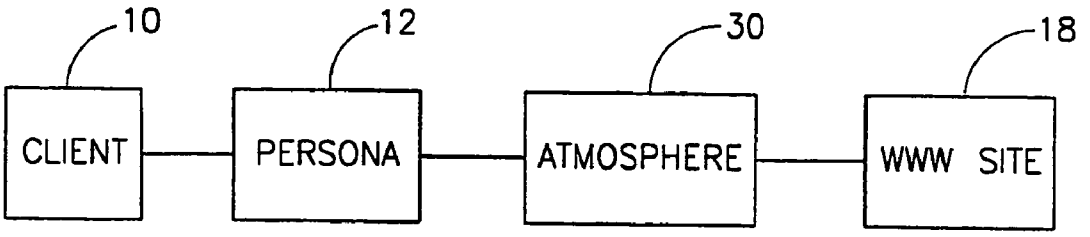

10
12
30
18
CLIENT
PERSONA
ATMOSPHERE
WWW SITE

FIG.3

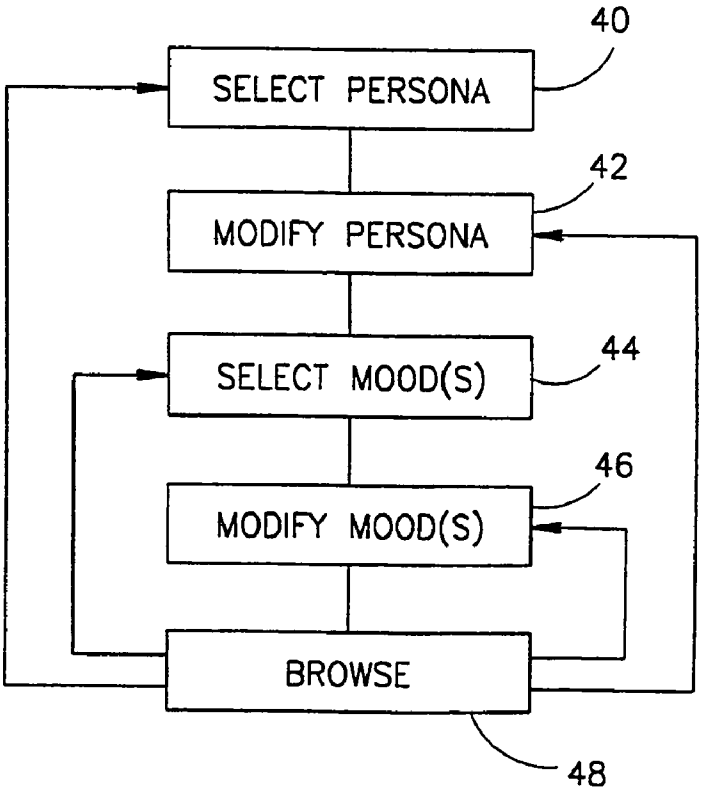

40
SELECT PERSONA
42
MODIFY PERSONA
44
SELECT MOOD(S)
46
MODIFY MOOD(S)
BROWSE
48

FIG.4

PERSONALIZED INTERNET INTERACTION BY ADAPTING A PAGE FORMAT TO A USER RECORD

RELATED APPLICATION

This application is a US national filing of PCT Application PCT/IL99/00056, filed Jan. 28, 1999.

FIELD OF THE INVENTION

The present invention is related to the field of Internet information searching and download and in particular to personalizing interaction with an Internet.

BACKGROUND OF THE INVENTION

The Internet contains an enormous amount of information. Several methods have been developed for searching the Internet for a particular piece of information. These include:

(a) Yellow pages. Each web site is hierarchically categorized by a subject matter and a user can browse sites by subject matter and/or perform a keyword search limited to a subject matter.

(b) White pages. The owners of web sites are listed in a geographical and/or alphabetical ordering. A user can browse web sites by owner name.

(c) Search engines. A user can perform a keyword search based on the content of WWW pages.

(d) Intelligent agents. A user defines a search criterion to an automated agent, which agent then searches for the information using search engines and/or web-crawling and/or alerts the user when new information comes to light.

(e) Indexes. Various services maintain indexes of interesting information and locations where such information may be found. In a WWW site of a particular company it is possible to determine a nearest outlet.

(f) Registry searches. A few new index services provide the possibility for a company or an individual to register an association between a slogan, trademark, product, subject and/or acronym and a WWW address and/or other company information. When a user enters such a keyword, the relevant information and/or addresses may be presented.

The Internet, in general, is an anonymous network and a particular search engine has no direct way of identifying a particular user. A mechanism called "cookies" is used by some WWW servers to store, at a user's computer, personalized information that is useful for accessing that site. Such information typically includes a preferred page layout and usage information, such as when the page was last accessed. In addition, such information may be used for targeting advertisements and/or for storing user preferences and/or previously entered data.

PCT publication 97/41673, the disclosure of which is incorporated herein by reference, describes a method of generating a psychographic profile of Internet users. It is suggested that the resulting profile be used, inter alia, for targeting of advertisements.

SUMMARY OF THE INVENTION

One object of some preferred embodiments of the invention is to provide a method of aiding information search and retrieval on an Internet. In a preferred embodiment of the invention, Internet searching is personalized to a particular user's profile. Alternatively or additionally, matching up of a supplier and a buyer, of a goods and/or a service, is facilitated, based on such personalization.

Another object of some preferred embodiments of the invention is to provide a "person" on an Internet with a persona and/or a mood, which affects the way the person interacts with other electronic entities on the Internet. In a preferred embodiment of the invention, the true identity of the person may remain anonymous.

One aspect of some preferred embodiments of the invention relates to generating an electronic person having a personality profile. The person preferably defines a personality for the Internet to interact with and/or be personalized to. Since the personality does not exist in a non-electronic form, it may be termed a "virtual personality". In a preferred embodiment of the invention, a user may switch between several personalities. In a preferred embodiment of the invention, a personality includes one or more of demographic information, geographic location, marketing information, subjects of interest to the user and/or other information, such as entertainment habits and ownership of a car. Each of the above subjects may include many sub-elements, for example, subjects of interest may include chess, checkers, baseball and swimming. Marketing information may include price preferences and buying habits. Geographical location may include one or more home locations, one or more shopping locations, one or more work locations and/or one or more vacation locations. In a preferred embodiment of the invention, the elements are hierarchically defined. In one example, a geographical location may include one or more of a country, city, neighborhood, street and house number levels. In another example, the area of interest "basketball" divides into a plurality of subject teams, and each team may be further subdivided into a plurality of players of interest. In a preferred embodiment of the invention, the elements include a relative weighting.

Another aspect of some preferred embodiments of the invention relates to providing such electronic persona with "moods", which define an instantaneous configuration of preferences and/or outlook. Typically, the mood modifies a persona. However, a mood may also operate without a persona. In one example, a particular persona may include a preference for difficult language. However, the persona's mood may be an "easy-going" mood, in which cases WWW sites having a simpler sentence structure and more graphics will be preferred. In another example, a mood may change between a "rush" mood, in which a user does not want to download large images and a leisure mood, where a user is willing to wait for long downloaded ad is willing to view advertisements if this makes his WWW access cheaper.

It should be appreciated that, technically speaking, both a mood and a persona may have a similar structure: preferences, weights and other aspects as described below. However, in a preferred embodiment of the invention, a persona is used to define a steady state personality which varies slowly, if at all. A mood is preferably used to emulate an instantaneous condition. In a preferred embodiment of the invention, the persona is defined as a structure and the mood defines changes in the structure, especially functional changes. For example, a "meticulous" persona which always desires complete downloads of images, may be modified by a "rush" mood, so that instantaneously it does not require complete downloads.

Another aspect of some preferred embodiments of the invention relates to using "persona" and/or "mood" (hereafter referred to together as "personality") to define a view of an Internet. In a preferred embodiment of the invention, one or more aspects of browsing and/or using the Internet may be

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

In one example, a search mechanism, such as yellow pages, white pages, indexes, search engines, intelligent agents and/or registry search, may filter and/or sort search results responsive to personality. In one example, a search may be limited to sites having a minimum percentage of graphics. In another example, a search result may be ordered by their average word length. Alternatively or additionally, the level of detail of the search results may be dependent on the mood, for example, in a rushed mood, only a line will be displayed for each search result. It should be appreciated that a persona is an indirect method of defining search criteria, as opposed to a usual method of defining search criteria, which precisely defines what type of information is desired in a site.

In a preferred embodiment of the invention, the browser itself, which servers as an interface to the Internet may also modify its functionality, responsive to the mood. Such modification may include one or more of menu length, help detail level, dialog boxes format, and response time vs. image quality.

It should be appreciated that a personality preferably includes both a persona and a mood. However, a personality may include only one of a personality and a mood. In one example, a personality may include only static preferences. In another example, a personality may include no static preferences but only dynamic mood parameters, such as being rushed or at leisure.

In a preferred embodiment of the invention, a personality defines the interpretation of key words and/or search terms. In one example, the word "U2" is interpreted as an aircraft designation for a "aeronautic" persona and as a name of a rock group for a "musical" persona. Another example is the word "chair" which can mean an academic position or an article of furniture. In one preferred embodiment of the invention, a single search may be applied to a plurality of different persona, yielding several sets of search results.

Additionally or alternatively, the personality may be used when entering any WWW site to provide personally tailored service. In one example, a news site will provide happy news for an "up-beat" persona and depressing news to a "pessimistic" persona. In another example, when entering a book-store or a library site, the site can tailor searches performed to the personality, for example, the regular interests of the user. In another example, a business mood will be greeted mainly with business news (and business related advertisements).

Another aspect of some preferred embodiments of the invention relates to personalizing advertisements responsive to a mood and/or a persona. This personalization of advertisements may be in addition to or alternatively to personalization responsive to a particular search and/or other actions performed by a user at a site. In a preferred embodiment of the invention, a site obtains information on a persona and/or a mood of the accessing user and then tailors services and/or advertisements based on the mood or persona. In a preferred embodiment of the invention, when a user enters a bookseller's web site, even if the user has never been at the bookseller, he may be offered books which match his persona and/or mood. It should be appreciated that, in some preferred embodiments of the invention, such a personality is not generally created and/or maintained by the site which uses the information for personalization.

In a preferred embodiment of the invention, a personality is stored as one or more cookies on a user's computer. Additionally or alternatively, the personalization information may be stored by a persona-service. Preferably, a user enters some type of identification, such as a code number, so that the service identifies the user. In some preferred embodiments of the invention, the persona are stored at a central location. Additionally or alternatively, the persona are stored in a distributed manner, such as locally to the users which use the persona. In a preferred embodiment of the invention, the personality and/or portions thereof may be stored as scripts to be executed and/or as parameters for pre-defined functions.

In a preferred embodiment of the invention, a persona is embodied by a program running on a user's computer, which program communicates with remote sites. The personalization information may be stored locally or may be acquired from a remote location, such as a persona server. Additionally or alternatively, the persona may be embodied by a proxy server, through which the Internet communication of a user must pass. In one example, a user access the Internet through a name server which translates nicknames, freeform text and/or search terms into URLs. The name server preferably also exhibits the user's personality to any site which requires it. Additionally or alternatively, a user's computer may include only an identification, such as in the form of a cookie. When a site requires personalization information, that information is downloaded from a persona server, using the identification cookie.

In a preferred embodiment of the invention, a persona and/or an identification of the persona is stored on a removable media, such as a diskette. Additionally or alternatively, it may be stored on a smart-card. Additionally or alternatively, it may be stored as a printed bar-code or bar codes. In a preferred embodiment of the invention, a persona may be used outside the Internet, for example in automated stores, for customizing the selection and/or offerings to such a customer.

In a preferred embodiment of the invention, charges are made for one or more of using, updating, accessing and/or exhibiting personalities. In some cases, the user may be charged. Additionally or alternatively, site owners and/or advertisers using the information are charged. Additionally or alternatively, the user may agree to allow himself to be targeted by advertisements and/or other types of promotions based on his profile instead of or in addition to being charged. In a preferred embodiment of the invention, the charges may depend on the personality. For example, in a rushed mood, a user will not desire to see advertisements. In a preferred embodiment of the invention, a user can ransom advertisements by paying a certain charge for not being presented with them. Preferably, the charges are consolidated in a monthly bill.

Another aspect of some preferred embodiments of the invention relates to security. In many cases a user will want to use his persona to affect his view of the Internet but will not desire such information to be freely available. In a preferred embodiment of the invention, only portions of a persona and/or a mood are provided to each site. In a preferred embodiment of the invention, the information provided is limited so that a particular user cannot be identified. In one example, geographical location and telephone number are not provided together. Additionally or alternatively, at least some of the personalization information is utilized as in a "black box." When a site desires to personalize information and/or

functionality, the site queries the persona server and the persona server responds for a particular situation. In one example, a search engine will transmit, to a persona server, a list of search results with their grading. The persona server may then respond with an order which is preferred by the requesting user. Preferably, the search site will not be provided with any identifying information about the user, except for an identification number randomly generated by the persona service for the particular search session.

Additionally or alternatively, personalization information may have a limited availability, for example, being provided only to registered and/or otherwise certified sites. Alternatively or additionally, a user may be queried to approve a particular site and/or a particular request. In a preferred embodiment of the invention, a user must enter a password to approve such a request. In one preferred embodiment of the invention, such a request is embodied as a WWW page sent by the persona server. Alternatively or additionally, the request is embodied as a pop-up window of the browser.

Another aspect of some preferred embodiments of the invention, relates to updating a mood and/or a persona. In a preferred embodiment of the invention, a user may manually modify a mood and/or a persona. Preferably, a mood and/or a persona is modified by changing parameters of the personality. Additionally or alternatively, a personality may be modified by selecting parameters and/or values for the parameters from existing definitions libraries. Optionally, such a selection may be modified by a user. Additionally or alternatively, a user may upload new personas and/or moods to a persona server, for personal use and/or for use by other users. Such personalization information, including libraries, moods, personas and/portions thereof may also be transmitted by e-mail to other users and/or to other locations.

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

In a preferred embodiment of the invention, a mood update also takes into account whether a WWW site is actually being viewed. In one example, the site may be downloaded by an invisible browser window. In another example, a user may be away from a computer (which may be determined if there are no user inputs and/or a screen-saver is activated).

The automatic updating may be performed at the user site, for example by tracking the activity of an Internet browser. Such tracking is preferably achieved using a standalone program which monitors the browser and/or TCP/IP connections. Alternatively or additionally, a dedicated TCP/IP stack and/or driver is used. Additionally or alternatively, the updating may be performed by a server, such as a proxy server, through which a significant portion of a user's requests and/or traffic, pass. In a preferred embodiment of the invention, such logging and/or tracking and/or persona modification require a user's explicit permission.

In a preferred embodiment of the invention, a user may require that certain WWW sites not be tracked, for example, sex related sites and Gambling related sites.

It should be appreciated that a mood is generally updated more often than a persona. In a preferred embodiment of the invention, a mood and/or a persona may be updated by modifying continuous parameters. Additionally or alternatively, such updating may include modifying discrete parameters. An example of a continuos parameters is "wait time" which indicates how long a user is willing to wait for a site to be downloaded. An example of a discrete parameter is a level of Parental Guidance rating of sites (PG-13, R, X). In a preferred embodiment of the invention, a plurality of personalities are predefined. Modifying a persona and/or a mood may include switching between such predefined personalities.

Another aspect of some preferred embodiments of the invention relates to interpretation of site functionality, responsive to a particular personality. In one example, a button marked "flowers" comprises a link to a site for ordering flowers. Responsive to the persona's geographical location and/or price preferences, connection to a different site may be made when the button is pressed, depending on the personality. Another example is a button marked "music", which downloads music to match the particular mood. Examples of moods to which music may be matched include: upbeat, rushed, loud, muted, and/or to mach color schemes.

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

In a preferred embodiment of the invention, parameters may include information, such as "subject of interest=chess". A parameter may also be negative, for example, a blacklisting: "reject=pornography" or "reject if pornography level >3". Additionally or alternatively, a persona may include weighing information, such as relative preference of subjects of interest, for example "baseball=5, basketball=3". Additionally or alternatively, a persona may include functional information, such as how to evaluate a particular parameter, the affect of a parameter and/or evaluate a grade for a particular site, in view of a parameter. Additionally or alternatively, a parameter may be reflexive towards the persona, for example defining how to modify the persona and/or a mood based on user activities. Additionally or alternatively, a parameter may define the traits which should be evaluated when determining a suitability of a site to a persona.

Additionally or alternatively, a parameter may define what type of site atmosphere (i.e., an ambiance) and/or other traits are suitable for different moods and/or based on values of

other parameters. In one example, a first happy persona will desire happy sites, while a second happy persona will desire bland sites, but good news. In another example, when a personality becomes happier the sites provided should be more morbid, thereby counteracting changes, while for another persona, the types of site provided should reinforce the changes in persona or mood.

Additionally or alternatively, to defining a persona using parameters, a persona may be defined by scripts. Preferably, such scripts are activated as result of events, such as "user entered a search request", "does attached site match the present mood" or "site asking for personalization information". Additionally or alternatively, a persona may be defined as a single program that accepts inputs and generates outputs. Additionally or alternatively, a persona may be defined by pre-selecting certain behaviors and setting parameter values for each selected behavior.

In a preferred embodiment of the invention, a group of personalities may be defined for related users. In one example, friends or members of a club may desire to share a small set of personalities or moods. In another example, family members may share many attributes, including address and financial situation.

In a preferred embodiment of the invention, a mood or a persona may be provided by an outside entity. In one example, an advertise may provide a persona and/or mood tailored for a particular product or group of products. This type of persona could be configured to receive advertisement, promotions and/or search results geared towards the product. Typically, using such a mood may result in a rebate on purchases, Internet fees and/or may involve a promotion including the product, for example a free sample.

In a preferred embodiment of the invention, a mood is defined as parameters with values that affect a persona. Such values may be, inter alia, single values, ranges of allowed values, scripts, continuous values and/or discrete values. A mood may replace certain parameter values, affect their value and/or affect their relative weighting. In some preferred embodiments of the invention, a persona may be embodied as a filter program which generates relative weights for a list of sites, based on the persona parameters and/or values. In a preferred embodiment of the invention, a mood may define filters to apply before the input and/or after the output of a persona.

Additionally or alternatively, a mood may define parameters that affect the execution of a persona's behavior. Additionally or alternatively, a mood may include segments of a persona to add to an existing persona. In a preferred embodiment of the invention, a single mood may be defined to fit more than one persona. Preferably, the different types of persona are grouped and/or hierarchically organized and each mood has a scope of personas it may affect.

Another aspect of some preferred embodiments of the invention relates to evaluating an atmosphere and/or other traits of a site. In a preferred embodiment of the invention, depending on a persona, several characteristics of a site may be defined, which may be used in filtering out and/or prioritizing such a site. Alternatively or additionally, such information may be used database of sites with their associated values is maintained, so that such characteristics do not need to be re-evaluated very often.

Another aspect of some preferred embodiments of the invention relates to associating traits and/or an atmosphere with a WWW site. The associations may be stored at a central location. Additionally or alternatively, the associations and/or trait-related keywords and/or values may be associated with each site. Additionally or alternatively, a site may include an identification number, which when used with a proper trait server, provides information about the sites traits and/or a match and/or grade with a particular personality. Additionally or alternatively, such associations may be stored in search indexes, preferably in a manner similar to the storage of key words.

In a preferred embodiment of the invention, a trait server provides the trait information for a fee. Additionally or alternatively, the information is provided in return for an agreement by the requester to be targeted for at least a limited number of advertisements and/or other promotions. Alternatively or additionally, the service may be provided for free There is therefore provided in accordance with a preferred embodiment of the invention, 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, responsive to said determined aspect,

wherein said modified interactions comprise site-content related interactions with a plurality of unrelated sites.

Preferably, said tracking comprises tracking at a computer at which said user accesses the Internet. Alternatively or additionally, said tracking comprises tracking at a tracking computer which tracks a plurality of users. Preferably, said tracking computer is physically remote from said plurality of sites.

In a preferred embodiment of the invention, said analyzing comprises analyzing previously acquired tracking data. Alternatively or additionally, said analyzing comprises analyzing of currently acquired tracking data. Alternatively or additionally, said determined aspect is modeled using a virtual personality, which is a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet. Preferably, said virtual personality comprises a persona, which is a static aspect of a personality. Alternatively or additionally, said virtual personality comprises a mood, which is a dynamic aspect of a personality. Preferably, said mood comprises a rush mood, which favors fast responses. Alternatively or additionally, said persona comprises a meticulous persona, which favors complete responses.

In a preferred embodiment of the invention, said personality comprises geographical information. Alternatively or additionally, said personality comprises demographic information. Alternatively or additionally, said personality comprises interests and preference information. Alternatively or additionally, said personality comprises marketing information. Alternatively or additionally, said personality comprises identification and contact information. Alternatively or additionally, said personality comprises relational information, which defines relations between various aspects of the personality. Alternatively or additionally, said personality comprises reflective information, which defines how a personality changes and/or interacts with other electronic entities. Alternatively or additionally, said user selects a particular virtual personality from a plurality of personalities to which to attribute said tracked interactions.

In a preferred embodiment of the invention, said future interactions comprise searching. Alternatively or additionally, said future interactions comprise viewing presented data. Preferably, modifying said interactions comprises changing a layout of data. Alternatively, modifying said interactions comprises changing which data is displayed.

In a preferred embodiment of the invention, said future interactions comprise downloading files. Alternatively or additionally, said future interactions comprise WW navigation.

There is also provided in accordance with a preferred embodiment of the invention, a method of user virtual personality maintenance, comprising:

interacting with an Internet via a virtual personality, which is a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet;

tracking at least one user activity of interaction with an Internet; and

modifying said virtual personality responsive to said user activity, wherein said virtual personality is user-selected for interaction with a plurality of different sites.

Preferably, modifying comprises modifying a mood of said virtual personality, wherein a mood is a dynamic aspect of a personality. Alternatively or additionally, the method comprises a user selecting said virtual personality to be modified.

There is also provided in accordance with a preferred embodiment of the invention, a method of user virtual personality maintenance, comprising:

providing first virtual personality, which is a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet;

copying at least a part of said first virtual personality into a second virtual personality; and

selecting said second virtual personality, by a user, to interact with an Internet.

Preferably, the method comprises further modifying said second virtual personality. Alternatively or additionally, providing said first virtual personality comprises:

providing a library of virtual personalities; and

selecting said first virtual personality from said library.

There is also provided in accordance with a preferred embodiment of the invention, a method of virtual personality interaction with an Internet, comprising:

providing a virtual personality, which is a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet, through which virtual personality an interaction with an Internet is meditated;

identifying at least one prospective site for the interaction;

automatically analyzing a content of said site to determine a match to said virtual personality; and

performing said interaction responsive to said analysis. Preferably, analyzing a content, comprises determining at least one trait of said site. Alternatively or additionally, analyzing a content comprises determining an ambiance of said site. Alternatively or additionally, analyzing comprises analyzing lexicographical characteristics of said site. Alternatively or additionally, analyzing comprises analyzing graphical characteristics of said site. Alternatively or additionally, identifying at least one site comprises identifying a plurality of sites. Preferably, identifying comprises searching using an Internet search engine.

In a preferred embodiment of the invention, said virtual personality comprises a mood, which is a dynamic aspect of a personality. Alternatively or additionally, said virtual personality comprises a persona, which is a static aspect of a personality. Alternatively or additionally, said interaction is performed to complement said virtual personality. Alternatively, said interaction is performed to match said virtual personality.

There is also provided in accordance with a preferred embodiment of the invention, a method of Internet interaction by a single user, comprising:

selecting, from a remote location, by the user, one of a plurality of virtual personalities available for interaction with a particular site, wherein a virtual personality comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet; and

interacting with the particular site using the selected virtual personality.

There is also provided in accordance with a preferred embodiment of the invention, a method of site ambiance provision, comprising:

receiving an identification of a site; and

providing an indication of an ambiance of said site, responsive to said identification.

Preferably, providing comprises retrieving said indication of an ambiance from a memory. Alternatively or additionally, providing comprises analyzing said site. Alternatively or additionally, providing comprises requesting an indication of said ambiance from said site. Alternatively or additionally, providing comprises requesting an indication of said ambiance from an ambiance server.

There is also provided in accordance with a preferred embodiment of the invention, a virtual personality server, comprising:

a connection to a user, through which said user indicates a desired Internet interaction;

a virtual personality adapter, which adapts said interaction utilizing a virtual personality for the user, wherein a virtual personality comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet; and

a connection to a WWW site, through which said virtual personality adapter interacts said modified interaction with said site.

Preferably, said connection to a user is operable to receive a selection of a particular virtual personality by said user. Alternatively or additionally, said server modifies said virtual personality responsive to said modified interaction. Alternatively or additionally, said virtual personality comprises a persona. Alternatively or additionally, said virtual personality comprises a mood.

There is also provided in accordance with a preferred embodiment of the invention, a method of virtual personality serving, comprises:

connecting to a WWW site, to request an interaction;

determining, at said WWW site, a desired virtual personality adaptation of said interaction, wherein a virtual personality comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet;

completing said interaction, by said WWW site, responsive to said determined virtual personality adaptation.

Preferably, determining comprises receiving an indication of a desired virtual personality from a virtual personality server. Preferably, said virtual personality server is located at a location remote from said WWW site and from a location at which said connection is initiated. Alternatively, said virtual personality server is located at a location from which said connection is initiated.

In a preferred embodiment of the invention, determining comprises reading virtual personality information from a computer at a location from which said connection is initiated.

In a preferred embodiment of the invention, said virtual personality server generates a one-time virtual personality for said interaction.

In a preferred embodiment of the invention, said desired virtual personality adaptation comprises a mood-responsive adaptation, wherein a mood is a dynamic aspect of a personality. Alternatively or additionally, said desired virtual personality adaptation comprises a persona-responsive adaptation, wherein a persona is a static aspect of a personality.

There is also provided in accordance with a preferred embodiment of the invention, a method of site matching to a virtual personality, comprising:

providing a list of relevant sites;

analyzing each of said sites to determine a match with said virtual personality, which is a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet; and

grading said sites responsive to said analysis.

Preferably, providing a list comprises executing a search on an Internet search engine to provide said list. Alternatively or additionally, providing a list comprises retrieving a plurality of matches from a name server. Alternatively or additionally, analyzing comprises analyzing at least one of said sites responsive to a presented ambiance. Alternatively or additionally, analyzing comprises analyzing at least one of said sites responsive to a presented trait. Alternatively or additionally, analyzing comprises analyzing a content of at least one of said sites. Alternatively or additionally, the method comprises displaying said graded list. Alternatively or additionally, the methdo comprises displaying only a highest graded site of said list.

There is also provided in accordance with a preferred embodiment of the invention, a method of advertisement personalization, comprising:

determining a present virtual personality of a human interactor, wherein a virtual personality comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet;

selecting at least one advertisement to match said virtual personality; and

presenting said advertisement to said interactor.

Preferably, said advertisement is presented through an Internet. Alternatively or additionally, said virtual personality comprises a persona, which is a static aspect of a personality. Alternatively or additionally, said virtual personality comprises a mood, which is a dynamic aspect of a personality. Alternatively or additionally, said virtual personality is selected and provided by said interactor.

There is also provided in accordance with a preferred embodiment of the invention, a method of WWW site modification, comprising:

detecting at the WWW a desired interaction from a particular virtual personality, which personality comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet;

modifying at least one characteristic of said site to match said virtual personality; and

responding to said desired interaction with a response indicating a match of said modified characteristic to said virtual personality.

Preferably, said modification comprises modifying a display layout. Alternatively or additionally, said modification comprises modifying a level of detail shown. Alternatively or additionally, said modification comprises selecting data to be displayed.

There is also provided in accordance with a preferred embodiment of the invention, a method of data directory display, comprising:

requesting a display of data from a data directory;

providing, in association with said request, a virtual personality for said request, which personality comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet; and

displaying said data, responsive to said virtual personality.

Preferably, said virtual personality is provided as part of said request. Alternatively or additionally, said displaying comprises filtering. Alternatively or additionally, said displaying comprises sorting. Alternatively or additionally, said displaying comprises controlling a level of detail. Alternatively or additionally, said displaying comprises controlling a spatial layout of said data.

There is also provided in accordance with a preferred embodiment of the invention, a method of data directory display, comprising:

requesting a search from a search engine, using at least one keyword, which request includes a virtual personality for said request, which personality comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet;

interpreting said key-word at said search engine, utilizing said virtual personality; and

performing said search request by said search engine, utilizing said interpreted key-word.

Preferably, said search engine comprises an Internet search engine.

There is also provided in accordance with a preferred embodiment of the invention, a method of Internet search, comprising:

connecting to an Internet search engine;

providing the search engine with search criteria;

performing a search for WWW sites by the search engine, utilizing said search criteria, to obtain search results; and

filtering said search results utilizing personal information.

Preferably, said filtering is performed at a different computer from said searching. Alternatively or additionally, said personal information is provided using a virtual personality, which comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet.

Preferably, said personal information is provided as a non-keyword input to said search engine.

There is also provided in accordance with a preferred embodiment of the invention, a method of interacting with a computer:

providing a software application having a user interface on said computer;

providing an electronic representation of at least part of a user's desired personality; and

said software modifying its interaction with said user, responsive to said representation of said personality.

Preferably, said software comprises an Internet Browser. Alternatively or additionally, said software modifies a visual display of said interface. Alternatively or additionally, said software modifies a behavior of said interface. Alternatively or additionally, said software modifies a menu length of said interface. Alternatively or additionally, said software modifies a help level of said software. Alternatively or additionally, said software modifies a level of detail presented by said software. Alternatively or additionally, said software modifies a display format of said software. Alternatively or additionally, said software modifies an image quality of said software. Alternatively or additionally, said software modifies a

response time of said software. Alternatively or additionally, said representation is generated by tracking a plurality of interactions of said user with an Internet. Alternatively or additionally, said representation comprises a representation of a persona, which is a static aspect of a personality. Alternatively or additionally, said representation comprises a representation of a mood, which is a dynamic aspect of a personality. Alternatively or additionally, said desired personality comprises a true personality of said user. Alternatively or additionally, said desired personality comprises a true personality of said user.

There is also provided in accordance with a preferred embodiment of the invention, a method of utilizing an electronic representation of a user's desired personality, comprising:

storing said representation on a computer-readable storage media; and

interacting with a computer using said representation, wherein said representation mediates the interaction.

Preferably, said computer comprises a remote computer connected to an Internet. Alternatively or additionally, said computer comprises a controller of an automated store. Alternatively or additionally, said mediation comprises varying a range of offered selection of products. Alternatively or additionally, said media comprises a diskette. Alternatively or additionally, said media comprises a smart card. Alternatively or additionally, said media comprises printed optically readable codes.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will be more clearly understood by reference to the following description of preferred embodiments thereof in conjunction with the figures, wherein identical structures, elements or parts which appear in more than one figure are labeled with the same numeral in all the figures in which they appear, in which:

FIG. **1** schematically illustrates the use of a persona while browsing an Internet, in accordance with a preferred embodiment of the invention;

FIG. **2** illustrates a persona server configuration, in accordance with a preferred embodiment of the invention;

FIG. **3** illustrates the use of both an atmosphere and a persona in accordance with a preferred embodiment of the invention; and

FIG. **4** is a flowchart of a process of persona and mood selection and update, in accordance with a preferred embodiment of the invention.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

FIG. **1** schematically illustrates the use of a persona **12** while browsing an Internet **16**, in accordance with a preferred embodiment of the invention. A client **10** views Internet **16** through a persona **12**. Persona **12** may be further modified by a mood **14**. In a preferred embodiment of the invention, a persona and/or a mood (hereafter "personality") can be used to filer information on Internet **16**, thereby making the amount information more manageable and better suited for client **10**. It should be appreciated that in many cases there are several possible correct responses to a client's need. However, one of these responses may better suit the frame of mind of the client. In a preferred embodiment of the invention, a persona and/or a mood may be used to have one or more of the following effects on the interaction between client **10** and Internet **16**:

(a) preferentially guide client **10** to certain sites;

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

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

(d) affect the display of information;

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

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

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

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

In a preferred embodiment of the invention, a client may have a plurality of personalities and manually select a particular persona for a session or a portion of a session. Thus, client **10** might start the day with a "rushed" personality and in the afternoon switch to a "leisurely" persona.

In a preferred embodiment of the invention, a persona includes one or more of the following types of information:

(a) Demographic information, possibly including one or more of: age, sex, religious affiliation, culture subgroup, socioeconomic status, martial state, educational state, number of telephones at home, distance from post office, family members information and/or pets. Alternatively or additionally, also personal information, such as birthdays and wedding anniversaries are included. In a preferred embodiment of the invention, an advertiser matches a promotion to the special date, for example, a gift on a birthday, suitable for the persona's professed interests (see "c" below).

(b) Geographical information, possibly including one or more of, home location(s), work location(s), shopping location(s), vacation location(s) and/or other user defined locations and/or areas. In some cases more than one geographical location may be defined, for example, for a person having two homes or where two shopping districts are applicable.

(c) Interests and preference information, possibly including one or more of hobbies (chess, sports, painting), academic interests, taste in music, taste in movies, preferred fashion designers, political views, membership in voluntary groups and/or fraternities, membership in a persona group, favorite color, set of beliefs, personality type, outlook (optimistic, pessimistic, believer in human nature being good), sexual preferences, density of information preferred, number of advertisements preferred, requirement for a seal of approval, requirement for a site review, requirement for a site being accessed often, number of links, download time, preference for sites that give out presents and/or other promotional merchandise. Alternatively or additionally, IDs of friends may be included. Preferably, the personality of a user may be estimated based on the types of friends him keeps in contact with. In a preferred embodiment of the invention, the persona server tracks frequency, length and/or content of electronic communications with the friends, for example by e-mail, by chat group, by Internet telephone or by computer-dialer, to evaluate an instantaneous mood and/or to assess the relative effect of these friends. Alternatively or additionally, the persona server may create a representation of the connection between friends, based for example on the electronic communication between them and/or on inputted connections. Using such a representation, a WWW site may simultaneously provide an offer to a group of friends, for example a jeep tour. Alternatively or additionally, an advertiser may user the information to identify trend setters. In some types of advertising campaigns such trend setters are targeted first.

(d) Marketing information, possibly including one or more of: credit rating, preferred price range, purchasing style (conservative or adventurous), preferred purchasing method, price point for various products, purchasing history, maintenance level desired for a product, willingness to pay for the use of a service, willingness to be identified, willingness to fill in forms type of preferred advertisements, preference for national chains vs. local chains, preference for large number of nearby stores vs. small number of nearby stores, belonging to a consumer's group and/or profile of advertisements previously viewed and/or judged by the client. Marketing information may depend on a type of product, for example, a particular persona may prefer cheap cars and expensive alcohol. Alternatively or additionally, more complex relationships may be defined. For example, a persona may agree to consider Swiss chocolate only if it costs less than $10/lb. or if it contains at least 33% chocolate solids.

(e) Identification and contact information, possibly including one or more of: mail address, WWW home page(s), e-mail address and/or financial information, such as bank account and/or credit card information.

(f) Relational information, which defines relations between various aspects of the persona, including, relative weights, interaction between parameters (something that is both chess and baseball may be 10 times as interesting if it matches only one of the subjects), dependencies (a client can prefer complex language in chess-related information but simple language in baseball-related information).

(g) Reflective information, which defines how a persona changes and/or interacts with other electronic entities, including, effect of mood on grading of sites, matching functions for evaluating sites and/or matching an information file to one or more parameters, default mood, and/or effect of browsing and/or user input and/or downloaded information on personality. In addition, reflective information may define limitations on allowed ranges, for example relative weights. As described below, a persona may change over time. These ranges may limit the allowed changes in a persona. Additionally or alternatively, a reflective information may define an activity to perform when a certain mood is reached, for example, if a mood becomes too happy, the persona may be adjusted to cause the mood to become less happy.

(h) Miscellaneous information. Any information may be associated with a persona. In one example, a persona may include examples of preferred sites.

The above described information about a persona may be stored in one or more of many ways, in accordance with preferred embodiments of the invention. In a preferred embodiment of the invention, the information is implicit in a program and/or a set of programs of scripts which carry-out the persona. Additionally or alternatively, the information is stored as parameters for predefined and/or user defined functions. Additionally or alternatively, parameter values comprise scripts or functions which generate values, responsive to input values, internal variables, global variables, other parameters of the persona and/or the parameters of the mood. Additionally or alternatively, the information is stored as values for preset persona parameters.

In a preferred embodiment of the invention, a mood may include any of the above pieces of information. The mood information may then be used to augment and/or to replace portions of the persona. In a preferred embodiment of the invention, the persona is defined as a structure and the mood defines changes in the structure, especially functional changes. For example, a persona may define subjects of interests and a mood defines relative weighting between the subjects.

In a preferred embodiment of the invention, the mood may be used to modify the identification information. In one example, play moods receive e-mail in which the subject line is prefixed with the phrase "mood=play", while work moods may be prefixed with the phrase "mood=work". Thus, a user can differentiate between the results of his various moods, using existing software. Alternatively or additionally, software, such as browser software may be adapted to present the results of the separate moods independently or comparatively.

In a preferred embodiment of the invention, a mood may be utilized to search for suitable personal contacts, for example in a chat group or in a dating system. It should be appreciated that matches, if any, are made based on the instantaneous mood and not based on some general profile which is entered at a single point in time. Thus, at one time, a user may be matched up with a somber chat member while at another time a user may be matched up with an upbeat chat member.

In a preferred embodiment of the invention, a mood and/or persona may be used to aid in polling. Preferably, when creating a list of persons to poll, the pollster may take advantage of the listing of the persons preferences, demographic information, preferences and other information associated with a person. In a preferred embodiment of the invention, a poll may be made more statistical significant by selecting certain slices of society based on the persona information. Alternatively or additionally, an existing poll may be analyzed based on the persona information. Alternatively or additionally, a poll may be directed only at segments of society which would have a meaningful response. This is especially true of advertising polls, where only a potential market should be polled.

FIG. **2** illustrates a persona server configuration **11**, in accordance with a preferred embodiment of the invention. A client **10** is connected to a WWW site **18**. A persona server **20** provides persona information to site **18**, to enable site **18** to personalize it's interaction with client **10**. Persona server **20** itself may include a database having persona and/or mood details for a plurality of users and Internet connections, for connecting to sites and/or a client computer. A CPU in persona server may be used to evaluate personalities and/or suitability of sites.

Alternatively, other configurations may be used, in accordance with preferred embodiments of the invention. In one alternative configuration, a persona server is situated between a client and the Internet and may also control access to various parts of the Internet. This configuration may be implemented if the persona server is associated with an ISP, a proxy server and/or a name sever, for example as described in U.S. Pat. No. 5,764,906, issued Jun. 9, 1998 or Israel patent application number 123,129, filed Jan. 30, 1998, to Aviv Refuah, the disclosures of which are incorporated herein by reference. In another alternative configuration, a persona server may be installed on a client computer.

In a preferred embodiment of the invention, the persona information is stored on at a centralized location. Additionally or alternatively, the persona information is stored using a distributed configuration, in which it is partially stored in a centralized location and partially stored at a client's computer, preferably, using cookies. Alternatively or additionally, multiple copies of a single persona and/or mood may be stored at different locations. Alternatively or additionally, at least part of the persona and/or mood may be stored at sites, such as search sites, which use the personality information. Alternatively or additionally, the persona information is stored on a client's computer, preferably using cookies. In a preferred embodiment of the invention, at least an identifica-

tion number is stored on a clients computer, so that persona information may be retrieved by the client's computer from a remote location using that identification number. In a preferred embodiment of the invention, a program running on a client's computer modifies the cookies to present a particular personality. Preferably, such a program stores the persona in a local database file.

In a preferred embodiment of the invention, the persona information is associated with the identification number using a database. Additionally or alternatively, a persona may be created ad-hoc, for use in a single WWW site access, a single search and/or for a short period of time.

In a preferred embodiment of the invention, a persona is statically stored at a first location and a mood is dynamically stored at a second location. In a preferred embodiment of the invention, a single persona may be shared between several users, with only the moods being different. The moods may also be selected from a table, such that what categorizes a particular personality is an instantaneous association of a predefined persona with one or more predefined moods.

In a preferred embodiment of the invention, a persona is used to personalize information retrieval. Such personalization can affect many methods of information retrieval, including search engines, name servers, intelligent agents, yellow pages, white pages, and searching inside a WWW site, such as searching for articles on Microsoft products inside the Microsoft WWW site. It should be noted in this context that search engines return matches for a particular query, while personality and mood are designed to affect the results of substantially any query, even though a personality does not specifically point out a desired piece of information.

The information retrieval may be filtered and/or sorted based on the persona. Additionally, the persona may decide how a particular search word is interpreted. Various combinations of thresholding, grading and sorting may be applied on search results, by comparing them to a persona. One or both of the following two techniques are preferably used to match a persona to a search result, namely key-words and evaluation. In the key-words technique, a search index includes a classification and/or key-words which match parameters such as those described above for a persona. For example, a site may be indexed as being related to a particular type of music, which type may match a preference of a client.

In the evaluation technique, a site is evaluated for suitability and/or for qualities which are preferred and/or match a particular persona. Example include: number of images in the site, expected download time and/or number of links from the site.

In a preferred embodiment of the invention, the presentation of search results may also be parameters of the persona. In one example, the persona can dictate whether or not to grade sites or information files and whether or not to limit the results using criteria such as geographical criteria. Thus, in one case, a strong match will be shown even if its associated geographical location is 1000 miles away. In the other case, only hits having an associated geographical location within 50 miles are shown. Additionally or alternatively, a mix between near and far results may be defined. One or more parameters of a persona may define matching requirements, for example exactness of match and allowed error. These parameters may depend on the preference being matched. In some cases, there is no way to match a parameter of a persona (e.g., if no geographical location is associated with the site). Default behavior in such cases may also be a parameter of the persona.

In one example, entering searching for a pizza store will generate a different web site connection, based on where the connection is from. For example, a user in Brighton, Mass. will be directed to a different pizza store from a user in downtown Boston, even if both stores belong to the same franchise.

Additionally or alternatively, a persona may define multiple response sets. In one example, one set includes low-cost book stores and a second set-includes high-cost book stores. This division may be the result of a preference for differentiating between high-cost and low-cost suppliers.

In a preferred embodiment of the invention, a far geographical location may be considered to have the effect of a near location, responsive to the availability of mail order and/or courier services. Preferably, such a translation is also a function of the transportation cost. This type of translation may also be governed by a parameter of a persona.

In a preferred embodiment of the invention, the personality is used as an input to an intelligent search agent. It should be appreciated that a persona is an indirect method of defining search criteria, as opposed to a usual method of defining criteria for an agent. Additionally or alternatively, the agent can dynamically modify his searching and/or presentation of results, responsive to changes in the personality.

When an interactive agent, such as an agent which performs negotiation is used, the persona may also be used to affect the agent's behavior. In one example, an agent is made more aggressive, as the persona's outlook becomes more optimistic. In a preferred embodiment of the invention, the interaction between the personality and the agent may also be defined as parameters of the persona. Some aspects of automated agents are described in PCT publication WO 97/26612, titled "Intelligent Agents for Electronic Commerce", filed Jan. 17, 1997 in the US Receiving office, the disclosure of which is incorporated herein by reference.

When white and/or yellow pages are used, the display of information from a database may be determined by the personality, for example, the display of listings of baby-sitters, handymen and car garages may all depend on a geographical distance.

In a preferred embodiment of the invention, the personality may be used to target advertisements to the client. Such targeting may also take into account previous advertisements viewed by the client. In one example, advertisements are matched to professed subjects of interest. Additionally or alternatively, advertisements are matched to an outlook, for example morbid or sunny. Alternatively or additionally, advertisements are selected from a set of suitable advertisements to match a persona. For example, if two soft drink advertisements are available, one which includes animals and one which includes cars, the "animal" advertisement will be selected for a persona which likes animals. Similarly, some advertisements are garish, while some are reasoned out. A somber mood will preferably be targeted with the reasoned out advertisement, since a garish advertisement might antagonize the client.

In a preferred embodiment of the invention, a WWW site may tailor its reactions to the client based on the persona. In one example, the content of links on a page may depend on the persona. Additionally or alternatively, the effect of a button may depend on the persona, for example a persona's geographical location. Additionally or alternatively, a Java applet and/or a JavaScript script may utilize persona information in their execution. For example, prior to rendering a button, a Java applet may check if to use a garish background for an upbeat persona or a muted background for a somber persona.

In a preferred embodiment of the invention, persona information is used to exercise parental control over a child's

browsing. In such a case the persona is "hardwired" or protected from changes to reject undesirable sites.

In a preferred embodiment of the invention, the persona is generated and/or maintained by a client. Additionally or alternatively, the persona may be generated by a gateway through which a client accesses the Internet. In some cases, the persona may be generated on a client's computer, in a manner which is transparent to the client. In a preferred embodiment of the invention, a client can define certain sites and/or site types as not being tracked. For example, a client may not desire that his excursions into gambling sites be of record. Alternatively or additionally, a client is preferably able to define what sites may obtain what information about him.

In some preferred embodiments of the invention, a client may be unable to access the contents of his own persona. Additionally or alternatively, the existence and/or usage of the persona is kept secret from the client and the persona is used to target the client with promotions and/or advertisements

In a preferred embodiment of the invention, a persona of a client may be automatically generated by tracking the way a client interacts with the Internet.

In a preferred embodiment of the invention, a persona is divided into several layers of privacy. Some requesters may be able top access some parameters of the persona but not others. In a preferred embodiment of the invention, at least some of the personalization information is utilized as in a "black box", in that the persona server generates behaviors in response to queries but only on an individual basis, without letting out on the values of its internal parameters.

In a preferred embodiment of the invention, a persona server can generate and/or print reports. Such reports preferably include distribution of persona and/or mood use, statistics of site access, user satisfaction, statistics of persona parameters and/or values and/or any other information relating to the use of personalities. In some cases, a persona server may pull the information from remote sites, for example, from user computers at which persona are stored.

In a preferred embodiment of the invention, the persona server generates reports for WWW sites. Such reports may includes statistics of visitors, for example number of visitors and duration of connection. The report may be sent on a periodic basis, such as once a month or the report may be sent on demand. Preferably, the report includes statistics of the types and/or parameters of personas and/or moods of visitors.

In a preferred embodiment of the invention, a persona may be made anonymous, utilizing a personality server. Thus, when a user access a site which requires or prefers a persona, the user performs the access through a persona server. The persona server preferably generates a new persona identification for each such interaction, so that the accessed site has no way of associating a particular client and/or tracking a particular client's preferences.

Additionally or alternatively, user names and passwords required to access various sites may also be parameters of a persona. Thus, when a user accesses a site, that site can obtain required access codes from the persona, without bothering the user. As can be appreciated, different moods for a single user may have different access codes for a same site, for example, to control the behavior of the site to match a mood. Alternatively or additionally, the persona may include credit card information, deposit account and/or other account information which may be used for billing. In a preferred embodiment of the invention, the persona server includes a charge account for each persona, so that a persona can purchase services and/or goods on the Internet without sending a credit card number. The WWW site at which the goods were purchased is then reimbursed by the persona server, which bills the "persona".

FIG. **3** illustrates the use of both an atmosphere **30** associated with a site and a persona **12** associated with a client in accordance with a preferred embodiment of the invention. While a client **10** may desire to present itself as a persona having various preferences, a WWW site **18** may also desire to present itself as meeting certain preferences and/or characteristics. Such a presentation of site **18** assists clients who are searching for the site. In a preferred embodiment of the invention, atmosphere **30** may define a complement of persona **12**, so that matching persona will have a high affinity for the site. The atmosphere may be implemented in search engines, in indexes, as information stored at the site and/or using an atmosphere server. The atmosphere server provides, possibly at a fee to site **18**, client **10** and/or an advertiser, the atmosphere of site **18**.

An atmosphere may include a plurality of traits, for example, political slant, garishness, reading grade level, subjects of interest and in general the complements of parameters of a persona.

The atmosphere of a site may be evaluated on the fly using various methods described below. Additionally or alternatively, a client may grade a site. Such grading may become publicly available or it may be limited to the client or a group of clients. Additionally or alternatively, the persona server or the atmosphere server may evaluate a site. Additionally or alternatively, a separate server may provide a site evaluation service.

In a preferred embodiment of the invention, a site may be automatically evaluated by tracing the personas and/or moods of clients who visit the site and/or remain at the site for a significant amount of time. In a preferred embodiment of the invention, such tracing is performed by the site server. Additionally or alternatively, the tracing is performed by a persona server and/or an atmosphere server.

In a preferred embodiment of the invention, a plurality of sites are evaluated using a web-crawler. The web-crawler browses the Internet and sends an e-mail to the operator of each site. The e-mail preferably includes a questionnaire having open-ended and/or close-ended questions about the traits and/or atmosphere of the sight. An e-mail address for the questionnaire may be determined by searching the page for e-mail addresses, especially those having the form "info@.", "postmaster@ . . . ", "webmaster@ . . . ". In some preferred embodiments of the invention, only one address per domain name is used, unless a page has a title such as "home page" which indicates that the page is a root of a site.

Additionally or alternatively, to a site having a static atmosphere, a site may have a dynamic atmosphere. Thus a client may have to recheck the atmosphere and/or traits of a site periodically or before every access. Additionally or alternatively, a site can modify itself to match a desired atmosphere and be more acceptable to certain personas and/or moods. In one example, a site may have the option of using one of four color schemes: muted, respectable, garish and art-deco. In a preferred embodiment of the invention, a site may automatically select one of these schemes responsive to a persona which requests access to it. Additionally or alternatively, a site may change its design and/or presentation responsive to statistics of personas and/or moods which access the site. In one example, a site may modify one or more parameters of its design to match the personas of its clientele and/or of a desired clientele. In a preferred embodiment of the invention, such modification of a site is at least partly manual, by providing a site manager with statistics regarding the accessing

of various persona and their parameters. Additionally or alternatively, a site may have predefined at least one parameter which automatically changes its display format and/or information filtering mechanism, for example using a predefined script, to match moods and/or personas which access the site.

In a preferred embodiment of the invention, an atmosphere of a site may be automatically evaluated by analyzing the content of a site, in addition to or instead of utilizing a client's reaction to the site or statistics of accessing the site. Various characteristics of a site may be automatically determined. Each of these characteristics and/or combinations thereof may be used to estimate values for traits and/or atmosphere. The characteristics preferably include one or more of:

(a) word length;

(b) whether certain words and/or phrases used by or associated with the site belong to certain groups, such as "academic words", "swear words", "adult words", "new-age words", "sports words", "baseball words";

(c) sentence complexity;

(d) density of displayed text;

(e) ratio between images and text;

(f) size of text;

(g) distribution of colors in image sand in background;

(h) number of links; number of links visited, date of last visit, by the client, by the persona, by the mood and/or by other moods, personas and/or clients;

(i) size of site;

(j) key-words presented by the site; and/or

(k) number of images; and/or

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

In a preferred embodiment of the invention, an atmosphere of a site is used to match advertisements even if a persona and/or mood are not known. For example, a macabre site will prompt the usage of dark advertisements. Alternatively or additionally, a complex site will prompt the use of complex advertisements.

FIG. **4** is a flowchart of a process of persona and mood selection and update, in accordance with a preferred embodiment of the invention.

In a preferred embodiment of the invention, a client creates a persona by selecting an existing "standard" persona form a library of personas (**40**). Typically, the "standard" persona is modified by the client (**42**) to better match the client's exact desires. One or more moods may then be select from a library of moods, for the selected persona (**44**). In a preferred embodiment of the invention, a list of matching moods is associated for each persona. The matching moods are preferably moods which modify a persona in a most natural and/or useful manner. These moods may also be modified (**46**).

During browsing (**48**) a user may select a different persona or a different mood to better match his present frame of mind (left arrows). Additionally or alternatively, the personality(s) may be automatically modified. As described above the type of automatic modification and/or its parameters may also be a function of the personality.

In a preferred embodiment of the invention, a client selects a persona and/or a mood when entering the Internet or by entering a special site ad/or by running a special program on hi computer. In a preferred embodiment of the invention, each connection location may have associated therewith a default personality, for example a "work" personality and a "home" personality. When a client connects from an unknown computer, he can elect to assume an existing personality, for example the "home" personality.

In a preferred embodiment of the invention, when a client selects a personality at login, the client may also elect to modify the personality and/or mood. In one example, a user can indicate to the persona server that he is not alone in the room, so that sensitive information is not displayed. Alternatively or additionally, a client may define a "cover" personality, for example in case a boss walks in and is interested in what the client is doing.

In a preferred embodiment of the invention, a personality are updated responsive to one or more of the types and/or contents of sites that a client accesses, the time spent at each site (preferably with a deduction for connection time), activities performed at the sites and/or data downloaded from the sites. In a preferred embodiment of the invention, a client can indicate to the persona server if he is pleased with a particular site and/or displeased. Such an indication may also be used to modify the personality.

Additionally or alternatively, a persona server may generate a persona and/or a mood by providing a client with a questionnaire and filling in various parameters for a personality using the contents of the questionnaire.

In a preferred embodiment of the invention, a client base may already exist and what is necessary is to generate persona and/or moods for them. In such a case, a standard questionnaire may be sent to all the client base. Alternatively or additionally, a program may be loaded to the individual client to track Internet usage patterns. The program may continuously update a central location. Alternatively or additionally, the program may store its tracking results and update the center periodically. Alternatively or additionally, the program generates a locally stored persona and does not update a central location and/or updates the central location only with the persona. Such tracking may be performed with or without a user's knowledge. Alternatively or additionally, such tracking may be performed with or without user participation with the process. Alternatively or additionally, a database of user site access may already exist, either on individual client computers or at a centralized locations. In such a case, personas may be generated off-line by analyzing the Internet usage.

In some cases, off-line analysis of Internet usage may suggest more than one persona and/or more than one mood and/or a range of values for parameters of the mood. These variations may be identified, for example, by clustering of Internet access statistics. In one example, a user access profile may include time periods where a "rush" mood is evident and other periods where a "leisure" mood is evident.

Additionally or alternatively, a persona and/or a mood may be generated by tracing the Internet activities of a client. In a preferred embodiment of the invention, a client indicates at a start of an Internet session which personality he is emulating at the time.

In a preferred embodiment of the invention, a persona may be defined in a parametric manner. A client may modify a persona and/or a mood by multiplying various parameter values by factors. In one example, a client may decide that on a certain day he is 15% more conservative and 20% less interested in chess. The values of the parameters may also be a function of the time of day, day of the week, date and/or length of time logged on to the Internet.

In a preferred embodiment of the invention, various tools are provided for maintaining personas and/or moods. In a preferred embodiment of the invention, personalities may be stored, retrieved and/or electronically transmitted. In a preferred embodiment of the invention, personalities may be stored in a smart card and/or on a diskette. In a preferred embodiment of the invention, a central repository of personal portions and/or moods may be established for sale and/or trading of personalities. In some cases a group of clients may desire to have similar or identical moods.

In a preferred embodiment of the invention, two or more moods and/or persona may be compared to determine differences between them. Preferably, such comparison includes comparing the parameters and/or comparing the effects of using the moods and/or persona in various test cases. Additionally or alternatively, a client may rewind his Internet activities and perform them again using a different and/or modified personality. Additionally or alternatively, a client may perform activities in parallel using several personalities and compare the effects of the personalities.

In an example of utilizing personas for Internet commerce, a client may be a married business man, having two children, on teenage and one a toddler, a dog, an company office in London and family living in London. Also, the client is an exercise freak. All of the above information is preferably part of the client's profile. The client is looking for a hotel in London for business meeting. When such a client connects to WWW sites of hotel chains, he is not required to reenter personal information. Each site can offer an hotel which best suites his needs (near the office, shopping and family). Alternatively or additionally, each site can personalize its response to his query, for example, informing that it does or does not have an exercise spa. Alternatively or additionally, each site can personalize its promotions, for example, offer a low-rate accommodation for an accompanying teenage family member. Alternatively or additionally, the site can personalize advertisements, for example display advertisements for quality dog food available in the London region and/or dog sitters.

It should be appreciated that the above description of personalization has been focused on an Internet application. However, it should be appreciated that the same principles may be applied to any search mechanism and/or large database. Nevertheless, various preferred embodiments of the invention are particularly useful for the Internet due to the type of indexing of the Internet, the type of browsing practiced in the Internet, the Internet's social implications and the very large number of authors in the Internet.

It will be appreciated that the above described methods of Internet personalization may be varied in many ways, including, changing the order of steps, which steps are performed on-line and which steps are performed off-line. In addition various distributed and/or centralized configurations may be used to implement the above invention, preferably utilizing a variety of software tools. In addition, a multiplicity of various features, both of method and of devices have been described. It should be appreciated that different features may be combined in different ways. In particular, not all the features shown above in a particular embodiment are necessary in every similar preferred embodiment of the invention. Further, combinations of the above features are also considered to be within the scope of some preferred embodiments of the invention. Also within the scope of the invention are computer readable media on which software, for performing part or all of a preferred embodiment of the invention, are written. It should also be appreciated that many of the embodiments are described only as methods or only as apparatus. The scope of the invention also covers hardware and/or software adapted and/or designed and/or programmed to carry out the method type embodiments. In addition, the scope of the invention includes methods of using, constructing, calibrating and/or maintaining the apparatus described herein.

It will be appreciated by a person skilled in the art that the present invention is not limited by what has thus far been described. Rather, the scope of the present invention is limited only by the following claims. When used in the following claims, the terms "comprises", "comprising", "includes", "including" or the like means "including but not limited to".

The invention claimed is:

1. A method of a user interacting with an Internet site managed by a first entity, comprising:

(a) receiving at the site, from a user of a remote computer connected to the Internet, a request for an Internet page;

(b) receiving at the site, a predefined site-independent user profile record of the user transmitting the request, provided by a second entity not associated with the first entity;

(c) obtaining, responsive to the request, information content of the Internet page;

(d) selecting through an atmosphere server associated with the site a trait defining characteristics for a display format of an Internet page to the user that is a complement to at least one attribute of the user profile record provided to the site and used by said atmosphere server; and

(e) responding of the site to the request, with a page presentation of the provided information content of the requested Internet page in a display format defined in accordance with the trait selected through the atmosphere server responsive to the user profile record.

2. A method according to claim **1**, wherein selecting the display format comprises modifying a display layout.

3. A method according to claim **1**, wherein selecting the display format comprises modifying a level of detail shown.

4. A method according to claim **1**, wherein selecting the display format comprises selecting data to be displayed.

5. A method according to claim **1**, wherein selecting the display format responsive to the user profile comprises modifying a display format of a browser of the client.

6. A method according to claim **1**, wherein the user profile records are managed at least partially by tracking interactions of the user with an Internet and analyzing the tracked interactions to determine at least a part of the user profile.

7. A method according to claim **6**, wherein said tracking comprises tracking at a computer from which the user accesses the Internet.

8. A method according to claim **6**, wherein said tracking comprises tracking at a tracking computer which tracks a plurality of users, accessing the Internet from different client stations.

9. A method according to claim **8**, wherein said tracking computer is physically remote from a computer hosting the site.

10. A method according to claim **6**, wherein said tracking comprises tracking at a tracking computer which tracks access to a plurality of different unrelated web sites.

11. A method according to claim **1**, wherein providing the information content of the Internet page comprises providing the information content of the Internet page without relation to the user profile of the user transmitting the request.

12. A method according to claim **1**, wherein the selected display format includes at least one non-information-content attribute selected responsive to the user profile.

13. A method according to claim **12**, wherein the at least one attribute of the display format determined responsive to the user profile comprises one or more display colors.

14. A method according to claim **12**, wherein the at least one attribute of the display format determined responsive to the user profile comprises an image quality.

15. A method according to claim **12**, wherein the at least one attribute of the display format determined responsive to the user profile comprises a size, amount or density of displayed text.

16. A method according to claim **1**, wherein transmitting the request comprises transmitting along with the user profile record.

17. A method according to claim 1, wherein responding to the request comprises responding by the site, which selects the display format.

18. A method according to claim 1, wherein responding to the request comprises providing the information content by the site, and selecting the display format at least partially by a persona server, separate from the site, which provides the presentations to the users.

19. A method according to claim 1, comprising repeating (a), (b), (c) and (d) for a plurality of unrelated sites using a single user profile.

20. A method according to claim 1, comprising repeating (a), (b), (c) and (d) for a plurality of user profile records and a single site, wherein at least some of the responses to the requests differ at least in their display formats.

21. A method according to claim 20, wherein providing the information content comprises providing the same information content for a plurality of the users.

22. A method according to claim 20, wherein providing the information content comprises providing different information content including different search results, for a plurality of the users.

23. A method according to claim 20, wherein the display formats selected responsive to the user profile records of at least two of the users differ in their level of detail.

24. A method according to claim 20, wherein the display formats selected responsive to the user profile records of at least two of the users differ in a percentage of image data that they include for one or more of their images.

25. A method according to claim 20, wherein the display formats selected responsive to the user profile records of at least two of the users differ in their colors.

26. A method according to claim 20, wherein the display formats selected responsive to the user profile records of at least two of the users differ in their spatial layout.

27. A method according to claim 20, wherein the display formats selected responsive to the user profile records of at least two of the users differ in a size, amount or density of displayed text.

28. A method according to claim 20, wherein the display formats selected responsive to the user profile records of at least two of the users differ in a number or percentage of non-textual objects.

29. A method according to claim 20, wherein the display formats selected responsive to the user profile records of at least two of the users differ in the type of words they use.

30. A method according to claim 20, wherein the display formats selected responsive to the user profile records of at least two of the users differ in a ratio between images and text in the page.

31. A method according to claim 1, wherein providing the user profile record of a user comprises providing a record at least partially determined for one or more other users.

32. A method according to claim 1, wherein providing the user profile record comprises providing by a proxy through which the request for the Internet page is provided to the site.

33. A method according to claim 1, wherein providing the user profile record comprises providing a user profile record generated in a manner transparent to the user.

34. A method according to claim 1, wherein the display format is selected responsive to at least one attribute of the user profile record, which is independent of the information content of the site.

* * * * *

Exhibit 9

#8/B
MM
4-14-04

P02/01664 A02

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	A. REFUAH, et al.
Serial Number:	09/601,385
Filed:	July 28, 2000
For:	PERSONALIZED INTERNET INTERACTION
Art Unit:	2155
Examiner:	DINH, Khanh Q

Honorable Commissioner of Patents and Trademarks
P.O. Box 1450
Alexandria, VA 22313-1450

AMENDMENT

Sir:

Further to an Office Action dated July 30, 2003, kindly amend the application as follows:

092/01664 A02

IN THE CLAIMS

Attached is a listing of claims in accordance with 37 C.F.R. §1.121 (Revised Amendment Practice.) Claims 1, 20-22, 32, 44, 49, 54, 62, 79, 87, 91, 101 and 102 have been currently amended, claims 23-31, 33, 69 and 88 have been currently cancelled, and claims 112-135 have been currently added.

1. (Currently Amended) 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 a part of a user profile;

transmitting by the user a request for data to an Internet site;

generating information content of the site to be provided responsive to the request; and

determining a display format of the generated information content, including at least one non-information-content attribute, responsive to the user profile.

~~modifying a plurality of future interactions of said user with computers on said Internet, responsive to said user profile, by modifying at least one of a presentation of information to said user or a functional response of a computer to input from user,~~

~~wherein said plurality of modified interactions comprise interactions with the site content of a plurality of unrelated sites.~~

2. (Original) A method according to claim 1, wherein said tracking comprises tracking at a computer at which said user accesses the Internet.

3. (Previously Amended) A method according to claim 1, wherein said tracking comprises tracking at a tracking computer which tracks a plurality of users.

4. (Original) A method according to claim 3, wherein said tracking computer is physically remote from said plurality of sites.

5. (Previously Amended) A method according to claim 1, wherein said analyzing comprises analyzing previously acquired tracking data.

6. (Previously Amended) A method according to claim 1, wherein said analyzing comprises analyzing of currently acquired tracking data.

7. (Previously Amended) A method according to claim 1, wherein said user profile is maintained as a virtual personality, which is a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet.

8. (Original) A method according to claim 7, wherein said virtual personality comprises a persona, which is a static aspect of a personality.

9. (Original) A method according to claim 7, wherein said virtual personality comprises a mood, which is a dynamic aspect of a personality.

10. (Original) A method according to claim 9, wherein said mood comprises a rush mood, which favors fast responses.

11. (Original) A method according to claim 8, wherein said persona comprises a meticulous persona, which favors complete responses.

B1

12. (Previously Amended) A method according claim 7, wherein said personality comprises geographical information.

13. (Previously Amended) A method according to claim 7, wherein said personality comprises demographic information.

14. (Previously Amended) A method according to claim 7, wherein said personality comprises interests and preference information.

15. (Previously Amended) A method according to claim 7, wherein said personality comprises marketing information.

16. (Previously Amended) A method according to claim 7, wherein said personality comprises identification and contact information.

092/01664 A02

17. (Previously Amended) A method according to claim 7, wherein said personality comprises relational information, which defines relations between various aspects of the personality.

18. (Previously Amended) A method according to claim 7, wherein said personality comprises reflective information, which defines how a personality changes and/or interacts with other electronic entities.

19. (Previously Amended) A method according to claim 7, wherein said user selects a particular virtual personality from a plurality of personalities to which to attribute said tracked interactions.

20. (Currently Amended) A method according to claim 1, wherein ~~said future interactions~~ generating the information content comprises~~comprise~~ searching.

B1

21. (Currently Amended) A method according to claim 1, wherein generating the information content comprises~~said future interactions comprise~~ providing the data content of a requested page of the site~~viewing presented data~~.

22. (Currently Amended) A method according to claim 21, wherein determining the display format responsive to the user profile~~modifying said interactions~~ comprises changing a layout of data.

23-31. (Cancelled)

32. (Currently Amended) A method of virtual personality interaction with an Internet, comprising:

providing a virtual personality, which is a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet~~, through which virtual personality an interaction with an Internet is meditated~~;

requesting an interaction, with said Internet, by a user with whom said virtual personality is associated;

identifying at least one prospective site for the interaction, from a plurality of unrelated sites;

automatically analyzing a content of said site, by a computer, to determine a match to said virtual personality, the analysis including determination of at least one trait of the site; and

electing to perform said interaction or modifying a performance of said interaction responsive to said analysis.

33. (Cancelled)

34. (Previously Amended) A method according to claim 32, wherein analyzing a content comprises determining an ambiance of said site.

35. (Previously Amended) A method according to claim 32, wherein analyzing comprises analyzing lexicographical characteristics of said site.

36. (Previously Amended) A method according to claim 32, wherein analyzing comprises analyzing graphical characteristics of said site.

B1

37. (Previously Amended) A method according to claim 32, wherein identifying at least one site comprises identifying a plurality of sites.

38. (Original) A method according to claim 37, wherein identifying comprises searching using an Internet search engine.

39. (Previously Amended) A method according to claim 32, wherein said virtual personality comprises a mood, which is a dynamic aspect of a personality.

40. (Previously Amended) A method according to claim 32, wherein said virtual personality comprises a persona, which is a static aspect of a personality.

41. (Previously Amended) A method according to claim 32, wherein said interaction is performed to complement said virtual personality.

42. (Previously Amended) A method according to claim 32, wherein said interaction is performed to match said virtual personality.

43. (Original) A method of Internet interaction by a single user, comprising:

selecting, from a remote location, by the user, one of a plurality of virtual personalities available for interaction with a particular site, wherein a virtual personality comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet; and

interacting with the particular site using the selected virtual personality.

44. (Currently Amended) A method of site ambiance provision, comprising:

requesting an ambiance of an Internet site, said request including an identification of the site; ~~and~~

determining an ambiance of said site, responsive to said identification; and

responding to said request with at least an indication of said ambiance.

45. (Original) A method according to claim 44, wherein determining an ambiance comprises retrieving said indication of an ambiance from a memory.

46. (Original) A method according to claim 44, wherein determining an ambiance comprises analyzing said site.

47. (Original) A method according to claim 44, wherein determining an ambiance comprises requesting an indication of said ambiance from said site.

48. (Original) A method according to claim 44, wherein determining an ambiance comprises requesting an indication of said ambiance from an ambiance server.

49. (Currently Amended) A virtual personality server, comprising:

a connection to a user, through which said user indicates a desired Internet interaction;

a connection to a WWW site, with which the user interacts, said connection adapted to connect to a plurality of unrelated WWW sites for interaction with by said user; and

a virtual personality adapter, which adapts said interaction by modifying at least one of a presentation of information from said site to said user or a functional response of said site to input from user, utilizing a virtual personality for the user, wherein a virtual personality comprises a

complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet,

wherein the virtual personality adapter modifies at least one interaction attribute based on a portion of the virtual personality which does not point out a desired piece of information.

50. (Original) A server according to claim 49, wherein said connection to a user is operable to receive a selection of a particular virtual personality by said user.

51. (Previously Amended) A server according to claim 49, wherein said server modifies said virtual personality responsive to said modified interaction.

52. (Previously Amended) A server according to claim 49, wherein said virtual personality comprises a persona, which is a static aspect of a personality.

B1

53. (Previously Amended) A server according to claim 49, wherein said virtual personality comprises a mood, which is a dynamic aspect of a personality.

54. (Currently Amended) A method of virtual personality serving, comprises:

connecting to a WWW site, to request an interaction;

determining, at said WWW site, a desired virtual personality adaptation of said interaction, wherein a virtual personality comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet; and

completing said interaction, by said WWW site, responsive to said determined virtual personality adaptation, wherein said desired adaptation comprises modifying at least one of an ambiance of ~~presentation of information from~~ said site or a functional response of said site to input.

55. (Original) A method according to claim 54, wherein determining comprises receiving an indication of a desired virtual personality from a virtual personality server.

56. (Original) A method according to claim 55, wherein said virtual personality server is located at a location remote from said WWW site and from a location at which said connection is initiated.

57. (Original) A method according to claim 55, wherein said virtual personality server is located at a location from which said connection is initiated.

58. (Original) A method according to claim 54, wherein determining comprises reading virtual personality information from a computer at a location from which said connection is initiated.

59. (Previously Amended) A method according to claim 55, wherein said virtual personality server generates a one-time virtual personality for said interaction.

60. (Previously Amended) A method according to claim 54, wherein said desired virtual personality adaptation comprises a mood-responsive adaptation, wherein a mood is a dynamic aspect of a personality.

61. (Previously Amended) A method according to claim 54, wherein said desired virtual personality adaptation comprises a persona-responsive adaptation, wherein a persona is a static aspect of a personality.

62. (Currently Amended) A method of site matching to a virtual personality, comprising:

providing a list of relevant sites. in response to a user request;

analyzing each of said sites to determine a match with said virtual personality, which is a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet; ~~and~~

grading said sites responsive to said analysis; and

automatically displaying data from a highest graded site on the list.

63. (Original) A method according to claim 62, wherein providing a list comprises executing a search on an Internet search engine to provide said list.

64. (Original) A method according to claim 62, wherein providing a list comprises retrieving a plurality of matches from a name server.

65. (Previously Amended) A method according to claim 62, wherein analyzing comprises analyzing at least one of said sites responsive to a presented ambiance.

66. (Previously Amended) A method according to claim 62, wherein analyzing comprises analyzing at least one of said sites responsive to a presented trait.

67. (Previously Amended) A method according to claim 62, wherein analyzing comprises analyzing a content of at least one of said sites.

68. (Previously Amended) A method according to claim 62, comprising displaying said graded list.

69. (Cancelled)

70. (Original) A method of advertisement personalization, comprising:

B1 determining an instantaneous virtual personality of a human interactor, wherein a virtual personality comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet, said virtual personality including a mood, which is a dynamic aspect of said personality;

selecting at least one advertisement to match said virtual personality; and

presenting said advertisement to said interactor.

71. (Original) A method according to claim 70, wherein said advertisement is presented through an Internet.

72. (Previously Amended) A method according to claim 70, wherein said virtual personality comprises a persona, which is a static aspect of a personality.

73. (Previously Amended) A method according to claim 70, comprising selecting said virtual personality from a plurality of available virtual personalities.

74. (Previously Amended) A method according to claim 70, wherein said virtual personality is provided by said interactor.

75. (Original) A method of WWW site modification, comprising:

detecting at the WWW site a desired interaction from a particular virtual personality, which personality comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet;

determining a suitable modification of at least one characteristic of said site to match said virtual personality;

responding to said desired interaction with a response indicating a match of said modified characteristic to said virtual personality; and

modifying said site in association with providing said site during said desired interaction to a user associated with said virtual personality.

76. (Original) A method according to claim 75, wherein said modification comprises modifying a display layout.

77. (Previously Amended) A method according to claim 75, wherein said modification comprises modifying a level of detail shown.

78. (Previously Amended) A method according to claim 75, wherein said modification comprises selecting data to be displayed.

79. (Currently Amended) A method of data directory display, comprising:

requesting a display of data from a data directory;

providing, in association with said request, a virtual personality for said request, which personality comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet; and

displaying said data, with at least one of a level of detail or a spatial layout selected responsive to said virtual personality.

80. (Original) A method according to claim 79, wherein said virtual personality is provided as part of said request.

81. (Previously Amended) A method according to claim 79, wherein said displaying comprises filtering.

82. (Original) A method according to claim 79, wherein said displaying comprises sorting.

83. (Original) A method according to claim 79, wherein said displaying comprises controlling a level of detail.

84. (Original) A method according to claim 79, wherein said displaying comprises controlling a spatial layout of said data.

85. (Original) A method of data directory display, comprising:

requesting a search from a search engine, using at least one keyword, which request includes a virtual personality for said request, which personality comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet;

interpreting said key-word at said search engine, utilizing said virtual personality; and

performing said search request by said search engine, utilizing said interpreted key-word.

86. (Original) A method according to claim 85, wherein said search engine comprises an Internet search engine.

87. (Currently Amended) A method of Internet search, comprising:

a user connecting to an Internet search engine;

providing the search engine with search criteria, by said user;

performing a search for WWW sites by the search engine, utilizing said search criteria, to obtain search results; ~~and~~

processing~~filtering~~ said search results utilizing personal information associated with said user, by a different computer from the search engine; and

presenting said processed~~filtered~~ search results to said user.

88. (Cancelled)

89. (Previously Amended) A method according to claim 87, wherein said personal information is provided using a virtual personality, which comprises a complex of characteristics that distinguishes an electronic person, for the purpose of interacting with an Internet.

90. (Original) A method according to claim 87, wherein said personal information is provided as a non-keyword input to said search engine.

91. (Currently Amended) A method of interacting with a computer, comprising:
providing a software application having a user interface on said computer;
providing an electronic representation of at least part of a user's desired personality; and
said software modifying its interaction with said user, responsive to said representation of said personality.

92. (Original) A method according to claim 91, wherein said software comprises an Internet Browser.

93. (Previously Amended) A method according to claim 91, wherein said software modifies a visual display of said interface.

94. (Previously Amended) A method according to claim 91, wherein said software modifies a behavior of said interface.

95. (Previously Amended) A method according to claim 91, wherein said software modifies a menu length of said interface.

96. (Previously Amended) A method according to claim 91, wherein said software modifies a help level of said software.

97. (Previously Amended) A method according to claim 91, wherein said software modifies a level of detail presented by said software.

98. (Previously Amended) A method according to claim 91, wherein said software modifies a display format of said software.

99. (Previously Amended) A method according to claim 91, wherein said software modifies an image quality of said software.

100. (Previously Amended) A method according to claim 91, wherein said software modifies a response time of said software.

101. (Currently Amended) A method of utilizing an electronic representation of a user's desired personality, comprising:

providing~~storing~~ said representation on a portable computer-readable storage media;

reading the representation by a terminal; and

interacting with a computer using said representation read by the terminal, wherein said representation mediates the interaction.

102. (Currently Amended) A method according to claim 101, wherein said computer is~~comprises~~ connected to the terminal ~~a remote computer connected to an~~ through the Internet.

103. (Previously Amended) A method according to claim 101, wherein said computer comprises a controller of an automated store.

104. (Original) A method according to claim 103, wherein said mediation comprises varying a range of offered selection of products.

105. (Previously Amended) A method according to claim 101, wherein said media comprises a diskette.

106. (Previously Amended) A method according to claim 101, wherein said media comprises a smart card.

107. (Previously Amended) A method according to claim 101, wherein said media comprises printed optically readable codes.

108. (Previously Amended) A method according to claim 91, wherein said representation is generated by tracking a plurality of interactions of said user with an Internet.

109. (Previously Amended) A method according to claim 91, wherein said representation comprises a representation of a persona, which is a static aspect of a personality.

B1

110. (Previously Amended) A method according to claim 91, wherein said representation comprises a representation of a mood, which is a dynamic aspect of a personality.

111. (Previously Amended) A method according to claim 91, wherein said desired personality comprises a true personality of said user.

---

112. (New) A method of virtual personality interaction with an Internet, comprising:
tracking interactions of the user with an Internet;
analyzing said tracked interactions to determine at least a part of a user profile;
requesting an interaction, with the Internet;
selecting a site to access responsive to the request and the user profile; and
providing the user with content of the selected site responsive to the request.

B2

113. (New) A method according to claim 112, wherein selecting the site to access comprises consulting a name server.

114. (New) A method according to claim 1, wherein the at least one attribute of the display format determined responsive to the user profile comprises one or more display colors.

115. (New) A method according to claim 1, wherein the at least one attribute of the display format determined responsive to the user profile comprises an image quality.

116. (New) A method according to claim 1, wherein the at least one attribute of the display format determined responsive to the user profile comprises a size, amount or density of displayed text.

117. (New) A method according to claim 1, wherein determining the display format responsive to the user profile comprises modifying a display format of a browser of the client.

118. (New) A method according to claim 117, wherein modifying the display format of the browser comprises modifying at least one of a help detail level, a menu length or a dialog box format.

119. (New) A method according to claim 32, wherein the plurality of unrelated sites are received from a name server.

120. (New) A method according to claim 32, wherein the at least one trait comprises an average word length.

121. (New) A method according to claim 32, wherein the at least one trait comprises a number or percentage of non-textual objects.

122. (New) A method according to claim 44, wherein determining the ambiance comprises determining the type of words used by the site.

123. (New) A method according to claim 44, wherein determining the ambiance comprises determining a ratio between images and text in the site.

124. (New) A method according to claim 44, wherein determining the ambiance comprises determining the colors used by the site.

125. (New) A method according to claim 44, wherein determining the ambiance comprises determining a number of links, images or multimedia files of the site.

126. (New) A method according to claim 87, wherein processing the search results comprises filtering the search results.

127. (New) A method according to claim 87, wherein processing the search results comprises ordering the search results.

128. (New) A method of Internet interaction, comprising:

requesting by a user a response from a web site;

generating a response to the request, by the web site;

providing the web site response to a persona server, which determines modifications to the generated response, responsive to personal information associated with the user; and

presenting the generated response, to the user, modified according to the determination of the persona server.

129. (New) A method according to claim 128, wherein the response of the web site comprises search results from a search engine.

130. (New) A method according to claim 101, wherein the terminal is implemented by the computer.

131. (New) A method of interacting with an Internet server, comprising:

providing a user profile of a user's desired personality on a user terminal, through which a user interacts with the Internet;

transmitting the user profile from the terminal to an Internet server, along with a request for Internet site data; and

providing the user terminal with the requested Internet site data, modified responsive to the user profile.

132. (New) A method according to claim 131, wherein providing the user profile comprises reading the user profile from a portable computer-readable storage media.

133. (New) A method according to claim 131, wherein the user profile comprises a representation of a persona, which is a static aspect of a personality.

134. (New) A method according to claim 91, wherein said software modifies its functionality responsive to said representation of said personality.

092/01664 A02

B2

135. (New) A method according to claim 1, wherein the at least one attribute comprises a visual attribute.

REMARKS

The present application includes claims 1-22, 32, 34-68, 70-87 and 89-135. Claims 112-135 are new. Claims 1, 20-22, 32, 44, 49, 54, 62, 79, 87, 91, 101 and 102 were amended. The amendments to claims 20, 21 and 22 are in order to conform to amended claim 1. Claims 44 and 91 were amended for clarity, without affecting their scope. Claim 44 was amended to remove an extra 'and' and to explicitly state what was implicit in the claim that 'a site' is 'an Internet site'. In claim 91 the word 'comprising' was added. The amendment to claim 101 finds support, for example, on page 30, lines 12-18. The amendment to claim 102 is in order to conform to amended claim 101.

The amendments to claims 1 and 49 find support at least in the examples of original claims 76, 83 and 84. The amendment to claim 32 finds support at least in original claim 33. The amendment to claim 49 finds support at least on page 22, lines 31-33. Claim 62 was amended to clarify that the list is in response to a user request. This amendment finds support at least in the example on page 3, lines 22-30. Claim 62 was also amended to incorporate the requirement of claim 69 that finds support thereon. The amendment to claim 79 finds support at least in original claims 83 and 84. Claim 87 was amended to include the limitation of claim 88 and to broaden the term 'filtering' to 'processing' based on the description on page 6, lines 2-9.

Claims 1-22 stand rejected under 35 USC §102(e) as being anticipated by Herz (US patent 6,029,195). Claim 1 was amended to require determining a display format including at least one non-information-attribute responsive to the user profile.

Herz describes identifying desirable objects, such as news articles, for a user, according to a target profile interest summary (abstract). The Herz application is directed solely to finding the information and not to determining a non-information attribute of a display format of the information responsive to the user profile, as required by amended claim 1. The dependent claims are allowable at least because they depend on an allowable claim. At least some of the dependent claims, however, add to the patentability of claim 1. New claim 114, for example, requires determining one or more display colors. New claim 115, for example, requires determining an image quality responsive to the user profile. New claim 116, for example, requires determining a size or quality of displayed text. Claim 22, for example, requires changing a layout of data.

Claims 32 and 34-42 stand rejected under 35 USC §102(e) as being anticipated by Herz (US patent 6,029,195). Claim 32 was amended to incorporate the requirement of claim 33 that the analysis includes determining at least one trait of the site.

Herz describes identifying WWW pages of interest based on their content (Col. 67, lines 30-35). Herz does not teach or suggest determining a trait of a site. In the office action, no citations regarding old claims 33-36 appear, possibly due to a formatting error of the office action. Absent specific references it is very hard to assure that a requirement does not appear anywhere in the 98 columns of Herz. Nonetheless, in a review of the entire text of Herz, applicants did not find any description of an analysis that includes determining a trait of a site and electing to perform an interaction or modifying a performance of the interaction responsive to the analysis.

Claim 43 stands rejected under 35 USC §102(e) as being anticipated by Herz (US patent 6,029,195). Applicants respectfully traverse the rejection. Claim 43 requires selecting, by the user, one of a plurality of virtual personalities available for interaction with a particular site. Herz does not teach or suggest a single user having a plurality of virtual personalities for a single site. Rather, Herz suggests having different virtual personalities for different sites (col. 43, lines 28-42). Furthermore, Herz states plainly that a user U requires a new pseudonym for use with a new and disjoint coalition (col. 44, lines 40-42) and does not teach or suggest any other use for additional pseudonyms. Column 39, lines 31-43, referred to by the Examiner, suggests different pseudonyms for different types of transactions from different service providers.

Claims 44-48 stand rejected under 35 USC §102(e) as being anticipated by Herz (US patent 6,029,195). Applicants respectfully disagree. Claim 44 was amended for clarity, by removing an extra "and" and clarifying that the site refers to an Internet site, which was inherent in the claim as filed.

Claim 44 requires requesting an ambiance of a site. Herz relates to finding objects most likely to be of interest to a user (abstract) and does not teach or suggest determining the ambiance of objects. A discussion of site ambiance appears in the present application on page 9, line 26 – page 10, line 14 and the description of Fig. 3 on pages 26-28.

Applicants do not understand the relevance of the passages referenced by the Examiner with regard to claim 44. Fig. 2 shows no more than a computer network. Column 37, lines 45-50 of Herz relate to using the WWW for content channels and do not mention or hint to determining an ambiance of a site, let alone to generating a request for an ambiance of a site. Columns 63 and 64 of Herz relate to search profiles and not to ambiance information.

The dependent claims add further patentability over Herz. Claim 45, for example, requires retrieving the requested ambiance indication from a memory. Claim 46, for example, requires determining the ambiance by analyzing the site. Neither of these requirements are taught or suggested by Herz.

New claims 120-123 which depend on claim 44 recite different specific attributes determined in determining the ambience. None of these attributes are taught or suggested by Herz.

Claims 49-53 stand rejected under 35 USC §102(e) as being anticipated by Herz (US patent 6,029,195). Claim 49 was amended to require that the virtual personality adapter modifies at least one interaction attribute based on a portion of the virtual personality which does not point out a desired piece of information. Herz is directed solely to determining information desired by the user (Abstract). Herz does not teach or suggest using a portion of the virtual personality which does not point out a desired piece of information in modifying attributes of the interaction.

Claims 54-61 stand rejected under 35 USC §102(e) as being anticipated by Herz (US patent 6,029,195). Applicants amended claim 54 to require that the desired adaptation comprises modifying an ambiance of the site or a functional response of the site.

Herz does not teach or suggest modifying an ambiance of a site or a functional response of the site.

Claims 62-68 stand rejected under 35 USC §102(e) as being anticipated by Herz (US patent 6,029,195). Applicants amended claim 62 to include the requirement of claim 69 that requires automatically displaying a highest graded site on the list. Herz, in contrast, relates to providing lists of highest matching world wide web pages (column 67, lines 34-35) using filtering methods and does not teach or suggest automatically displaying a single highest graded site, rather than a list of sites.

The dependent claims add further patentability over Herz. Claim 65, for example, requires analyzing a site responsive to a presented ambiance. As discussed above, Herz does not teach or suggest relating to an ambiance.

New independent claim 112 requires selecting a site to access responsive to a request and a user profile and providing the user with content of the selected site responsive to an interaction request. As discussed above, Herz describes providing information on available web pages and does not teach or suggest providing a specific web page responsive to an interaction request.

Claims 70-74 stand rejected under 35 USC §102(e) as being anticipated by Herz (US patent 6,029,195). Applicants respectfully traverse the rejection. Claim 70 requires determining an instantaneous virtual personality including a mood and selecting an advertisement accordingly. Herz describes matching advertisements to a user profile (column 73, lines 10-43) but does not teach matching advertisements to an instantaneous personality including a mood.

Dependent claim 72 adds further patentability over Herz in requiring that the virtual personality include also a static aspect in addition to the mood.

Applicants do not understand the relevance of the passages cited by the Examiner. These passages do not even relate to advertisements or to dynamic aspects of a user profile. Therefore, the Examiner did not establish a *prima facie* case of anticipation with regard to claim 70.

Claims 75-78 stand rejected under 35 USC §102(e) as being anticipated by Herz (US patent 6,029,195). Applicants respectfully traverse the rejection. Claim 75 requires detecting at a WWW site a desired interaction from a particular virtual personality and determining a suitable modification to match the virtual personality. Herz describes identifying WWW sites of interest that match a user profile (column 67, lines 34-35), but does not teach or suggest modifying a site to match a personality. Herz is directed at filtering data that is of interest to a user and not to modifying WWW sites to user preferences.

The dependent claims further differentiate over Herz. Claim 76 requires modifying a display layout and claim 77 requires modifying a level of detail shown. Neither of these is taught or suggested by Herz.

The rejection of the Examiner does not explain the relevance of the long passages referred to and applicants did not find the relevance of these passages.

Claims 79-84 stand rejected under 35 USC §102(e) as being anticipated by Herz. Applicants amended claim 79 to include the requirements of claims 83 and 84 in the alternative, i.e., to require that the data is displayed with a level of detail or spatial layout selected responsive to the virtual personality. As discussed above, Herz relates to finding data and not to the display format of the data. Applicants did not find controlling the level of detail or the spatial layout in any of the passages cited by the Examiner.

Claims 85-86 stand rejected under 35 USC §102(e) as being anticipated by Herz (US patent 6,029,195). Applicants respectfully traverse the rejection. Claim 85 requires requesting a search using at least one keyword and interpreting the keyword utilizing a virtual personality. Herz does not teach or suggest interpreting words according to a virtual personality and does not relate to interpreting keywords at all.

Claims 87-90 stand rejected under 35 USC §102(e) as being anticipated by Herz. Applicants amended claim 87 to include the limitation of claim 88. Claim 87 as amended requires processing the search results utilizing personal information by a different computer from the search engine. Herz suggests a post-process that filters and ranks the many target objects found by a conventional search (col. 67, lines 54-59). Herz, however, does not teach or suggest performing the post process on a different computer from the search engine. Putting the post-process on a

separate computer would not be obvious since it would be counter productive, requiring placement of the target objects on the separate computer.

New independent claim 128 requires providing a web site response to a persona server which determines modifications to the generated response responsive to personal information. Herz does not teach or suggest such a persona server.

Claims 91-100 stand rejected under 35 USC §102(e) as being anticipated by Herz (US patent 6,029,195). Applicants respectfully traverse the rejection. Claim 91 was amended to add the term comprising which was inadvertently left out. The amendment does not affect the patentability of the claim.

Claim 91 requires a software having a user interface that modifies its interaction with a user, responsive to a personality representation. Such software that modifies its interaction is not taught or suggested by Herz. In fact, Herz teaches away from modifying a software having a user interface, such as a browser, responsive to a personality presentation. Herz relates to a privacy problem (column 38) which does not exist if the modifications are in the software used for interfacing with a user. Column 63, line 30, referred to by the Examiner, describes pacing the user profile on a server which does not include a user interface.

The dependent claims add further patentability over Herz. Claim 92, for example, requires modifying an Intrernet browser. Claim 93, for example, requires that the software modifies a visual display of the interface. Claim 95, for example, requires modifying a menu length of the interface. Claim 96, for example, requires modifying a help level. None of these requirements is taught or suggested by Herz.

Claims 101-111 stand rejected under 35 USC §102(e) as being anticipated by Herz (US patent 6,029,195). Applicants amended claim 101 to clarify that the storage media is a portable media read by a terminal. In addition, the term "storing" was replaced by the more general term "providing". Herz does not teach or suggest providing the user profile on a portable storage media. In Herz, the user profile is stored at a server (column 63, line 30).

The only thing that Herz suggests storing on a portable media is user identity information, equivalent to passwords, which allow the user to enter the system (column 34, line 64 and column 37, lines 57-62).

New claim 131 requires transmitting the user profile from a terminal to an Internet server, along with a request for Internet site data. In contrast, Herz suggests storing the user profile at the network vendor (column 34, lines 13-15) or on proxy servers (col. 34, lines 21-32). Herz does not teach or suggest storing the user profile on the user terminal.

092/01664 A02

In view of the above remarks and amendments, the present application is believed to be allowable. Allowance of the application is respectfully awaited.

Respectfully submitted,
A. REFUAH et al.

Yaakov Schatz

Yaakov Schatz
Reg. No. 44,320

December 30, 2003
William H. Dippert, Esq.
Reed Smith LLP
599 Lexington Avenue, 29th Floor
New York, NY 10022-7650

Tel: (212) 521-5400

2155

COMBINED AMENDMENT & PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a) (Small Entity)	Docket No. 092/01664

In Re Application Of: Aviv REFUAH et al.

Serial No.	Filing Date	Examiner	Group Art Unit
09/601385	July 28, 2000	DINH, Khanh Q.	2155

Invention: PERSONALIZED INTERNET INTERACTION

TO THE COMMISSIONER FOR PATENTS:

This is a combined amendment and petition under the provisions of 37 CFR 1.136(a) to extend the period for filing a response to the Office Action of July 30, 2003 (Date) in the above-identified application.

The requested extension is as follows (check time period desired):

☐ One month ☒ Two months ☐ Three months ☐ Four months ☐ Five months

from: October 30, 2003 (Date) until: December 30, 2003 (Date)

A verified statement of small entity status as a small entity under 37 CFR 1.27:

☐ is enclosed.
☒ has already been filed in this application.

The fee for the amendment and extension of time has been calculated as shown below:

CLAIMS AS AMENDED

	CLAIMS REMAINING AFTER AMENDMENT		HIGHEST # PREV. PAID FOR		NUMBER EXTRA CLAIMS PRESENT	RATE	ADDITIONAL FEE
TOTAL CLAIMS	123	-	111	=	12	x $9.00	$108.00
INDEP. CLAIMS	17	-	16	=	1	x $43.00	$43.00
						FEE FOR AMENDMENT	$151.00
						FEE FOR EXTENSION OF TIME	$210.00
						TOTAL FEE FOR AMENDMENT AND EXTENSION OF TIME	**$361.00**

01/14/2004 SNAJARRO 00000092 033419 09601385
01 FC:2615 108.00 DA
02 FC:2614 43.00 DA

01/14/2004 SNAJARRO 00000092 033419 09601385
03 FC:2252 210.00 DA

P28SMALL/REV03

COMBINED AMENDMENT & PETITION FOR EXTENSION OF TIME UNDER 37 CFR 1.136(a) (Small Entity)	Docket No. 092/01664

The fee for the amendment and extension of time is to be paid as follows:

☐ A check in the amount of **$0.00** for the amendment and extension of time is enclosed.

☒ Please charge Deposit Account No. **03-3419** in the amount of **$361.00**

☒ The Director is hereby authorized to charge payment of the following fees associated with this communication or credit any overpayment to Deposit Account No. **03-3419**

- ☒ Any additional filing fees required under 37 C.F.R. 1.16.
- ☒ Any patent application processing fees under 37 CFR 1.17.

☒ If an additional extension of time is required, please consider this a petition therefor and charge any additional fees which may be required to Deposit Account No. **03-3419**

Paul Fenster
Signature

Dated: **December 30, 2003**

Paul FENSTER, Reg. No. 33,877

William H. Dippert, Esq.
Reed Smith LLP
599 Lexington Avenue, 29th Floor
New York, NY 10022-7650

Tel: (212) 521-5400

cc:

I certify that this document and fee is being deposited on December 30, 2003 with the U.S. Postal Service as first class mail under 37 C.F.R. 1.8 and is addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.

William H. Dippert
Signature of Person Mailing Correspondence

WILLIAM H. DIPPERT
Typed or Printed Name of Person Mailing Correspondence

P28SMALL/REV03

Exhibit 15

REDACTED

IN ITS

ENTIRETY

Exhibit 16

REDACTED

IN ITS

ENTIRETY

Exhibit 17

REDACTED

IN ITS

ENTIRETY

Exhibit 18

REDACTED

IN ITS

ENTIRETY

Exhibit 26

Beyond the Commons: Investigating the Value of Personalizing Web Search

Jaime Teevan[1], Susan T. Dumais[2], and Eric Horvitz[2]

[1] MIT, CSAIL, 32 Vassar St., G472,
Cambridge, MA, 02139 USA
teevan@csail.mit.edu
http://www.csail.mit.edu/~teevan
[2] Microsoft Research, One Redmond Way,
Redmond, WA, 94114, USA
{sdumais,horvitz}@microsoft.com

Abstract. We investigate the diverse goals people have when they issue the same query to a Web search engine, and the ability of current search tools to address such diversity, in order to understand the potential value of personalizing search results. Great variance was found in the results different individuals rated as relevant for the same query—even when those users expressed their underlying informational goal in the same way. The analysis suggests that, while current Web search tools do a good job of retrieving results to satisfy the range of intentions people may associate with a query, they do not do a very good job of discerning an individual's unique search goal. We discuss the implications of this study on the design of search systems and suggest areas for additional research.

1 Introduction

Traditional search engines are designed to return a set of documents that match a query. Studies of search engine quality have tended to be based on the ability of search engines to return the set of results that its users want as a population, as opposed to the results that match each individual's unique search goal. For example, at the DARPA Text REtrieval Conference (TREC), relevant documents to a particular query are identified by an expert judge, based on a detailed description of an information need. Ideally the description is explicit enough and the rater skilled enough that the documents selected as relevant are the same ones that another rater would consider relevant.

However, Web search behavior suggests that providing results to an unambiguous query might not be the most appropriate design target for a search engine. Web queries are very short, and it is unlikely that a two- or three-word query can unambiguously describe a user's informational goal. What one person considers relevant to a query like "jaguar" is not necessarily the same as what someone else considers relevant to the same query. Even a seemingly precise query like "PIA 2005" returns Web pages about the Personal Information Access workshop, the Parachute Industry Asso-

ciation, Professional Insurance Agents, the Pacific Institute of Aromatherapy, etc. Further, if Web searchers are not skilled at stating their goal, even longer descriptions may not reliably disambiguate intent.

We report on a study of the ability of current Web search engines to provide relevant documents to users, in order to understand how future search tools can be built to best meet the needs of their users. Understanding relevance is a complex problem [11, 13], and we address only a small portion of it in our work. Our analysis is aimed at assessing the relationship between the rank of a search result as returned by a Web search engine and the individual's perceived relevancy of the result. We find a considerable mismatch due to a variation in the informational goals of users issuing similar queries. The study suggests personalization of results via re-ranking would provide significant benefit for users. We conclude with a discussion of how the results of this study should triage future research.

2 Methods

We conducted a study in which 15 participants evaluated the top 50 Web search results for approximately 10 queries of their choosing. Participants were employees of a large corporation. Their job functions included administrators, program managers, software engineers and researchers. All were computer literate and familiar with Web search.

Web search results were collected from a "Top Choice" search engine, as listed by Search Engine Watch. For each search result, the participant was asked to determine whether they personally found the result *highly relevant*, *relevant*, or *irrelevant*. So as not to bias the participants, the results were presented in a random order.

The queries evaluated were selected in two different manners, at the participants' discretion. In one approach (*self-selected queries*), users were asked to choose a query to mimic a recently performed search, based on a diary of searches they were asked to keep during the day. Thus, we believe that the self-selected queries closely mirrored the searches that the participants conducted in the real world.

In another approach (*pre-selected queries*), users were asked to select a query from a list of queries that were formulated to be of general interest (e.g., *cancer, Bush, Web search*). Although users did not generate these queries themselves, they were free to choose the pre-selected queries they found most interesting, and thus presumably only chose queries that had some meaning to them. By using pre-selected queries, we were able to explore the consistency with which different individuals evaluated the same results. Such data would have been difficult to collect using only self-selected queries, as it would have required us to wait until different participants coincidentally issued the same query on their own. We validate the conclusions drawn from pre-selected queries with data from the self-selected queries.

For both the self-selected queries and the pre-selected queries, participants were asked to write a more detailed description of the informational goal or *intent* they had in mind when they issued the query. Because the pre-selected queries were given to the user, the user had to create some intent for these queries. However, by allowing

them to decide whether or not they wanted to evaluate a particular query, we sought to provide them with a query and associated results that would have some meaning for them.

We collected a total of 137 queries. Of those, 53 were pre-selected queries and 85 were self-selected. The number of users evaluating the same set of results for the pre-selected query ranged from two to nine. Thus we had evaluations by different people for the same queries drawn from the pre-selected set of queries, as well as a number of evaluations for the searches that users had defined themselves.

3 Rank and Rating

We used the data we collected to study how the results that the Web search engine returned matched our participants' search goals. We expected them to match relatively closely, as current search engines seem to be doing well, and in recent years satisfaction with result quality has climbed.

Fig. 1 shows the average result's relevancy score as a function of rank. To compute the relevancy score, the rating *irrelevant* was given a score of 0, *relevant* a score of 1, and *highly relevant* a score of 2. Values were averaged across all queries and all users. Separate curves are shown for the pre-selected (solid line) and self-selected (dashed line) queries. Clearly there is some relationship between rank and relevance. Both curves show higher than average relevance for results ranked at the top of the result list. The correlation between rank and relevance is -0.66. This correlation coefficient is significantly different from 0 ($t(48) = 6.10$, $p < 0.01$). However, the slope of the curves flattens out with increasing rank. When considering only ranks 21-50, the correlation coefficient is -0.07, which is not significantly different from 0. Importantly, there are still many relevant results at ranks 11-50, well beyond what users typically see. This suggests the search result ordering could be improved.

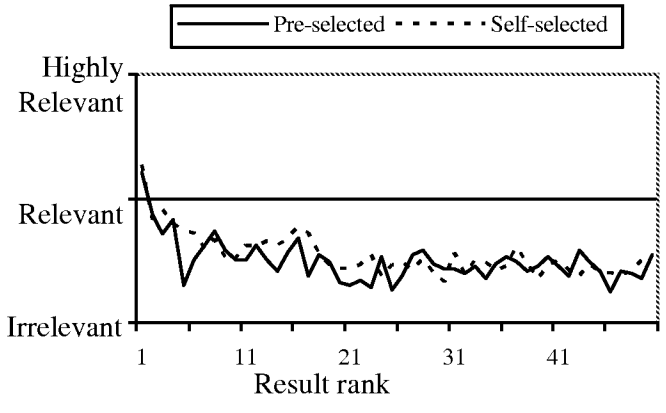


Fig. 1. Average ratings for Web search engine results as a function of rank. There are many relevant results that do not rank in the top ten

The general pattern of results seen in Fig. 1 is not unique to our sample of users or queries. A reanalysis of data from the TREC Web track [4] shows a similar pattern. In the TREC-9 Web track, the top 100 results from 50 Web queries were rated using a similar three-valued scale, *highly relevant*, *relevant* and *not relevant*. Results for one top-performing search systems, uwmt9w10g3, yielded an overall correlation between rank and relevance of -0.81, which drops off substantially to -0.30 for positions 21-50.

4 Same Query, Different Intents

Our analysis shows that rank and rating were not perfectly correlated. While Web search engines do a good job of ranking results to maximize their users' global happiness, they do not do a very good job for specific individuals. If everyone rated the same currently low-ranked documents as highly relevant, effort should be invested in improving the search engine's algorithm to rank those results more highly, thus making everyone happier. However, despite the many commonalities among our participants (*e.g.*, all were employees of the same company, lived in the same area, and had similar computer literacy), our study demonstrated a great deal of variation in their rating of results.

As will be discussed in the following sections, we found that people rated the same results differently because they had different information goals or intentions associated with the same queries. This was evidenced by the variation in the explicit intents our participants wrote for their queries. Even when the intents they wrote were very similar, we observed variation in ratings, suggesting that the participants did not describe their intent to the level of detail required to distinguish their different goals.

4.1 Individuals Rate the Same Results Differently

Participants did not rate the same documents as relevant. The average inter-rater agreement for queries evaluated by more than one participant evaluated was 56%. This disparity in ratings stands in contrast to previous work. Although numbers can't be directly compared, due to variation in the number of possible ratings and the size of the result set evaluated, inter-rater agreement appears to be substantially higher for TREC (*e.g.*, greater than 94% [8]) and previous studies of the Web (*e.g.*, 85% [3]). The differences we observed are likely based in our focus on understanding personal intentions; instead of instructing our participants to select what they thought was "relevant to the query," we asked them to select the results they would want to see personally.

The ratings for some queries agreed more than others, suggesting some queries might be less ambiguous to our population than others. Similarly, some participants gave ratings that were similar to other participants' ratings. It might be possible to cluster individuals, but even the most highly correlated individuals showed significant differences.

4.2 Same Intent, Different Evaluations

We found that our participants sometimes used the same query to mean very different things. For example, the explicit intents we observed for the query *cancer* ranged from "information about cancer treatments" to "information about the astronomical/astrological sign of cancer". This was evident both for the pre-selected, where the user had to come up with an intent based on the query, and self-selected queries, where the query was generated to describe the intent. Although we did not observe any duplicate self-selected queries, many self-selected queries, like "rice" (described as "information about rice university"), and "rancho seco date" (described as "date rancho seco power plant was opened") were clearly ambiguous.

Interestingly, even when our participants expressed the same intent for the same query, they often rated the query results very differently. For example, for the query *Microsoft*, three participants expressed these similar intents:

- "information about microsoft, the company"
- "Things related to the Microsoft corporation"
- "Information on Microsoft Corp"

Despite the similarity of their intent, only one URL (www.microsoft.com) was given the same rating by all three individuals. Thirty-one of the 50 results were rated *relevant* or *highly relevant* by one of these three people, and for only six of those 31 did more than one rating agree. The average inter-rater agreement among these three users with similar intentions was 62%.

This disparity in rating likely arises because of ambiguity; the detailed intents people wrote were not very descriptive. Searches for a simple query term were often elaborated as "information on *query term*" ("UW" → "information about UW", leaving open whether they meant the University of Washington or the University of Wisconsin, or something else entirely). It appears our participants had difficulty stating their intent, not only for the pre-selected queries, where we expected they might have some difficulty creating an intent (mitigated by the fact that they only rated pre-selected queries by choice), but also for the self-selected queries.

Although explicit intents generally did not fully explain the query term, they did provide some additional information. For example, "trailblazer" was expanded to "Information about the Chevrolet TrailBlazer", clarifying the participant was interested in the car, as opposed to, for example, the basketball team. Further study is necessary to determine why people did not include this additional information in their original query, but it does suggest that they could perhaps be encouraged to provide more information about their target when searching. However, even if they did this, they would probably still not be able to construct queries that expressed exactly what wanted. For example, the Trailblazer example above did not clarify exactly what kind of information (*e.g.*, pricing or safety ratings) was sought. This suggests searchers either need help communicating their intent or that search systems should try to infer it.

5 Search Engines are for the Masses

The previous sections showed that our participants ranked things very differently, in ways that did not correspond closely with the Web search engine ranking. We now describe analyses that show that the Web ranking did a better job of satisfying all of our participants than any individual.

5.1 Web Ranking the Best for the Group

In this section, we investigate the best possible ranking we could construct based on the relevance assessments we collected, and compare this ideal ranking with the original Web ranking. For scoring the quality of a ranking, we use *Discounted Cumulative Gain* (DCG), a measure of the quality of a ranked list of results commonly used in information retrieval research [5]. DCG measures the result set quality by counting the number of relevant results returned. It incorporates the idea that highly-ranked documents are worth more than lower-ranked documents by weighting the value of a document's occurrence in the list inversely proportional to its rank (i). DCG also allows us to incorporate the notion of two relevance levels by giving *highly relevant* documents a different gain value than *relevant* documents.

$$\mathrm{DCG(i)} = \begin{cases} \mathrm{G}(1) & \text{if } i = 1, \\ \mathrm{DCG}(i-1) + \mathrm{G}(i)/\log(i) & \text{otherwise.} \end{cases} \quad (1)$$

For *relevant* results, we used G(i) = 1, and for *highly relevant* results, G(i)=2, reflecting their relative importance.

The best possible ranking for a query given the data we collected is the ranking with the highest DCG. For queries where only one participant evaluated the results, this means ranking *highly relevant* documents first, *relevant* documents next, and *irrelevant* documents last. When there are more than one set of ratings for a result list, the best ranking ranks first those results that have the highest collective gain.

We compared how close the best possible rankings were to the rankings the search engine returned. To measure "closeness," we computed the Kendall-Tau distance for partially ordered lists [1]. The Kendall-Tau distance counts the number of pair-wise disagreements between two lists, and normalizes by the maximum possible disagreements. When the Kendall-Tau distance is 0, the two lists are exactly the same, and when it is 1, they are in reverse order. Two random lists have, on average, a distance of 0.5.

We found that for eight of the ten queries where multiple people evaluated the same result set, the Web ranking was more similar to best possible ranking for the group than it was, on average, to the best possible ranking for each individual. The average individual's best ranking was slightly closer to the Web ranking than random (0.5), with a distance of 0.469. The average group ranking was significantly closer ($t(9) = 2.14$, $p < 0.05$) to the Web ranking, with a distance of 0.440. The Web rankings seem to satisfy the group better than they do the individual.

5.2 Gains of Personalization via Re-ranking

Again taking DCG as an approximation of user satisfaction, we found a sizeable difference between our participants' satisfaction when given exactly what they wanted rather than the best group ranking for that query. On average, the best group ranking yielded a 23% improvement in DCG over what the current Web ranking, while the best individual ranking led to a 38% improvement.

The graph depicted in Fig. 2 shows the average DCG for group (dashed line) or personalized (solid line) rankings. These data were derived from the five pre-selected queries for which we collected six or more individual evaluations of the results, although the pattern held for other sets of queries. To compute the values shown, for each query we first randomly selected one person and found the DCG for that individual's best ranking. We then continued to add the additional people, at each step recomputing the DCG for each individual's best rankings and for the best group ranking. As can be seen in Fig. 2, as additional people were added to the analysis, the gap between user satisfaction with the individualized rankings and the group ranking grew. Our sample is small, and it is likely that the best group ranking for a larger sample of users would result in even lower DCG values.

These analyses underscore the promise of providing users with better search result quality by personalizing results. Improving core search algorithms has been difficult, with research leading typically to very small improvements. We have learned that, rather than improving the results to a particular query, we can obtain significant boosts by working to improve results to match the intentions behind it.

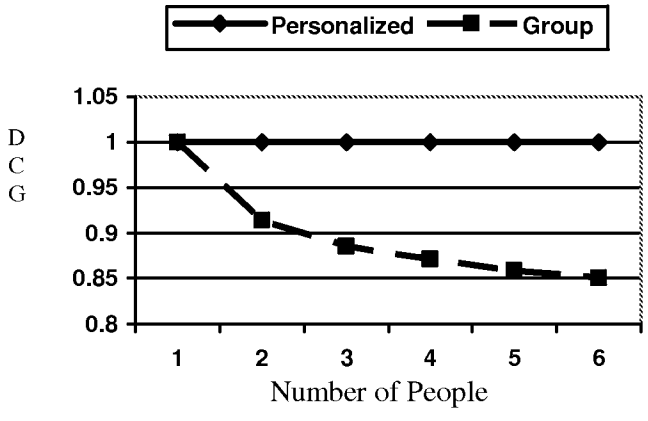


Fig. 2. As more people are taken into account, the average DCG for each individual drops for the ideal group ranking, but remains constant for the ideal personalized ranking

6 Directions in Personalized Search

We believe that Web search tools could be enhanced significantly by considering the variation in relevancy of results for users. We shall now touch on several directions for doing such personalization suggested by the above analysis.

We observed that our participants rated the results to the same queries differently because they had different intents. One solution to ambiguity is to aid users in better specifying their interests and intents. As an example, Google Personal [4] asks users to build a profile of themselves by specifying their interests. Other search systems have tried to help users better express their informational goals through techniques such as relevance feedback or query expansion. While it appears people can learn to use these techniques [2, 8], in practice, on the Web they do not appear to improve overall success [2, 3], and such features have been found to be used rarely. We agree with Nielsen [10], who cites the importance of not putting extra work on the users for personalization. Also, even with additional work, it is not clear that users can be sufficiently expressive. Participants in our study had trouble fully expressing their intent even when asked explicitly to elaborate on their query. In related work, people were found to prefer long search paths to expending the effort to fully specify their query [11].

We believe that another promising approach to personalizing search is to infer users' information goals automatically. Kelly and Teevan [7] give an overview of research done in information retrieval on how implicit measures can be used to help search, highlighting prior contributions focused on helping to improve results for individuals, versus for the general population. In a related paper [15], we describe a search personalization prototype that we have developed which builds on the lessons learned from the study described in this paper. The prototype, named *PS*, uses a person's prior interactions with a wide variety of content to personalize that person's current Web search in an automated manner.

Our study suggests that the results returned by Web search engines represent a range of intentions that people associate with queries. Thus, we believe that personalized search systems could take current Web search results as a starting point for user-centric refinement via re-ranking (e.g., [9, 15]). The original ranking of results by a Web search engine is a useful source of information for a more personalized ranking, and, as we discovered, the first several results are particularly likely to be relevant.

We found that not all queries should be handled in the same manner. For example, we observed that some queries appeared less ambiguous than others and showed less variation among individuals. For such queries, the group ranking (i.e., the current Web search ranking) might be sufficient. A search system that allows users to control how much personalization they receive would improve search relevance while following Neilson's [10] suggestion that users be given control of their content instead of having personalization imposed on them.

7 Conclusion

We have found that there is promise in building tools that perform personalization via re-ranking the results currently provided by current search engines. We have not discussed specific methods to automatically identify users' intentions. Instead we have worked to characterize the range of informational goals associated with queries, and investigated the potential value that can be seen by users via methods that re-rank the list of results provided by search engines.

References

1. Adler, L. M.: A modification of Kendall's tau for the case of arbitrary ties in both rankings. Journal of the American Statistical Society, Vol. 52 (1957) 33–35
2. Anick, P.: Using terminological feedback for Web search refinement: a log-based study. In Proceedings of WWW '04 (2004) 88–95
3. Eastman, C. M. and Jansen, B. J.: Coverage, relevance and ranking: The impact of query operators on Web search engine results. TOIS, Vol. 21(4) (2003) 383–411
4. Google Personal: http://labs.google.com/personalized
5. Hawking, D.: Overview of the TREC-9 Web track. In Proceedings of TREC '00 (2000) 87–102
6. Järvelin, K. and Kekäläinen, J.: IR evaluation methods for retrieving highly relevant documents. In Proceedings of SIGIR '00 (2000) 41–48
7. Jeh, G. and Widom, J.: Scaling personalized Web search. In Proceedings of WWW '03 (2003) 271–279
8. Kelly, D. and Teevan, J.: Implicit feedback for inferring user preference: A bibliography. SIGIR Forum, Vol. 37(2) (2003) 18–28
9. Koenmann, J. and Belkin, N.: A case for interaction: A study of interactive information retrieval behavior and effectiveness. In Proceedings of CHI '96 (1996) 205–212
10. Kritikopoulos, A. and Sideri, M.: The Compass Filter: Search engine result personalization using Web communities. In Proceedings of ITWP '03 (2003)
11. Mizzaro, S.: Relevance: The whole history. JASIST, Vol. 48(9) (1997) 810–832
12. Nielsen, J.: Personalization is overrated. In Jakob Nielsen's Alertbox for October 4 (1998) http://www.useit.com/alertbox/981004.html
13. Schamber, L.: Relevance and information behavior. ARIST, Vol. 29 (1994) 3–48
14. Teevan, J., Alvarado, C., Ackerman, M. S. and Karger, D. R.: The perfect search engine is not enough: A study of orienteering behavior in directed search. In Proceedings of CHI '04 (2004) 415–422
15. Teevan, J., Dumais, S.T. and Horvitz, E.: Personalizing search via automated analysis of interests and activities. To appear in Proceedings of SIGIR '05 (2005)

Exhibit 27

Tags

bing **bing cashback** **Bing for Mobile** **bing maps** **Bing Rewards** **Bing Shopping** **bing travel** **BingIt** **events** **Facebook** **instant answers** **Kari Dilloo** **live search** **Stefan Weitz** **Ted Roduner**

Search Blog

Making search yours

The Bing Team
2/10/2011 9:00 AM
Comments (10)

For years people have talked about **personalized search** as the next evolution of our amazing technology. Over the years, the biggest obstacle facing search engineers is the simple fact that human behavior is not predictable. Don't get us wrong, people can be creatures of habit and we can build functions that enable us to display different results based on logical assumptions made in the aggregate. For example, 'traffic' at 6pm on a Friday likely refers to road conditions, not the movie or the band. So we can generally use some math magic to make really good predictions about what you mean when you type 'traffic' at 6pm tonight, but what if we're wrong? It's easy to see how that could happen especially if you, say, walk to work.

In that case a *more personal* search would benefit you; having more detailed information about the person doing the search can make results more effective.

We think one of the challenges with delivering results which are truly individualized is that, to date, personalized search "can't see the forest for the trees". In other words everyone is collecting everything and trying to figure out the foibles of human behavior from a mass of digital bits. To an extent, we've all been looking at the wrong inputs which in turn haven't given us the output we want.

We've found something interesting: a person's history or profile often does not necessarily deliver better

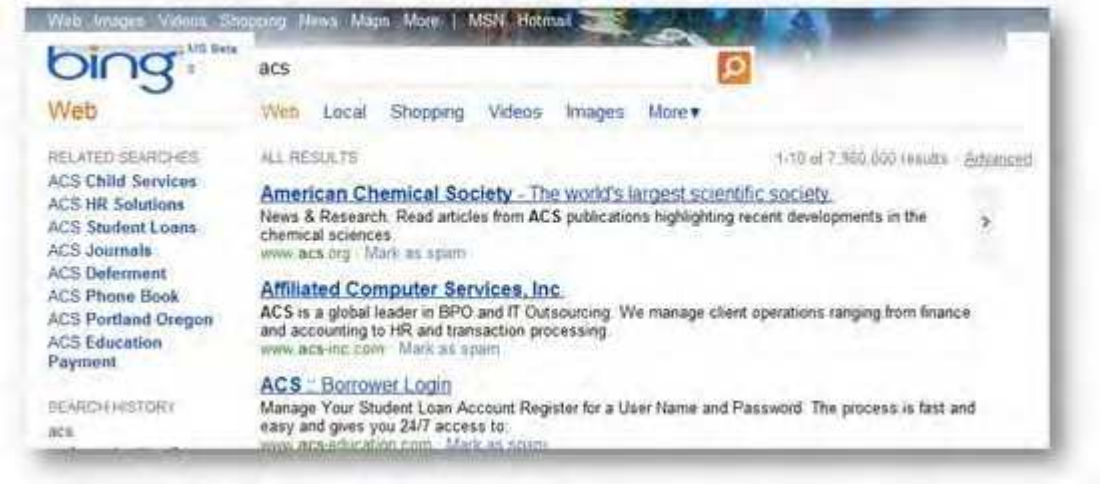
Even as we continue to develop more relevant search through smart personalization, we are very focused on maintaining an industry-leading privacy stance. For more information, see here.

We're currently 'flighting' (or "testing", for non search-geeks) a raft of experiments to see which techniques deliver the best results for a given user behavior, but today we want to talk about two we've recently put out there for you all! First, something relatively simple: **automatically tailoring search results based on your physical location**

As 76% of people use search engines to plan trips, events or social gatherings, Bing Local has always provided you with maps to nearby business listings, authoritative reviews and areas of interest. Starting

relevant result for that user is not necessarily the same as that for the majority of people in the U.S. To numerous users with an interest in pursuing a career in chemistry, the most relevant result may be the *American Chemical Society*, but to someone interested in how they can get involved in the fight against cancer, the most relevant result is more likely to be the *American Cancer Society.*

Suppose, in this latter case, the chemistry fan selects *American Chemical Society* . Our research shows that users commonly re-issue such navigational queries and the intent of that user rarely changes. This new personal search feature uses this human behavior as its core premise – if Bing thinks a user is trying to "re-find" a site, the relevant result is promoted to the top position on the page:



The beauty of thinking differently about personalized searching is that it enables us to construct elegant solutions that require a minimal amount of personal information and, frankly, often exhibit better results than a more computationally complex predictive model alone.

There is much more to come, but take Bing for spin and tell us what you think!

- Aidan Crook & Sanaz Ahari, Bing Search

Join Bing Community

2/14/2011 4:51 AM

http://bollywoodmasala.info/


sunny.seo15
2/16/2011 9:27 PM

I like very much...Pizza.....:)


kuriositee
2/17/2011 6:40 AM

Will you offer a way to turn personalized search results off like Google does? Or manually change your location?

Create an account to comment.

Exhibit 28

REDACTED

IN ITS

ENTIRETY