Personalized User Model LLP v. Google Inc.

Doc. 584 Att. 2

EXHIBIT 5

Dockets.Justia.com

IN THE UNITED STATES DISTRICT COURT

FOR THE DISTRICT OF DELAWARE

PERSONALIZED USER)
MODEL, LLP,)
Plaintiff,)
vs.) CA number 09-525 (LPS)
GOOGLE, INC.,)
Defendant.)

\- - - - -

VIDEOTAPED DEPOSITION OF JAIME CARBONELL

WASHINGTON, D.C.

NOVEMBER 27, 2012

The videotaped deposition of JAIME CARBONELL was convened on Tuesday, November 27, 2012, commencing at 10:05, at the law offices of SNR Denton, located at 1301 K Street, Northwest, in Washington, D.C., before Paula G. Satkin, Registered Professional Reporter and Notary Public.

\- - - - -

Job No. CS1565706

A P P E A R A N C E S

ON BEHALF OF THE PLAINTIFF:
JENNIFER BENNETT, ATTORNEY AT LAW
SNR DENTON
1530 Page Mill Rd.
Suite 200
Palo Alto, CA 94304-1125
650.798.0300

ON BEHALF OF THE DEFENDANT:
DAVID PERLSON, ATTORNEY AT LAW
JOSH SOHN, ATTORNEY AT LAW
QUINN EMANUEL
50 California Street
22nd Floor
San Francisco, CA 94111
415.875.6600
and
MARC FRIEDMAN, ATTORNEY AT LAW
QUINN EMANUEL
51 Madison Avenue
22nd Floor
New York, NY 10010
212.849.7000

ALSO PRESENT:
T.J. O'TOOLE, Videographer

C O N T E N T S

E X H I B I T S

P R O C E E D I N G S
(10:05 a.m.)

THE VIDEOGRAPHER: On the record with disk number one of the video deposition of Dr. Jaime Carbonell taken by the Defendant in the matter of Personalized User Model LLP versus Google Inc. and Google Inc. versus Personalized User Model LLP, both cases being heard before the United States District Court for the District of Delaware, Civil Action Number 09-525 LPS.

This deposition is being held at the law offices of SNR Denton, located at 1301 K Street, Northwest, in Washington, D.C., on November 27th, 2012, at approximately 10:05 a.m.

My name is T.J. O'Toole. I am the certified legal video specialist. The court reporter is Paula Satkin. We are both here representing Veritext of New Jersey.

Will counsel please introduce themselves and indicate which parties they represent.

MS. BENNETT: Jennifer Bennett representing Plaintiff Personalized User Model and the witness, and with me today I have Marc

Friedman.

MR. PERLSON: David Perlson from Quinn Emanuel representing Defendant Google.

MR. SOHN: And Josh Sohn of Quinn Emanuel also representing the Defendant.

THE VIDEOGRAPHER: Thank you. Will the court reporter please swear in the witness.

Whereupon--

JAIME CARBONELL

a witness, called for examination, having been first duly sworn, was examined and testified as follows:

EXAMINATION BY COUNSEL FOR THE DEFENDANT

BY MR. PERLSON:

Q. Good morning. Could you state and spell your name for the record?
A. It's Jamie Carbonell, J-A-I-M-E, C-A-R-B-O-N-E-L-L.
Q. And do you go by Dr. Carbonell or Mr. Carbonell?
A. Dr. Carbonell.
Q. Okay. And you've been deposed before; correct?
A. Yes.

Q. And you understand that you are testifying under oath as if you were testifying before a jury; correct?
A. Yes.
Q. And I'm going try to be as clear as I can today and -- but if I ask a question that you do not understand, please let me know and I will do my best to make it more clear. Okay?
A. Okay.
Q. Now, Dr. Carbonell, you -- you co-authored a book called Machine Learning and Artificial Intelligence Approach; is that right?
A. There were three of them in that series.
Q. Okay. And when -- when was the first one?
A. I believe it was 1983.
Q. When was the second one?
A. 1986.
Q. And how about after that?
A. There was one more. I don't recall the date. A year or two afterwards.
Q. It was about 1990; does that sound right?

A. It could be, or it could have been a little earlier.
Q. You've been publishing in the machine learning field since then?
A. Yes, I have.
Q. When was the last time?
A. This year.
Q. What -- what did you publish this year?
A. The latest paper is one at the Association of Computing Machinery on learnability of DNF, disjunctive normal form, expressions.
Q. What's that?
A. Disjunctive normal form is a -- learning when you have different expressions of the target concept. So maybe an example is the clearest way to explain it.
Q. Sure.
A. If there is bank fraud, there are different ways of defrauding the bank. For example, by pretending to be a customer when you really aren't. By pretending you have a lot more in an account than you really do and withdrawing it. Insider transactions that are

be safe in here.
Q. That's good to know.
A. A Bayesian world, you can use information like that without data based on priors which can be updated if you have other observations. If you observe that a meteor has struck somewhere else and a second one has struck, then the probability that a third one will strike might be higher than it would have been had there been no other meteor strikes. In the frequentist case, you're not allowed to use the equivalent of a prior. You base it only on the data. And if there is no data, you basically cannot provide an estimate.
Q. And -- but mathematically, is -- is that probability expressed as a number between 0 and 1?
MS. BENNETT: Objection. Form.
THE WITNESS: In the frequentist approach, it is.
BY MR. PERLSON:
Q. I'm sorry.
A. I was trying to answer your earlier question.
Q. The -- okay. Go ahead.

A. It's not just frequentist and Bayesian. If you go a little broader, there are things interpreted like fuzzy logic and other forms of reasoning with degrees of belief, of belief propagation, that do not require the values of the interval to be between 0 and 1.
Q. Okay. So fuzzy logic doesn't require a number between 0 and 1?
A. Some types of fuzzy logic do not require that; others do.
Q. Okay.
A. Fuzzy logic is a broad term for introducing numbers into logic -- degree of belief into logic.
Q. But in order for there to be a degree of belief, there has to be some sort of scale of the -- the degree of likelihood of interest?
A. Yes, sir. That's right.
Q. So if I had a -- if I had a -- numbers that went from 1 to 10 and assigning something a number of 2 was not -- well, in order for something to be a probability, would -- let's say I have a -- I can assign numbers 1 through 10, and if I assign something a number

of 1, that is -- and that does not mean that a number of 2 is twice as likely to be of interest to the user as 1, would that be a probability?
MS. BENNETT: Objection. Form.
THE WITNESS: Sorry. I did not quite grasp the -- the premise of your question.
If you're talking about if you had a scale that went from 1 to 10, 1 was the lowest value and 10 was the highest value --
BY MR. PERLSON:
Q. Correct.
A. -- 2 then would not represent twice as likely as 1 because if 1 is the low end of the scale, 1 means it's not going to happen.
Q. Okay. Okay. Let's say it's between 0 and 10?
A. Okay.
Q. And if I assign something a number 1 and in order for that range to be a range of probabilities, wouldn't it be the case that a number of 2 would have to be twice as likely to be -- show the interest of a user in a document in order for it to be a probability?
MS. BENNETT: Objection. Form.
THE WITNESS: Okay. So, first of

all, let's call it a likelihood rather than a probability.
BY MR. PERLSON:
Q. Okay.
A. Technically speaking, it's hard to think about probability other than 0-to-1 interval. That's how the math works out.
In probability theory, what you're saying is correct in the sense that it's a linear scale. If something has twice the probability value of another event so long as it's not 0, it means it's twice as likely to happen.
If you use likelihoods, the typical interpretation is the same. So if a 0-to-10 scale, an event has a probability or likelihood of 1 and the second event has a likelihood of 2, the second event would be twice as likely to happen as the first.
It is not required that the scale be linear, but by convention you assume linearity unless told otherwise. So anybody's scheme of likelihood is either linear or they inform you how to calculate it.
Q. How to -- what do you mean, inform

you how to calculate it?
A. If it's -- if the scale were not linear, it could be, for example, on a -- based on a sigmoid function or something else, then they would have to provide you that sigmoid function that says if something has a value of 2 and something else has a value of 1, here's how you estimate how much more likely the value of 2 is over the value of 1. So in the absence of providing a function, and I used sigmoid function as an example of one that is sometimes used, it would be exactly as you say. It would be linear.
Q. And then let's say that I had a -- a -- likelihood numbers of 1 through 4 where 1 was somewhat likely, 2 was very likely, 3 was extremely likely, and 4 was a certainty of likelihood. Would that be -- would those numbers 1 through 4 be probabilities?
MS. BENNETT: Objection. Form.
THE WITNESS: Okay. So, first of all, you didn't define the bottom end of the range -- 0 means unlikely or 0 means impossible?
BY MR. PERLSON:
Q. Let's say 0 means highly unlikely?

A. So where's the point that it means -- so neither of the two ends are definitive, so that cannot be converted directly into a probability.
Q. So you --
A. A probability requires both end points to be nailed down, to be defined. The impossible versus the certain.
Q. Now, the patent talks about estimating parameters. Are you familiar with that?
A. Yes.
Q. And what does it mean to estimate a parameter?
MS. BENNETT: Objection. Form.
THE WITNESS: It means to compute the value of that parameter based on the information available. That computation can be inexact. It can be an approximation because the amount of information available is finite. It's not all possible likes or dislikes by a user. It's a finite set of those documents they have already seen. Given that it's based on partial observations of how a person would react to a document rather than the totality, that's why

it's called an estimation. It's an approximate calculation of the value.
Now, a model can have multiple parameters. It can have parameters that represent whether they like certain terms, certain concepts, whether they like certain sources of documents, whether they like certain topics within the documents, whether they like to see documents about the same area they've seen before and so forth.
The collection of all these parameters together with a mathematical function that combines them is the model. And estimating the parameters is finding or estimating a value, approximating a value, for each one of these inputs to the model, as it were, one of these variables in the model. A parameter is like a variable. It has a value. And you're estimating the values.
BY MR. PERLSON:
Q. Is the parameter the value or the -- or the variable?
A. It's used to mean both, and that is a cause of confusion, I'm afraid. I wish that my colleagues had been, let's say, more

discriminating in using it to mean only one of the two. That would have avoided future -- future confusion, but a parameter is used to mean the value and the parameter is also used to be the variable.
Q. And is that -- is that how it's used in -- in the patents, too?
MS. BENNETT: Objection. Form.
THE WITNESS: The patent talks about estimating the parameters. It really talks about estimating the values of variables.
BY MR. PERLSON:
Q. Sorry. Were you done?
A. Yeah. I'm done.
Q. And the -- in order to -- to estimate the values of the variables, is that done by a calculation?
MS. BENNETT: Objection. Form.
THE WITNESS: It is done -- everything is done by a calculation. So an estimation is a calculation based on the available data.
BY MR. PERLSON:
Q. And that's the -- that's -- the -- you mentioned a Dr. Jordan earlier. Is Michael

THE VIDEOGRAPHER: Counsel passed me a note asking me how much time we had left, and I told her that we've been on the record 6 hours and 10 minutes. I have no idea how much time is left.

MR. PERLSON: Okay.

MS. BENNETT: 50 minutes.

MR. PERLSON: That's all I need to know.

BY MR. PERLSON:

Q. The -- okay. So let's go to -- okay. And so on page 59, you see it says "representation for web navigation"?

A. Yes.

Q. And then underneath it, it says, "The probability distribution of the pages to be accessed is based on collecting the visiting patterns of many users."

A. Yes.

Q. And what do you understand the probability distribution of the pages is that's referred to there?

A. That is the probability of navigating from one web page to another web page by following a link between these pages. In

some cases some links are followed by many users. Other links may be followed by few users. Some links may be followed by no users.

Q. And how is that information used in the -- in Wasfi?

A. The main use that he puts to it is he builds what's called an order M model. So is there any problem with the recording?

Q. No.

A. We can just continue?

Q. He said that there was 6 hours 10 minutes, and then he passed a note that said 5 minutes. We were just chuckling -- it seemed inconsistent, but nothing to do -- sorry.

A. So an order -- an order M model decides how far back in the sequence of navigation you look to. So an order 1 model means that you look at the current page and where else you go next. An order 2 model is where you came from, the current page. An order 3 model, where you came from before the last one you came from and so on. The higher the order of the model, the more information you have, but then again, the less generalization because you must have traversed this particular sequence of

-- of hyperlinks in order to be in the same current state.

So he trades off the order of the model in order to balance accuracy with -- and generality, and he mostly does an order M equals 2 model as he states on the second -- in the middle of the second column. He goes through an example that I have no need to repeat here.

And so this is essentially a navigation process, and he shows in a finite state diagram in Figure 1 on the next page -- he does that illustration so he can calculate the probability that you will traverse a certain link, a certain hyperlink, from one page to another based on what others have traversed before.

Q. And that probability is used to determine the variable TIJ; is that right?

A. That probability may be used to initialize the variable TIJ. He has -- there's two parts to this -- to this paper, the part that we are talking about now and the entropy-based part which is just before it.

In the entropy-based part that he -- excuse me. E-N-T-R-O-P-Y. In the

entropy-based part that comes before it, he defines TIJ in a different way as a negative log of probability, rather than the probability itself. That negative log of the probability is bounded from -- let's see. The probability is zero -- is bounded from zero to infinity. So it's not really a probability in -- in that particular definition of TIJ on page 59, column 1.

MR. PERLSON: Okay. I think we need to take a break.

THE VIDEOGRAPHER: This ends disk number 4 of the Carbonell deposition. The time is 6:01:58. Off the record.

(A brief recess was taken.)

THE VIDEOGRAPHER: On the record with disk number five of the testimony of Dr. Jamie Carbonell in the matter of Personalized User Model versus Google. The date is November 27th, 2012. The time is 6:11:21.

BY MR. PERLSON:

Q. So now, we were discussing the variable TIJ in Wasfi?

A. Yes.

Q. What is that?

A. TIJ is meant to be measure of importance or interestingness of the page -- of the Ith page the Jth user. Ith page to the Jth user. In fact, I believe Wasfi says so explicitly -- let me find it. Yes, column 1, page 59, just above the formula. However, that statement is not exactly consistent with his -- with his formula. This is sometimes called stochastic entropy rather than the more traditional or more commonly used Shannon entropy. Shannon entropy is minus P log P, and that is bounded on both ends. This is unbounded, at infinity.

Q. Which is unbounded?

A. TIJ, the H of PR, which is the same thing.

Q. So something that is unbounded cannot be a probability; is that right?

A. That's right. It cannot be normalized into a probability.

Q. The TIJ variable, that indicates how much weight a new page should get in a user's profile when that user accesses that page; right?

A. That's right.

Q. And then --

A. So a page that would have zero probability in this case would have an infinite value.

Q. Okay. But what -- I don't understand what the --

A. The negative -- the logarithm of zero is minus infinity. And so if you take minus the minus infinity, it becomes positive infinity.

Q. You agree that a probability can't be a negative number?

A. That's right.

Q. So --

A. It also cannot be infinite.

Q. Does the fact that the -- the TIJ is calculated based on a probability distribution of pages based on collecting the visiting patterns of many users affect your view of whether the -- the variable TIJ is a parameter of a learning machine or user model?

A. No, not really. The -- the fact that there are many -- information is collected about many users means that it's less often the probability of a particular transition will be

zero. And so this pathological case will not occur with nearly as much frequency as it would occur if it was just an individual user who had not traversed that link.

Q. Can the weights of a user-specific learning machine be -- let me start over.

Can the parameters of a user-specific learning machine be set based on formulas that take into account the activity of other users?

MS. BENNETT: Objection. Form.

THE WITNESS: You're talking about a learning machine for an individual user or a learning machine for all users?

BY MR. PERLSON:

Q. A learning machine for an individual user?

A. A learning machine for an individual user can take account of behaviors of other users, but it must also take account of behavior by this specific user so at least some of the parameters must be estimated from data specific to this user, not necessarily all of them.

Q. So a user-specific learning

machine must at least have some parameters that are specific to that user?

A. Yes.

Q. What if the parameters that are specific to that user -- well, let me give you -- let me give you an example of something.

If a -- the system creates a parameter for the user interest in sports and it determines by the fact that users -- in reference to all users, that if you've clicked on sports pages five times, that that indicates that you should get a weight of .5 for the variable interest in sports. Would that -- would that be a user-specific parameter?

MS. BENNETT: Objection. Form.

THE WITNESS: So how did you determine that it should have a weight of .5?

BY MR. PERLSON:

Q. Because you look to see it was assigned based on the activity of all users, that if in observing all the users, they saw that if a user clicked on sports pages five times, that an appropriate weight was .5.

A. Okay. And then this specific user also clicked on it exactly five times?

298

THE WITNESS: No. Estimating the parameters is -- always means estimating the values or weights with those parameters. It is the case, as I mentioned in answers to both of you, that the field uses the word "parameter" more loosely sometimes to mean the variables of the -- and sometimes to mean the values. And Refuah does that as well because -- and the way that is consistent with the Court's construction is consistent with my report and is consistent with the claim language is "parameters" mean the values that are being estimated.

BY MR. PERLSON:

Q. Right. And if you look at 1E, it refers to estimating a probability PUD that an unseen document D is of interest to the user U. Then it goes on to say, "wherein the probability PUD is estimated by applying the identified properties of the document to the learning machine having the parameters defined buy the User Model." Do you see that?

A. Yes.

Q. So that requires that the learning machine must actually have the values of the variables that are defined by the user model; right?

299

MS. BENNETT: Objection. Form.

THE WITNESS: That's right.

MR. PERLSON: I don't have any further questions.

MS. BENNETT: Okay. And we reserve the right to review the transcript and provide an errata.

THE VIDEOGRAPHER: This ends disk number 5 and concludes the testimony of Dr. Jamie Carbonell in the matter of Personalized User Model versus Google. The date is November 27th, 2012. The time is 7:08:47. Off the record.

MR. FRIEDMAN: Ms. Satkin, you did a stellar job.

(Signature not waived.)

(Whereupon, at 7:08 p.m., the deposition was concluded.)

\- - - - -

300

ACKNOWLEDGMENT OF DEPONENT

I do hereby acknowledge that I have read and examined the foregoing of the transcript of my deposition and that:

(Check appropriate box):

( ) the same is a true, correct and complete transcription of the answers given by me to the questions therein recorded.

( ) except for the changes noted in the attached errata sheet, the same is a true, correct and complete transcription of the answers given by me to the questions therein recorded.

________________ ________________________

DATE SIGNATURE

301

CERTIFICATE OF NOTARY PUBLIC

I, Paula G. Satkin, the officer before whom the foregoing proceedings were taken, do hereby certify that the witness whose testimony appears in the foregoing proceeding was duly sworn by me; that the testimony of said witness was taken by me in stenotype and thereafter reduced to typewriting under my direction; that said proceedings is a true record of the testimony given by said witness; that I am neither counsel for, related to, nor employed by any of the parties to the action in which these proceedings were taken; and, further, that I am not a relative or employee of any attorney or counsel employed by the parties hereto, nor financially or otherwise interested in the outcome of the action.

My commission expires November 14, 2015.

____

PAULA G. SATKIN
Notary Public in and for the
District of Columbia

ERRATA SHEET
VERITEXT CORPORATE SERVICES
800-567-8658
ASSIGNMENT NO. CS1565706
CASE NAME: Personalized User Model v. Google
DATE OF DEPOSITION: 11/27/2012
WITNESS' NAME: Jaime Carbonell

PAGE/LINE(S)/ CHANGE REASON
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________
____/_______/_________________/__________

____
Jaime Carbonell

SUBSCRIBED AND SWORN TO
BEFORE ME THIS______DAY
OF_______________, 2012.

____
NOTARY PUBLIC
MY COMMISSION EXPIRES__________________

Veritext Legal Solutions
25B Vreeland Road - Suite 301
Florham Park, New Jersey 07932
Toll Free: 800-227-8440 Fax: 973-629-1287

_____________, 2012

To: JENNIFER BENNETT, Esq.

Case Name: Personalized User Model v. Google
Veritext Job Number: 1565706
Witness: Jaime Carbonell
Deposition Date: 11/27/2012

Dear Ms. Bennett:

Enclosed please find a deposition transcript. Please have the witness review the transcript and note any changes or corrections on the included errata sheet, indicating the page, line number, change, and the reason for the change. Have the witness' signature at the bottom of the sheet notarized and forward errata sheet back to us at the address shown above.

If the jurat is not returned within thirty days of your receipt of this letter, the reading and signing will be deemed waived.

Sincerely,

Production Department

Encl.
Cc: DAVID PERLSON, Esq.

EXHIBIT 6

O.I.P.E. JC109
MAR 0 8 2004
PATENT & TRADEMARK OFFICE

#11
RWB
3-17-04

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application No.: 09/597,975 Docket No.: UTO-101

Filing Date: 06/20/2000 Art Unit: 2157

Applicants: Konig *et al.* Examiner: Barbara N. Burgess

Title: Automatic, Personalized Online Information and Product Services

CERTIFICATE OF MAILING
I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as First Class Mail in an envelope addressed to: Commissioner of Patents, Alexandria, VA 22313-1450

on 3/4/04
Date

Signature

Aga Kleszcz
Type or print name of person signing

RECEIVED
MAR 1 0 2004
Technology Center 2100

Reply under 37 CFR 1.111

Commissioner for Patents
Mail Stop Non-Fee Amendment
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

In reply to the Office Action mailed by the USPTO on January 29, 2004, the Applicants respectfully submit the following remarks.

REMARKS

CLAIM REJECTION, 35 USC Paragraph 103

Claims 1-62 were rejected under U.S.C. 103(a) as being unpatentable over *Breese et al.* (U.S. Patent No. 6,006,218) in view of *Hertz et al.* (U.S. Patent No. 5,754,939).

In reply, the Applicants respectfully disagree.

A. GENERAL COMMENTS

What does the present invention teach and claim in independent claims 1 and 32?

The present invention is a method for **predicting user interests** in documents and products using a **learning machine** and **probability measures**. The steps are among others (See claim 1 and 32):

- **transparently monitoring** user interactions;
- using the monitored user actions (note: transparently monitored) for user-specific files;
- **estimate parameters of a learning machine** to define a user model based on user specific files;
- **using the learning machine** (i.e. with user estimated parameters) to **estimate the probability that a document is of interest to a user** (i.e. probability estimates);
- **using the estimated probability** to provide **personalized information** to user.

The Applicants would like to respectfully note that learning can be divided into two parts: (1) **memorization** and (2) **generalization or prediction**.

Ad 1. Memory

Memory refers to what happened in the **past**. A model could be developed that keeps track or score of what happened. For instance, a **user model** could be developed of the scored/tracked items (e.g. which websites **were** visited or which documents **were** looked at). Items could be correlated or similarities could be established (See e.g. *Hertz* Col. 8, line 49; *Hertz* Claim 3).

Using such a model (called knowledge or memory model) one could determine the probability that a user **has seen** or **knows about** an item. Based on this memory, one could determine correlations/similarities/matches (See e.g. *Hertz* Fig. 10 item **1103**; *Hertz* Col. 78 lines 51-52 "... cluster articles based on similarity ... ") with items obtained through a search query. Note such a model is **only applicable** to determine the probability for:

(1) an **individual user**, and

(2) **for that particular item**.

There is no carry over and **no generalization** to other users or other items. Memorization could also be referred to as low-level learning (or limited learning).

More specifically to *Breese*, who teaches that one could determine the **probability that a user knows** about an item (*Breese*: Column 7, lines 1-10, 31-36) – i.e. the user has seen that item in the past. Note knowledge probability (i.e. memory) as in *Breese* IS NOT the same as probability that documents are of interest (i.e. generalization/estimate probability) as in the present application as an artisan would readily appreciate.

In a model one could further make the distinction between application-dependent or application-independent learning. An example of application-dependent learning could be "choose all relevant NY Times articles". An example of application-independent learning could be "choose all relevant NY Times articles **and** find the most important emails, provide personalized search results, etc.". The Applicants assert that *Hertz* teaches the application-dependent approach, whereas the present application is application-independent as defined by elements 1(e) and 1(f) (same for our claim 32).

Classification as an application-**in**dependent approach requires at least two criteria:

(i) "*cross fertilization*" (see present application), i.e. feedback or learning in one application is used to serve all applications. Neither *Hertz* nor *Breese* teach cross-fertilization.

(ii) a user-model can be used for a new personalized application, without the need for application specific learning or initialization. Neither *Hertz* nor *Breese* teach such a generic user model.

To illustrate the application-dependency of *Hertz*, see for instance column 10, lines 10-24 and column 11, lines 3-16. *Hertz* also teaches different sets of attributes for different applications, which makes it obvious that *Hertz* **can't conceive** an application-independent user model. It is again further noted that the present application **does not teach memorization**. Rather, the present invention teaches a learning model to estimate probabilities to predict personalized information that is of interest to the user.

Ad 2.Generalization

Neither *Breese* nor *Hertz* teach any type of generalization; there is no learning involved other than keeping score or tracking what happened in the past. Please note that there is no learning or generalization in these prior art references and could therefore not suggest the present invention to render it obvious.

For example could *Breese* or *Hertz* use a user-model for apples to predict if the user is interested in pears? The answer is no, since the user-model for apples has no knowledge or generalization power related to pears. The teachings of *Breese* and *Hertz* are knowledge-based without any teaching on how to use that knowledge model to generalize beyond that or become application independent – independent from the apples and extend to pears. It is one of the objectives of the present invention to overcome this shortcoming; i.e. a **learning machine in the probability domain** and **cross-fertilization of learning in one mode to another mode**.

Generalization **predicts beyond** items in the past and even beyond the user itself; it **estimates probability** of something to happen in the future. It is exactly this generalization that is claimed in claims 1 and 32 by:

(1) using the monitored actions to **estimate parameters of a learning machine**, and

(2) using the learning machine **to estimate the probability** that a document is of interest to a user.

As clearly taught in the present application, **generalization** is made possible by defining a model in the **probability domain**, which **decouples particular feature vectors** and learns to make the model **application/item independent**. The user model of the learning machine in the present invention represents user interests **independent** of any specific (note: specific is application dependent) user information. In other words, the present invention is not related to a specific query. There is therefore no need to distinguish between seen or unseen documents.

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

Accordingly, the Applicants submit that the present claims 1-62 are **NOT obvious** with respect to *Breese* in view of *Hertz*. A ***prima facie*** case of obviousness (See MPEP 2143) has **not been established** as discussed *supra*.

B. SPECIFIC COMMENTS

Claims 1 and 32

1. The Office Action asserts that column 5, lines 25-38 of *Breese* discloses, "transparently monitoring user interactions with data while the user is engaged in normal use of a computer."

In reply, the Applicants assert that the cited passages **do not** specify **nor imply** that the user is engaged in normal use of the computer, nor that the monitoring is transparent. In fact, the cited passage includes obtaining information from questionnaire results, which are certainly not transparently obtained when the user is engaged in normal use of a computer.

2. The Office Action asserts that column 8, lines 33-36, 44-46 of *Breese* discloses, "updating user-specific data files, wherein the user-specific data files comprise the monitored user interactions with the data and a set of documents associated with the user."

In reply, the Applicants assert that if the step in element (a) "transparently monitoring user interactions .." is not taught or implied, then there can not be a teaching or implication of step (b) that follows (a). Note it is updating (step b) with the monitored user interactions (step a).

3. The Office Action asserts that element, "analyzing a document to identify properties of the document," is described in column 8, lines 15-26 of *Breese*.

In reply, the cited section of *Breese* does not discuss any analysis of documents and are irrelevant to the claim element.

4. The Office Action asserts that several sections of *Hertz* discloses steps (c), (e) and (f).

In reply, the Applicants respectfully disagree and refer to the arguments made *supra* (general comments). The Applicants would like to respectfully point out that the Office Action fails to **clearly point out** where *Hertz* teaches steps (c), (e) and (f) since reviewing these sections the Applications are unable to identify the relevant teachings. Perhaps the Examiner could assist and be more precise by pointing to the selective sentences instead of an *aggregate* of independent sections/paragraphs/words.

In addition, *Hertz*:

(i) teaches **memorization**, we don't,

(ii) teaches an **application specific** user model without any generalization power, we have an application-independent learning model,

(iii) does not teach or imply any learning to **estimate probability** of user interests, we do,

(iv) does not teach or imply any information theory to **determine probability measures**, we do,

(v) does not teach **probability measures** if whether an item is of interest to a user (See also *infra*), we do, and/or

(vi) teaches **clusters of documents** (See Hertz Col. 78, lines 51-53) and does not teach **clusters of user models** like we do (which is a big difference).

None of the sections (either *individually* or *combined*) of *Hertz* referred to in the Office Action discusses, teaches or implies steps (either *individually* or *combined*) (c), (e) and (f). Accordingly, the Applicants submit, as submitted *supra*, that the present claims 1-62 are **NOT obvious** with respect to *Breese* in view of *Hertz*. A ***prima facie*** case of obviousness (See MPEP 2143) has **not been established**.

CLAIMS 2-31 and 33-62

The Applicants believe that the significant differences discussed above between the claimed invention and *Breese* in view of *Hertz* make the claimed invention novel and *non*-obvious. Because all other claims depend from either claim 1 or claim 32, the Applicants believe that all pending depending claims are also novel and *non*-obvious. In addition to their dependency on claims 1 or 32, the Applicants incorporate herewith **all previous arguments** made on the record in the previous reply to the first Office Action.

In addition, the Applicants have trouble comprehending the relevant teaches pointed out by the Examiner related to *Hertz* that would render the present claims obvious. As a side note, *Hertz* in Column 7, lines 47-67 to Column 8 1-9 teaches "truly passive" and "browsing and filtering", which shows that *Hertz* does not have the intention to suggest its teachings to be a basis for predicting user interests for personal search and services. This is in contrast to claim 1 and 32 of the present application.

Furthermore, Applicants would like to point out that *Hertz* does not teach nor imply probability measures, or how to define probability measures in either formula or wordings. A simple word search on the word probability in *Hertz* doesn't return a favorable answer. Note the word "probability" can be found e.g. in *Hertz* Col. 50 line 28 it refers to " ... probability that a user will access target object T". **However**, this probability is based on a memorized user model (see *supra*) and not the probability that the document is of interest to a user (which is based on a learning model of estimated probabilities and not memories). Furthermore, a description or implication of the necessary information theory to establish probability measures as claimed in claim 1 and 32 is missing in *Hertz*. Accordingly, the Applicants are puzzled to why the Office Action asserts that *Hertz* teaches or renders our claims obvious in combination with *Breese*.

CONCLUSION

Applicants respectfully submit that the present claims 1-62 are **NOT obvious** with respect to *Breese* in view of *Hertz*. A ***prima facie*** case of obviousness (MPEP 2143) has **not been established** as discussed *supra*. Even *if* at the time the invention (i.e. hindsight is impermissible, See MPEP 2141.01 III) was made one skilled in the art would be motivated to combine *Breese* and *Hertz*, the resulting method would still not possess the capability to provide automated and personalized information services to a user that uses machine learning including memorization and generalization defined in the probability domain simply because neither *Breese* or *Hertz* teach or suggest anything beyond memorization models.

Therefore, the Applicants submit that claims 1-62 are novel and unobvious over the closest prior art of record. Accordingly, allowance of the claims now in the application is kindly requested.

Respectfully submitted,

Dr. Ron Jacobs
Reg. No. 50,142
LUMEN Intellectual Property Services
2345 Yale Street, 2nd Floor
Palo Alto, CA 94306-1429

Phone: (650) 424-0100
Fax: (650) 424-0141
Email: ron@lumen.com