EXHIBIT A

EXHIBIT A – PUM's Claim Construction Comparison Chart

I. AGREED TERMS/PHRASES

Claim Language -- Terms/Phrases in Issue are in Bold and Italics	Claims	Terms/Phrases for Construction	PUM's Revised Proposed Constructions	Google's Revised Proposed Constructions
automatic, personalized information services to the user	'040 patent: 1, 32	automatic	"without human intervention"	"without human intervention"
central computer	'040 patent: 32	central computer	"computer on the server side of a client-server relationship"	"computer on the server side of a client-server relationship"
set of documents associated with the user	'040 patent: 1, 32	set of documents associated with the user	"group or collection of documents associated with the user"	"group or collection of documents associated with the user"

II. DISPUTED TERMS/PHRASES

Claim Language -- Terms/Phrases in Issue are in Bold and Italics	Claims	Terms/Phrases for Construction	PUM's Proposed Constructions	Google's Proposed Constructions
THE LEARNING MACHINE TERMS/PHRASES: **parameters, estimated parameters, learning machine, User Model specific to the user, and user-specific learning machine**				
c) ***estimating parameters of a learning machine***, wherein the ***parameters*** define a ***User Model specific to the user*** and wherein the parameters are estimated in part from the ***user-specific data files***; e) estimating a probability P(u|d) that an unseen document d is of interest to the user u, wherein the probability P(u|d) is estimated by applying the identified properties of the document to the ***learning machine*** having the ***parameters*** defined by the ***User Model***; and	'040 patent: 1, 32	parameters	*"values or weights of the variables of the learning machine"*	"variables, having a value or weight, that are used by the learning machine to calculate a probability"
		estimating parameters	*"estimating values or weights of the variables of the learning machine"*	"estimating a value or weight of each of the variables that are used by the learning machine to calculate a probability"

		learning machine	PUM's Original Construction: "a model and/or mathematical function that is used to make a prediction or intelligent decision that attempts to improve performance in part by altering the values/weights given to its variables depending upon past observations or experiences" *PUM's Revised Construction: "a model and/or mathematical function that is used to make a prediction that attempts to improve performance in part by altering the values/weights given to its variables depending upon past observations or experiences"*	Google's Original Construction: "program that contains parameters used to calculate a probability, and where the predictive ability of the program improves over time with the addition of new data" Google's Revised Construction: "program that contains parameters used to calculate a probability, and where the predictive ability of the program improves over time with the addition of monitored user interactions" Google's Further Revised Construction: "program that contains parameters used to estimate a probability and attempts to improve its predictive ability over time and as the program receives new data by altering the values/weights given to its parameters"
		User Model specific to the user	*"an implementation of a learning machine updated in part from data specific to the user"*	Google's Original Construction: "model unique to the user, that is created an updated by the learning machine and stored in a data structure" Google's Revised Construction: "model restricted to the user, that is created and updated by the learning machine and stored in a data structure"
[c] ***estimating parameters of a user-specific learning machine*** based at least in part on the documents of interest to the user;	'276 patent: 1, 23	user-specific learning machine	PUM's Original Construction: "a model and/or mathematical function that is used to make a prediction or intelligent decision that attempts to improve performance in part by altering the values/weights given to its variables depending upon past	Google's Original Construction: "learning machine unique to the user"

[f] for each retrieved document of said plurality of retrieved documents: identifying properties of the retrieved document, and applying the identified properties of the retrieved document to the ***user-specific learning machine*** to estimate a probability that the retrieved document is of interest to the user; and			observations or experiences specific to the user" *<u>PUM's Revised Construction</u>: "a learning machine specific to the user"*	<u>Google's Revised Construction</u>: "learning machine restricted to a particular user"
THE PROBABILITY TERMS/PHRASES: **probability P(u|d)/probability, posterior probability P(u|d,q), and the larger phrases**				
e) ***estimating a probability P(u|d) that an unseen document d is of interest to the user u***, wherein the ***probability P(u|d)*** is estimated by applying the identified properties of the document to the learning machine having the parameters defined by the User Model; and [f' for each retrieved document … applying the identified properties of the retrieved document to the user-specific learning machine to ***estimate a***	'040 patent: 1, 32 '276 patent: 1, 23	estimat[ing]	*"approximating or roughly calculating"*	"calculating"
		probability/probability P(u|d)	<u>PUM's Original Construction</u>: "degree of belief or likelihood" *<u>PUM's Revised Construction</u>: "numerically-based degree of belief or likelihood"*	"percentage chance"
		estimating a probability P(u|d) that an unseen document d is of interest to the user u	<u>PUM's Original Construction</u>: "approximating or roughly calculating a degree of belief or likelihood that an unseen document d is of interest to the user u given the information that is known about the unseen document" *<u>PUM's Revised Construction</u>: "approximating or roughly calculating a*	"calculating the percentage chance that an unseen document d is of interest to the user u given the information that is known about the unseen document"

probability that the retrieved document is of interest to the user; and			*numerically-based degree of belief or likelihood that an unseen document d is of interest to the user u given the information that is known about the unseen document"*	
11. The method of claim 1 further comprising ***estimating a posterior probability P(u|d,q) that the document d is of interest to the user u, given a query q submitted by the user****.*	'040 patent: 11	posterior probability; posterior probability P(u|d,q)	PUM's Original Construction: "the degree of belief or likelihood that a document d is of interest to the user u given the information that is known about the document and given a query" *PUM's Revised Construction: "the numerically-based degree of belief or likelihood that a document d is of interest to the user u given the information that is known about the document and given a query"*	"the percentage chance of the user u being interested, taking into account what is previously known about that user's interests in general, given new knowledge of the document d the user is considering and a search query q submitted by the user"
		[estimating a] posterior probability P(u|d,q) that the document d is of interest to the user u, given a query q submitted by the user.	PUM's Original Construction: "approximating or roughly calculating the degree of belief or likelihood that a document d is of interest to the user u given the information that is known about the document and given a query q" *PUM's Revised Construction: "approximating or roughly calculating the numerically-based degree of belief or likelihood that a document d is of interest to the user u given the information that is known about the document and given a query q"*	"calculating the percentage chance of the user u being interested, taking into account what is previously known about that user's interests in general, given new knowledge of the document d the user is considering and a search query q submitted by the user"

THE USER AND USER-SPECIFIC DATA FILES TERMS/PHRASES: **user, user-specific data files, monitored user interactions with the data, and set of documents associated with the user**				
user	passim	user	*“a person operating a computer as represented by a tag or identifier”*	“person operating a computer”
b) 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***;	‘040 Patent: 1, 32	user-specific data files	*“the monitored user interactions with data and a set of documents associated with the user”*	<u>Google’s Original Construction</u>: “data files unique to the user” <u>Google’s Revised Construction</u>: “data files restricted to a particular user”
		monitored user interactions with the data	*“the collected information about the user’s interactions with data”*	“user interactions with the data obtained from the monitoring step of 1(a)”
		set of documents associated with the user	agreed	agreed
DOCUMENT AND UNSEEN DOCUMENT				
document	passim	document	<u>PUM’s Original Construction</u>: “text or any type of media” *<u>PUM’s Revised Construction</u>: “electronic text or any type of electronic media”*	<u>Google’s Original Construction</u>: “electronic file” <u>Google’s Proposed Revised Construction</u>: “an electronic file including text or any type of media”
unseen document	‘040 patent: 1, 32	unseen document	*“document not previously seen by the user”*	“document not previously seen by any user”
PRESENT/PRESENTING				
present/presenting	‘276 patent: 1, 23, 24	present	*“to provide or make available”*	“display”
		presenting	*“providing or making available”*	“displaying”
THE DEFINITENESS/INDEFINITENESS PHRASES				
document of interest to the user	‘276 patent: 1, 23	document of interest to the user	<u>PUM’s Original Construction</u>: “text or media for which the user has a positive response”	Indefinite

			<u>PUM's Revised Construction</u>: "electronic text or media for which the user has a positive response"	
document not of interest to the user	'276 patent: 5	document not of interest to the user	<u>PUM's Original Construction</u>: "text or media for which the user has a negative response or has ignored" *<u>PUM's Revised Construction</u>: "electronic text or media for which the user has a negative response or has ignored"*	Indefinite
user interest information derived from the User Model	'040 patent: 21	user interest information derived from the User Model	*"interests or other information inferred from the User Model"*	Indefinite
ORDER OF STEPS AND ANTECEDENT BASIS				
ORDER OF STEPS	'040 patent: 1, 32 '276 patent: 1, 23		<u>PUM's Original Construction - '040 Patent, claims 1 and 32</u>: "steps may be performed in a consecutive, non-consecutive, repeating, and/or overlapping order, except that step (d) must be performed (at least once) before step (e) is performed" *<u>PUM's Revised Construction - '040 Patent, claims 1 and 32</u>: "steps may be performed in a consecutive, non-consecutive, repeating, and/or overlapping order, except that step (d) must be performed (at least once) before step (e) is performed, and step (f) is performed after step (e)"* <u>PUM's Original Construction - '276 Patent, claims 1 and 23</u>: "steps may be performed in a consecutive, non-consecutive, repeating, and/or overlapping order, except that step [d] must be performed before step [e] and the	<u>Google's Original Construction -- '040 Patent, claims 1 and 32</u>: "steps (a), (b), and (c) must be performed in that order and before steps (e) and (f); step (d) must be performed before steps (e) and (f); and step (e) must be performed before step (f)" <u>Google's Proposed Revised Construction</u>: "the claim steps must be performed in order, but individual steps may be repeated" <u>Google's Original Construction - '276 patent</u>: (1): "steps (a), (b), and (c) in that order; step (d) before step (e); step (f) after steps (c) and (e); and step (g) must be performed after step (f)" '276 patent, 23: step (a), (b), (c), (d), (e),

			'applying' portion of step [f] must be performed before step [g]" <u>*PUM's Revised Construction - '276 Patent, **claim 1***</u>*: "steps may be performed in a consecutive, non-consecutive, repeating, and/or overlapping order, except that step [d] must be performed before step [e] and the 'applying' portion of step [f] must be performed before step [g]; the 'identifying' portion of step [f] must be done at least once before the applying portion of step [f], but need not occur after steps [d] and [e]"* <u>*PUM's Revised Construction - '276 Patent, **claim 23***</u>*: "steps may be performed in a consecutive, non-consecutive, repeating, and/or overlapping order, except that step [d] must be performed before the 'applying' portion of step [e], which must be performed before step [f], which must be performed before step [g]; the 'identifying' portion of step [e] must be done at least once before the applying portion of step [e], but need not occur after step [d]"*	and (f) in that order" <u>Google's Proposed Revised Construction - '276 patent, claims 1 and 23</u>: "the claimed steps must be performed in order, but individual steps may be repeated"
ANTECEDENT BASIS TERMS				
	'040 Patent: 1, 11, 21, 32	"User u"/"the user" and "the user u"	*No construction necessary*	"A user u" and "the user" / "the user u" refer to the same user.
	'276 Patent: 1, 6, 21, 23	"user" / "the user"	*No construction necessary*	"A user u" and "the user" / "the user u" refer to the same user.

	‘040 Patent: 1, 32	“user-specific data files” / “the user-specific data files”	*No construction necessary*	“user-specific data files” and “the user-specific data files” refer to the same files
	‘276 Patent: 6	“a document d” / “the document”	*No construction necessary; alternatively, “the document” in step (e) of the ‘040 patent refers to “the unseen document d” earlier in that claim element*	“a document” and “the document” refer to the same document
	‘040 Patent: 1, 32	“a learning machine” / “the learning machine”	*No construction necessary*	“a learning machine” and “the learning machine” refer to the same learning machine.
	‘276 Patent: 1, 5, 23	“a user-specific learning machine” / “the user-specific learning machine”	*No construction necessary*	“a user-specific learning machine” and “the user-specific learning machine” refer to the same user-specific learning machine
	‘040 Patent: 1, 32	“a probability P(u|d) that an unseen document d is of interest to the user u” / “the probability P(u|d)” / “the estimated probability”	*No construction necessary*	“a probability P(u|d) that an unseen document d is of interest to the user u,” “the probability P(u|d),” and “the estimated probability” refer to the same probability.
	‘040 Patent: 1, 32	“parameters of a learning machine” / “the parameters”	*No construction necessary*	“a user model” and “the user model” refer to the same user model.
	‘040 Patent 1, 21, 32	“a user model” / “the user model”	*No construction necessary*	“a user model” and “the user model” refer to the same user model
	‘276 Patent: 1, 21	“a search query” / “the search query”	*No construction necessary*	“a search query” and “the search query” refer to the same search query

4045015

EXHIBIT B

Level 3 (upper level)

Bayesian probability theory

Reverend Thomas Bayes (1702-1761)

Meaning of probability

There are two schools about the interpretation of probability. In classical statistics, probability is interpreted as a limiting frequency when an experiment is repeated infinitely many times. For instance in throwing a dice, the probability of having three is one out of six (exactly so only if the dice is ideal).

In everyday language the probability is, however, understood is a wider sense. One can, for example, speak about the probability of rain tomorrow, even though the event is unique and there is no way its frequency could be measured by repeated experiments. Moreover, different people can give the same event different probability. This is natural since different people have different background knowledge and beliefs.

The interpretation of Bayesian probability theory is very close to everyday language. Probability expresses how strongly someone believes in something. Belief is always subjective and depends on background knowledge. Notation P(A | B) means: how true A seems if B is assumed. Often all the background knowledge is denoted and P(A) can thus mean different things depending on which background assumptions are used. It is good to remember, however, that according to Bayesian interpretation there is no absolute probability since there doesn't exist an absolutely correct set of background assumptions.

Sometimes the interpretation of probability has no effect on how the actual computations are conducted or what is the result. For the probabilities in dice throwing, for example, the interpretation has no significance. However, from the point of view of learning and intelligent systems, the difference in interpretation is significant.

Boolean algebra (George Boole 1854)

Propositions, for which the probabilities are defined, obey the rules of Boolean algebra. It is defined for elements which have two binary operations, sum and product, and an unary operation, complement, which will be denoted here by ¬. The set of axioms defining the Boolean algebra is

There exist elements 0 and 1, which are not equal.		[A1]
AB = BA	A+B = B + A	[A2]
A(B+C) = (AB)+(AC)	A+(BC) = (A+B)(A+C)	[A3]
1A = A	0+A = A	[A4]
A¬A = 0	A+¬A = 1	[A5]

The axioms on the same row are dual. If the product and sum, and 0 and 1 are exchanged, one can transform between the dual axioms. Let's denote the axioms on the left hand column by a and right hand by b, i.e., A2b means the axiom AB = BA. From the axioms one can derive the following lemmas

¬¬A = A		[L1]
AA = A	A+A = A	[L2]

¬1 = 0	¬0 = 1	[L3]
AB = 0 & A+B = 1 => B = ¬A		[L4]
0A = 0	1+A = 1	[L5]
A(A+B) = A	A+AB = A	[L6]
A(BC) = (AB)C	A+(B+C) = (A+B)+C	[L7]
¬A(AB) = 0	¬A+(A+B) = 1	[L8]
¬(AB) = ¬A+¬B	¬(A+B) = ¬A¬B	[L9]
AB = 1 => A = 1	A+B = 0 => A = 0	[L10]

Boolean logic will be obtained when only the elements 0 and 1 are taken in the algebra. Zero is interpreted as false and one as truth. Product means the and, sum the or and complement the negation operation.

The basic rules of Bayesian probability theory

The Bayesian probability theory can be based on a few simple rules. It is evident that a proposition and its negation are related. According to the sum rule their probabilities sum up to one.

Sum Rule: P(A | B) + P(¬A | B) = 1

If one wishes to verify the truth of AB, one can first verify A and then verify B assuming A. Hence P(AB | C) is evidently a function of P(A | C) and P(B | AC). The product rule states that this function is a product.

Product Rule: P(AB | C) = P(A | C) P(B | AC)

Probability is a real number between zero and one. The probability is not defined if the background assumptions, premisses, conflict. P(A | B¬B), for example, is undefined.

Other rules

Using the rules of arithmetics and Boolean algebra, all other rules of Bayesian probability theory can be derived from the sum and product rule. Let's take the derivation of the generalised sum rule for example. In what follows, the rule that will be applied is denoted at each step, unless only the rules of basic arithmetics are applied.

P(A+B | C) =	[L1]
P(¬¬(A+B)) | C) =	[L7b]
P(¬(¬A¬B) | C) =	[Sum Rule]
1 - P(¬A¬B | C) =	[Product Rule]
1 - P(¬A | C) P(¬B | ¬AC) =	[Sum Rule]
1 - P(¬A | C) [1 - P(B | ¬AC)] =	
1 - P(¬A | C) + P(¬A | C) P(B | ¬AC) =	[Sum Rule]
P(A | C) + P(¬A | C) P(B | ¬AC) =	[Product Rule]
P(A | C) + P(¬AB | C) =	[A2a]
P(A | C) + P(B¬A | C) =	[Product Rule]
P(A | C) + P(B | C) P(¬A | BC) =	[Sum Rule]
P(A | C) + P(B | C) [1 - P(A | BC)] =	

$P(A \mid C) + P(B \mid C) - P(B \mid C)\, P(A \mid BC) =$ [Product Rule]
$P(A \mid C) + P(B \mid C) - P(BA \mid C) =$ [A2a]
$P(A \mid C) + P(B \mid C) - P(AB \mid C)$

Usually, of course, not all the intermediate results are presented. From the sum and product rule, also the equations $P(1 \mid A) = 1$ and $P(A \mid B) > 0 \Rightarrow P(A \mid AB) = 1$ can be derived. Let's denote $x = P(1 \mid A)$. Then

$$1 - x = 1 - P(1 \mid A) = P(0 \mid A) = P(10 \mid A) = P(1 \mid A)\, P(0 \mid 1A) = x(1 - x) \Rightarrow x^2 - 2x + 1 = 0,$$

whose only solution is $x = 1$. On the other hand,

$$P(A \mid B) = P(AA \mid B) = P(A \mid B)\, P(A \mid AB),$$

and it follows that $P(A \mid AB) = 1$ if $P(A \mid B) > 0$.

Marginalisation principle

Let's assume that $B_1, B_2, \ldots, B_n$ are n propositions, one of which is true. Thus $B_1 + B_2 + \ldots + B_n = 1$ and $B_iB_j = 0$, unless $i = j$. The generalised sum rule yields

$$P(AB_1+AB_2 \mid C) = P(AB_1 \mid C) + P(AB_2 \mid C) - P(AB_1AB_2 \mid C) = P(AB_1 \mid C) + P(AB_2 \mid C).$$

This follows from $AB_1AB_2 = A(B_1B_2) = A0 = 0$. Adding AB_3 gives

$$P(AB_1+AB_2+AB_3 \mid C) = P(AB_1 \mid C) + P(AB_2 \mid C) + P(AB_3 \mid C) - P((AB_1 + AB_2)AB_3 \mid C) = P(AB_1 \mid C) + P(AB_2 \mid C) + P(AB_3 \mid C).$$

Continuing to AB_n results in

$$P(AB_1 + AB_2 + \ldots + AB_n \mid C) = P(AB_1 \mid C) + P(AB_2 \mid C) + \ldots + P(AB_n \mid C).$$

On the other hand, since $AB_1 + AB_2 + \ldots + AB_n = A(B_1 + B_2 + \ldots + B_n) = A1 = A$, we have

$$P(A \mid C) = P(AB_1 \mid C) + P(AB_2 \mid C) + \ldots + P(AB_n \mid C).$$

By applying the product rule we get the marginalisation principle

$$P(A \mid C) = P(A \mid B_1C)\, P(B_1 \mid C) + \ldots + P(A \mid B_nC)\, P(B_n \mid C).$$

The significance of the principle become clear, then the propositions B_i are interpreted as possible explanations for A. The probability of A is thus the sum of probabilities which different explanations give for A weighed by the probabilities of the explanations.

Bayes' rule

The Bayes' rule can be derived from the product rule. It tells how the probabilities of explanantions change, when A is observed.

$$P(B_i \mid AC) = P(B_i \mid C)\, P(A \mid B_iC) / P(A \mid C)$$

$P(B_i \mid C)$ is the probability before the knowledge about A and it is called the prior probability of B_i. Correspondingly, $P(B_i \mid AC)$ is called the posterior probability of B_i. One can see from the Bayes' rule that the posterior probabilities of explanations B_i which explain A well are higher than the prior probabilities and vice versa.

An example hopefully illuminates the use of Bayes' rule. A = I have fever, B_1 = I have a flu and B_2 = no flu = $\neg B_1$. Let's assume that I know the probabilities $P(A \mid B_1C)$, $P(A \mid B_2C)$ and $P(B_1 \mid C)$, i.e., the probabilities of having fever when having flu, of having fever without having flu and of having flu in the first place. Let's assing them the numerical values $P(A \mid B_1C) = 0.95$, $P(A \mid B_2C) = 0.05$ and $P(B_1 \mid C) = 0.1$. According to the marginalisation principle, the probability of having fever is

$$P(A \mid C) = P(A \mid B_1C)\, P(B_1 \mid C) + P(A \mid B_2\, C)\, P(B_2 \mid C) = 0{,}95 * 0{,}1 + 0{,}05 * 0{,}9 = 0{,}095 + 0{,}045 = 0{,}14.$$

The probability of having flu is originally fairly small, only one in 10. If it now turns out that I have fever, the probability of flu increases

$$P(B_1 \mid AC) = P(B_1 \mid C)\, P(A \mid B_1C) / P(A \mid C) = 0{,}1 * 0{,}95 / 0{,}14 = 0{,}68..$$

Together the marginalisation principle and the Bayes' rule tell how the belief in a hypothesis changes when observations are made and how the beliefs in hypotheses are taken into account when making predictions based on them.

Probability density

With real valued quantities, the probability of any particular value is usually 0. If, for instance, according to a measurement the length of a pencil is about 16 cm, the probability of the length being *exactly* 16 cm is zero. The probability that the length is between 15 cm and 17 cm can, in contrast, easilly be very close to one.

The phenomenon is tha same as in measuring a mass. If one takes a single point of an object, it doesn't have any mass. If one takes a volume instead, the mass differs from zero. Just like the density of an object equals to the mass divided by volume, the probability density is the probability of a range divided by its length.

The Bayes' rule will remain the same also when using probability densities.

Often probability mass is denoted by capital P and density by lower case p, but usually it becomes clear from the contex whether probability mass or density is ment.

Level 3 (upper level)

Last updated 15.10.1998

Harri Lappalainen
<Harri.Lappalainen@hut.fi>

EXHIBIT C

FULLY REDACTED

EXHIBIT D

FULLY REDACTED