Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US2006/025441

International filing date: 28 June 2006 (28.06.2006)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/694,661
Filing date: 28 June 2005 (28.06.2005)

Date of receipt at the International Bureau: 24 August 2006 (24.08.2006)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

1506784

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 18, 2006

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: ***60/694,661***
FILING DATE: ***June 28, 2005***
RELATED PCT APPLICATION NUMBER: ***PCT/US06/25441***

THE COUNTRY CODE AND NUMBER OF YOUR PRIORITY APPLICATION, TO BE USED FOR FILING ABROAD UNDER THE PARIS CONVENTION, IS ***US60/694,661***

UNITED STATES PATENT AND TRADEMARK OFFICE
DEPARTMENT OF COMMERCE

Certified by

Jon W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

Attorney Docket No.: CHS05-01[illegible]

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant(s): Jayendu Patel and Dinesh Gopinath
Serial No.: Unassigned
For: STATISTICAL SYSTEM FOR TARGETING ADS
Filing Date: June 28, 2005

PROVISIONAL PATENT APPLICATION COVER SHEET AND TRANSMITTAL LETTER

Certificate of Express Mail Under 37 C.F.R. §1.10

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as Express Mail in an envelope addressed to: Commissioner for Patents, PO Box 1450, Alexandria, Virginia 22313-1450 on:

Date: June 28, 2005

By: Mary A. Maietta
(Typed or printed name of person mailing Document, whose signature appears below)

Express Mail Mailing
Label No.: EV 403 711 385 US

Signature: Mary A. Maietta

Commissioner for Patents
PO Box 1450
Alexandria, Virginia 22313-1450

Sir:

Transmitted herewith for filing is a Provisional Patent Application entitled:

STATISTICAL SYSTEM FOR TARGETING ADS

Named Inventor(s) / Inventor Address(es) / Inventor Citizenship(s):

Jayendu Patel 11 Bolton Street Somerville, MA 02143	Citizenship: India
Dinesh Gopinath 3 Arlington Street, #41 Cambridge, MA 02140	Citizenship: India

WRITTEN ASSERTION OF SMALL ENTITY STATUS

The undersigned Registered Patent Attorney hereby asserts that the Applicant(s) are entitled to small entity status under 37 C.F.R. 1.27.

Enclosed is/are:

[x] Provisional Patent Application Cover Sheet and Transmittal Letter including fee calculations (this form, 3 pages, in duplicate), Total Pages: 6;
[x] Provisional Patent Application, figures disbursed throughout, Total Pages: 11;
[x] Return Receipt Pre-paid Postcard (in duplicate), Total Postcards: 2;
[x] Check in the amount of: $100.00, with fee amounts calculated as follows:

SMALL ENTITY Basic Filing Fee =	$100.00
Each Additional 50 Sheets that Exceeds 100 Sheets Fee (0 X $125.00) =	$0.00
TOTAL FEE FOR THIS PATENT APPLICATION =	$100.00

Please direct all telephone calls and address all correspondence to:

Barry W. Chapin, Esq.
U.S.P.T.O. Registration No.: 39,934
Attorney for Applicant(s)
CHAPIN & HUANG, L.L.C.
Westborough Office Park
1700 West Park Drive
Westborough, Massachusetts 01581
Telephone: (508) 366-9600
Facsimile: (508) 616-9805
Customer No.: 022468

Applicant(s) hereby petition(s) for any extension of time which is required to maintain the pendency of this case. If there is a fee occasioned by this response, including an extension fee, that is not covered by an enclosed check, please charge any deficiency to Deposit Account No. 50-0901.

If any questions concerning this application remain, the Mail Room and/or the Application Branch is respectfully requested to contact the undersigned collect at (508) 366-9600, in Westborough, Massachusetts.

Respectfully submitted,

By: ______________________

Barry W. Chapin, Esq.
U.S.P.T.O. Registration No.: 39,934
Attorney for Applicant(s)
CHAPIN & HUANG, L.L.C.
Westborough Office Park
1700 West Park Drive
Westborough, Massachusetts 01581
Telephone: (508) 366-9600
Facsimile: (508) 616-9805

Attorney Docket No.: CHS05-01p

Dated: June 28, 2005

Certificate of Express Mail Under 37 C.F.R. §1.10

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as Express Mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on:

Date: June 28, 2005 Express Mailing Label No.: EV 403 711 385 US

Signature: Mary A. Maietta

Typed or Printed Name: Mary A. Maietta

Inventors: Jayendu Patel and Dinesh Gopinath

Attorney Docket No.: CHS05-01p

Statistical System for Targeting Ads

Jayendu Patel, Chief Scientist, ChoiceStream
Dinesh Gopinath, VP of Analytics, ChoiceStream
First Draft: January 2005
Last Revised: June 2005

A novel practical system optimally targets a set (portfolio) of advertisements (ads) or sponsored links (SLs) or promotions based on (1) attributizing URLs associated with the ads/links, (2) inferring preferences of users who are to be shown a set of sponsored links from their activity/demographics/recent-searches, and (3) combining attributes and preferences in the user's context to identify the desired portfolio to be displayed. In the attached exhibits, we illustrate with the case of a particular initial application where we surface a personalized portfolio of four sponsored links to be displayed at the foot of an email browser, the use of the browser itself providing no insight about the user's interests/moods. However our novel method is not limited to such particular cases and we sketch in the concluding section and elsewhere several other situations where the method applies readily.

Our system maximizes the expected revenue from clicks on sponsored links, which can also be straightforwardly adapted to goals such as maximizing click-rate or maximizing user interest in the served links/ads. We explicitly cover two practical cases: (A) The eligible choiceset of SLs may be directly examined and accessed (ChoiceSet A) such as when the client using our system has a direct relationship with the SL owners; and (B) The eligible choiceset of sponsored links is only accessible indirectly via keywords passed to a SL-Server that externally returns links appropriate to key-words (ChoiceSet B). ChoiceSet B applies, for instance, when the client using our system has a revenue-sharing arrangement with Google to use its SL Server to obtain SLs to display on the client's site. For specificity in the discussion below with ChoiceSet B, the sponsored link sets that are returned by the SL Server are assumed to be required to be displayed to the user without further filtering. If further filtering is permitted, then aspects of the techniques associated with the case of ChoiceSet A can help improve the relevance of the SLs received from the SL Server.

A title for the frame in which the SLs are presented may be available to be personalized and is an example of another feature that can be usefully addressed by our system. In the discussion below, such a personalized title for the frame is subsumed in the general "*Context*" term that also generally is a place-holder for user behaviors/interests that predictably vary with the particular nature of the application/site in which the SLs are presented.

In this document, we use the term "ad" to stand in for all types of advertising and related marketing content that lends itself to targeting and which includes "normal advertisements", "banner ads", "sponsored links", "promotions", and "discount pricing".

1. Information Used or Inferred

Examples of sponsored link attributes: clothing-retail, travel, music, promotional, mainstream v eclectic, prestige-appeal, titillating, gaming.

Examples of user-profile information (such as available on members of portals like AOL, Yahoo! or of shopping services like Amazon, eBay, BestBuy, etc.): demographics (such as age and gender), visits to areas (such as sports, celebrities, news, music, shopping), recent search queries.

Example of context: time-of-day, day-of-week, purpose of area where sponsored links/ads are being served.

Example of information that must be tracked on ad/link serving and activity: Track each ad served to each member – non-clicks as well as clicks. Retain information on complete available context (such as date-time of SL, prevailing user profile, SQ that surfaced SL, ID of surfacing method, ...).

Example of targeted framing of a personalized selection of sponsored links: "for the hip girl", "Best Bets for Value Shopping", ...

2. KW-SQ-SL System

We first introduce notation and make preliminary remarks on the variables. Then we discuss the general system. Finally, we lay out two specific implementations of the general system in detail.

2.1 Notation and Variables

Notation:

Pr(.) = probability of event in parentheses

SL = sponsored link (stand-in for ads, promotions, coupons, etc.)

KW = key word used to fetch sponsored links from Sponsored-Link Server as necessary

SQ = vector of search queries made recently by user

U = vector of user's profile beside information on user's search queries

c(U) = user's cohort based on U, possibly latent

A = vector of relevant-to-user attributes of SL

X = vector of context attributes, where context is one in which links/ads are being served, etc.

Rev() = revenue to portal or site from click (or other success outcome)

Note that X (*Context*) includes attention to information on application where the ads/links are to be displayed (such as on a travel site versus a finance site versus a health site) as well as information on date-of-display (such as weekday, holidays or weekend) and time-of-display (such as workday hours or evening), i.e., all measurable factors besides general attributes of the user that predict variations in propensity to click. For example, the user's interests and click behavior in the run-up to Valentine's Day is likely to be different from that around SuperBowl. And late-night usage entails different moods than usage during the work-day.

The relevant attributes, A, of any SL can be imputed by an attributizer that analyzes the associated webpage/websiteURL or by explicit information provided by the creator of the link/ad. The attributizer can be an automated system or use human scorers or a combination – for example, see ChoiceStream's attributizer of webpages/websites.

Relevant information of the user is the U-vector. In practice, we address measurement errors for U by introducing latent cohorts and Bayesian exchangeability.

2.2 General Set-Up

The typical set-up of our targeting system seeks to maximize expected revenue by choice of a portfolio of SLs. Consider the simpler case where we desire to find the best single SL for a user:

(1) $$SL^* = \underset{SL}{\arg\max} \Pr(click \mid A,U,X) \times Rev(SL)$$

A conditional logit model (or similarly a probit model) for the click probabilities is:

(2) $$\Pr(click \mid U,A,X) = \frac{\exp(I_{A,X|U})}{1 + \exp(I_{A,X|U})}$$

where the index $I_{A,X|U} = A b_{1c(U)} + X b_{2c(U)} + AX b_{3c(U)}$ has cohort-specific coefficients and allows for needed interactions between A and X. In practice, we find cohort differences such as cohorts based on gender, age, and recent visit-area information.

We model the latent-cohort membership model, typically necessary, in a parallel way to equation (2). Thus,

(3) $$\Pr(c \mid U) = \frac{\exp(U g_c)}{\sum_{c'=1}^{K} \exp(U g_{c'})}$$

where we have K latent cohorts (-- typically three to five latent cohorts proved adequate in our estimations for targeted sponsored links).

The coefficients of the latent-cohort click-choice model described in equations (2) and (3) are estimated by maximum likelihood or by Bayesian methods, where the latter proving more robust. The latent-cohort conditional logit model for the targeting of SLs is estimated from data of observed user-clicks (***and non-clicks***) on the SLs that are served up. The click data are from similar contexts to the use of the application (or adjusted otherwise). In practice, the click rate on SLs can be low (often below 1%); in such cases, we find that using all data with the rare click-events, say N observations, can be combined with a random sample of 10N of non-click observations to obtain efficient unbiased estimates of the desired slope coefficients.

The system lends itself to straightforwardly integrate out terms to accommodate users for whom U is only known incompletely. Thus,

(4) $$\Pr(click \mid A,U_1,X) = \int \Pr(click \mid A,U,X) g(U \mid U_1) dU$$

where U_1 is an incomplete profile.

2.3 Case of No Direct Information on Key Words Relevant to User and Only Indirect Access to Pool of Sponsored Links

Consider the case where the sponsored link serving system does not have direct access to the candidate set of sponsored links, such as when passing key words (KWs) to a Sponsored Link Server which returns links

(ChoiceSet B). The task is to find good KWs to submit to the SL Server given a user with profile U. Our practical system is shown in Exhibit 1 for this scenario. The optimizer formally addresses:

$$(5) \qquad KW^* = \arg\max_{KW} \Pr(click \mid KW, U, X)' \; Rev(KW)$$

Here $Rev(KW)$ denotes the expected revenue if the top sponsored link associated with the KW is clicked. For cases like the Google's Sponsored Links Server, the $Rev(KW)$ function is straightforwardly estimated by directly looking up recent prices associated with the top position for the KW.

To make equation (5) operational, we appear to need A of the SLs that will typically returned by the KW submissions. We can examine sponsored links associated with KWs using historical data of KW→SL mappings for this. In fact, the direct attributization of the KW based on the top ten webpages/websites returned from an internet search based on the KW provides a useful proxy for A. The latter task is simple to implement using ChoiceStream's SQ attributizer.

Exhibit 1 outlines the key steps and logical framework for the generation of personalized sponsored links in this setting. For Exhibit 1, we have:

- *Keyword Pool*: At any point in time, we have a choice set of eligible keywords constructed from a previous pool of keywords with new keyword additions and drops based on what the system learned during the previous time period, say last week. Additions of keywords stem from editorially selected keywords, search queries with high incidence and click rates in the previous time period, etc. Keywords with low clickrates in the previous period are candidates for keyword drops. This process ensures that the keyword pool is relevant over time and across users.
- *Keyword Click Rate Model*: The keyword clickrate model predicts the clickrate for the keyword for each user given user's profile and past click activity and keyword attributes. Keyword attributes include attributes of search results and sponsored links generated by the keyword, associated positive and negative token buckets from clickability considerations, and how often the keyword was served up for the user in the recent past.
- *Keyword Portfolio Optimization:* If revenue per click is available, then we optimize for revenue per KW impression subject to portfolio considerations to induce "variety" in the sponsored links returned. In this optimization step we also incorporate potential drop-off in clickrate for lower ranking sponsored links from the same KW. For example, depending on KW clickrates, revenue per click and drop-off in clickrates over sponsored link position, we may server one or more sponsored links from one or more optimized keywords. If the return from optimized keywords is lower than an acceptable threshold, typically set based on context and the application, then "default" keywords or banner ads are shown.

2.4 Case of Direct Information on KeyWords Relevant to User (such as history of user's recent searches)

We attend to information from last SQ for user's propensity to click. The recent search queries of the user serve two purposes:

- Candidates for the choice set from which the optimal portfolio of key words can be used to fetch sponsored links
- Inform the system pertaining to the current *ephemeral* interests of the user as inferred by the attributes of the person's recent search queries.

Exhibit 2 outlines the key steps and logical framework for the generation of personalized sponsored links with information of user's recent search queries, SQ.

SQ is initially grouped through a fast clustering algorithm to identify clusters of similar search queries. Depending on the application context, rules and policies of the website or customer may preclude a subset of the recent search queries from being used as candidate search query pool. Such queries are filtered and the click rates are predicted for the remainder of the query pool that forms the candidate set of KWs. The attributes used in predicting click rates include:

- *Attributes based on a pre-computations for a large pool of likely search queries*: We maintain a cache of millions of common search queries along with their attributes so that we can quickly predict the click-rates of most search queries in a particular user's SQ.
- *Positive and negative token counts*: Based on historical SQ impressions and click data, a repository of positive and negative tokens of search queries is constructed and used to generate positive and negative tokens as attributes for the click rate model.
- *Staleness*: The relevance of the search query drops with the time since the query; consequently, "staleness" proves to be an important attribute.
- *Cluster size*: The relative ephemeral interests in a search query may be related to the size of the cluster of the search query.

Similar to the keyword portfolio optimization step, we generate a portfolio of recent search queries for generating sponsored links for the user. If the return from optimized search queries is lower than an acceptable threshold, typically set based on context and the application, and then we invoke the optimization process with keywords as in section 2.3

2.5 Portfolio considerations

In the illustrations of sections 2.3 and 2.4, we indicate a portfolio optimization step. Our targeting system induces variety in the set of presented sponsored links through the following types of mechanisms:

- *Clustering of attributes of keywords (applies to section 2.3)*: Given the taxonomy that is used to attributize ads/sponsored links, we may induce variety in the sponsored links by diversifying over attributes. For example, if the top candidate KWs for a user are "baseball cap", "basketball", and "50 cent", then we use "baseball cap" and "50 cent" to obtain sponsored links; the system drops "baseball" and "basketball" since they belong to the "Sports" cluster from which "baseball cap" is the highest value KW.
- *Clustering of recent search queries (see discussion of section 2.4)*: Recent search queries are tokenized and passed through a clustering algorithm to identify clusters of search queries. These clusters server two goals:
  - Induce variety in the search queries chosen to generate sponsored links by skipping over clusters. For example, if the user's history of search queries had "baseball cap,", "baseball", "50 cent" in the search history, then we keep only one from the Sports cluster.
  - Identify the intensity of the user's current interest in a particular area/category and which is positively related to the likelihood of the user's click to sponsored links in the area.

2.6 Spanning Pr(Click) Models

While we discuss a specific latent-cohort conditional logit model that we have used in our targeting SL approach, there are alternative statistical/machine-learning/engineering models that may be useful complementary predictors of clicks across the user base. In such situations, we update the model's suitability to the user based on her on clicks or lack thereof; i.e., if the user does not click on ads relying on her model being M , assign her ads based on model $M¢$that is now the best model capturing her Pr(Click).

3. Extensions for Constraints

The setup herein allows many useful constraints to be straightforwardly introduced, such as accommodating guaranteed number of displays by ad per period (day/week). For the latter case, consider the strategy of maximizing the expected value to users (or advertisers) subject to showing the ads to the required number of users in the period.

Let ad attributes be indexed by j=1,...,J, i.e., each ad belongs to one of $A_1,..A_j,...,A_J$. The problem of decomposing the ad-type probability into that for K cohorts with preferences for specific ads within the ad-type is a straightforward extension. User profiles are indexed by c=1,...,K, i.e., each user belong to one of K cohorts (possibly latent). The expected number of users from cohort c in the period of interest is n_c. We seek probabilities for serving ad-types to users based on their cohort-type. This is available applying the methods described in section 2 above. Let $k_c(A_j)$ denote the probability with which ad j is to be served to cohort c. Of course, k is bounded in [0,1] and $\sum_j k_c(A_j) = 1$. The contractual constraints are expressed stochastically as $N_j = \sum_c n_c k_c(A_j)$ where N_j is the total number of times that the ad-type j is required to be shown. The problem is then:

(6) find $k_c(A_j)$ to $\max \sum_c n_c \sum_j f_c(V \mid A_j) k_c(A_j)$ subject to N_j constraints.

Here, parallel to the f-function in the previous section, $f_c(V \mid A_j)$ captures the value of the ad-type for a cohort c. The solution demands $(J - 1) \times K$ probabilities, though many of them are likely zero. The solution is non-trivial in general but numerical Monte Carlo optimization methods can help (-- for example, see chapter 5 of Robert & Casella, 2004).

4. Optimizing Advertiser's Bids for Key Words

Till now, the targeting system was enabled the targeting of a set of SLs/Ads/Promotions to a user such that we maximize the benefits accruing to the owner of a site of portal where the SLs are being served. However the system can be straightforwardly modified to help advertisers find ***underpriced*** KWs. These applications arise in contexts where search engines (such as Google and Yahoo!) allow advertisers to buy KWs such that an advertiser's SLs are surfaced whenever a user enters a KW to initiate a search. Similarly, some sites (or group of sites) allow advertisers to buy KWs such that the advertiser's SLs/Ads become the click-through site for the user who clicks on the purchased KW on the seller-site. This extension to optimize the advertiser's purchases of KWs follows.

The expected value to the advertiser from a user depends on the user's profile, U, and the context in which the user is shown the SL/ad (i.e., all clicks-through are not equal to the advertiser). Let $W(click \mid U, KW, X)$ denote the value to the advertiser of a click by user U who has interest in the keyword KW in context X. The specification of W is an input for our system. We then compute the expected benefit-to-cost ratio, BC hereafter, for any KW:

$$(7) \qquad BC = \int \Pr(click \mid A,U,X) \frac{W(click \mid U,KW,X)}{R(KW)} f(U \mid KW,X) dU$$

where $f(U \mid KW, X)$ is the distribution of the user profiles interested in keyword KW in context X. Here A are the user-relevant attributes of the KW, and are obtained following the discussion of section 2.3 above. Other notation is as applied in section 2. Note that R() is a revenue to the site selling the KW and the object of optimization before, but is now a cost to the advertiser buying the KW. Note how we reuse the Pr() function whose estimation was laid out in section 2 for targeting ads. Clearly an advertiser can use estimates of equation (7) by itself to weed out poor-to-the-advertiser KWs.

The expected expenditure with the purchase of a key word is estimable as:

$$(8) \qquad Cost(KW) = N(KW,X) \int \Pr(click \mid A,U,X) R(KW) f(U \mid KW,X) dU$$

where N is the expected number of users for the KW in context X.

Assume for simplicity that the same ad is surfaced across different KWs purchased by the advertiser. Further assume that there are no covariances in benefits across user-types attracted to the advertiser. In this case, the optimal portfolio of KWs for the advertiser can be constructed sequentially. Buy the KW that has the highest benefit-to-cost ratio (establishing a cap if the expected number of clicks exceed the budget). If the budget is not expected to be exhausted by the purchase of the first KW, buy the KW with the second highest benefit-to-cost ratio, etc.

Quite generally, a model which sheds light on the demand function is valuable not only to distributors but also to producers.

4.1 Advertiser's Surplus

Advertisers expect to spend, on average, 39 percent more on all search marketing programs (organic search engine optimization, paid placement, paid inclusion, and SEM technology) in 2005 compared to 2004; smaller firms projected 32 percent more, while larger firms (larger than 500 employees) projected a 43 percent year-over-year increase. SEM agencies projected a bullish budget, with overall gross revenue increases for 2005 of 79 percent on average. Online ad spending was $9-10 billion in 2004. For many advertisers, "Brand awareness is #1 objective with search marketing programs."

In this world, the surfacer of sponsored links relying on inventory from a SL Server maintained by a Search Engine will generally wind up providing an uncompensated advertiser's surplus. This arises because the expected click rate outside the search-word context for which the advertiser originally signed up for is estimated in practice by our system to be (much) lower. The advertiser's click-through payment to the search engine is really paying for the bundle of general brand-building via display and the expected value from a click-through in the search-engine context. For example, suppose the click-through-rate is 5%, the value per such click-through is 10 cents to the advertiser, and the brand-awareness value per exposure is 1 cent. Then the advertiser can sensibly bid up to 30 cents per click-through [$= (100\times1 + 5\times10)/5$]. As click-through rate drops, the maximum payable price per click-through rises (-- for example, the max price with the illustrative numbers in this paragraph rises to $1.10 per click-through with a click-through rate of 1%). While this indicates that firms with meaningful brands in play should pay more for ad-words than just the expected benefit of the transaction associate with a click-through, it also indicates that surfacers of sponsored links encounter under-pricing because of their indirect access via SL Servers owned by Search Engines. For an advertiser, the analysis surrounding equations (7) and (8) can be straightforwardly extended to account for brand-awareness benefits, both direct in the Search Engine context but also in the context of indirect surfacers of the SLs.

5. Other Applications For Our Method

The listing of scenarios heretofore of the applications of our targeting methods is indicative but not exhaustive. As before, we use the term ad to stand in for all types of advertising and related marketing content that lends itself to targeting and which includes "normal advertisements", "banner ads", "sponsored links", "promotions", and "discount pricing". Thus, our invention applies to the following cases where ads may be surfaced:

1. Ads on web portals, web sites, web pages. Such ads can be drawn from the sites own inventory or from calls to an external ad server.
2. Ads via the incorporation of our system on user's set-top boxes to merge in with content for television display.
3. Ads at physical sites where user experiencing ad can be identified.
4. Ads targeted to users who may have complete or partial taste profiles.
5. Ads where context is uninformative about user's current interests, where only information on the permanent interests is available or inferrable.
6. Ads where context is informative about user's current interests, such as when a history of the user's recent search history is available or where history of recent clicks is available or where history of recent areas of visits is available.

7. More generally, surfacing of ads where some information of user is known and where there is a choiceset of ads to draw from (including choosing default of "no ad" when better than "inappropriate ad").

Our invention also applies to optimizing the purchase of keywords (tags quite generally) by advertisers from search engines or other sites selling keywords as discussed in section 4.

Our invention also applies to the optimal design of ads by advertisers who assess the cost-benefit applicable to each design following the approach of this invention outlined in section 4.

Exhibit 1. Case of no information of particular-interest-keywords to user

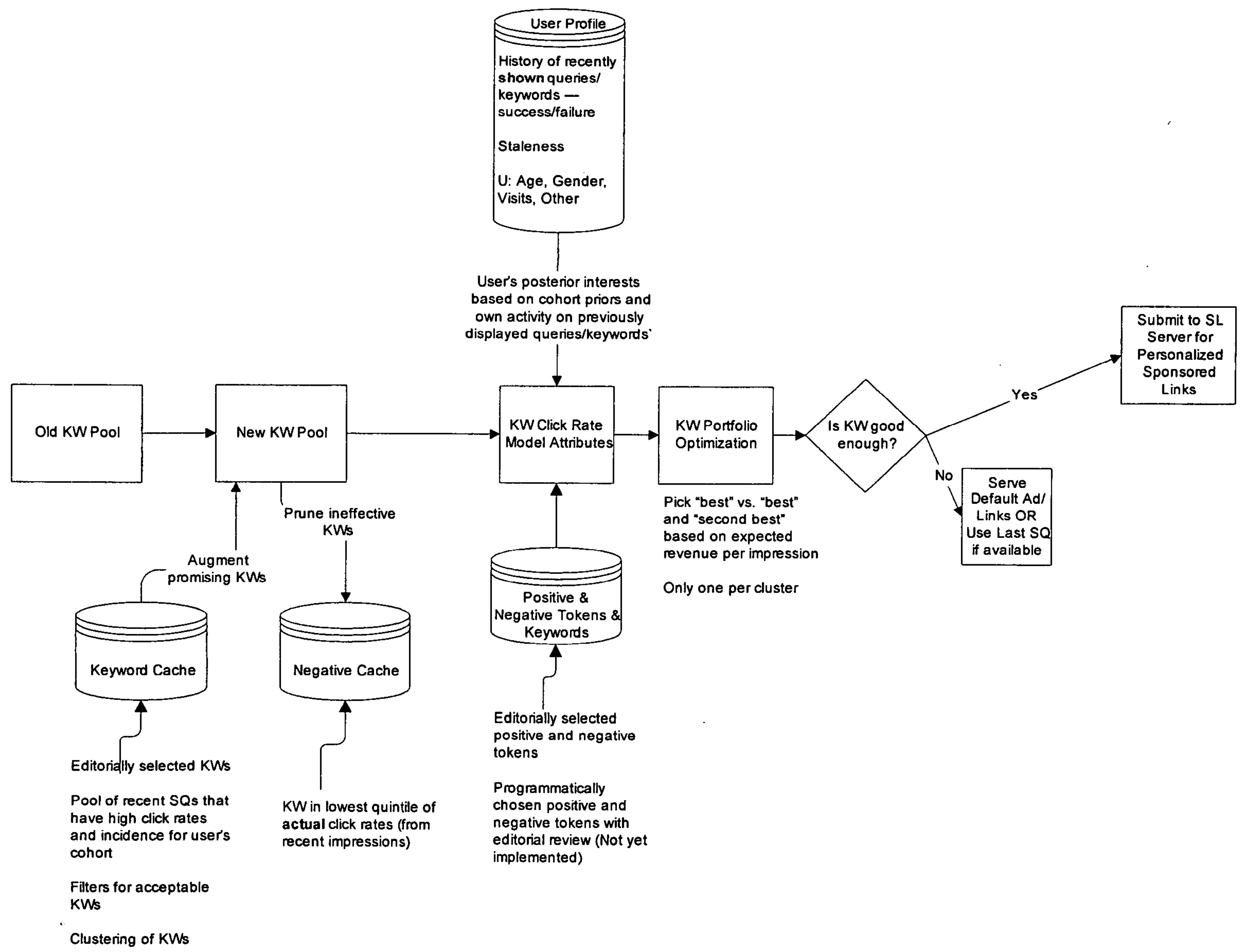
User Profile
History of recently shown queries/ keywords — success/failure
Staleness
U: Age, Gender, Visits, Other
User's posterior interests based on cohort priors and own activity on previously displayed queries/keywords`
Old KW Pool
New KW Pool
KW Click Rate Model Attributes
KW Portfolio Optimization
Is KW good enough?
Yes
No
Submit to SL Server for Personalized Sponsored Links
Serve Default Ad/ Links OR Use Last SQ if available
Pick "best" vs. "best" and "second best" based on expected revenue per impression
Only one per cluster
Prune ineffective KWs
Augment promising KWs
Keyword Cache
Negative Cache
Positive & Negative Tokens & Keywords
Editorially selected KWs
Pool of recent SQs that have high click rates and incidence for user's cohort
Filters for acceptable KWs
Clustering of KWs
KW in lowest quintile of actual click rates (from recent impressions)
Editorially selected positive and negative tokens
Programmatically chosen positive and negative tokens with editorial review (Not yet implemented)

Exhibit 2. Case of information on particular-interest-keywords to user such as recent search queries (SQ)