

(19) World Intellectual Property  
Organization  
International Bureau



(43) International Publication Date  
8 April 2004 (08.04.2004)

PCT

(10) International Publication Number  
**WO 2004/029827 A1**

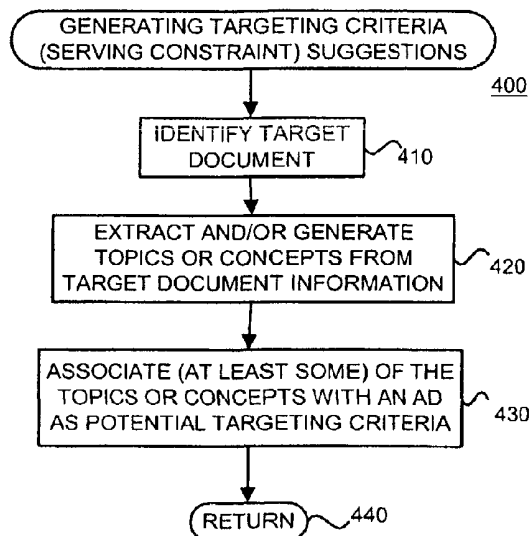
- (51) International Patent Classification<sup>7</sup>: **G06F 17/21**, 17/30, 17/60
- (21) International Application Number: PCT/US2003/030234
- (22) International Filing Date: 24 September 2003 (24.09.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/413,536 24 September 2002 (24.09.2002) US  
10/314,427 6 December 2002 (06.12.2002) US  
10/389,688 14 March 2003 (14.03.2003) US
- (71) Applicant: **GOOGLE, INC.** [US/US]; 2400 Bayshore Parkway, Mountain View, CA 94043 (US).
- (72) Inventors: **DEAN, Jeffrey, A.**; 884 Fifteenth Avenue, Menlo Park, CA 94025 (US). **HARIK, Georges, R.**; 950 High School Way, Apt. #3135, Mountain View, CA 94041 (US). **BUCHHEIT, Paul**; 900 High School Way, Apt. #2324, Mountain View, CA 94041 (US).
- (74) Agent: **SUN, Karl**; Google, Inc., 2400 Bayshore Parkway, Mountain View, CA 94043 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Declarations under Rule 4.17:**

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,*

[Continued on next page]

(54) Title: SUGGESTING AND/OR PROVIDING AD SERVING CONSTRAINT INFORMATION



(57) Abstract: Targeting information (also referred to as ad "serving constraints") or candidate targeting information for an advertisement is identified (1). Targeting information may be identified (410) by extracting topics or concepts (420) from, and/or generating topics or concepts based on, ad information, such as information from a Web page to which an ad is linked (or some other Web page of interest to the ad or advertiser) (400). The topics or concepts may be relevant queries associated with the Web page of interest, clusters, etc.



KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SK, SL, TJ, TM, TN, TR,

TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

**Published:**

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

## **SUGGESTING AND/OR PROVIDING AD SERVING CONSTRAINT INFORMATION**

### RELATED APPLICATIONS

- 5           This application claims the benefit of U.S. Provisional Application Serial  
No. 60/413,536, (incorporated herein by reference) entitled "METHODS AND  
APPARATUS FOR SERVING RELEVANT ADVERTISEMENTS", filed on  
September 24, 2002 and listing Jeffrey Dean, Georges Harik, Paul Bucheit as  
inventors, U.S. Patent Application Serial No. 10/314,427 (incorporated herein by  
10   reference), entitled "METHODS AND APPARATUS FOR SERVING RELEVANT  
ADVERTISEMENTS", filed on December 6, 2002 and listing Jeffrey Dean,  
Georges Harik, Paul Bucheit as inventors, and U.S. Patent Application Serial No.  
10/\_\_\_\_,\_\_\_\_ (incorporated herein by reference), entitled "SERVING  
ADVERTISEMENTS BASED ON CONTENT", filed on February 26, 2003, and  
15   listing Darrell Anderson, Paul Bucheit, Alex Carobus, Claire Cui, Jeffrey A. Dean,  
Georges R. Harik, Deepak Jindal, and Narayanan Shivakumar as inventors.

### BACKGROUND OF THE INVENTION

#### A.   Field of the Invention

- 20           The present invention concerns advertising, such as on-line advertising.  
More particularly, the present invention concerns helping on-line advertisers to  
target their ads.

#### B.   Description of Related Art

- 25           Advertising using traditional media, such as television, radio, newspapers  
and magazines, is well known. Advertisers have used these types of media to  
reach a large audience with their advertisements ("ads"). To reach a more  
responsive audience, advertisers have used demographic studies. For example,  
advertisers may use broadcast events such as football games to advertise beer  
30   and action movies to a younger male audience. However, even with  
demographic studies and entirely reasonable assumptions about the typical

audience of various media outlets, advertisers recognize that much of their ad budget is simply wasted because the target audience is not interested in the ad they are receiving.

Interactive media, such as the Internet, has the potential for better  
5 targeting of advertisements. For example, some websites provide an information search functionality that is based on query keywords entered by the user seeking information. This user query can be used as an indicator of the type of information of interest to the user. By comparing the user query to a list of keywords specified by an advertiser, it is possible to provide some form of  
10 targeted advertisements to these search service users. An example of such a system is the Adwords system offered by Google, Inc.

While systems such as Adwords have provided advertisers the ability to better target ads, advertisers would like to have help so that they can target their ads more effectively and/or more easily.

15 It would be useful, therefore, to have methods, apparatus and data structures for helping advertisers to target their ads more effectively and/or more easily.

#### SUMMARY OF THE INVENTION

20 The present invention addresses this and other needs by identifying targeting information (also referred to as ad "serving constraints" or "targeting criteria") for an advertisement. Targeting information may be identified by extracting topics or concepts from, and/or generating topics or concepts based on, ad information, such as for example advertising creatives, information from a  
25 Web page to which an ad is linked, etc. Targeting information may also be identified by extracting topics or concepts from, and/or generating topics or concepts based on, other information relevant to the delivery of an ad, such as for example documents (e.g., Web pages) on which an ad is likely to be shown, documents (e.g., Web pages) on which an advertiser is or may be interested in  
30 advertising, documents (e.g., Web pages) on which competitive, analogous, or complementary products or services are advertised or to which such are linked,

documents (e.g., Web pages) on which competitive, analogous or complementary products or services are offered or described, information related to the geographic targeting or temporal targeting of an ad, documents (e.g., Web pages) similar to any of the above identified documents or sources of information, etc. or in general any other document (e.g., Web page) of interest to the advertiser.

In one embodiment of the invention, the topics or concepts may be words or relevant queries associated with a Web page of interest. In another embodiment of the invention, the topics or concepts may be clusters. Thus, in one embodiment of the present invention, keyword suggestions may be provided to advertisers for keyword triggered advertisements based on topical or conceptual clusters extracted from one or more documents (e.g., Web pages) of interest. Alternatively, or in addition, concept, topic and/or cluster suggestions may be provided to advertisers for concept, topic and/or cluster triggered advertisements.

Additional aspects of the present invention are directed to computer systems and to machine-readable media having features relating to the foregoing aspects.

## BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate an embodiment of the invention and, together with the description, explain the invention. In the drawings,

Fig. 1 is a diagram illustrating an environment within which the invention may be implemented;

Fig. 2 is a diagram functionally illustrating an advertising system consistent with the invention;

Fig. 3 is a diagram illustrating apparatus with which the invention may be implemented;

Fig. 4 is a flow diagram of an exemplary method for identifying (e.g., candidate) targeting criteria for an advertiser, consistent with the present invention;

Fig. 5 is an exemplary apparatus that may be used to perform various acts or store various information in a manner consistent with principles of the present invention.

Fig. 6 is an example in which a Web page is a target document.

#### DETAILED DESCRIPTION

The following detailed description of the invention refers to the accompanying drawings. The detailed description does not limit the invention.

The present invention involves methods, apparatus and data structures for determining (e.g., candidate) targeting criteria from a target document or documents. The targeting criteria may then be provided to an advertiser as suggested targeting criteria. In one implementation, the target document is a Web page and the advertisements are electronic files that are capable of being rendered on, or in association with, that Web page. A set, such as a list, of concepts or topics corresponding to the Web page may be generated by analyzing the content of the Web page. There are a variety of techniques by which this may be performed, one of which is by computing a term vector for the Web page and selecting the top N terms from that vector. The list of concepts or topics may then be provided as targeting criteria associated with an advertisement (e.g., keywords specified for the advertisements).

Those skilled in the art will recognize that many other implementations are possible, consistent with the present invention.

#### **A. Exemplary Environment and Architecture**

Fig. 1 is a diagram illustrating an exemplary environment within which the invention may be used. The exemplary environment includes an advertiser 110, an advertising system 120, an advertisement consumer 130, and an advertising target 140.

Advertiser 110 may be the party that directly sells the goods or services being advertised (e.g., Amazon.com) or an agent acting on the advertiser's behalf. The advertisement desired by advertiser 110 may exist in a variety of forms. For example, the advertisement may be a standard print advertisement,  
5 an online advertisement, a text advertisement, and image advertisement, an audio advertisement, a video advertisement, audio/visual advertisements, or any other type of sensory message, or combination of the foregoing desired.

Advertising system 120 interfaces with both the advertiser 110 and the advertisement consumer 130. It may perform a variety of functions, as explained  
10 in more detail below in reference to Fig. 2. This invention may be used with such an advertising system 120.

Advertisement consumer 130 is the entity that will issue a request for advertisements to advertising system 120, obtain the advertisements from advertising system 120, and present the advertisement to the advertising target  
15 140. Typically, the advertisement consumer is the entity that provides the content with which the advertisement is to be associated. In one implementation, the advertising consumer 130 is a search engine, such as that employed by Google, Inc. at [www.google.com](http://www.google.com).

Advertising target 140 is the individual (or set of individuals) who ultimately  
20 receive the advertisement. In the case of visual advertisements, for example, the advertisement target 140 is the person who views the advertisement.

Fig. 2 is a diagram functionally illustrating an advertising system consistent with the invention. The system includes an ad campaign entry and management component 210, a tools component 220, a billing component 230,  
25 one or more databases 240, an ad consumer interface component 250, an ad selection component 260, an ad ordering component 270, an ad serving component 280, and a statistics engine component 290. If the present invention is to be used with such an advertising system, it will primarily concern tools component 220. To help understand the invention, other components of the  
30 advertising system will be explained below. Furthermore, although Fig. 2 shows a particular arrangement of components constituting advertisement system 120,

those skilled in the art will recognize that not all components need be arranged as shown, not all components are required, and that other components may be added to, or replace, those shown.

Ad entry and management component 210 is the component by which the advertiser enters information required for an advertising campaign and manages the campaign. In one embodiment, an ad campaign contains one or more advertisements that are related in some manner. For example, the Ford Motor Company may have an ad campaign for zero percent financing, which could contain a series of advertisements related to that topic. Among the other things that could be provided by an advertiser through ad entry and management component 210 are the following: one or more advertising creatives (simply referred to as "ads" or "advertisements"), one or more set of keywords or topics associated with those creatives (which may be used as targeting information for the ads), geographic targeting information, a value indication for the advertisement, start date, end date, etc. The data required for, or obtained by, ad entry and management component 210 resides in one of the databases 240. In one embodiment of the ad system, the ad entry and management component 210 may perform operations such as those described in: (i) U.S. Patent Application Serial No. 10/340,867 (incorporated herein by reference), entitled "PAUSING ONE OR MORE ADS, ONE OR MORE AD GROUPS, AND/OR ON OR MORE AD CAMPAIGNS", filed on January 10, 2003 and listing Salar Kamangar and Amit Patel as inventors; and/or (ii) U.S. Patent Application Serial No. 10/340,193 (incorporated herein by reference), entitled "PRICING ACROSS KEYWORDS ASSOCIATED WITH ONE OR MORE ADVERTISEMENTS", filed on January 10, 2003 and listing Eric Veach as the inventor.

Tools component 220 contains a variety of tools designed to help the advertiser 110 create, monitor, and manage its campaigns. For example, tools component 220 may contain a tool for helping advertiser 110 estimate the number of impressions an ad will receive for a particular keyword or topic. Similarly, tools component 220 may be used to help advertiser 110 define targeting criteria, such as a list of keywords, for a given advertisement. Other



possible tools may be provided as well. Depending on the nature of the tool, one or more databases 240 may be used to gather or store information. In one embodiment of the ad system, the tools component 220 may perform operations such as those described in U.S. Provisional Application Serial No. 60/\_\_\_\_, \_\_\_\_  
5 (incorporated herein by reference), entitled "METHOD AND APPARATUS FOR ESTIMATING ELECTRONIC ADVERTISING INVENTORY", filed on January 10, 2003 and listing Magnus Sandburg, Eric Veach, John Bauer and Zhe Qian as inventors.

Billing component 230 helps perform billing-related functions. For  
10 example, billing component 230 generates invoices for a particular advertiser 110 or ad campaign. In addition, billing component 230 may be used by advertiser 110 to monitor the amount being expended for its various campaigns. The data required for, or obtained by, billing component 230 resides in a database 240. In one embodiment of the ad system, the billing component 230 may perform  
15 operations such as those described in: (i) U.S. Provisional Application Serial No. 60/424,792 (incorporated herein by reference), entitled "AUTOMATED PRICE MAINTENANCE FOR USE WITH A SYSTEM IN WHICH ADVERTISEMENTS ARE RENDERED WITH RELATIVE PREFERENCE BASED ON PERFORMANCE INFORMATION AND PRICE INFORMATION", filed on  
20 November 8, 2002 and listing Eric Veach as the inventor; (ii) U.S. Patent Application Serial No. 10/340,543 (incorporated herein by reference), entitled "AUTOMATED PRICE MAINTENANCE FOR USE WITH A SYSTEM IN WHICH ADVERTISEMENTS ARE RENDERED WITH RELATIVE PREFERENCES", filed on January 10, 2003 and listing Eric Veach and Salar Kamangar as inventors;  
25 and/or (iii) U.S. Patent Application Serial No. 10/340,542 (incorporated herein by reference), entitled "AUTOMATED PRICE MAINTENANCE FOR USE WITH A SYSTEM IN WHICH ADVERTISEMENTS ARE RENDERED WITH RELATIVE PREFERENCE BASED ON PERFORMANCE INFORMATION AND PRICE INFORMATION", filed on January 10, 2003 and listing Eric Veach and Salar  
30 Kamangar as inventors.

Databases 240 contain a variety of data used by advertising system 120. In addition to the information mentioned above in reference to ad entry and management system 210, databases 240 may contain statistical information about what ads have been shown, how often they have been shown, the number of times they have been selected, who has selected those ads, how often display  
5 of the ad has led to consummation of a transaction, etc. Although the databases 240 are shown in Fig. 2 as one unit, one of ordinary skill in the art will recognize that multiple databases may be employed for gathering and storing information used in advertising system 120.

10 Ad consumer interface 250 is a component that interfaces with ad consumer 130 to obtain or send information. For example, ad consumer 130 may send a request for one or more advertisements to ad consumer interface 250. The request may include information such as the site requesting the advertisement, any information available to aid in selecting the advertisement,  
15 the number of ads requested, etc. In response, ad consumer interface 250 may provide one or more advertisements to ad consumer 130. In addition, ad consumer 130 may send information about the performance of the advertisement back to the ad system via the ad consumer interface 250. This may include, for example, the statistical information described above in reference to a database  
20 240. The data required for, or obtained by, ad consumer interface component 250 resides in a database 240.

Ad selection component 260 receives a request for a specified number of advertisements, coupled with information to help select the appropriate advertisements. This information may include, for example, a search query  
25 specified by an end user. Alternatively, or in addition, as described in more detail below, this information may include data related to the content of the page for which the advertisements are being requested. In one embodiment of the ad system, the ad selection component 260 may perform operations such as those described in: (i) U.S. Provisional Application Serial No. 60/413,536 (incorporated  
30 herein by reference), entitled "METHODS AND APPARATUS FOR SERVING RELEVANT ADVERTISEMENTS", filed on September 24, 2002 and listing

Jeffrey Dean, Georges Harik and Paul Bucheit as inventors; (ii) U.S. Patent Application Serial No. 10/314,427 (incorporated herein by reference), entitled "METHODS AND APPARATUS FOR SERVING RELEVANT

ADVERTISEMENTS", filed on December 6, 2002 and listing Jeffrey Dean,

- 5 Georges Harik and Paul Bucheit as inventors; and/or (iii) U.S. Patent Application Serial No. 10/\_\_\_\_,\_\_\_\_ (incorporated herein by reference), entitled "SERVING ADVERTISEMENTS BASED ON CONTENT", filed on February 26, 2003 and listing Darrell Anderson, Paul Bucheit, Alex Carobus, Claire Cui, Jeffrey Dean, Georges Harik, Deepak Jindal, and Narayanan Shivakumar as inventors.

- 10 Ad ordering component 270 receives a list of relevant ads from ad selection component 260 and determines a preference order (or some other preference attribute such as size, position, volume, etc.) in which they should be rendered to an end user. For example, relevant ads may be ordered based on a value indication associated with each ad. These ordered ads may be provided to
- 15 an ad serving component 280. In one embodiment of the ad system, the ad ordering component 270 may perform operations such as those described in: (i) U.S. Provisional Application Serial No. 60/317,847 (incorporated herein by reference), entitled "METHODS AND APPARATUS FOR ORDERING ADVERTISEMENTS", filed on September 6, 2001 and listing Jane Manning,
- 20 Salar Kamangar and Eric Veach as inventors; (ii) U.S. Patent Application Serial No. 10/112,656 (incorporated herein by reference), entitled "METHODS AND APPARATUS FOR ORDERING ADVERTISEMENTS BASED ON PERFORMANCE INFORMATION", filed on March 29, 2002 and listing Jane Manning, Salar Kamangar, Eric Veach and Lawrence Page as inventors; and/or
- 25 (iii) U.S. Patent Application Serial No. 10/112,654 (incorporated herein by reference), entitled "METHODS AND APPARATUS FOR ORDERING ADVERTISEMENTS BASED ON PERFORMANCE INFORMATION AND PRICE INFORMATION", filed on March 29, 2002 and listing Salar Kamangar, Ross Koningstein and Eric Veach as inventors.

- 30 Ad serving component 280 receives an ordered list of ads from ad ordering component 270, and formats that list into a manner suitable for

presenting to ad consumer 130. This may involve, for example, rendering the ads into hypertext markup language (HTML), into a proprietary data format, etc.

Statistics engine 290 contains information pertaining to the selection and performance of advertisements. For example, statistics engine 290 may log the information provided by ad consumer 130 as part of an ad request, the ads selected for that request by ad selection component 260, the order selected by ad ordering component 270, and the presentation of the ads by ad serving component 280. In addition, statistics engine 290 may log information about what happens with the advertisement once it has been provided to ad consumer 130. This includes information such as on what location the ad was provided, what the response was to the advertisement, what the effect was of the advertisement, etc.

In one embodiment of the ad system, optimization operations (not shown) may be used to optimize ad spot inventory, revenue, or both. Such operations (or operations of one of the other components described above) may include those described in: (i) U.S. Provisional Application Serial No. 60/439,174 (incorporated herein by reference), entitled "SETTING MINIMUM PRICES IN AN ADVERTISING SYSTEM", filed on January 10, 2003 and listing Eric Veach as the inventor; (ii) U.S. Patent Application Serial No. 10/350,898 (incorporated herein by reference), entitled "DETERMINING A MINIMUM PRICE", filed on January 24, 2003 and listing Eric Veach as the inventor; and/or (iii) U.S. Patent Application Serial No. 10/340,553 (incorporated herein by reference), entitled "GOVERNING THE SERVING OF ADVERTISEMENTS BASED ON A COST TARGET", filed on January 10, 2003 and listing Eric Veach as the inventor.

Online ads, such as those used in the exemplary systems described above with reference to Figures 1 and 2, or any other system, may have various features. Such features may be specified by an application and/or an advertiser. These features are referred to as "ad features" below. For example, in the case of a text ad, ad features may include a title line, ad text, executable code, an embedded link, etc. In the case of an image ad, ad features may additionally include images, etc. Depending on the type of online ad, ad features may

include one or more of the following: text, a link, an audio file, a video file, an image file, executable code, embedded information, etc.

When an online ad is served, one or more parameters may be used to describe how, when, and/or where the ad was served. These parameters are referred to as "serving parameters" below. Serving parameters may include, for example, one or more of the following: features of (including information on) a page on which the ad is served (including one or more topics or concepts determined to be associated with the page, information or content located on or within the page, information about the page such as the host of the page (e.g. AOL, Yahoo, etc.), the importance of the page as measured by e.g. traffic, freshness, quantity and quality of links to or from the page etc., the location of the page within a directory structure, etc.), a search query or search results associated with the serving of the ad, a user characteristic (e.g., their geographic location, the language they use, the type of browser used, previous page views, previous behavior), a host or affiliate site (e.g., America Online, Google, Yahoo) that initiated the request that the ad is served in response to, an absolute position of the ad on the page on which it is served, a position (spatial or temporal) of the ad relative to other ads served, an absolute size of the ad, a size of the ad relative to other ads, a color of the ad, a number of other ads served, types of other ads served, time of day served, time of week served, time of year served, etc. Naturally, there are other serving parameters that may be used in the context of the invention.

Although serving parameters may be extrinsic to ad features, they may be associated with an ad as conditions or constraints. When used as serving conditions or constraints, such serving parameters are referred to simply as "serving constraints". Serving constraints may also be referred to as "targeting criteria". For example, in some systems, an advertiser may be able to specify that its ad is only to be served on weekdays, no lower than a certain position, only to users in a certain location, etc. As another example, in some systems, an advertiser may specify that its ad is to be served only if a page or search query includes certain keywords or phrases. As yet another example, in some

systems, an advertiser may specify that its ad is to be served only if a Web page being served includes certain topics or concepts, or falls under a particular cluster or clusters, or some other classification or classifications.

“Ad information” may include any combination of ad features, ad serving  
5 constraints, information derivable from ad features or ad serving constraints (referred to as “ad derived information”), and/or information related to the ad (referred to as “ad related information”), as well as an extensions of such information (e.g., information derived from ad related information).

A “document” is to be broadly interpreted to include any machine-readable  
10 and machine-storable work product. A document may be a file, a combination of files, one or more files with embedded links to other files, etc.; the files may be of any type, such as text, audio, image, video, etc. Parts of a document to be rendered to an end user can be thought of as “content” of the document. Ad spots in the document may be defined by embedded information or instructions.  
15 In the context of the Internet, a common document is a Web page. Web pages often include content and may include embedded information (such as meta information, hyperlinks, etc.) and/or embedded instructions (such as Javascript, etc.). In many cases, a document has a unique, addressable, storage location and can therefore be uniquely identified by this addressable location. A universal  
20 resource locator (URL) is a unique address used to access information on the Internet.

“Document information” may include any information included in the document, information derivable from information included in the document (referred to as “document derived information”), and/or information related to the  
25 document (referred to as “document related information”), as well as an extensions of such information (e.g., information derived from related information). An example of document derived information is a classification based on textual content of a document. Examples of document related information include document information from other documents with links to the  
30 instant document, as well as document information from other documents to which the instant document links.

Content from a document may be rendered on a "content rendering application or device". Examples of content rendering applications include an Internet browser (e.g., Explorer or Netscape), a media player (e.g., an MP3 player, a Realnetworks streaming audio file player, etc.), a viewer (e.g., an Adobe Acrobat pdf reader), etc.

Fig. 3 is a diagram illustrating an architecture in which the present invention may be implemented. The architecture includes multiple client devices 302, a server device 310, and a network 301, which may be, for example, the Internet. Client devices 302 each include a computer-readable medium 309, such as random access memory, coupled to a processor 308. Processor 308 executes program instructions stored in memory 309. Client devices 302 may also include a number of additional external or internal devices, such as, without limitation, a mouse, a CD-ROM, a keyboard, and a display. Thus, as will be appreciated by those skilled in the art, the client devices may be personal computers, personal digital assistants, mobile phones, content players, etc.

Through client devices 302, requestors 305 can communicate over network 301 with each other and with other systems and devices coupled to network 301, such as server device 310. Requestors 305 may, for example, be advertisers 110, advertisement consumer 130, or advertising target 140.

Similar to client devices 302, server device 310 may include a processor 311 coupled to a computer readable memory 312. Server device 310 may additionally include a secondary storage element, such as a database 240.

Client processors 308 and server processor 311 can be any of a number of well known micro-processors, such as processors from Intel Corporation, of Santa Clara, California. In general, client device 302 may be any type of computing platform connected to a network and that interacts with application programs, such as a digital assistant or a "smart" cellular telephone or pager. Server 310, although depicted as a single computer system, may be implemented as a network of computer processors.

Memory 312 may contain a number of programs, such as the components described above in reference to Fig. 2.

**B. Exemplary Methods**

Fig. 4 is a flow diagram of an exemplary method 400 for determining (e.g., candidate) targeting criteria based on a target document, consistent with the present invention. For purposes of illustrating the invention in the context of an ad system such as the one depicted in Figure 2, it may be understood that the method 400 described herein may take place as part of the tools component 220, although those skilled in the art will recognize that it need not take place in that component 220, or that component 220 alone.

The exemplary method 400 is not limited by the order shown in the flow diagram. The method 400 identifies one or more target documents. (Block 410) The target document(s) may be the Web page linked with an advertisement, Web pages on which an advertiser is interested in showing its ads, or any other document as previously described. Alternatively, or in addition, the target document or documents may be determined algorithmically, based on ad information. Alternatively, or in addition, the target documents may be specified by an advertising user.

Concepts or topics are extracted and/or generated from document information associated with the identified target documents. (Block 420) At least some of the concepts or topics extracted and/or generated are then associated with the ad at issue as targeting criteria and/or potential targeting criteria. (Block 430) In one embodiment, potential targeting criteria are simply suggestions, though such potential targeting criteria may also be used to automatically target ads. In one embodiment, potential targeting criteria is used as targeting criteria unless removed (e.g., by the advertiser). In another embodiment, potential targeting criteria is not used as targeting criteria unless selected for inclusion (e.g., by the advertiser). In yet another embodiment, a first set of the potential targeting criteria is only used as targeting criteria if it (or constituents of the set) is selected for inclusion, while another set of the potential targeting criteria is used unless it (or constituents of the set) is removed. Although not shown, the topics or concept may also be useful in generating or suggesting ad information, such as text in an ad creative.



***Extracting and/or Generating Concepts or Topics***

One way to identify a topic corresponding to the target document is by analyzing some or all text within the target document. It could be determined that one or more of each term (word or phrase) that appears in the title of the target document corresponds to a topic of the target document. Alternatively, it could be determined that one or more of each term that appears in the body of the target document corresponds to a topic of the target document. In a simple case, each term within the target document could be identified as a topic. A slightly more complex approach would be to identify a term as a topic if it appears in the target document more than N times, such as  $N=2$  (and indeed such a threshold-based approach could be used whenever terms within text are being analyzed). Even more complex analysis could be performed, such as by using a term vector for the target document which assigns weights to each term. For example, terms that appear frequently in the target document may be assigned a relatively higher weight than those that appear less frequently.

In addition, the weighting could be adjusted to give higher weight to terms that appear less frequently in a collection, such as a collection to which the document belongs or the general collection of documents. In any situation where terms within text are assigned weights or scores, those resulting scores may be used to determine which terms will be identified as topics for the target document. For example, it may be determined that only the top scoring term would constitute a topic for the target document. Alternatively, or in addition, it may be determined that the top Z terms (or a subset thereof) will constitute topics for the target document, with Z being some defined number. Alternatively, or in addition, it may be determined that terms having a score that exceeds Y (or a subset thereof) will constitute topics for the target document, with Y being some defined number. Thus, topics may be determined based on absolute and/or relative criteria.

Alternatively, or in addition to using text or other information within the target document, meta-information associated with the target document may be

used. In the context of a Web page, this may be anchor text. One or more such brief descriptions may be used to revise (figuratively) the target document by supplementing or replacing some or all of its content with the brief descriptions. So, for example, the topic could be identified from the combination of the target  
5 document's title and the brief descriptions of the target document.

Alternatively, or in addition to the brief descriptions from these references, the references themselves may be used. For example, a reference from another document to the target document may be used as an indication that the two documents are similar. Alternatively, or in addition, a reference from the target  
10 document to another document may be used as an indication that the two documents are similar. On this basis, the content (or meta-information) of the other document may be used to revise (figuratively) the target document by supplementing or replacing its content with that of the other document. The revised target document's content may then be analyzed using the techniques  
15 described above to identify one or more topics.

Alternatively, or addition to using the content (including perhaps meta-data) associated with a target document, other techniques may be used to identify one or more topics for the target document. For example, the top N queries (or subset thereof) that result in a reference to the target document could  
20 be determined to constitute a topic for the target document, with N being some defined number. These may be, for example, text queries in a search engine that yield a result that links to the target document or web page. Alternatively, or in addition, the content of other similar documents (e.g., in the same collection as the target document, in the same category as the target document, etc.) may be  
25 used to revise (figuratively) the content of the target document. Any of the techniques described above may then be used to analyze the target document to identify one or more topics. In the context of Web pages, this may be other Web pages that are stored within a subdirectory of related pages on the same host as the target Web page. Alternatively, or in addition, any technique for classifying  
30 the target document into a set of one or more topics or categories may be used. Even the search query history of one or more users who visit the target

document (or target Web page) may be used to identify a topic for the target document, on the theory that a visit to the target document that is temporally proximate to that search query history indicates that the user thought the concepts were related.

5           Using one or more of the various techniques described above, or other techniques, one or more topics may be identified for the target document. Once these topics have been identified, a variety of techniques may be used to determine other topics that are related to those identified topics. For example, a thesaurus could be used to determine other topics (e.g., synonyms) that are  
10   closely related to the identified topics, or that are conceptually similar to the identified topics.

          For the sake of clarity, the foregoing references to "revising" the target document are a figurative aid in understanding the use of additional information that is not literally within the target document. Those skilled in the art will  
15   recognize that the target document need not be actually revised to make use of this additional information.

          U.S. Provisional Application Serial No. 60/416,144, entitled "Methods and Apparatus for Probabilistic Hierarchical Inferential Learner" filed on October 3, 2002 (incorporated herein by reference) describes exemplary ways to determine  
20   one or more concepts or topics of information that may be used consistent with the principles of the present invention.

### **C.   Exemplary Apparatus**

          Figure 5 is high-level block diagram of a machine 500 that may effect one  
25   or more of the operations discussed above. The machine 500 basically includes one or more processors 510, one or more input/output interface units 530, one or more storage devices 520, and one or more system buses and/or networks 540 for facilitating the communication of information among the coupled elements. One or more input devices 532 and one or more output devices 534 may be  
30   coupled with the one or more input/output interfaces 530.

The one or more processors 510 may execute machine-executable instructions (e.g., C or C++ running on the Solaris operating system available from Sun Microsystems Inc. of Palo Alto, California or the Linux operating system widely available from a number of vendors such as Red Hat, Inc. of Durham,  
5 North Carolina) to effect one or more aspects of the present invention. At least a portion of the machine executable instructions may be stored (temporarily or more permanently) on the one or more storage devices 520 and/or may be received from an external source via one or more input interface unit s530.

In one embodiment, the machine 500 may be one or more conventional  
10 personal computers. In this case, the processing units 510 may be one or more microprocessors. The bus 540 may include a system bus. The storage devices 520 may include system memory, such as read only memory (ROM) and/or random access memory (RAM). The storage devices 520 may also include a hard disk drive for reading from and writing to a hard disk, a magnetic disk drive  
15 for reading from or writing to a (e.g., removable) magnetic disk, and an optical disk drive for reading from or writing to a removable (magneto-) optical disk such as a compact disk or other (magneto-) optical media.

A user may enter commands and information into the personal computer through input devices 532, such as a keyboard and pointing device (e.g., a  
20 mouse) for example. Other input devices such as a microphone, a joystick, a game pad, a satellite dish, a scanner, or the like, may also (or alternatively) be included. These and other input devices are often connected to the processing unit(s) 510 through an appropriate interface 530 coupled to the system bus 540. The output devices 534 may include a monitor or other type of display device,  
25 which may also be connected to the system bus 540 via an appropriate interface. In addition to (or instead of) the monitor, the personal computer may include other (peripheral) output devices (not shown), such as speakers and printers for example.

30

**D. Illustrative Operations of an Exemplary Embodiment**

Figure 6 illustrates operations of an exemplary embodiment of the invention used in the context of suggesting keywords and other targeting criteria in the context of text ads with a link to an advertiser's Web page. As shown in Figure 6, a text-based ad 610 includes title text 620, body text 630, and a link 640 to a desired Web page 660. One or more targeting criteria or serving constraints 650 may be associated with the text-based ad 610.

As further illustrated in this example, the Web page 660 linked from the text-based ad 610 includes further links (indicated by dashed arcs) to further Web pages 662 and 664. Relevant queries 675, such as past queries that have returned a link to Web page 660 and/or potential queries that would return a link to Web page 660 may be available (for example, from a search engine).

In this example, target documents 660, 662, 664 were identified (Recall, e.g., block 410.) from the ad itself 610. In this example, topics or concepts may have been extracted and/or generated (Recall block 420.) from some or all of the information 670, and/or information from the ad 610 itself. Finally, some or all of these concepts or topics may be provided as candidate targeting criteria or serving constraints 680 (Recall block 430.), such as suggested keywords, geo-location serving constraints, and/or temporal serving constraints.

**E. Conclusions**

The foregoing description of preferred embodiments of the present invention provides illustration and description, but is not intended to be exhaustive or to limit the scope of the invention to the precise form disclosed.

Modifications and variations are possible in light of the above teachings or may be acquired from practice of the invention.

## WHAT IS CLAIMED:

- 1 1. A method for generating serving constraints for an advertisement, the method  
2 comprising:
  - 3 a) identifying a target document;
  - 4 b) generating concepts from information associated with the target  
5 document; and
  - 6 c) associating at least some of the concepts with the advertisement as  
7 serving constraints or as candidate serving constraints.
- 1 2. The method of claim 1 wherein the target document is a Web page.
- 1 3. The method of claim 1 wherein the target document is identified based on a  
2 link in the advertisement.
- 1 4. The method of claim 1 wherein the target document is identified using a user  
2 input.
- 1 5. The method of claim 1 wherein the target document is identified based on  
2 concepts associated with the target document.
- 1 6. The method of claim 1 wherein the target document is a document  
2 associated with the advertisement.
- 1 7. The method of claim 6 wherein the advertisement is rendered with the  
2 target document.
- 1 8. The method of claim 1 wherein the information associated with the target  
2 document includes text extracted from the target document.

- 1 9. The method of claim 1 wherein the information associated with the target  
2 document includes a further document linked with the target document.
- 1 10. The method of claim 1 wherein the information associated with the target  
2 document includes search engine queries that have returned a link to the target  
3 document.
- 1 11. The method of claim 1 wherein the information associated with the target  
2 document includes search engine queries that would return a link to the target  
3 document.
- 1 12. The method of claim 1 wherein the serving constraints include keywords.
- 1 13. The method of claim 1 wherein the serving constraints include geo-location  
2 information.
- 1 14. The method of claim 1 wherein the serving constraints include temporal  
2 information.
- 1 15. The method of claim 1 further comprising:  
2 d) sending the serving constraints to a user for presentation;  
3 e) determining an intention of the user with regard to the presentation of the  
4 serving constraints; and  
5 f) accepting or rejecting at least some of the serving constraints responsive  
6 to the determined intention of the user.
- 1 16. Apparatus for generating serving constraints for an advertisement, the  
2 apparatus comprising:  
3 a) an input for accepting a target document;  
4 b) means for generating concepts from information associated with the target  
5 document; and

6 c) means for associating at least some of the concepts with the  
7 advertisement as serving constraints or as candidate serving constraints.

1 17. The apparatus of claim 16 wherein the target document is a Web page.

1 18. The apparatus of claim 16 wherein the target document is identified based  
2 on a link in the advertisement.

1 19. The apparatus of claim 16 wherein the target document is identified using a  
2 user input.

1 20. The apparatus of claim 16 wherein the information associated with the target  
2 document includes text extracted from the target document.

1 21. The apparatus of claim 16 wherein the information associated with the target  
2 document includes a further document linked with the target document.

1 22. The apparatus of claim 16 wherein the information associated with the target  
2 document includes search engine queries that have returned a link to the target  
3 document.

1 23. The apparatus of claim 16 wherein the information associated with the target  
2 document includes search engine queries that would return a link to the target  
3 document.

1 24. The apparatus of claim 16 wherein the serving constraints include keywords.

1 25. The apparatus of claim 16 wherein the serving constraints include  
2 geo-location information.



1 26. The apparatus of claim 16 wherein the serving constraints include temporal  
2 information.

1 27. The apparatus of claim 16 further comprising:  
2 d) a user interface adapted to  
3 - send the serving constraints to a user for presentation,  
4 - determine an intention of the user with regard to the presentation of the  
5 serving constraints, and  
6 - accept or reject at least some of the serving constraints responsive to  
7 the determined intention of the user.

1 28. The apparatus of claim 16 wherein the target document is identified based  
2 on concepts associated with the target document.

1 29. The apparatus of claim 16 wherein the target document is a document  
2 associated with the advertisement.

1 30. The apparatus of claim 29 wherein the advertisement is rendered with the  
2 target document.

1 31. A machine-readable medium storing machine-executable instructions which,  
2 when executed by a machine, effect the method of claim 1.

1 32. A method for generating content for an advertisement, the method  
2 comprising:  
3 a) identifying a target document;  
4 b) generating concepts from information associated with the target  
5 document; and  
6 c) associating at least some of the concepts with the advertisement as  
7 content or as candidate content.

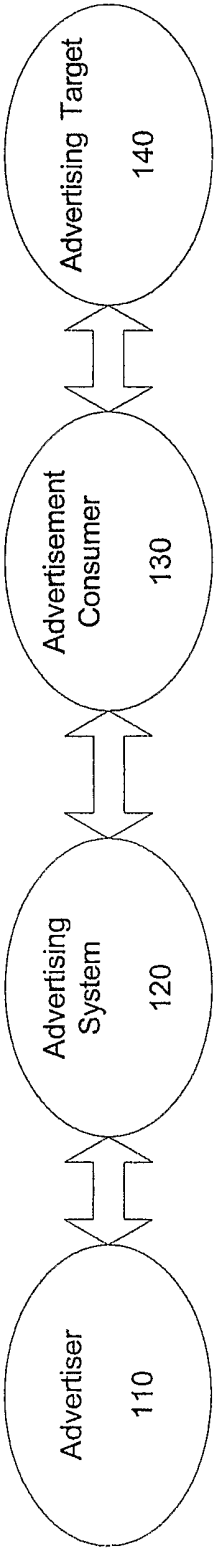


FIG. 1

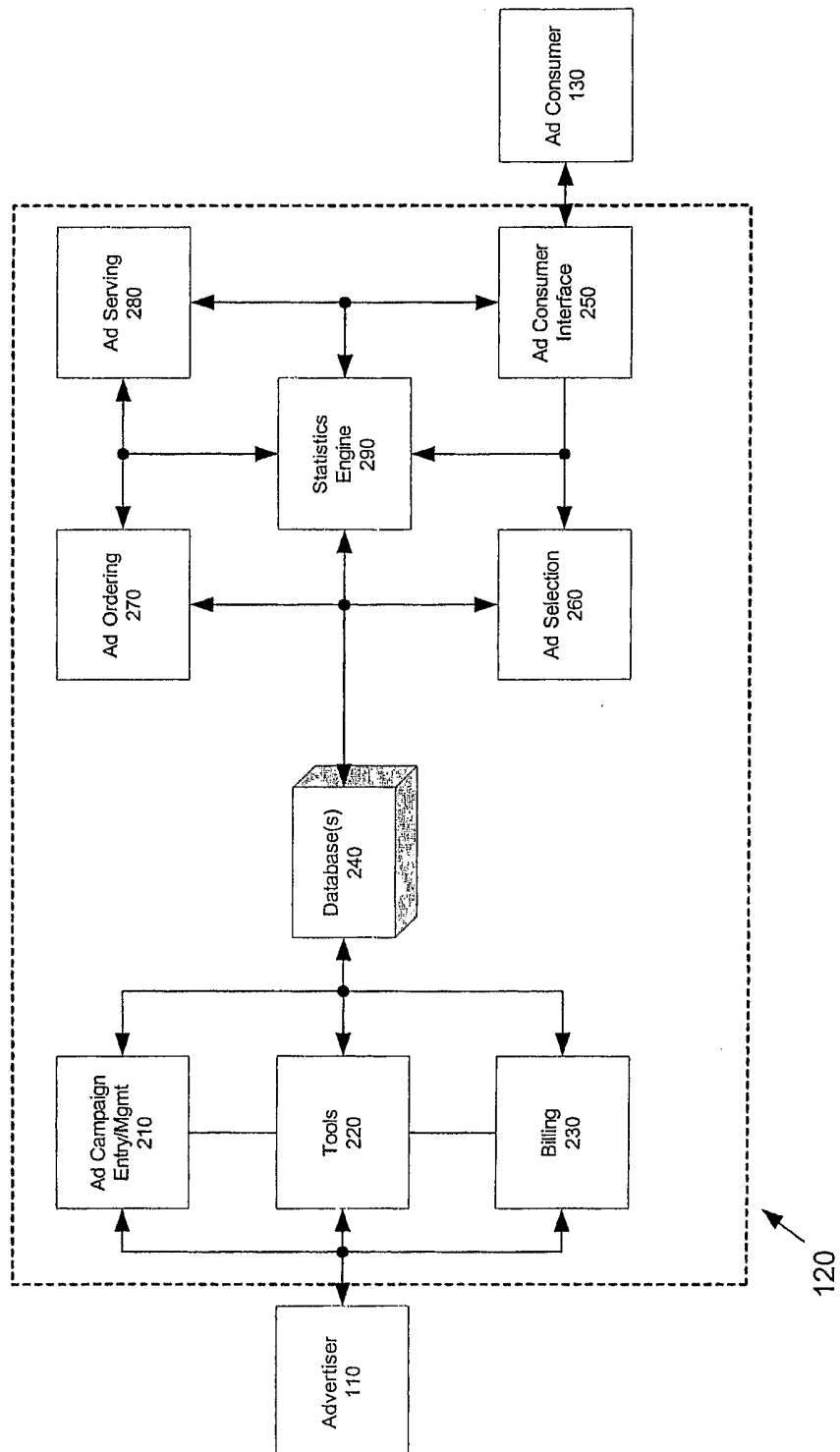


FIG. 2

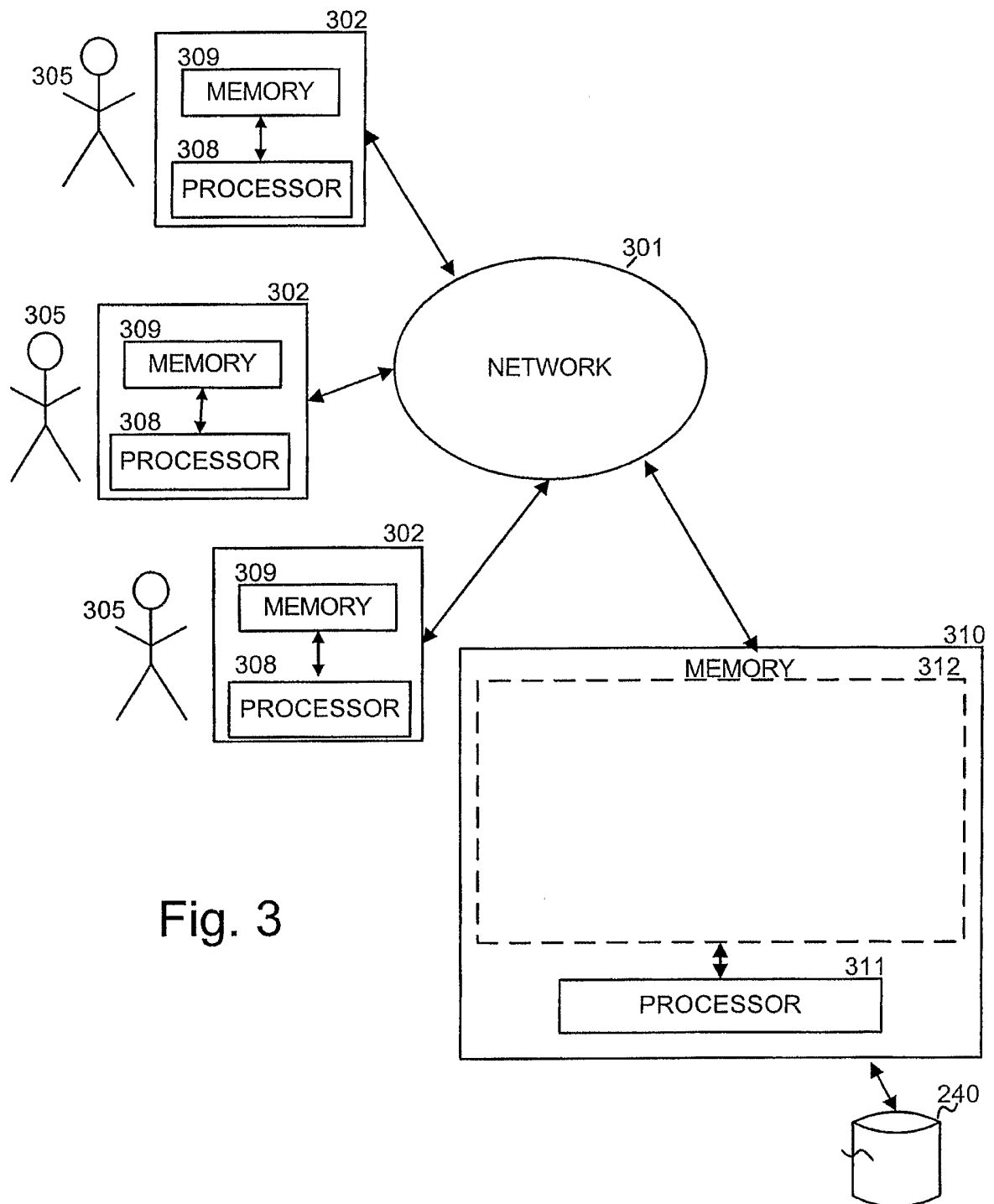


Fig. 3

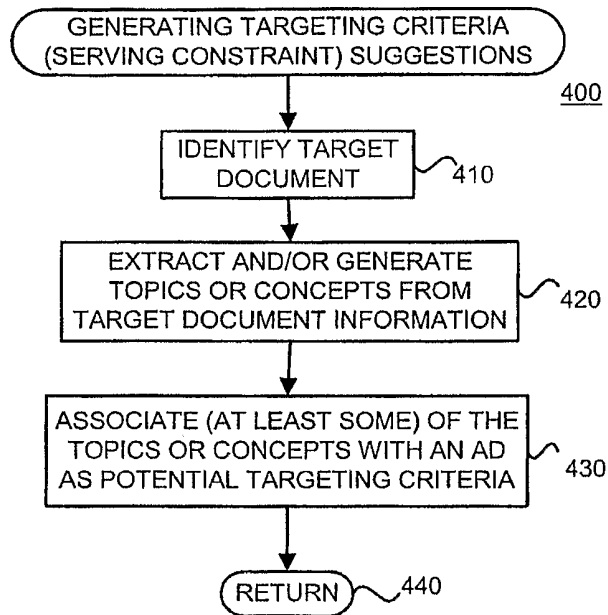


FIGURE 4

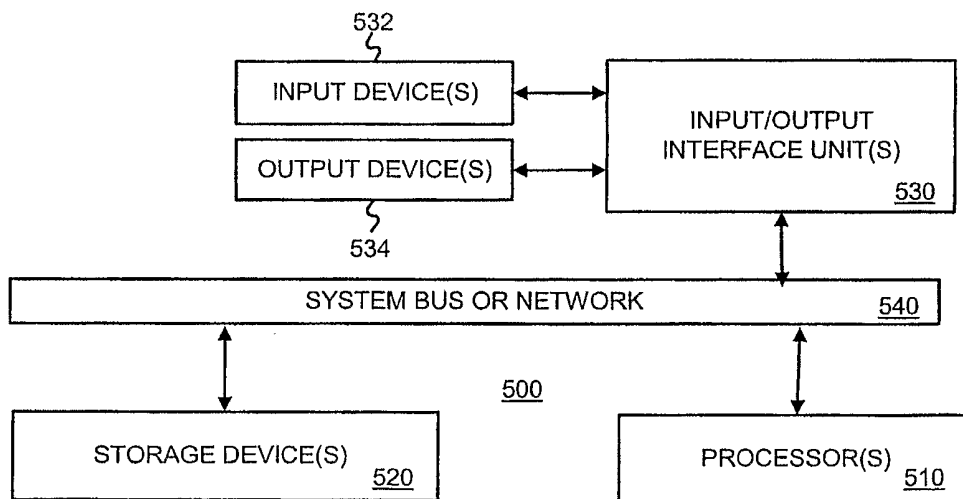


FIGURE 5

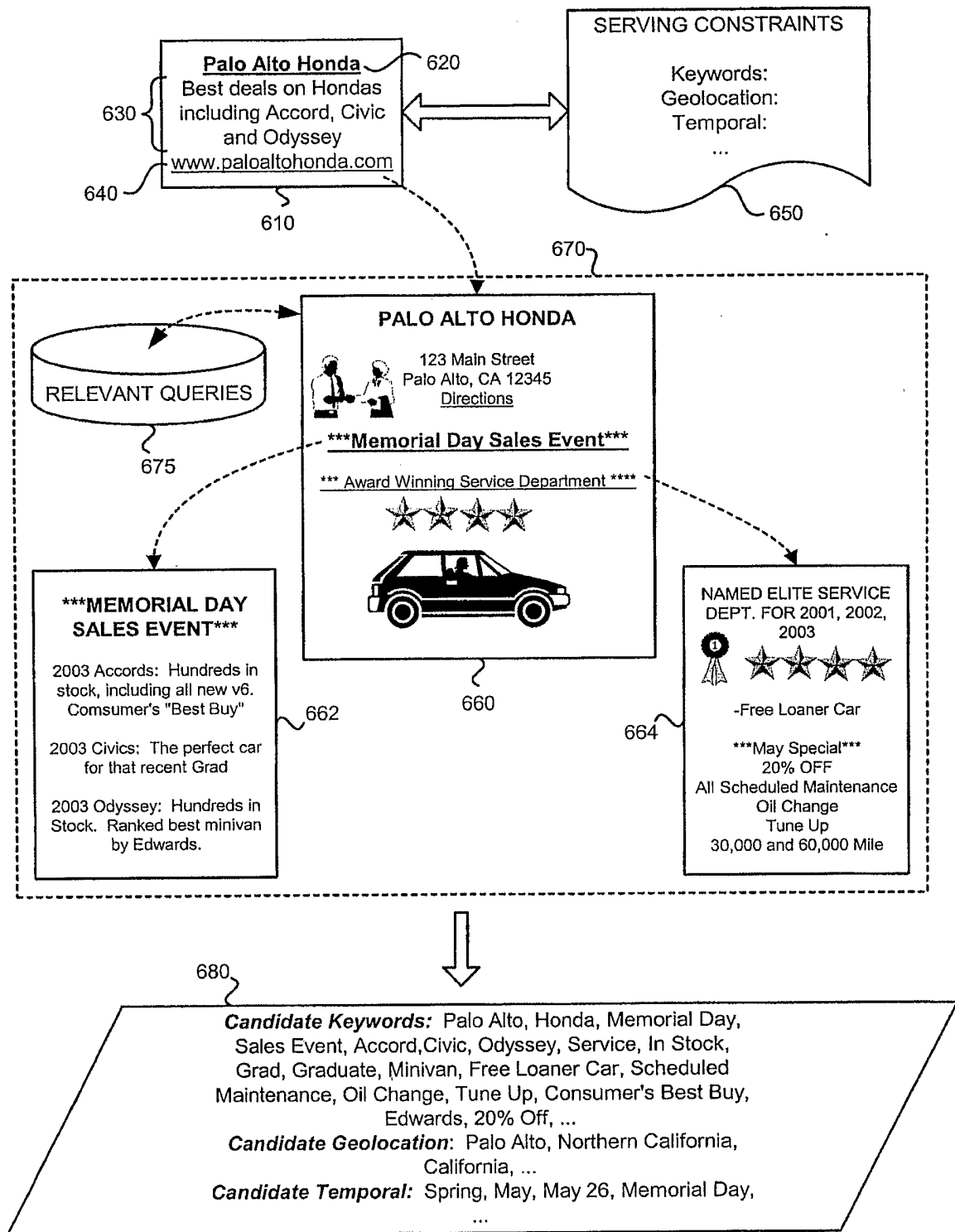


FIGURE 6

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US03/30234

<b>A. CLASSIFICATION OF SUBJECT MATTER</b>																										
IPC(7) : G06F 17/21, 17/30, 17/60																										
US CL : 705/14; 707/2, 5, 10; 709/203; 715/501.1																										
According to International Patent Classification (IPC) or to both national classification and IPC																										
<b>B. FIELDS SEARCHED</b>																										
Minimum documentation searched (classification system followed by classification symbols) U.S. : 705/14; 707/2, 5, 10; 709/203; 715/501.1																										
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched																										
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EAST, GOOGLE.COM																										
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>																										
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.																								
Y, E	US 6,665,838 B1 (BROWN et al) 16 December 2003 (16.12.2003), abstract, Figure 5 and 8, column 1, lines 33 - 57; column 6, lines 5 - 18; column 7, lines 17 - 28 and 44 - 67; column 8, lines 46 - 59 and column 12, lines 18 - 28	1 - 32																								
Y	US 6,385,592 B1 (ANGLES et al) 7 May 2002 (07.05.2002), abstract, Figures 1, 2 and 4; column 10, line 53 - column 11, line 23; column 13, lines 34 - 67.	1 - 32																								
A	US 5,887,133 A (BROWN et al) 23 March 1999 (23.03.1999), abstract.	1 - 32																								
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.																										
<table border="0"> <tr> <td colspan="2">* Special categories of cited documents:</td> <td>"T"</td> <td>later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</td> </tr> <tr> <td>"A"</td> <td>document defining the general state of the art which is not considered to be of particular relevance</td> <td>"X"</td> <td>document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</td> </tr> <tr> <td>"E"</td> <td>earlier application or patent published on or after the international filing date</td> <td>"Y"</td> <td>document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</td> </tr> <tr> <td>"L"</td> <td>document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</td> <td>"&amp;"</td> <td>document member of the same patent family</td> </tr> <tr> <td>"O"</td> <td>document referring to an oral disclosure, use, exhibition or other means</td> <td></td> <td></td> </tr> <tr> <td>"P"</td> <td>document published prior to the international filing date but later than the priority date claimed</td> <td></td> <td></td> </tr> </table>			* Special categories of cited documents:		"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention	"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone	"E"	earlier application or patent published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art	"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family	"O"	document referring to an oral disclosure, use, exhibition or other means			"P"	document published prior to the international filing date but later than the priority date claimed		
* Special categories of cited documents:		"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention																							
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone																							
"E"	earlier application or patent published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art																							
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family																							
"O"	document referring to an oral disclosure, use, exhibition or other means																									
"P"	document published prior to the international filing date but later than the priority date claimed																									
Date of the actual completion of the international search 20 February 2004 (20.02.2004)		Date of mailing of the international search report 09 MAR 2004																								
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703) 305-3230		Authorized officer <i>Michelle L. Sison</i> Shahid Al Alam Telephone No. (703) 305-3900																								