



(19) **United States**

(12) **Patent Application Publication**  
Nance et al.

(10) **Pub. No.: US 2008/0243780 A1**

(43) **Pub. Date: Oct. 2, 2008**

(54) **OPEN PROFILE CONTENT IDENTIFICATION**

(73) Assignee: **GOOGLE INC.**, Mountain View, CA (US)

(75) Inventors: **Megan Nance**, Sunnyvale, CA (US); **Mayur Datar**, Santa Clara, CA (US); **Julie Tung**, Mountain View, CA (US); **Bahman Rabii**, Berkeley, CA (US); **Jason C. Miller**, San Francisco, CA (US); **Mike Hochberg**, Sunnyvale, CA (US); **Jeremiah Harmsen**, Mountain View, CA (US); **Tomasz J. Tunguz-Zawislak**, San Francisco, CA (US); **Andres S. Perez-Bergquist**, Mountain View, CA (US)

(21) Appl. No.: **11/694,307**

(22) Filed: **Mar. 30, 2007**

**Publication Classification**

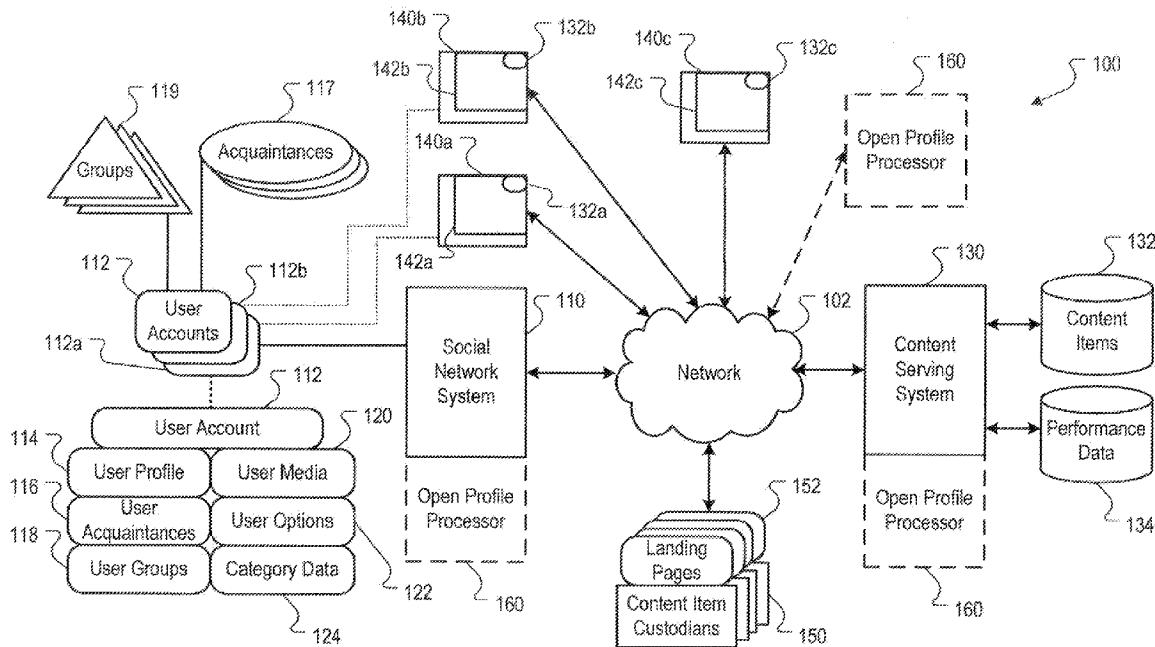
(51) **Int. Cl.**  
**G06F 7/06** (2006.01)  
**G06F 17/30** (2006.01)

(52) **U.S. Cl.** ..... **707/3; 707/E17.014**

(57) **ABSTRACT**

Open profile data in a user profile, e.g., free-form fields in a user profile, are processed to identify interests and preferences of the user. The interests and preferences are utilized to identify categories associated with the user profile, and content items, e.g., advertisements, can be identified based on the categories.

Correspondence Address:  
**FISH & RICHARDSON P.C.**  
**PO BOX 1022**  
**MINNEAPOLIS, MN 55440-1022 (US)**



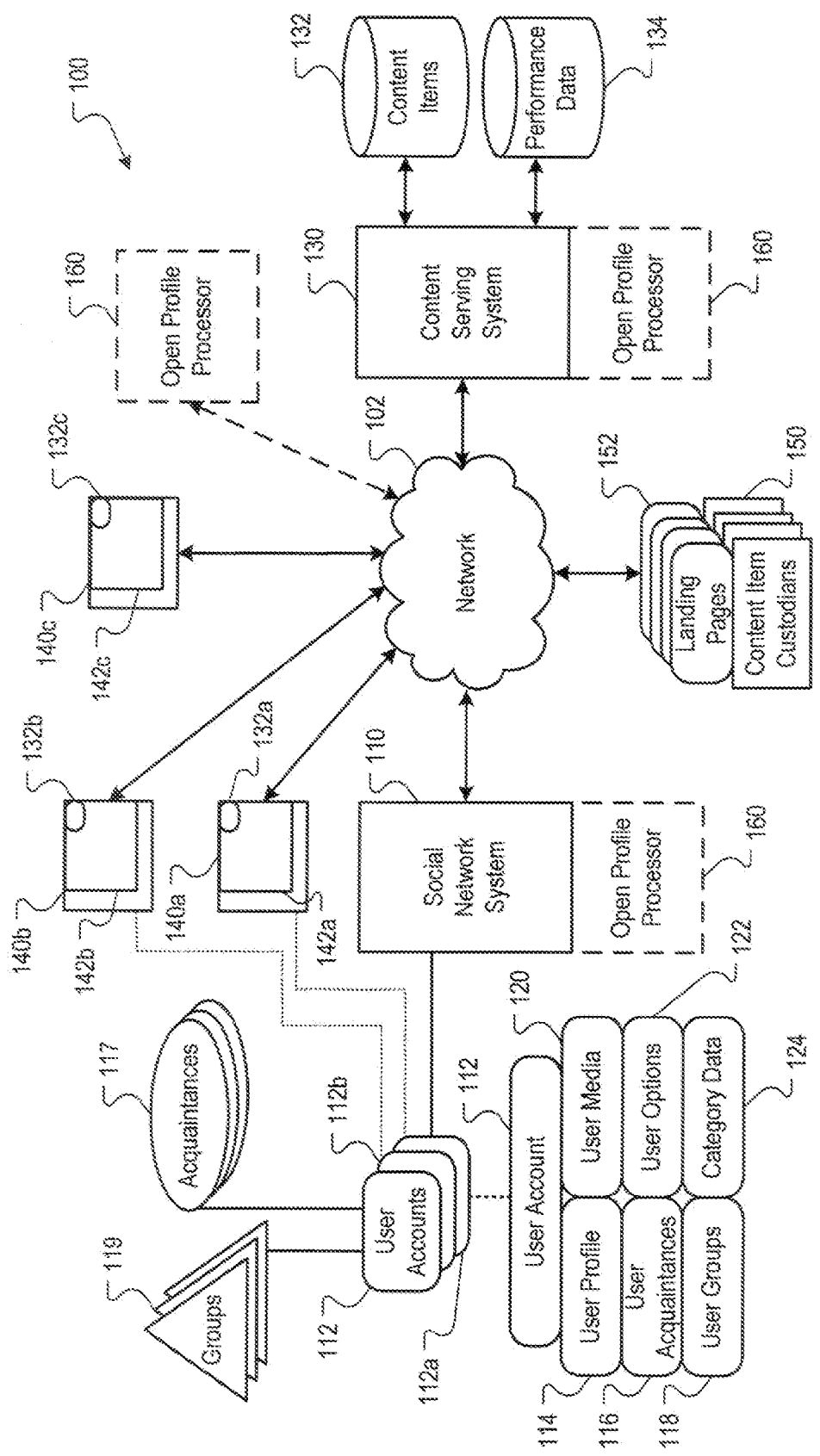


FIG. 1

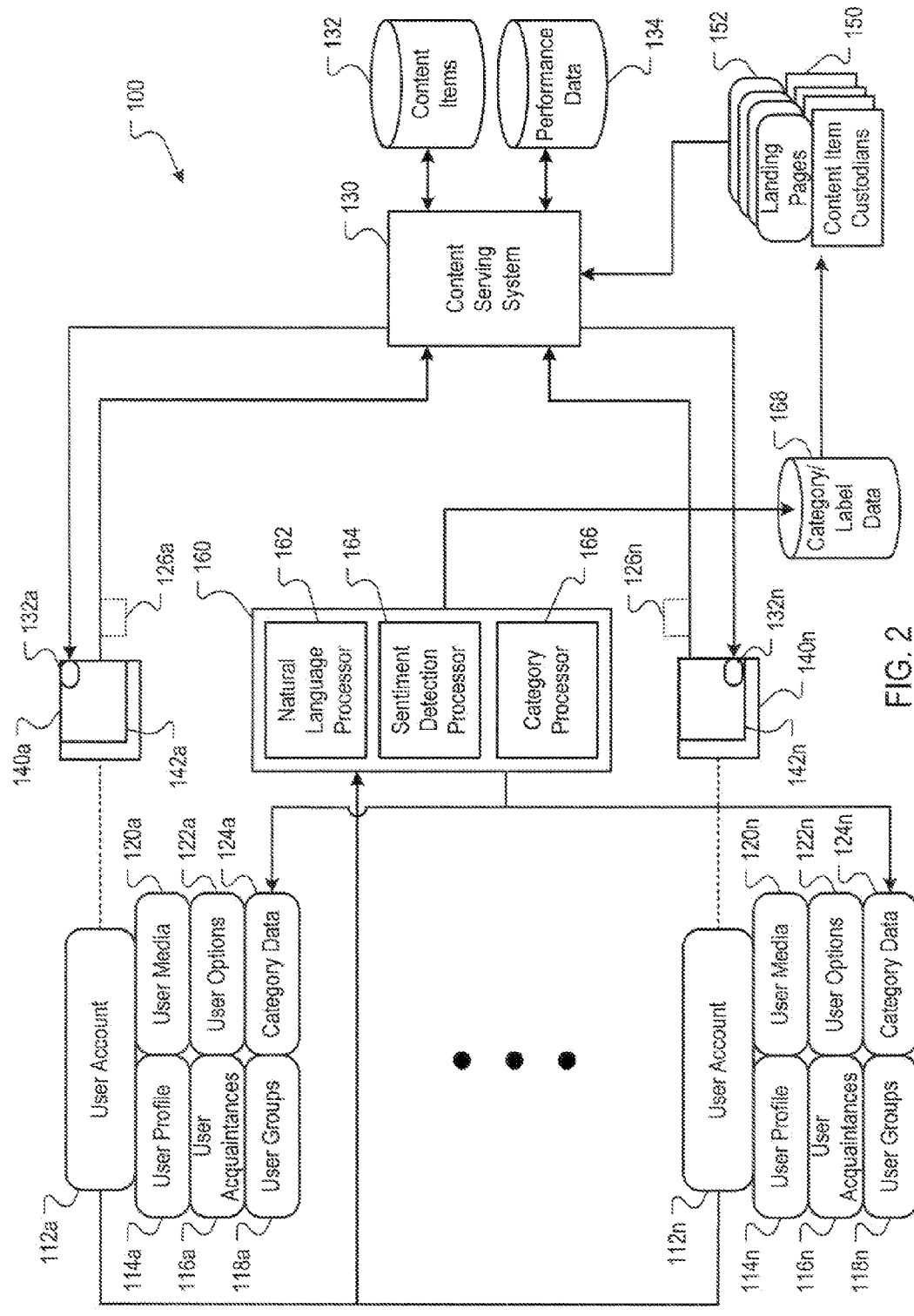


FIG. 2

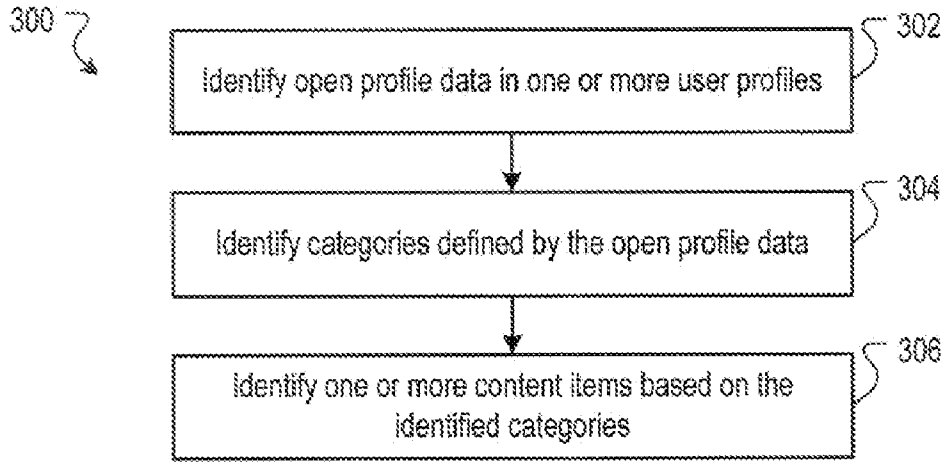


FIG. 3

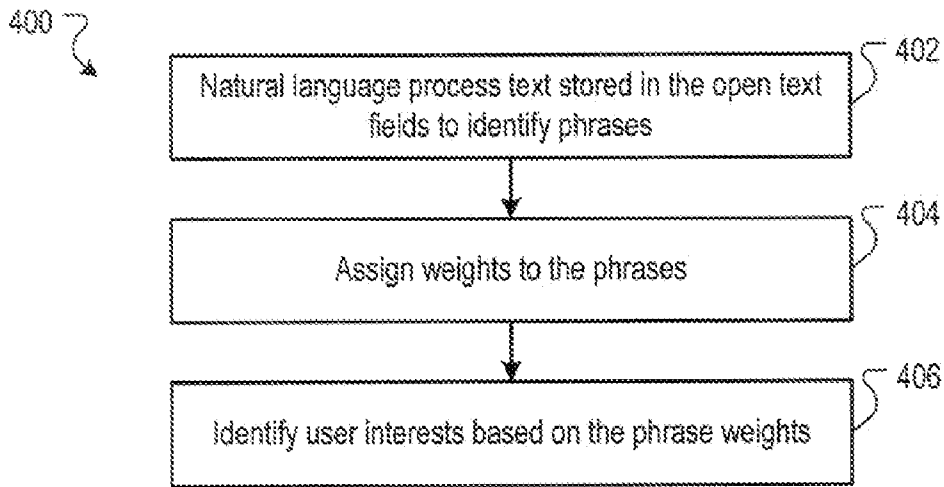


FIG. 4

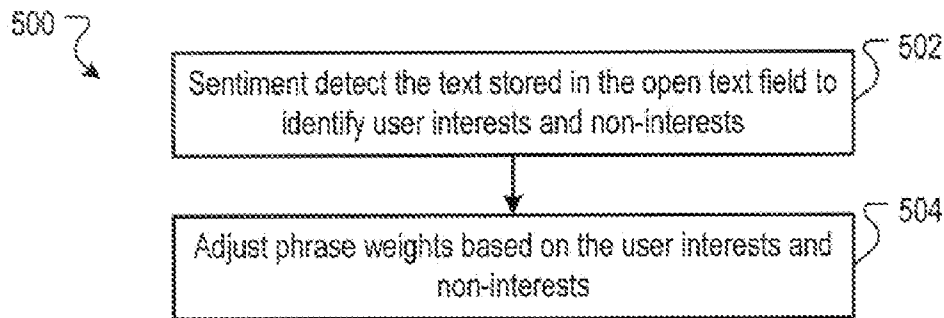


FIG. 5

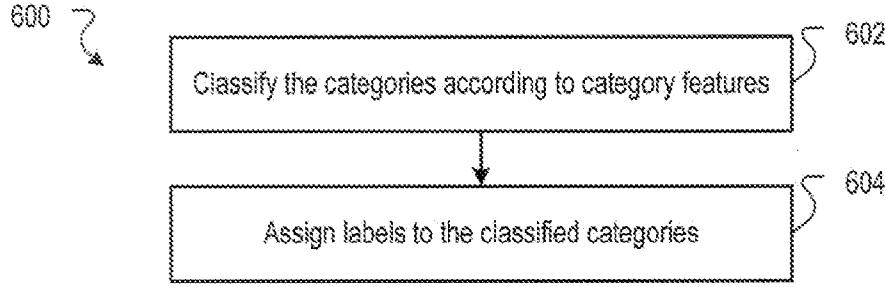


FIG. 6

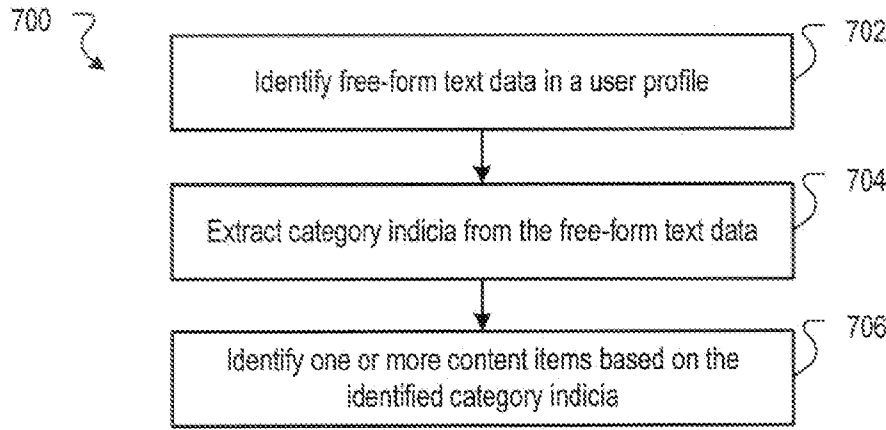


FIG. 7

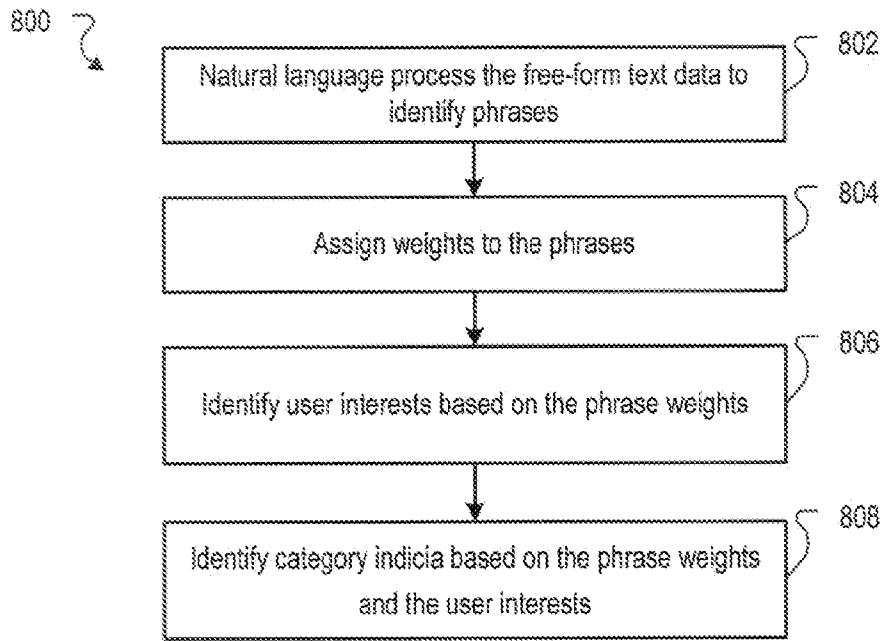


FIG. 8

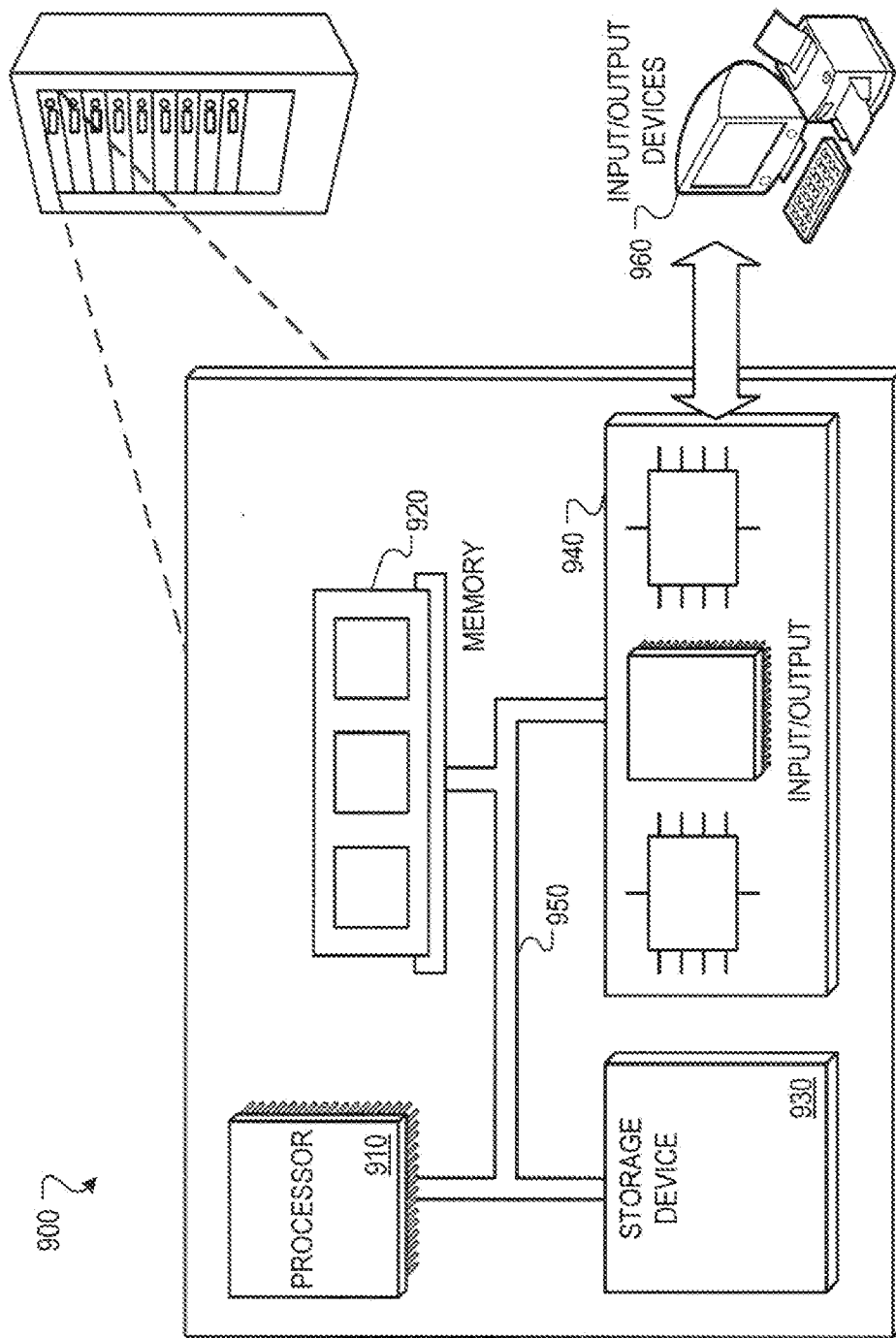


FIG. 9

## OPEN PROFILE CONTENT IDENTIFICATION

[0001] This application is related to U.S. patent application Ser. No. 11/\_\_\_\_\_, filed on \_\_\_\_\_, 2007, entitled “Custodian Based Content Identification,” and Ser. No. 11/\_\_\_\_\_, filed on \_\_\_\_\_, 2007, entitled “Related Entity Content Identification” the disclosures of which are incorporated herein by reference.

### BACKGROUND

[0002] This application relates to content identification.

[0003] Online social networks have become popular for professional and/or social networking. Some online social networks provide content items that may be of interest to users, e.g., digital advertisements targeted to a user, or identification of other users and/or groups that may of interest to a user. The content items can, for example, be selected based on content of a user account, e.g., based on keywords identified from a crawl of a user’s page. Such content item identification schemes, however, may not identify optimum content items if the user page includes only short, ambiguous messages, misspelled words, or is primarily non-textual content, e.g., photograph collections, that present unique challenges for machine-based relevance analysis.

[0004] Additionally, such content identification schemes do not readily facilitate advertiser targeting of publishers that may have a broad range of visiting users, e.g., social networks. For example, social networking sites have users that have many different interests, and thus advertisers may not readily identify particular verticals for the social networking site. Accordingly, some of the content items, e.g., advertisements directed to particular products, may not be of interest to many users of an online social network.

### SUMMARY

[0005] Described herein are systems and methods for facilitating open profile content identification. In one implementation, open profile data in one or more user profiles are identified, and categories defined by the open profile data are identified. One or more content items based on the categories can then be identified.

[0006] In another implementation, free-form text data in a user profile is identified, and category indicia is extracted from the free-form text data. One or more content items based on the category indicia can then be identified.

[0007] In another implementation, a system includes a natural language processor, a sentiment detection processor, and a category processor. The natural language processor is configured to extract one or more phrases from free-form text data in a user profile and assign phrase weights to the extracted phrases. The sentiment detection processor is configured to identify user interests and non-interests from the free-form text data in the user profile. The category processor is configured to associate labels with a user profile based on the extracted one or more phrases, assigned phrase weights, and the identified user interests and non-interests.

[0008] In another implementation, open profile data in user profiles are identified and categories defined by the open profile data are identified. The categories are associated with corresponding user profiles.

### BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. 1 is a block diagram of an example system for identifying content items based on open profile data.

[0010] FIG. 2 is a more detailed block diagram of the example system for identifying content items based on open profile data.

[0011] FIG. 3 is a flow diagram of an example process for identifying content items based on open profile data.

[0012] FIG. 4 is a flow diagram of an example process for identifying user interests based on open profile data.

[0013] FIG. 5 is a flow diagram of an example process for identifying user interests and non-interests based on open profile data.

[0014] FIG. 6 is a flow diagram of an example process for classifying categories and labels.

[0015] FIG. 7 is a flow diagram of an example process for identifying content items based on free-form text data.

[0016] FIG. 8 is a flow diagram of an example process for identifying categories based on free form text data.

[0017] FIG. 9 is a block diagram of an example computer system that can be utilized to implement the systems and methods described herein.

### DETAILED DESCRIPTION

[0018] FIG. 1 is a block diagram of an example system 100 for identifying content items based on open profile data. Open profile data in one or more user profiles, e.g., free-form fields in a user profile, are identified and processed to identify interests and preferences of the one or more users. The interests and preferences are utilized to identify categories associated with the user profiles, and content items, e.g., advertisements, can be identified based on the categories.

[0019] In an implementation, a social network system 110 can, for example, host numerous user accounts 112. An example social network system can include Orkut, hosted by Google, Inc., of Mountain View, Calif. Other social networks can, for example, include school alumni websites, an internal company web site, dating networks, etc.

[0020] Each user account 112 can, for example, include user profile data 114, user acquaintance data 116, user group data 118, user media data 120, user options data 122, and category data 124. Other user data can also be included in the user accounts 112.

[0021] The user profile data 114 can, for example, include general demographic data about an associated user, such as age, sex, location, interests, etc. In some implementations, the user profile data 114 can also include professional information, e.g., occupation, educational background, etc., and other data, such as contact information. In some implementations, the user profile data 114 can include open profile data, e.g., free-form text that is typed into text fields for various subjects, e.g., “Job Description,” “Favorite Foods,” “Hobbies,” etc., and constrained profile data, e.g., binary profile data selected by check boxes, radio buttons, etc., or predefined selectable profile data, e.g., income ranges, zip codes, etc. In some implementations, some or all of the user profile data 114 can be classified as public or private profile data, e.g., data that can be shared publicly or data that can be selectively shared.

Profile data **114** not classified as private data can, for example, be classified as public data, e.g., data that can be viewed by any user accessing the social network system **110**.

[0022] The user acquaintances data **116** can, for example, define user acquaintances **117** associated with a user account **112**. In an implementation, user acquaintances **117** can include, for example, users associated with other user accounts **112** that are classified as “friends,” e.g., user accounts **112** referenced in a “friends” or “buddies” list. Other acquaintances **117** can also be defined, e.g., professional acquaintances, client acquaintances, family acquaintances, etc. In an implementation, the user acquaintance data **116** for each user account **112** can, for example, be specified by users associated with each user account **112**, and thus can be unique for each user account **112**.

[0023] The user group data **118** can, for example, define user groups **119** to which a user account **112** is associated. In an implementation, user groups **119** can, for example, define an interest or topic, e.g., “Wine,” “Open Source Chess Programming,” “Travel Hints and Tips,” etc. In an implementation, the user groups **119** can, for example, be categorized, e.g., a first set of user groups **119** can belong to an “Activities” category, a second set of user groups **119** can belong to an “Alumni & Schools” category, etc.

[0024] The user media data **120** can, for example, include one or more content items, such as user documents, e.g., web pages. A document can, for example, comprise a file, a combination of files, one or more files with embedded links to other files, etc. The files can be of any type, such as text, audio, image, video, hyper-text mark-up language documents, etc. In the content of the Internet, a common document is a Web page. Other content items can also be defined by the user media data **120**.

[0025] The user options data **122** can, for example, include data specifying user options, such as e-mail settings, acquaintance notification settings, chat settings, password and security settings, etc. Other option data can also be included in the user options data **122**.

[0026] The category data **124** can, for example, include data that can be utilized by a content serving system **130** to identify one or more content items **132** that are determined to be relevant to a user account **112**, e.g., advertisements for products and/or services that may be of interest to a user associated with a user account **112**.

[0027] In an implementation, a content serving system **130** can directly, or indirectly, enter, maintain, and track content items **132**. The content items **132** can, for example, include a web page or other content document, or text, graphics, video, audio, mixed media, etc. In one implementation, the content items **132** are advertisements. The advertisements **132** can, for example, be in the form of graphical ads, such as banner ads, text only ads, image ads, audio ads, video ads, ads combining one or more of any of such components, etc. The advertisements **132** can also include embedded information, such as links, meta-information, and/or machine executable instructions.

[0028] In an implementation, user devices **140a**, **140b** and **140c** can communicate with the social network system **110** over a network **102**, such as the Internet. The user devices **140** can be any device capable of receiving the user media data **120**, such as personal computers, mobile devices, cell phones, personal digital assistants (PDAs), television systems, etc. The user devices **140** can be associated with user accounts **112**, e.g., the users of user devices **140a** and **140b** can be

logged-in members of the social network system **110**, having corresponding user accounts **112a** and **112b**. Additionally, the user devices **140** may not be associated with a user account **112**, e.g., the user of the user device **142c** may not be a member of the social network system **110** or may be a member of the social network system **110** that has not logged in.

[0029] In one implementation, upon a user device **140** communicating a request for media data **120** of a user account **112**, or a request for some other social network media data, to the social network system **110**, the social network system **110** can, for example, provide the user media data **120** or the other requested media data to user device **140**. In one implementation, the user media data **120** or other requested media data can include an embedded request code, such as Javascript code snippets. In another implementation, the social network system **110** can insert the embedded request code into the user media data **120** or the other requested media data when served to a user device **140**.

[0030] The user device **140** can render the requested data, e.g., a user web page, or some other social network **110** page, in a presentation environment **142**, e.g., in a web browser application. Upon rendering the requested data, the user device **140** executes the request code, which causes the user device **140** to issue a content request, e.g., an advertisement request, to the content serving system **130**. In response, the content serving system **130** can provide one or more content items **132** to the user device **140**. For example, the content items **132a**, **132b** and **132c** can be provided to the user devices **140a**, **140b** and **140c**, respectively. In one implementation, the content items **132a**, **132b** and **132c** are presented in the presentation environments **142a**, **142b** and **142c**, respectively.

[0031] In an implementation, the content items **132a**, **132b** and **132c** can be provided to the content serving system **130** by content item custodians **150**, e.g., advertisers. The advertisers **150** can, for example, include web sites having “landing pages” **152** that a user is directed to when the user clicks an advertisement **132** presented on page provided from the social networking system **110**. For example, the content item custodians **150** can provide content items **132** in the form of “creatives,” which are advertisements that may include text, graphics and/or audio associated with the advertised service or product, and a link to a web site.

[0032] In one implementation, the content serving system **130** can monitor and/or evaluate performance data **134** related to the content items **132**. For example, the performance of each advertisement **132** can be evaluated based on a performance metric, such as a click-through rate, a conversion rate, or some other performance metric. A click-through can occur, for example, when a user of a user device, e.g., user device **140a**, selects or “clicks” on an advertisement, e.g. the advertisement **132a**. The click-through rate can be a performance metric that is obtained by dividing the number of users that clicked on the advertisement or a link associated with the advertisement by the number of times the advertisement was delivered. For example, if advertisement is delivered 100 times, and three persons clicked on the advertisement, then the click-through rate for that advertisement is 3%.

[0033] A “conversion” occurs when a user, for example, consummates a transaction relates to a previously served advertisement. What constitutes a conversion may vary from case to case and can be determined in a variety of ways. For example, a conversion may occur when a user of the user

device **140a** clicks on an advertisement **132a**, is referred to the advertiser's Web page, such as one of the landing pages **152**, and consummates a purchase before leaving that Web page. Other conversion types can also be used. A conversion rate can, for example, be defined as the ratio of the number of conversions to the number of impressions of the advertisement (i.e., the number of times an advertisement is rendered) or the ratio of the number of conversions to the number of selections. Other types of conversion rates can also be used.

**[0034]** Other performance metrics can also be used. The performance metrics can, for example, be revenue related or non-revenue relates. In another implementation, the performance metrics can be parsed according to time, e.g., the performance of a particular content item **132** may be determined to be very high on weekends, moderate on weekday evenings, but very low on weekday mornings and afternoons, for example.

**[0035]** It is desirable that each of the content items **132** be related to the interests of the users utilizing the user devices **140a**, **140b** and **140c**, as users are generally more likely to select, e.g., click through, content items **132** that are of particular interest to the users. In one implementation, relevant content items **132** are identified by identifying category data **124** for each user account **112**, and providing the category data **124** to the content serving system **130**. Such category data **124** can, for example, be based on open-profile data stored in corresponding user profiles **114**, e.g., free-form text fields. The category data **124** of a particular user account, e.g. user account **112a** can, for example, be provide to the content serving system **130** by a user device **140** associated with the particular user account **112**, e.g., the user device **140a** upon which a logged in session for the particular user account **112a** has been established.

**[0036]** In another implementation, category data related to the user accounts **112** can be provided to advertisers, e.g., the content item custodians **150**. Such category data can, for example, be based on the category data **124** identified for each of the user accounts **112**. Such category data can facilitate the targeting of content items **132**, e.g., advertisements, to social networking sites that have users that have many different interests. For example, the category data related to the category data **124** of each user account may identify that the users of the social network **110** are predominately interested in extreme sports, sport utility vehicles, and fitness. Accordingly, one or more content item custodians **150** could target such advertisements to users of the social network system **110**.

**[0037]** In one implementation, an open profile processor **160** can be integrated into the social network system **110**. In another implementation, the open profile processor **160** can be integrated into the content server system **130**. In another implementation, the open profile processor **160** can be a separate system in data communication with the social network system **110** and/or the content server system **130**.

**[0038]** The open profile processor **160** can be implemented in software and executed on a processing device, such as the computer system **900** of FIG. 9. Example software implementations include C, C++, Java, or any other high-level programming language that may be utilized to produce source code that can be compiled into executable instructions. Other software implementations can also be used, such as applets, or interpreted implementations, such as scripts, etc.

**[0039]** FIG. 2 is a more detailed block diagram of the example system **100** for identifying content items based on

open profile data. In the example system **100** of FIG. 2, the open profile processor **160** can include a natural language processor **162**, a sentiment detection processor **164**, and a category processor **166**. The open profile processor **160** can utilize open profile data of the user profile data **114a-114n** of the user accounts **112a-112n** to generate corresponding category data **124a-124n**.

**[0040]** In one implementation, the open-profile data within the user profile data **114** is processed to identify psychographic information, e.g., traits such as personality, values, attitudes, interests, lifestyles, etc. of the users associated with the user accounts **112a-112n**. For example, the user profile data **114a** may include a free-form text field entitled "Hobbies," and the text field may include the sentence "I really enjoy hiking, especially long hikes when you can camp out for a few days. Indoor activities don't interest me at all, and I really don't like boring outdoor activities like gardening." The natural language processor **162**, the sentiment detection processor **164**, and the category processor **166** can, for example, process the open profile data and identify the categories of "Hiking" and "Camping" as categories of interest, and identify the category of "Gardening" as a category of non-interest. Accordingly, in one implementation, the category data **124a** can include the categories of "Hiking" and "Camping." In another implementation, the category data **124a** can include both interest categories and non-interest categories. Thus, for the above example, the category data **124** can include the interest categories of "Hiking" and "Camping," and can include the non-interest category of "Gardening."

**[0041]** In one implementation, an example process flow for identifying categories related to the interests of one or more users begins when the natural language processor **162** receives the open profile data of one or more user profiles **114**, e.g., free form text of one or more user profiles. The natural language processor **162** can, for example, be configured to extract phrases from the open profile data and assign a weight to each extracted phrase. For example, the natural language process **162** can utilize a part-of-speech (POS) tagger and a dependency relation parser to extract noun phrases. For each noun phrase, the nouns and any related adjectives can be retained.

**[0042]** For each of the extracted noun phrases, the natural language processor **162** can, for example, assign a default weight to the phrase that depends on the number of nouns in the phrase. In one implementation, the weight can be in substantial proportion to the quantity of nouns in the phrase, e.g., more nouns will result in a higher weight. In another implementation, the natural language processor **162** can also adjust the weight of the phrase in response to the number of adjectives in the noun phrase. In one implementation, the weight can be in substantial proportion to the quantity of adjectives in the phrase, e.g., more adjectives will result in a higher weight.

**[0043]** Other weighting schemes can also be used, e.g., other parts of speech may decrease the weight of the noun phrase; adjectives can be weighted less than nouns, etc. In one implementation, the natural language processor **162** can, for example, implement a term frequency-inverse document frequency (TF-IDF) weighting to discount phrases that occur above a threshold frequency in the open profile data of one or more of the user profiles **112a-112n**.

**[0044]** In another implementation, the sentiment detection processor **164** can identify sentiments, e.g., opinions, interests and/or non-interests related to identified subject matter,

such as nouns or noun phrases. In one implementation, the sentiment detection processor 164 can implement polarity analysis, e.g., detection of a negative or positive verb phrase related to an object noun, e.g. “don’t like,” and “like,” may be identified as being related to respective nouns “gardening” and “hiking.” The sentiment detection processor 164 can, for example, discount nouns or noun phrases for which a user has expressed disinterest, and can appreciate nouns or noun phrases for which a user has expressed an interest. Other sentiment analysis schemes can also be used.

[0045] In one implementation, the category processor 166 can, for example, determine whether the phrases weighted by the natural language processor 162 and/or the sentiment detection processor 164 exceeds a threshold. For the phrases having weights that exceed the threshold, the category processor 166 can identify one or more categories or interests based on the phrases. In another implementation, the category processor 166 can, for example, identify one or more categories or interests based on a set of the highest weighted phrases that remain after discounting phrases of non-interest.

[0046] In one implementation, the category processor 166 can identify the categories by clustering based on the phrase and weight pairs for phrases identified from the open profile data in a plurality of user profiles 114a-114n. In another implementation, the categories can be identified by keywords defined by the phrases. Other category identifications schemes can also be used.

[0047] In another implementation, the identified categories can be further categorized according to category labels. A category label can categorize specific categories into a more general category, e.g., the categories of “hang gliding” and “rock climbing” can be categorized under the category label of “extreme sports.”

[0048] The identified categories and/or category labels for each user account 112a-112n can be stored in the corresponding category data 124a-124n. In one implementation, the category data 124 of a particular user account, e.g. user account 112a can, for example, be provide to the content serving system 130 by a user device 140 associated with the particular user account 112, e.g., the user device 140a upon which a logged in session for the particular user account 112a has been established. For example, if the user device 140a renders a web page from the social network system 110 that includes a content item request instruction, a corresponding content item request 126a that includes the category data 124a can be provided to the content serving system 130. The content serving system 130 can identify and serve one or more content items 132a for presentation on the user device 140a in response to the request 126a and the category data 124a.

[0049] Other schemes can also be used to provide the category data 124a to the content server system 130. For example, the content serving system 130 may partner with the social network system 110 to receive user account 112 identifiers in corresponding content requests 126a-126n. In response to receiving the user account 112 identifiers, the content serving system 130 can transmit the user account 112 identifiers to the social network system 110 in a request for corresponding category data 124. The social network system 110 can, in response, provide the corresponding category data 124 to the content serving system 130.

[0050] In another implementation, aggregated category and label data 168 related to the user accounts 112 can be provided to advertisers, e.g., the content item custodians 150.

Such category data and label data 168 can, for example, facilitate the targeting of content items 132, e.g., advertisements, to social networking sites that have users having many different interests. For example, the category and label data 168 for the social network system 110 may identify a set of highly relevant marketing verticals for which many of the users of the social network system 110 may have a strong interest. Accordingly, one or more content item custodians 150 can target advertisements related to the verticals to users of the social network system 110.

[0051] In one implementation, the open profile processor 160 can process the user accounts 112a-112n periodically, e.g., monthly, to revise the category data 124 and/or the category and label data 168. Other processing triggers, e.g., changes in the user account 112 corpus, can also be used. In one implementation, the social open profile processor 160 can process the user accounts 112a-112n in an offline batch process. The processing results can, for example, be stored and accessed during the serving of content. In another implementation, the open profile processor 160 can process the user accounts 112a-112n in an online process, e.g., in response to a user device 140 submitting a content request to the social network system 110.

[0052] FIG. 3 is a flow diagram of an example process 300 for identifying content items based on open profile data. The process 300 can, for example, be implemented in the open profile processor 160. In one implementation, the open profile processor 160 can include the natural language processor 162, the sentiment detection processor 164, and the category processor 166. Other implementations, however, can also be used. In one implementation, the open profile processor 160 can be integrated into the social network system 110. In another implementation, the open profile processor 160 can be integrated into the content server system 130. In another implementation, the open profile processor 160 can be a separate system in data communication with the social network system 110 and/or the content server system 130.

[0053] Stage 302 identifies open profile data in one or more user profiles. For example, the open profile processor 160 and/or the natural language processor 162 can identify open profile data, e.g., free form text data, in the user profiles 112a-112n.

[0054] Stage 304 identifies categories defined by the open profile data. For example, the open profile processor 160 and/or the category processor 166 can identify categories by clustering or keyword identification. The identified categories can, for example, be associated with corresponding user profiles 112.

[0055] Stage 306 identifies one or more content items based on the identified categories. For example the content serving system 130 can identify content items 132 based on the category data 124 and/or the category and label data 138.

[0056] Other processes for identifying content items based on open profile data can also be used.

[0057] FIG. 4 is a flow diagram of an example process 400 for identifying user interests based on open profile data. The process 400 can, for example, be implemented in the open profile processor 160. In one implementation, the open profile processor 160 can include the natural language processor 162, the sentiment detection processor 164, and the category processor 166. Other implementations, however, can also be used. In one implementation, the open profile processor 160 can be integrated into the social network system 110. In another implementation, the open profile processor 160 can

be integrated into the content server system 130. In another implementation, the open profile processor 160 can be a separate system in data communication with the social network system 110 and/or the content server system 130.

[0058] Stage 402 performs natural language processing of text stored in the open text field to identify phrases. For example, the open profile processor 160 and/or the natural language processor 162 can perform natural language processing to identify noun phrase in the open profile data of the user profiles 114a-114n.

[0059] Stage 404 assigns weights to the phrases. For example, the open profile processor 160 and/or the natural language processor 162 can assign weights to the identified noun phrases.

[0060] Stage 406 identifies user interests based on the phrase weights. For example, the open profile processor 160 and/or the natural language processor 162 can identify user interests based on the phrase weights, e.g., phrases having weights exceeding a threshold can be utilized to define user interests, or clustering based on phrase and weight pairs can be performed to define user interests.

[0061] Other processes for identifying user interests based on open profile data can also be used.

[0062] FIG. 5 is a flow diagram of an example process 500 for identifying user interests and non-interests based on open profile data. The process 500 can, for example, be implemented in the open profile processor 160. In one implementation, the open profile processor 160 can include the natural language processor 162, the sentiment detection processor 164, and the category processor 166. Other implementations, however, can also be used. In one implementation, the open profile processor 160 can be integrated into the social network system 110. In another implementation, the open profile processor 160 can be integrated into the content server system 130. In another implementation, the open profile processor 160 can be a separate system in data communication with the social network system 110 and/or the content server system 130.

[0063] Stage 502 performs sentiment detection on text stored in the open text field to identify user interests and non-interests. For example, the open profile processor 160 and/or the sentiment detection processor 164 can perform polarity detection to identify user interests and non-interests.

[0064] Stage 504 adjusts phrase weights based on the user interests and non-interests. For example, the open profile processor 160 and/or the sentiment detection processor 164 can increase or decrease phrase weights based on the user interests and non-interests, respectively.

[0065] Other processes for identifying user non-interests based on open profile data can also be used.

[0066] FIG. 6 is a flow diagram of an example process 600 for classifying categories and labels. The process 600 can, for example, be implemented in the open profile processor 160. In one implementation, the open profile processor 160 can include the natural language processor 162, the sentiment detection processor 164, and the category processor 166. Other implementations, however, can also be used. In one implementation, the open profile processor 160 can be integrated into the social network system 110. In another implementation, the open profile processor 160 can be integrated into the content server system 130. In another implementation, the open profile processor 160 can be a separate system in data communication with the social network system 110 and/or the content server system 130.

[0067] Stage 602 classifies the categories according to category features. For example, the open profile processor 160 and/or the category processor 166 can classify categories according to a category feature, e.g., a general sports category, a general movie category, etc.

[0068] Stage 604 assigns labels to the classified categories. For example, the open profile processor 160 and/or the category processor 166 can assign labels to the classified categories, e.g., "Extreme Sports," "Comedy Movies," etc.

[0069] Other processes for classifying categories and labels can also be used.

[0070] FIG. 7 is a flow diagram of an example process 700 for identifying content items based on free-form text data. The process 700 can, for example, be implemented in the open profile processor 160. In one implementation, the open profile processor 160 can include the natural language processor 162, the sentiment detection processor 164, and the category processor 166. Other implementations, however, can also be used. In one implementation, the open profile processor 160 can be integrated into the social network system 110. In another implementation, the open profile processor 160 can be integrated into the content server system 130. In another implementation, the open profile processor 160 can be a separate system in data communication with the social network system 110 and/or the content server system 130.

[0071] Stage 702 identifies free-form text data in a user profile. For example, the open profile processor 160 and/or the natural language processor 162 can identify open profile data, e.g., free form text data, in the user profiles 112a-112n.

[0072] Stage 704 extracts category indicia from the free-form text data. For example, the open profile processor 160 and/or the category processor 166 can extract category indicators, e.g., keywords, etc., from the free-form text data.

[0073] Stage 706 identifies one or more content items based on the identified category indicia. For example, the content serving system 130 can identify one or more content items 132 based on the extracted keywords.

[0074] Other processes for identifying content items based on free-form text data can also be used.

[0075] FIG. 8 is a flow diagram of an example process 800 for identifying categories based on free form text data. The process 800 can, for example, be implemented in the open profile processor 160. In one implementation, the open profile processor 160 can include the natural language processor 162, the sentiment detection processor 164, and the category processor 166. Other implementations, however, can also be used. In one implementation, the open profile processor 160 can be integrated into the social network system 110. In another implementation, the open profile processor 160 can be integrated into the content server system 130. In another implementation, the open profile processor 160 can be a separate system in data communication with the social network system 110 and/or the content server system 130.

[0076] Stage 802 performs natural language processing on the free-form text data to identify phrases. For example, the open profile processor 160 and/or the natural language processor 162 can perform natural language processing to identify noun phrase in the open profile data of the user profiles 114a-114n.

[0077] Stage 804 assigns weights to the phrases. For example, the open profile processor 160 and/or the natural language processor 162 can assign weights to the identified noun phrases.

[0078] Stage 806 identifies user interests based on the phrase weights. For example, the open profile processor 160 and/or the natural language processor 162 can identify user interests based on the phrase weights, e.g., phrases having weights exceeding a threshold can be utilized to define user interests, or clustering based on phrase and weight pairs can be performed to define user interests.

[0079] Stage 808 identifies category indicia based on the phrase weights and the user interests. For example, the open profile processor 160 and/or the category processor 166 can extract category indicators, e.g., keywords, clusters, etc., based on the phrase weights and the user interests.

[0080] Other processes for identifying categories based on free form text data can also be used.

[0081] FIG. 9 is a block diagram of an example computer system 900 that can be utilized to implement the systems and methods described herein. The system 900 includes a processor 910, a memory 920, a storage device 930, and an input/output device 940. Each of the components 910, 920, 930, and 940 can, for example, be interconnected using a system bus 950. The processor 910 is capable of processing instructions for execution within the system 900. In one implementation, the processor 910 is a single-threaded processor. In another implementation, the processor 910 is a multi-threaded processor. The processor 910 is capable of processing instructions stored in the memory 920 or on the storage device 930.

[0082] The memory 920 stores information within the system 900. In one implementation, the memory 920 is a computer-readable medium. In one implementation, the memory 920 is a volatile memory unit. In another implementation, the memory 920 is a non-volatile memory unit.

[0083] The storage device 930 is capable of providing mass storage for the system 900. In one implementation, the storage device 930 is a computer-readable medium. In various different implementations, the storage device 930 can, for example, include a hard disk device, an optical disk device, or some other large capacity storage device.

[0084] The input/output device 940 provides input/output operations for the system 900. In one implementation, the input/output device 940 can include one or more of a network interface devices, e.g., an Ethernet card, a serial communication device, e.g., and RS-232 port, and/or a wireless interface device, e.g., and 802.11 card. In another implementation, the input/output device can include driver devices configured to receive input data and send output data to other input/output devices, e.g., keyboard, printer and display devices 960.

[0085] The apparatus, methods, flow diagrams, and structure block diagrams described in this patent document may be implemented in computer processing systems including program code comprising program instructions that are executable by the computer processing system. Other implementations may also be used. Additionally, the flow diagrams and structure block diagrams described in this patent document, which describe particular methods and/or corresponding acts in support of steps and corresponding functions in support of disclosed structural means, may also be utilized to implement corresponding software structures and algorithms, and equivalents thereof.

[0086] This written description sets forth the best mode of the invention and provides examples to describe the invention and to enable a person of ordinary skill in the art to make and use the invention. This written description does not limit the invention to the precise terms set forth. Thus, while the invention has been described in detail with reference to the

examples set forth above, those of ordinary skill in the art may effect alterations, modifications and variations to the examples without departing from the scope of the invention.

What is claimed is:

1. A computer-implemented method, comprising: identifying open profile data in one or more user profiles; identifying categories defined by the open profile data; and identifying one or more content items based on the identified categories.
2. The method of claim 1, wherein: the one or more content items comprise advertisements.
3. The method of claim 2, wherein: identifying open profile data in one or more user profiles comprises identifying open text fields in the one or more user profiles.
4. The method of claim 3, comprising: natural language processing text stored in the open text fields to identify phrases; and assigning weights to the phrases.
5. The method of claim 4, wherein: identifying categories defined by the open profile data comprises identifying user interests based on the phrase weights.
6. The method of claim 4, comprising: sentiment detecting the text stored in the open text field to identify user interests.
7. The method of claim 6, wherein: identifying categories defined by the open profile data comprises identifying categories based on the phrase weights and the user interests.
8. The method of claim 7, wherein: identifying categories defined by the open profile data comprises generating category clusters based on the phrase weights.
9. The method of claim 4, wherein: identifying categories defined by the open profile data comprises discounting high frequency phrases.
10. The method of claim 4, comprising: sentiment detecting the text stored in the open text field to identify user non-interests; and discounting phrases associated with the user non-interests.
11. The method of claim 2, comprising: classifying the categories according to category features; and assigning labels to the classified categories.
12. The method of claim 1, wherein: the user profiles comprises social network user profiles.
13. The method of claim 1, comprising: serving the one or more identified content items to a user associated with the identified categories.
14. A computer-implemented method, comprising: identifying free-form text data in a user profile; extracting category indicia from the free-form text data; and identifying one or more content items based on the identified category indicia.
15. The method of claim 14, wherein: the one or more content items comprise advertisements.
16. The method of claim 15, wherein: extracting category indicia from the free-form text data comprises: natural language processing the free-form text data to identify phrases; and assigning weights to the phrases.

**17.** The method of claim **16**, wherein:  
extracting category indicia from the free-form text data  
comprises identifying user interests based on the phrase  
weights.

**18.** The method of claim **16**, comprising:  
sentiment detecting the free-from text data to identify user  
interests.

**19.** The method of claim **18**, wherein:  
extracting category indicia from the free-form text data  
comprises identifying category indicia based on the  
phrase weights and the user interests.

**20.** A system, comprising:  
a natural language processor configured to extract one or  
more phrases from free-from text data in a user profile  
and assign phrase weights to the extracted phrases;

a sentiment detection processor configured to identify user  
interests and non-interests from the free-from text data  
in the user profile; and

a category processor configured to associate labels with a  
user profile based on the extracted one or more phrases,  
assigned phrase weights, and the identified user interests  
and non-interests.

**21.** A computer-implemented method, comprising:  
identifying open profile data in user profiles;  
identifying categories defined by the open profile data; and  
associating the identified categories with corresponding  
user profiles.

\* \* \* \* \*