US 20140067535A1

(19) **United States**

(12) **Patent Application Publication**

Rezaei et al.

(10) **Pub. No.: US 2014/0067535 A1**

(43) **Pub. Date: Mar. 6, 2014**

(54) **CONCEPT-LEVEL USER INTENT PROFILE EXTRACTION AND APPLICATIONS**

(71) Applicant: **NetSeer, Inc.**, Santa Clara, CA (US)

(72) Inventors: **Behnam A. Rezaei**, Santa Clara, CA (US); **Vwani Roychowdhury**, Los Angeles, CA (US); **Sanjiv Ghate**, Sunnyvale, CA (US); **Nima Khajehnouri**, Los Angeles, CA (US); **Riccardo Boscolo**, Culver City, CA (US); **John Mracek**, Los Altos, CA (US)

(21) Appl. No.: **14/017,123**

(22) Filed: **Sep. 3, 2013**

Related U.S. Application Data

(60) Provisional application No. 61/695,877, filed on Aug. 31, 2012.

Publication Classification

(51) **Int. Cl.**

G06F 17/30	(2006.01)
G06Q 50/00	(2006.01)
G06Q 30/02	(2006.01)

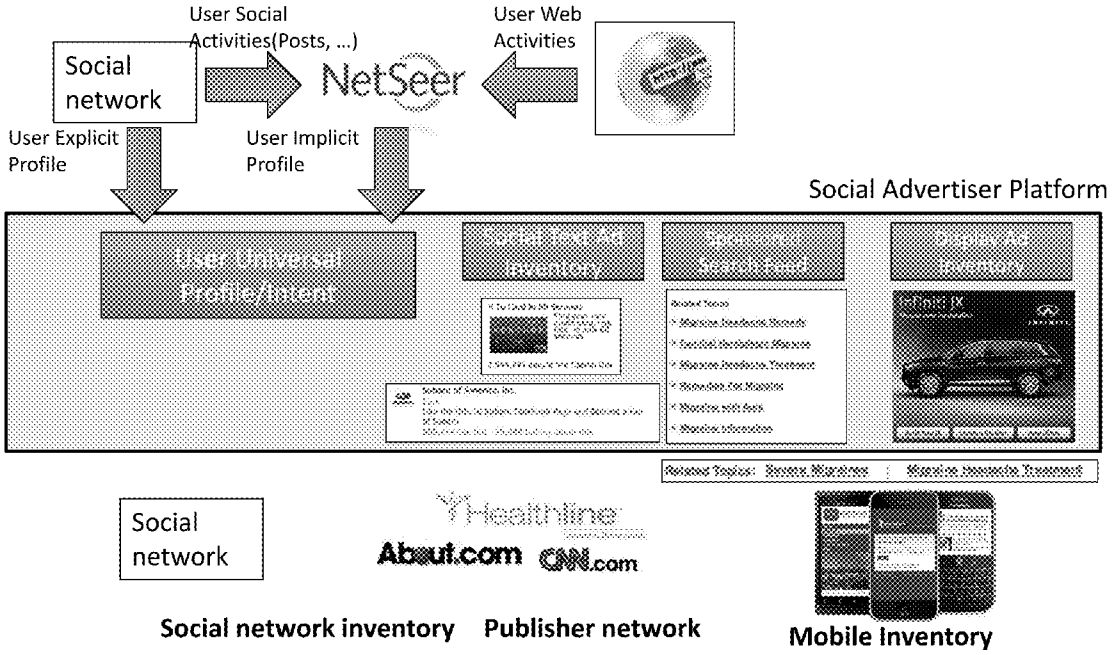
(52) **U.S. Cl.**

CPC ***G06F 17/30598*** (2013.01); ***G06Q 30/0256*** (2013.01); ***G06Q 50/01*** (2013.01)

USPC **705/14.54**; 707/748; 707/740

(57) **ABSTRACT**

Methods and systems for extracting intents and intent profiles of users, as inferred from the different activities they execute and data they share on social media sites, and then (i) monetization of such intents via targeted advertisements, and (ii) enhancement of user experience via organization of their contact lists and conversations and posts based on their content and conceptual context.



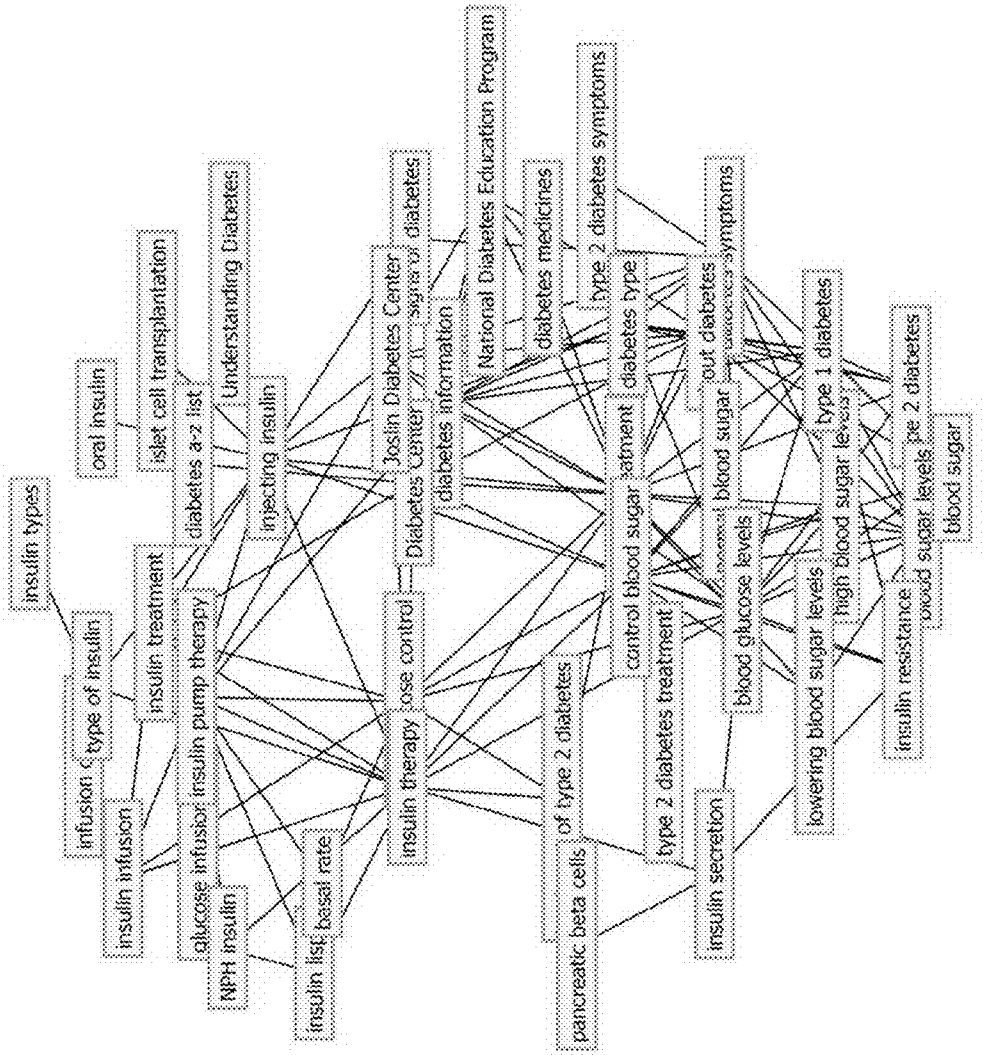

insulin types
oral insulin
islet cell transplantation
infusion
type of insulin
diabetes a-z list
Understanding Diabetes
insulin infusion
insulin treatment
injecting insulin
Josiin Diabetes Center
glucose infusion
insulin pump therapy
Diabetes Center
diabetes information
National Diabetes Education Program
diabetes medicines
NPH insulin
insulin therapy
ose control
type 2 diabetes symptoms
insulin list
basal rate
control blood sugar
diabetes type
of type 2 diabetes
out diabetes
blood sugar
pancreatic beta cells
type 2 diabetes treatment
blood glucose levels
type 1 diabetes
insulin secretion
lowering blood sugar levels
high blood sugar
insulin resistance
pe 2 diabetes
blood sugar levels
blood sugar

FIG. 1


http://diabetes.webmd.com/default.htm
User "31b23njh31abcds23j2h123213" visited webMD on 2012-08-01 5:31pm

FIG. 2A

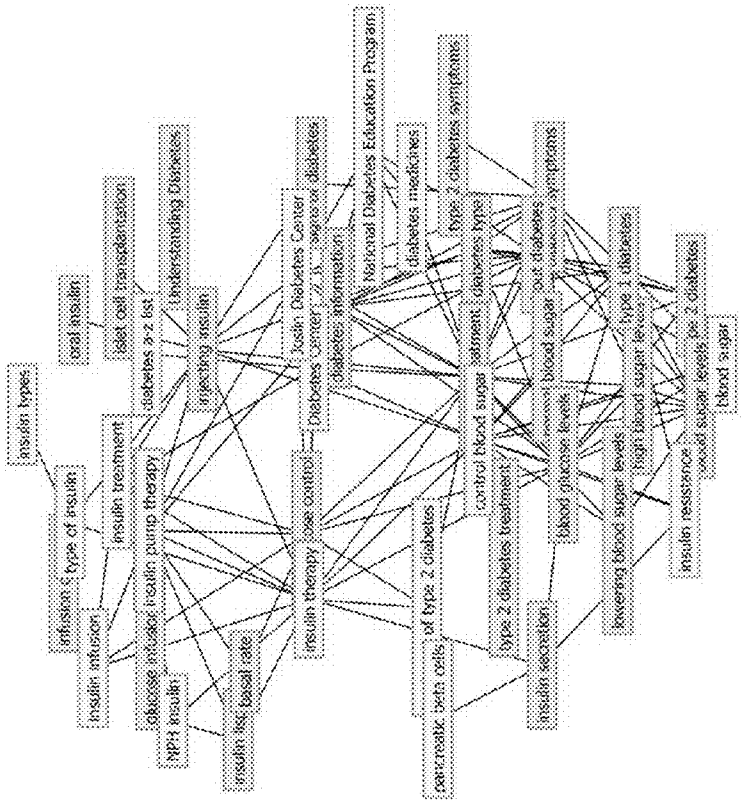

insulin types
infusion t type of insulin
insulin infusion
oral insulin
insulin treatment
islet cell transplantation
glucose infusion insulin pump therapy
diabetes a+z list
Understanding Diseases
NPH insulin
injecting insulin
insulin 70/30 basal rate
Joslin Diabetes Center
insulin therapy dose control
Diabetes Center
Diabetes information
National Diabetes Education Program
diabetes medicines
pancreatic beta cells
type 2 diabetes symptoms
control blood sugar
of type 2 diabetes
type 2 diabetes treatment
insulin secretion
blood glucose levels
lowering blood sugar levels
high blood sugar levels
type 1 diabetes
insulin resistance
blood sugar levels
pre 2 diabetes
blood sugar

FIG. 2B

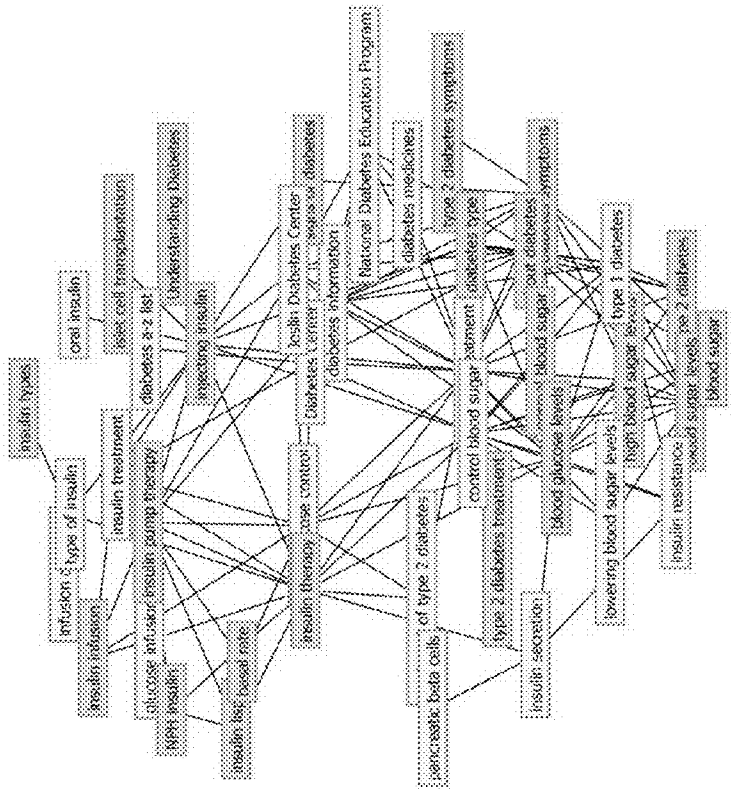

insulin types
oral insulin
islet cell transplantation
Understanding Diabetes
diabetes a-z list
injecting insulin
Joslin Diabetes Center
diabetes information
National Diabetes Education Program
diabetes medicines
type 2 diabetes symptoms
type 1 diabetes
blood sugar
insulin resistance
high blood sugar levels
lowering blood sugar levels
blood glucose levels
insulin secretion
type 2 diabetes treatment
control blood sugar
pancreatic beta cells
of type 2 diabetes
insulin therapy
insulin pump therapy
insulin treatment
type of insulin
insulin infusion
NPH insulin

FIG. 3B

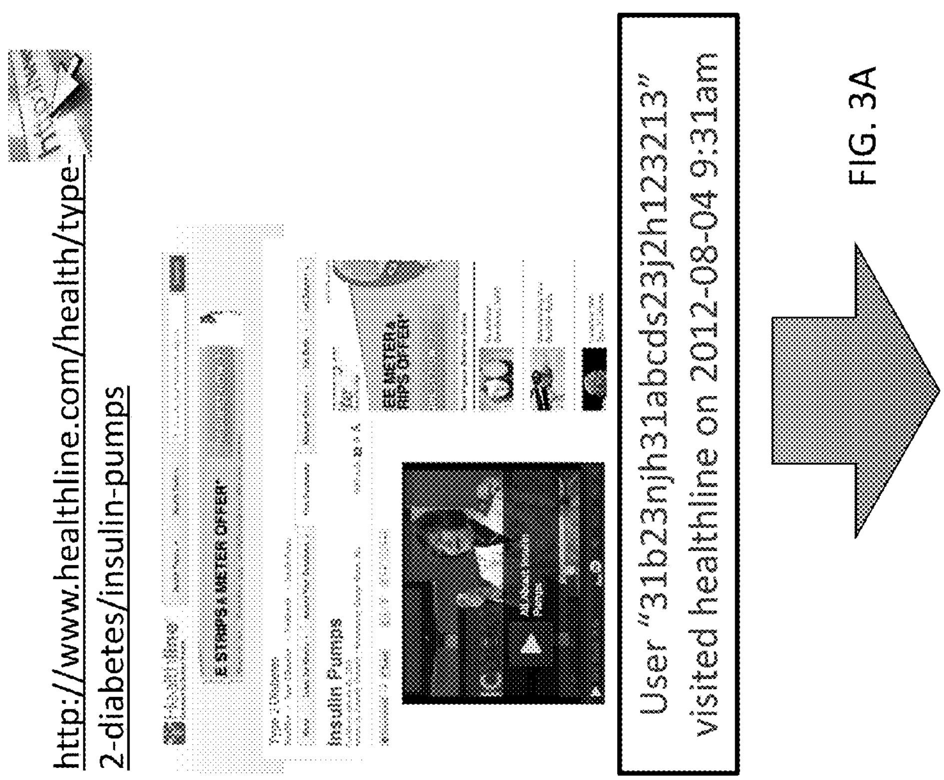

http://www.healthline.com/health/type-2-diabetes/insulin-pumps
E STRIPS & METER OFFER*
Insulin Pumps
User “31b23njh31abcds23j2h123213” visited healthline on 2012-08-04 9:31am

FIG. 3A

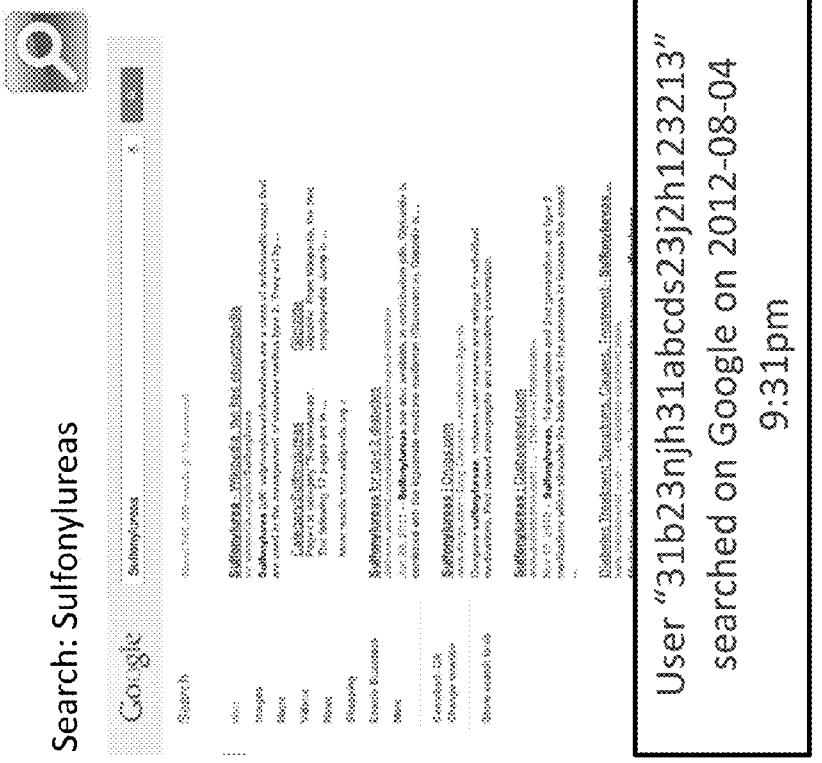

Search: Sulfonylureas
Google
Sulfonylureas
User "31b23njh31abcds23j2h123213" searched on Google on 2012-08-04 9:31pm

FIG. 4A

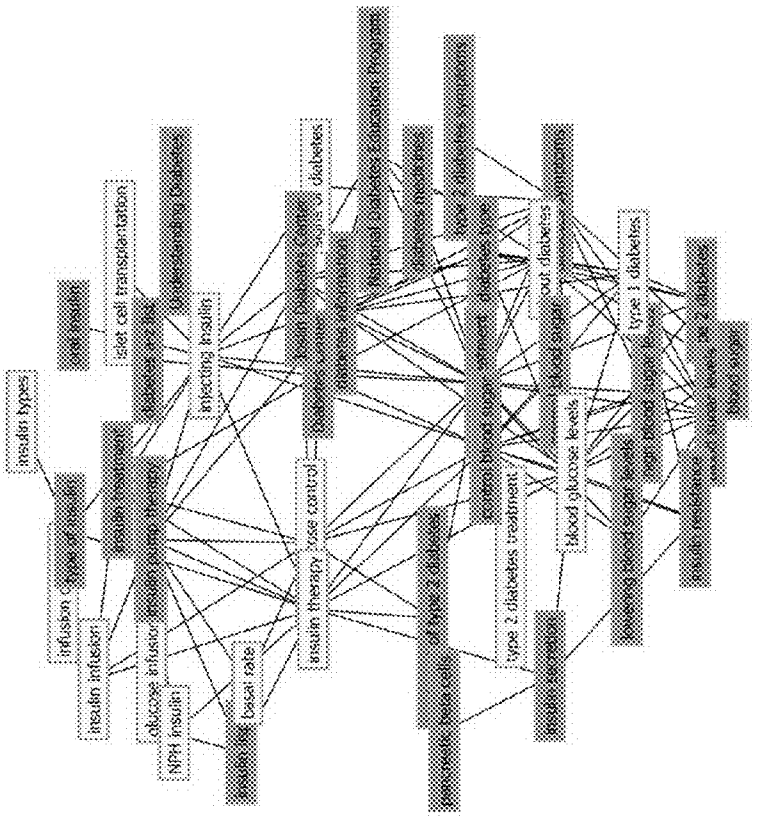

insulin types
insulin infusion
glucose infusion
NPH insulin
basal rate
insulin therapy
type 2 diabetes treatment
blood glucose levels
type 1 diabetes

FIG. 4B

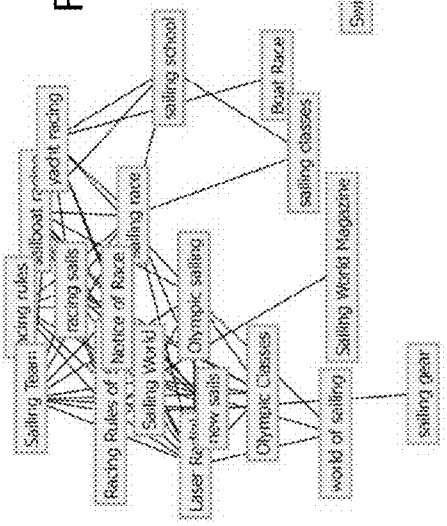


FIG. 5A

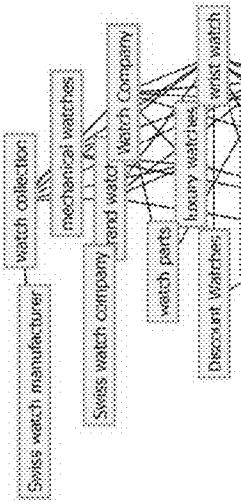


FIG. 5B

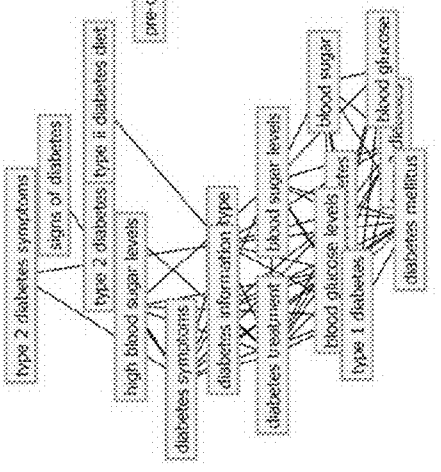


FIG. 5C

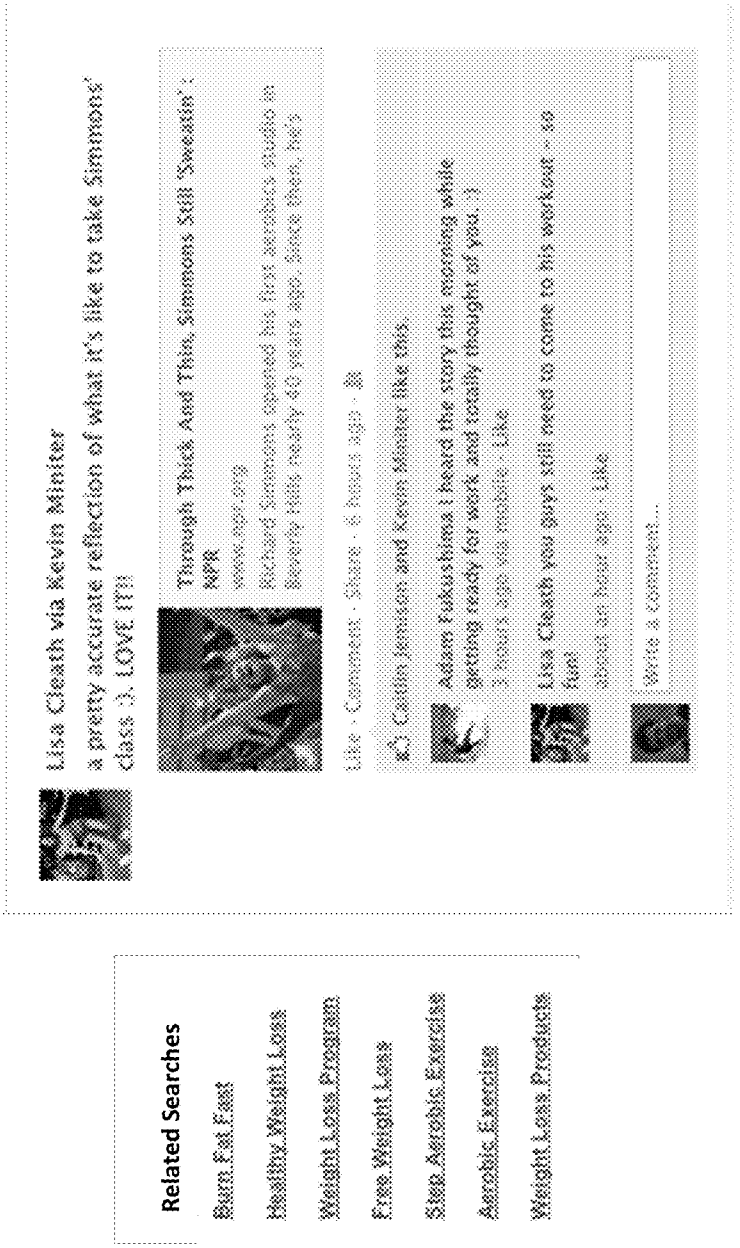
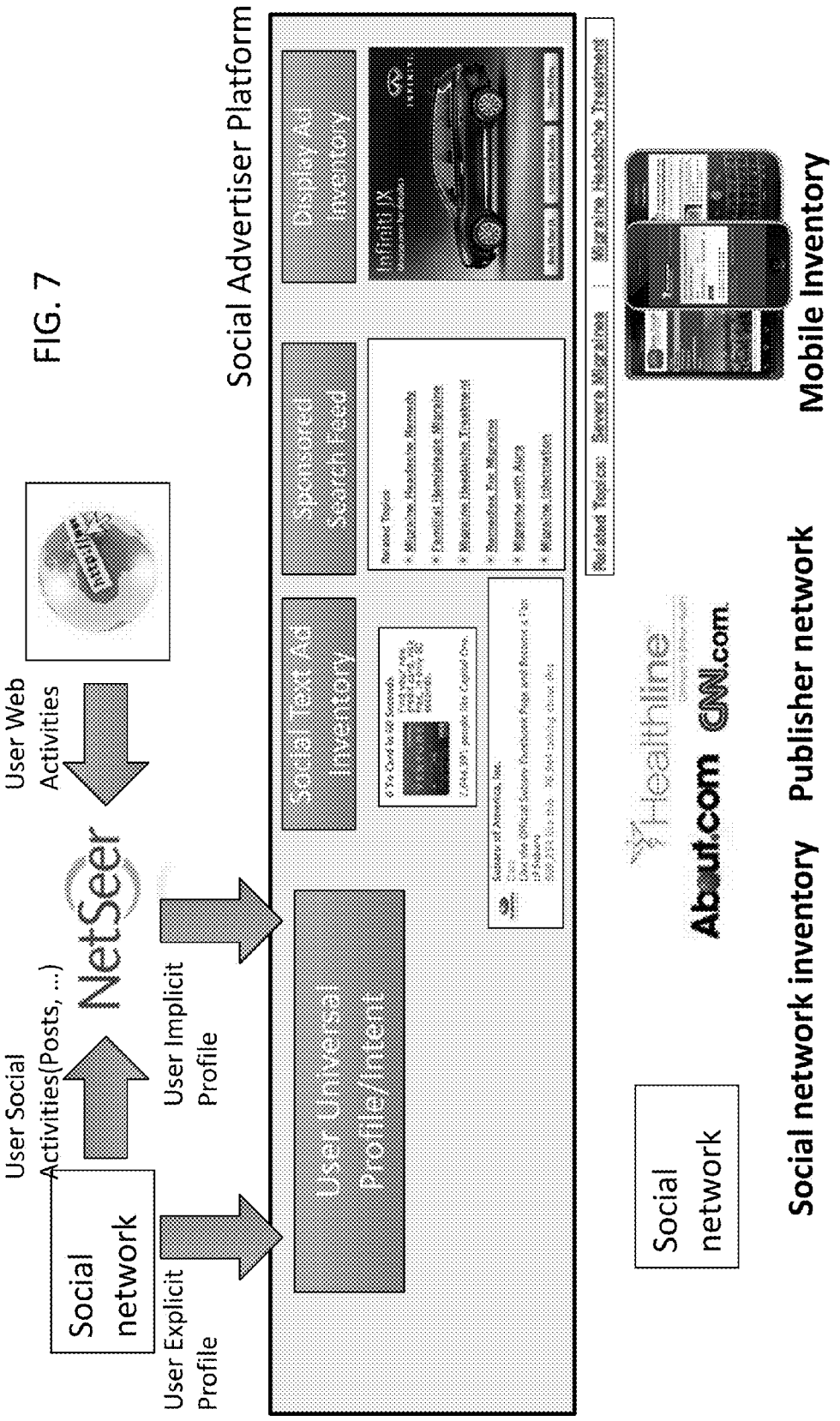

Lisa Cleath via Kevin Minter
a pretty accurate reflection of what it's like to take Simmons' class :). LOVE IT!!
Through Thick And Thin, Simmons Still 'Sweatin'
NPR
Like · Comment · Share · 6 hours ago · 38
Adam Fukushima I heard the story this morning while getting ready for work and totally thought of you. :)
Lisa Cleath you guys still need to come to his workout - so fun!
Write a comment...
Related Searches
Burn Fat Fast
Healthy Weight Loss
Weight Loss Program
Free Weight Loss
Step Aerobic Exercise
Aerobic Exercise
Weight Loss Products

FIG. 6

FIG. 7


Social Advertiser Platform
Display Ad Inventory
Sponsored Search Text
Social Ad Inventory
User Universal Profile/Intent
User Implicit Profile
User Explicit Profile
User Social Activities(Posts, ...)
User Web Activities
NetSeer
Social network
Social network
Social network inventory
Publisher network
Mobile Inventory
About.com
CNN.com
Healthline

Mr. X has 6 posts, 3 of them about the baseball team "Giants", 3 about online ad targeting.

Mr. Y likes 2 of the posts about Giants

Social signal mode (Only): Mr. Y likes posts from Mr. X (Higher priority)

Contextual Social Signal: Mr. Y likes posts from Mr. X about sports, specifically baseball, and specifically giants.

FIG. 8

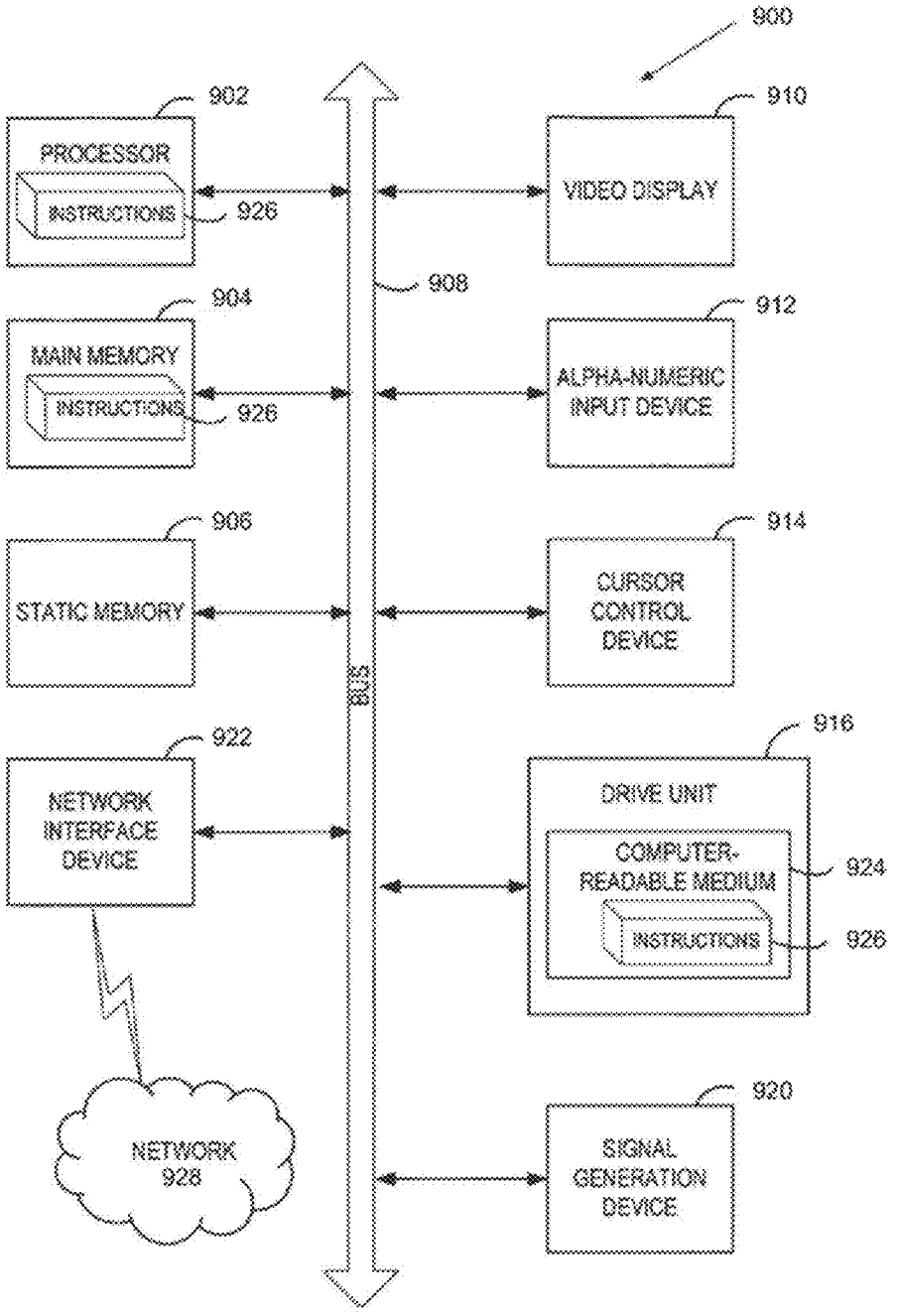

900
902
PROCESSOR
INSTRUCTIONS
926
904
MAIN MEMORY
INSTRUCTIONS
926
906
STATIC MEMORY
922
NETWORK INTERFACE DEVICE
NETWORK 928
908
BUS
910
VIDEO DISPLAY
912
ALPHA-NUMERIC INPUT DEVICE
914
CURSOR CONTROL DEVICE
916
DRIVE UNIT
COMPUTER-READABLE MEDIUM
924
INSTRUCTIONS
926
920
SIGNAL GENERATION DEVICE

FIG. 9

CONCEPT-LEVEL USER INTENT PROFILE EXTRACTION AND APPLICATIONS

PRIORITY

[0001] The present invention claims priority to U.S. Provisional Application No. 61/695,877, entitled "Concept-Level User Intent Profile Extraction and Applications to Monetization and User-Engagement Enhancement in Large-Scale Social Media Platforms," filed Aug. 31, 2012, the entirety of which is hereby incorporated by reference.

BACKGROUND

[0002] 1. Field

[0003] The present disclosure relates generally to concept-level user intent profile extraction and applications to monetization and user-engagement enhancement in large-scale social media platforms.

[0004] 2. Related Art

[0005] In the online world there is a major need to be able to understand and create temporally evolving profiles of users, and how they interact with the various institutions and activities, both online and in the real world. If such understanding and profiling, both at the individual user level and at the collective level of groups of users, can be achieved, then the various service providers (e.g., social media sites, online advertisers, offline stores and organizations) can use automated algorithms to serve the right information, content, and services to every individual and organizations (i.e., groups of users) in the right context and at the right time. The only kinds of information available online are the individual user actions, and the kind of structured data they share with various social media and other sites that they register with voluntarily. The structured data shared, e.g., one's place of residence, education level and degrees obtained, professional credentials, and their explicitly stated friends, email contact lists, and followers on social media and news sites, etc. is easy to categorize and collect and is being stored and heavily utilized and mined by various online entities such as social networking and media sites, including Facebook, Twitter, LinkedIn, Google+ etc.

[0006] The majority of user actions, however, are unstructured and when aggregated, comprises of billions of atomic or elementary actions, per day such as (i) user's Votes or Likes for articles, posts, or other users' posts and activities, (ii) searches done at major search engines and at individual sites, (iii) articles and web pages browsed, and (iv) posts on social media and networking sites and other interactions made among friends on such sites. For example, not all friends are created equal, and one shares different types of information and activities with different sets of friends and colleagues. Such preferences are not explicitly expressed and defined, but rather can only be inferred from the content of the posts shared and liked, and the locations visited together and can evolve over time.

[0007] One computationally challenging problem is how to make sense of individual users, and of groups of users collectively, from the billions of such seemingly diverse elementary actions and the available structured data. Is it possible to create a unified informational and functional view of individual users and groups of users that is granular enough to capture all aspects of behavior and preferences, and can evolve over time to be able to track a user's evolving needs and interests? Others have tried to accomplish such a task at different levels of granularity and with varying success, but a comprehensive and a computationally scalable solution has not been proposed.

[0008] For example, in the existing art detailed structured databases are created based on the explicitly stated attributes of users. This may include, age, gender, place of residence, education and schools attended, favorite institutions, such as sports teams, favorite, TV shows, music and music artists, celebrities, preferred types of food etc. These are valuable information but the expressive capabilities of such explicitly stated categories are known to be very limited in characterizing a user's intent and profile accurately. Moreover, often such information is outdated and is incorrectly entered making them prone to be highly noisy. Once entered in a database they cannot be easily updated or corrected.

[0009] The main way to deal with unstructured activities has been to use taxonomies with predefined categories organized in various data structures, such as a tree. For example if a person visits a sports page talking about the Los Angeles Lakers then that activity could be categorized as an activity related to Sports/Basketball/Lakers. These categories are then aggregated to create user profiles. The major drawbacks of such an approach are two-fold: (i) taxonomies have to be defined manually and can comprise only a limited number of categories in them. The manual nature of the process makes it less expressive, and user actions cannot be captured comprehensively and at the right granularity by such necessarily limited sets of categories. (ii) Every action and content has to be classified as belonging to one of the categories in a taxonomy and this process of classification is highly error prone. The only ways to achieve such classification is via (i) extensive training, which means providing examples of known pages or content for each category and (ii) providing a set of keywords or terms for each category and a classification is done based on how many or what sets of such keywords appear in a document. Both of these methods are highly manual and have computational problems associated with them, including (i) the accuracy of the underlying classification engine is only as good as the training sets provided to them; it can lead to over training quite easily and thereby poor generalization capabilities on new content, (ii) the bigger the taxonomy the more is the manual and supervised part of the training process, (iii) keywords are notoriously ambiguous and lead to highly inaccurate classifications, and finally (iv) often documents or content belong to multiple categories at the same time, and training for such cases that involves classifying documents as belonging to more than one category at the same time leads to a combinatorially intractable problem.

SUMMARY

[0010] The following summary of the invention is included in order to provide a basic understanding of some aspects and features of the invention. This summary is not an extensive overview of the invention and as such it is not intended to particularly identify key or critical elements of the invention or to delineate the scope of the invention. Its sole purpose is to present some concepts of the invention in a simplified form as a prelude to the more detailed description that is presented below.

[0011] To address the above mentioned limitations we present a system and methodology for creating unified intent profiles based on a collective/global concept graph, comprising nodes that are concepts, and edges that are relationships among such concepts. Naturally occurring and overlapping

sub-graphs or communities in such a concept graph organically define potential intent space of users. Each activity, such a search or a Like, is mapped to the concept graph and a weighted subgraph is identified in the underlying concept graph as a record of the activity. Such weighted subgraphs are then aggregated over all activities of a user to create a temporally evolving intent space profile of a user. Such a shared concept graph, with highly granular concept level expressiveness, allows one to not only capture individual users' intent with unprecedented accuracy and detail, but also allows one to compare and group similar users. It is as if an all-knowing human goes through and catalogs all the activities of every user and creates a summary in her mind, which allows comparison of individuals across both time and geo-locations.

[0012] According to one embodiment, a method is disclosed that includes extracting an intent profile of a user based on activities executed and information shared on a social media site by the user. Extracting the intent profile includes identifying structured user data from the social media site; identifying user activities exclusively on the social media site; identifying user activities involving the Internet; identifying incoming likes, sharing, recommendation on the social media site; identifying user searches; determining a social graph and connections of the user; and determining a weighted cluster and sub-graph of a global concept graph, wherein the global concept graph comprises nodes that are concepts, and edges that are relationships among such concepts; and determining a user intent from the intent profile of the user.

[0013] The method may further include targeting an advertisement to the user based on the determined user intent.

[0014] The method may further include generating a score for each vertical in a targeting vertical list. The score may include a time factor and an interest factor. The score may be generated based on user activities exclusively on the social media site, user activities involving the Internet, incoming likes, sharing, recommendation on the social media site, and user searches.

[0015] The method may further include modifying a user experience of the user by organizing a contact list of the user and conversations and posts of the user based on their content and conceptual context.

[0016] The method may further include performing an aggregation scoring.

[0017] The structured data may include an email identification, phone number, geo-location, friends and links.

[0018] The concepts may include phrases that represent entities (e.g. people, companies, drugs, diets, films, shows, events etc.), domain-specific terms (e.g., sports and medical terminologies, specific treatments, procedures etc.), and common expressions that are used to convey information. The relationships may be identified by annotated edges among concepts. The relationships may be measures of closeness among the concepts, including at least one of co-occurrence statistics and explicit semantic relationships.

[0019] The method may include tagging the user intent profiles with temporal data.

[0020] Determining the weighted sub-graph may be based on a number of the user's friends that share the same nodes or edges.

[0021] The method may further include tagging unstructured data using collective activities of users at the social media site.

[0022] The method may further include prioritizing unstructured data using collective activities of users at the social media site.

[0023] The method may further include post-processing and tagging the user profile post-processed and tagged with weighted category scores defined over a structured taxonomy of interest.

[0024] The method may further include assigning each such advertiser a set of advertiser target profiles.

[0025] The advertiser target profile may include one or more weighted and time-tagged sub-graphs of the global concept graph.

[0026] The advertiser target pro file may include a weighted list of categories picked from a structured taxonomy.

[0027] The method may further include determining for each user a weighted and prioritized list of advertisers that best match the user intent.

[0028] Matching user intent to an advertiser may include computing an overlap or distance between the user intent profile and the advertiser target profile by computing a measure of the distance between the two corresponding weighted sub-graphs in the global concept graph.

[0029] Selecting a final set of advertisement units may be completed by an optimization process that maximizes objective functions of interest, including revenue for the social media site, value and Return-On-Investment (ROI) for the advertisers, while considering the device and media that the user is on at the time of the impression.

[0030] Modifying a user engagement may include organizing friends or contact lists of a user into potentially overlapping groups by computing similarity between a user's profile and the profiles of those of his friends and contacts.

[0031] Organizing friends or contact lists of a user into potentially overlapping groups may include computing similarity between a user's profile and the profiles of those of the user's friends and contacts.

[0032] Modifying a user engagement may include organizing posts, comments and social interactions between a user and his friends based on an underlying context.

[0033] The linear list of posts on page of the social media site can be organized into categories by automatically classifying the posts by mapping the posts to categories in the global concept graph.

[0034] The method may further include post-processing and tagging the user profile with weighted category scores defined over a structured taxonomy of interest.

[0035] The method may further include determining a suggested search term based on the intent profile.

[0036] The method may further include grouping content based on communal user actions; and dividing the grouped content into clusters. The method may further include generating a collective content profile; and tagging the corresponding users. The method may further include prioritizing the grouped content based on page and domain statistics.

[0037] Computer systems having memory and a processor for implementing the method are also disclosed. A computer readable storage medium for executing the method is also disclosed.

BRIEF DESCRIPTION OF THE DRAWINGS

[0038] The accompanying drawings, which are incorporated into and constitute a part of this specification, illustrate one or more examples of embodiments and, together with the

description of example embodiments, serve to explain the principles and implementations of the embodiments.

[0039] FIG. **1** is a graphical view of a sample concept cluster around diabetes. FIG. **1** is an example sub-graph of a concept graph where nodes comprise concepts and edges their relationships. Such sub-graphs in the concept graph naturally define different topics or categories in knowledge space. Because of the nature of the graphical representation with edge and node weights, the categories are overlapping and can evolve as the nature of the network and the relative weights change over time.

[0040] FIGS. **2**A and **2**B illustrate mapping a user action to a global concept graph. The user activity is illustrated in FIG. **2**A by the user, who is tagged with an ID, who visits a page on diabetes. The user action in FIG. **2**A is mapped to the globally shared Concept Graph by finding the concepts that are most important to the page visited by the user as shown in FIG. **2**B. Nodes in grey are subset of nodes in the graph neighborhood related to the user action of visiting the WebMD page, around the concept diabetes.

[0041] FIGS. **3**A and **3**B further illustrate tracking user actions. As shown in FIG. **3**A, the same user, i.e., the user with the same ID as in FIG. **2**A, visits another page related to insulin pumps. As shown in FIG. **3**B, the user action is mapped to the global concept graph. In FIG. **3**B, the nodes in grey are subset of nodes in the graph neighborhood related to this user action. As shown in FIG. **3**B, the Concept graph provides one globally shared template for tracking user actions

[0042] FIGS. **4**A and **4**B further illustrate tracking user actions. As shown in FIG. **4**A, the same user as in FIGS. **2** and **3** queries for "Sulfonylureas". This is mapped to nodes in the Concept Graph, as shown in FIG. **4**B. As shown in FIG. **4**B, the nodes in dark grey are a subset of nodes in the graph neighborhood in the Concept Graph related to the search user action. As shown in FIGS. **2**B-**4**B, the three different actions (FIGS. **2**A-**4**A) may be mapped to the same neighborhood and highly connected region of the underlying concept graph.

[0043] FIG. **5** illustrates combing a social graph and a concept graph. A social networking site can use the improvements described herein to create powerful User-Intent Profiles. For example, a hypothetical Mr. X can have structured information such as: Male, 34, Santa Clara, Calif., Widowed, Working Organization Affiliations: Sharif University, Iran, UCLA and NetSeer. The user's explicit Intent may include: Likes: Classical Music, Chopin, Peruvian Food. The user's social Graph(Friends) may include: CEO's of Yahoo!, Facebook, and NetSeer; Michael Jackson, and Mario Batali. From the unstructured actions, such as likes and browsing, one can represent the person as: Mr X is a diabetic that likes sailing and Rolex watches. These characterizations are captured as the weighted sub-graphs in the globally shared Concept Graph, which are shown in FIGS. **5**A-**5**C.

[0044] FIG. **6** illustrates how a social media site such as Facebook can predict users' next searches, based on the complete user profile and posts. This can enable such social networking sites to get search market share from Search Engines by integrating search experience into social interactions. This can have a further advantage in being able to bootstrap and provide price support for their existing advertisement network. On a user clicking on one of these search links, the user could be directed to the current sponsored search/ad feed of the social media entity. In FIG. **6**, a post by a user about aerobics is matched to search suggestions related to exercise, dieting and weight loss.

[0045] FIG. **7** illustrates a schematic of an Intent Based Display Market that could be powered by the inventions stated in this application. As shown in FIG. **7**, an entity such as the assignee, NetSeer, can take all the inputs from social networks, web browsing, and other activities and then empower a Social Advertiser Platform. The user intent profiles can be used to create personalized advertisement creatives that are suitable for any format and media, including social networking posts, mobile devices, and web pages. The demand side feeds that can be used are search ad feed, display advertisement, as well as feeds meant for social sites. The supply side inventories could come from social networking inventory, the conventional web pages or mobile inventory. In mobile inventory, one form of advertisement could be search suggestions targeted for the individual user in the right context, i.e, the search suggestions not match the intent profile of the user but also match the context of the page being viewed or the App being used.

[0046] FIG. **8** illustrates one exemplary application of the invention to enhancing user experience on a social networking site. Currently the signal is mostly social, that is, for example, who likes whose posts. However this leads to a cluttered user experience. As illustrated in FIG. **8**, embodiments of invention enables addition intent/context signal to the social signal. Thus, in FIG. **8**, posts from Mr. X on baseball and specifically for the Giants could be given higher priority when listed on the wall of Mr. Y.

[0047] FIG. **9** is a schematic diagram of an exemplary computer system according to one embodiment of the invention.

DETAILED DESCRIPTION

[0048] Embodiments of the invention teaches methods and systems for extracting intents and intent profiles of users, as inferred from the different activities they execute and data they share on social media sites, and then (i) monetization of such intents via targeted advertisements, and (ii) enhancement of user experience via organization of their contact lists and conversations and posts based on their content and conceptual context.

[0049] The user activities and data processed in embodiments of the invention include, but not limited to, structured data, e.g. email-id, phone number, geo-location, friends/links, etc., as well as, unstructured data—e.g. searches, web-browsing (both on and off the social media properties), posts, comments, content of web pages that receive Likes or Links etc. The user intent and profiles extracted in embodiments of the invention are captured and expressed in terms of weighted sub-graphs of a collective/global concept graph. The collective/global concept graph includes nodes that are concepts, and edges that are relationships among such concepts. Concepts in embodiments of the invention include, but not limited to, phrases that represent entities (e.g. people, companies, drugs, diets, films, shows, events etc.), domain-specific terms (e.g., sports and medical terminologies, specific treatments, procedures etc.), or common expressions that are used to convey information. Relationships, as captured by annotated edges among concepts, include, but not limited to, measures of closeness among the concepts, e.g., co-occurrence statistics, or explicit semantic relationships (e.g., "acted in", "father of", "part of" etc.).

[0050] In some embodiments of the invention, the user intent profiles are determined and generated using weighted clusters and sub-graphs of the global concept graph. Furthermore, the user intent profiles can be tagged with temporal data. For example, certain subsets of a user's profile could be considered recent and time-sensitive (e.g., in the market to buy a product or go on a trip), whereas, other aspects of user's profile could be longer lasting (e.g., regions of the concept graph that represent an antique aficionado or a diabetes patient).

[0051] Similarly, one of the parameters in computing the weights in the sub-graph (representing a user's profile) can be based on how many of his/her friends also share the same nodes or edges. The group or collective activities of users at the social media site can be further used to tag and prioritize both the unstructured data used to extract the user profile, as well as, the distilled user profile. For example, in one embodiment of the invention, pages and content sites are grouped based on communal user actions, such as Likes or posts made to the page/site; thus, pages/sites belonging to the same group have been all visited/liked by a distinct group of users. Such a group of content sites and pages can then be collectively analyzed to create a collective content profile (again in terms of the global concept graph), which can then be used to tag the corresponding users.

[0052] In yet another embodiment of this invention, a user profile is post-processed and tagged with weighted category scores defined over a structured taxonomy of interest. For example, in one embodiment, a standard taxonomy used by the advertisement industry can be used to categorize the user profile (i.e., the sub-graph of the concept graph) in terms of the various categories in the taxonomy and compute a score for each such category. Since a number of advertisers define their campaigns based on categories, rather than the finer specificity of concept graphs, such coarse-grained tagging of user profiles can be used to more easily and directly match advertisements to users.

[0053] The advertisers, in embodiments of the invention, include but are not limited to, entities that carry out display and brand advertisement campaigns or perform search keyword auctions, or are lead-generators who aggregate demand across different platforms and media for merchants. Embodiments of the invention also allows one to assign each such advertiser a set of advertiser target profiles, which again could be expressed as (i) weighted and time-tagged sub-graphs of the global concept graph or (ii) a weighted list of categories picked from a certain structured taxonomy. Such advertiser target profiles serve the same purpose as user intent profiles. In other words, the advertiser target profiles provide a description of their target audience in terms of a common and globally-shared database, which is the concept graph. The act of targeting advertisement includes, but not limited to, determining for each user a weighted and prioritized list of advertisers that best match the user intent. One method for matching user intent to an advertiser is based upon computing the overlap or distance between the user intent profile and the advertiser target profile, e.g., by computing a measure of the distance (e.g. weighted shortest paths, or sum of the weights of the shared edges) between the two corresponding weighted sub-graphs in the global concept graph. The selection of the final set of advertisement units is completed by an optimization process that maximizes objective functions of interest, including revenue for the social media site, value and Return-On-Investment (ROI) for the advertisers, while considering the device and media that the user is on at the time of the impression. For example, in one embodiment of the invention, contextually relevant search-suggestions (i.e., keywords feed from a search engine) may be provided to users while they are engaged in a particular conversation on the social media site. In another embodiment of the invention, contextually relevant text or display ads are targeted to a user while the user is browsing a third-party web page (i.e., a web page not owned by the social media site). In yet another embodiment, an advertisement (both display and text units) is provided to a user on a mobile device by targeting the user profile.

[0054] In embodiments of the invention, user profiles and the underlying concept graph are used to enhance user engagement at a social media site. Such user engagement enhancement includes, but not limited to, (i) organizing friends or contact lists of a user into potentially overlapping groups, by computing similarity between a user's profile and the profiles of those of his friends and contacts. An active user working in the finance sector, for example, will have a lot of activities on the social media site (or activities visible from the site), such as searches or browsing history, related to finance and investment concepts; these concepts would be shared with his friends/contacts from the finance sector, leading to overlaps in their intent profiles and allowing the invention to define a common profile and a group comprising these individuals. Typically, a user has friends that are from different contexts of one's life, e.g., work related, hobby related, high-school friends, college friends etc. and the invention provides a method for identifying and labeling these groups in an automated fashion. (ii) Organizing posts, comments and social interactions between a user and his friends in terms of the underlying context. For example, the linear and a long list of posts on wall or the home page of a social media site can be organized into different subjects and categories by automatically classifying them into categories such as sports, sports teams, travel etc. by mapping them into the concept graph. This will also enable one to determine the dominant types of social interactions that happen among friends, which can then be used for targeting advertisement. For example, if posts from a particular friend have a predominant theme of electronic gadgets (as determined by the invention's automated classification method), and he then recommends/likes a new phone, then such a recommendation would have a lot higher relevance than that coming from a cooking-enthusiast.

[0055] The user activities and data processed in embodiments of the invention include, but are not limited to, structured data, e.g. email-id, phone number, geo-location, friends/links, etc., as well as, unstructured data—e.g. searches, web-browsing (both on and off the social media properties), posts, comments, content of web pages that receive Likes or Links etc. For example, some of the data would include:

[0056] 1. User activities exclusively on the Social Media Site:

[0057] a. Post Likes

[0058] b. Page and Group Likes

[0059] c. Posts

[0060] d. Subscriptions

[0061] 2. User activities involving the rest of the Internet:

[0062] a. Incoming likes, sharing, recommendation coming from WEB through for example, social-site's plugin on publisher sites

[0063] b. Searches

[0064] 3. Social graph and connections

[0065] These data sources are processed to extract user intent and profiles in terms of weighted sub-graphs of a collective/global concept graph, comprising nodes that are concepts, and edges that are relationships among such concepts. Concepts in embodiments of the invention include, but not limited to, phrases that represent entities (e.g. people, companies, drugs, diets, films, shows, events etc.), domain-specific terms (e.g., sports and medical terminologies, specific treatments, procedures etc.), or common expressions that are used to convey information. Relationships, as captured by annotated edges among concepts, include, but not limited to, measures of closeness among the concepts, e.g., co-occurrence statistics, or explicit semantic relationships (e.g., "acted in", "father of", "part of" etc.). For the purpose of demonstrating the invention, an exemplary concept graph can be used to capture and represent user profiles. The global concept graph is constructed by using the structure of the web (analyzing trillions of words) and an embodiment of it comprises of over 50 million concepts and 2 billion relationships, as disclosed, for example, in U.S. Pat. No. 7,958,120 and co-pending U.S. patent application Ser. No. 11/625,279, entitled "SYSTEMS AND METHODS FOR CREATING, NAVIGATING, AND SEARCHING INFORMATIONAL WEB NEIGHBORHOODS", U.S. patent application Ser. No. 12/436,748, entitled "METHODS AND APPARATUS FOR DISCOVERING RELEVANT CONCEPT AND CONTEXT FOR CONTENT SPECIFIC NODE", and U.S. patent application Ser. No. 12/906,051, "GENERATING A CONCEPTUAL ASSOCIATION GRAPH FROM LARGE-SCALE LOOSELY-GROUPED CONTENT", the entireties of which are hereby incorporated by reference. FIG. **1** shows a sub-graph of the global concept graph for the concept "Diabetes".

[0066] In embodiments of the invention, a sequence of page views and search activities can be followed on the concept graph and can be used to distill a subgraph that represents the intent as expressed by the sequence of page views. This is represented, for example, in FIGS. **2-4**, where a User "31b23njh31abcds23j2h123213" visited webMD on 2012-08-01 5:31 pm (FIG. **2**A), then the same user visited healthline on 2012-08-04 9:31 am (FIG. **3**A), and then searched on Google on 2012-08-04 9:31 pm (FIG. **4**A). All three activities are tracked as group of concepts in the same neighborhood around the concept "Diabetes" in the underlying concept graph (FIGS. **2**B, **3**B and **4**B).

[0067] FIG. **5** illustrates an example of a user profile based on content data that is typically found on a social media site: A Mr. X that is diabetic, enjoys or interested in sailing and in Rolex watches.

[0068] The data is processed to generate such profiles (i.e., the posts, page content, etc. that is processed for content). The data is selected based on a number of social media signals. The data includes, for example, Post Likes, Page and Group Likes, pages viewed/browsed on and off the social media site. The group or collective activities of users at the social media site can be further used to tag and prioritize both the unstructured data used to extract the user profile, as well as, the distilled user profile. For example, in one embodiment of the invention, pages and content sites are grouped based on communal user actions, such as Likes or posts made to the page/site. Thus, pages/sites/Groups can be divided into clusters, such that pages/sites belonging to the same clusters have been all visited/liked by a distinct group of users. Such a cluster of content sites and pages and Groups can then be collectively analyzed to create a collective content profile (again in terms of the global concept graph), which can then be used to tag the corresponding users. Moreover, usual page and domain statistics, such as page-rank and com-score ratings can be used to prioritize.

[0069] Furthermore, the user intent profiles (which include weighted clusters and sub-graphs of the global concept graph) can be tagged with temporal data, for example, certain subsets of a user's profile can be considered recent and time-sensitive (e.g., in the market to buy a product or go on a trip), whereas other aspects of user's profile could be longer lasting (e.g., regions of the concept graph that represent an antique aficionado or a diabetes patient). Similarly, one of the parameters for computing the weights in the sub-graph (representing a user's profile) can be based on how many of his/her friends also share the same nodes or edges. In yet another embodiment of this invention, a user profile is post-processed and tagged with weighted category scores defined over a structured taxonomy of interest. For example, a standard taxonomy can be used by the advertisement industry, and then be used to categorize the user profile (i.e., the sub-graph of the concept graph) in terms of the various categories in the taxonomy and compute a score for each such category. Since a number of advertisers define their campaigns based on categories, rather than the finer specificity of concept graphs, such coarse-grained tagging of user profiles can be used to more easily and directly match advertisements to users.

[0070] The user intent profiles then can be used to, for example, target the right advertisement to the right user at the right moment. The advertisers, in embodiments of the invention, include but not limited to, entities that carry out display and brand advertisement campaigns or perform search keyword auctions, or are lead-generators who aggregate demand across different platforms and media for merchants. Embodiments of the invention relate to assignment of a set of advertiser target profiles to each such advertiser, which again could be expressed as (i) weighted and time-tagged sub-graphs of the global concept graph or (ii) a weighted list of categories picked from a certain structured taxonomy. Such advertiser target profiles serve the same purpose as user intent profiles, i.e., provide a description of their target audience in terms of a common and globally-shared database, which is the concept graph. The act of targeting advertisement includes, but not limited to, determining for each user a weighted and prioritized list of advertisers that best match the user intent. One method for matching user intent to an advertiser is based upon computing the overlap or distance between the user intent profile and the advertiser target profile, e.g., by computing a measure of the distance (e.g. weighted shortest paths, or sum of the weights of the shared edges) between the two corresponding weighted sub-graphs in the global concept graph. The selection of the final set of advertisement units is completed by an optimization process that maximizes objective functions of interest, including revenue for the social media site, value and Return-On-Investment (ROI) for the advertisers, while considering the device and media that the user is on at the time of the impression.

[0071] For example, in one embodiment described in this invention, contextually relevant search-suggestions (i.e., keywords feed from a search engine) are shown to users while they are engaged in a particular conversation on the social media site.

[0072] In another embodiment of the invention, contextually relevant text or display advertisements are targeted to a

user while browsing a third-party web page (i.e., a web page not owned by the social media site). Yet another embodiment teaches how to show advertisement (both display and text units) to a user on a mobile device by targeting user profile.

[0073] Embodiments of the invention are also directed to use of user profiles and the underlying concept graph to enhance user engagement at a social media site. Such user engagement enhancement includes, but is not limited to, (i) organizing friends or contact lists of a user into potentially overlapping groups, by computing similarity between a user's profile and the profiles of those of his friends and contacts. An active user working in the finance sector, for example, will have a lot of activities on the social media site (or activities visible from the site), such as searches or browsing history, related to finance and investment concepts; these concepts would be shared with his friends/contacts from the finance sector, leading to overlaps in their intent profiles and allowing the invention to define a common profile and a group comprising these individuals. Typically, a user has friends that are from different contexts of one's life, e.g., work related, hobby related, high-school friends, college friends etc. and the invention provides a method for identifying and labeling these groups in an automated fashion. (ii) Organizing posts, comments and social interactions between a user and his friends in terms of the underlying context. For example, the linear and a long list of posts on wall or the home page of a social media site can be organized into different subjects and categories by automatically classifying them into categories such as sports, sports teams, travel etc. by mapping them into the concept graph.

[0074] This enables determination of the dominant types of social interactions that happen among friends, which can then be used for targeting advertisement. For example, if posts from a particular friend have a predominant theme of electronic gadgets (as determined by the invention's automated classification method), and he then recommends/likes a new phone, then such a recommendation would have a lot higher relevance than that coming from a cooking-enthusiast.

[0075] Embodiments of the invention are also directed to a Temporal Commercial Intent Profile (Collective Retargeting) that can be used to target advertisement and create commercially useful user intent profiles in a computationally efficient way from large-scale user data.

[0076] The goal is to generate a profile for each user that shows the level of urgency and interest in each of the standard commercial verticals. So, this can be used as an important factor in placement score of an ad on a page based on the cross relevancy of the page and ad to the top verticals of interests.

[0077] Output:

[0078] For each user we generate a scored version of a list, which is the standard targeting vertical used in the advertisement industry. Since this list is a standard, it makes it very easy to match against intention of the advertiser when we have such a profile for users.

[0079] Score could be separated into two parts as well:

[0080] 1—Urgency (time factor)

[0081] 2—Interest (has generic interest factor regardless of time as well)

[0082] Example: It is winter time and Mr. X is interested in Travel>Tourist Destinations>Mountain & Ski Resorts, the score of this vertical for my profile should only last until the season lasts and I'm still interested in taking or retaking such a trip based on my activities.

[0083] Inputs:

[0084] 1. Internal user activities:

[0085] a. Posts and Likes

[0086] b. Pages and groups Likes

[0087] c. Posts

[0088] d. Subscriptions

[0089] 2. External activities:

[0090] a. Incoming likes, sharing, recommendation coming from WEB through for example FaceBook social-plugin on publisher sites

[0091] 3. Social graph and connections

[0092] 4. Using social-plugin visits (not actions) as a retargeting pixel

[0093] 5. Concept Graph and all the related meta data

[0094] Besides being able to process content accurately, the main challenge would be filtering out large percentage of the collective data and only stick to the part which has relevant commercial intent hidden in it. For example if Mr. X has liked a page related to a poet, months back, that should be considered with a very low commercial intent.

[0095] Exemplary Implementations:

[0096] Add commercial intent score to concepts; this could be done using third party data, so we will have a DB of <CONCEPT,SCORED VERTICALS>. This db should be temporal as well, and we need to find a good trending method to keep this db updated. For example, Shopping>Toys should be boosted at Christmas time, or News>Politics should be boosted around election times.

Join <VISITED EXTERNAL URL,CONCEPTS> to <CONCEPT,SCORED VERTICALS> and create <URL,SCORED VERTICALS>

Join <USERS,VISITED EXTERNAL URL > to <URL,SCORED VERTICALS> and create <USERS,SCORED VERTICALS BASED ON EXTERNAL PAGES>

Do the same and join <FB PAGE,CONCEPT> to <CONCEPT,SCORED VERTICALS> and create <FB PAGE,SCORED VERTICALS>

Join <USERS,POSTS> to <POSTS,CONCEPTS> and create <USERS,CONCEPTS>

Join <USERS, CONCEPTS > to <CONCEPT,SCORED VERTICALS> and create <USERS,SCORED VERTICALS BASED ON POSTS>

Join <USERS, FB PAGES > to <CONCEPT,SCORED VERTICALS> and create <USERS,SCORED VERTICALS BASED ON FB PAGES>

[0097] We can use the social graph and cross correlation (overlap of users) on different FB pages to boost up or filter out any of the above data. The aggregation scoring can be done based above steps.

[0098] At the end of this process, we get mapping of <USER,AGGREGATED SCORED VERTICALS> which can then used to target advertisement as illustrated in FIG. **7**.

[0099] FIG. **9** shows a diagrammatic representation of machine in the exemplary form of a computer system **900** within which a set of instructions, for causing the machine to

perform any one or more of the methodologies discussed herein, may be executed. In alternative embodiments, the machine operates as a standalone device or may be connected (e.g., networked) to other machines. In a networked deployment, the machine may operate in the capacity of a server or a client machine in server-client network environment, or as a peer machine in a peer-to-peer (or distributed) network environment. The machine may be a personal computer (PC), a tablet PC, a set-top box (STB), a Personal Digital Assistant (PDA), a cellular telephone, a web appliance, a network router, switch or bridge, or any machine capable of executing a set of instructions (sequential or otherwise) that specify actions to be taken by that machine. Further, while only a single machine is illustrated, the term "machine" shall also be taken to include any collection of machines that individually or jointly execute a set (or multiple sets) of instructions to perform any one or more of the methodologies discussed herein.

[0100] The exemplary computer system **900** includes a processor **902** (e.g., a central processing unit (CPU), a graphics processing unit (GPU) or both), a main memory **904** (e.g., read only memory (ROM), flash memory, dynamic random access memory (DRAM) such as synchronous DRAM (SDRAM) or Rambus DRAM (RDRAM), etc.) and a static memory **906** (e.g., flash memory, static random access memory (SRAM), etc.), which communicate with each other via a bus **908**.

[0101] The computer system **900** may further include a video display unit **910** (e.g., a liquid crystal display (LCD) or a cathode ray tube (CRT)). The computer system **900** also includes an alphanumeric input device **912** (e.g., a keyboard), a cursor control device **914** (e.g., a mouse), a disk drive unit **916**, a signal generation device **920** (e.g., a speaker) and a network interface device **922**.

[0102] The disk drive unit **916** includes a computer-readable medium **924** on which is stored one or more sets of instructions (e.g., software **926**) embodying any one or more of the methodologies or functions described herein. The software **926** may also reside, completely or at least partially, within the main memory **904** and/or within the processor **902** during execution thereof by the computer system **900**, the main memory **904** and the processor **902** also constituting computer-readable media.

[0103] The software **926** may further be transmitted or received over a network **928** via the network interface device **922**.

[0104] While the computer-readable medium **924** is shown in an exemplary embodiment to be a single medium, the term "computer-readable medium" should be taken to include a single medium or multiple media (e.g., a centralized or distributed database, and/or associated caches and servers) that store the one or more sets of instructions. The term "computer-readable medium" shall also be taken to include any medium that is capable of storing, encoding or carrying a set of instructions for execution by the machine and that cause the machine to perform any one or more of the methodologies of the present invention. The term "computer-readable medium" shall accordingly be taken to include, but not be limited to, solid-state memories, and optical and magnetic media.

[0105] One or more of the methodologies or functions described herein may be embodied in a computer-readable medium on which is stored one or more sets of instructions (e.g., software). The software may reside, completely or at least partially, within memory and/or within a processor during execution thereof. The software may further be transmitted or received over a network.

[0106] It should be noted that the intent profile extraction systems, methods and applications are illustrated and discussed herein as having various modules which perform particular functions and interact with one another. It should be understood that these modules are merely segregated based on their function for the sake of description and represent computer hardware and/or executable software code which is stored on a computer-readable medium for execution on appropriate computing hardware. The various functions of the different modules and units can be combined or segregated as hardware and/or software stored on a computer-readable medium as above as modules in any manner, and can be used separately or in combination.

[0107] It should be noted that the invention is illustrated and discussed herein as having various modules which perform particular functions and interact with one another. It should be understood that these modules are merely segregated based on their function for the sake of description and represent computer hardware and/or executable software code which is stored on a computer-readable medium for execution on appropriate computing hardware. The various functions of the different modules and units can be combined or segregated as hardware and/or software stored on a computer-readable medium as above as modules in any manner, and can be used separately or in combination.

[0108] The term "computer-readable medium" should be taken to include a single medium or multiple media that store the one or more sets of instructions. The term "computer-readable medium" shall also be taken to include any medium that is capable of storing, encoding or carrying a set of instructions for execution by a machine and that cause a machine to perform any one or more of the methodologies of the present invention. The term "computer-readable medium" shall accordingly be taken to include, but not be limited to, solid-state memories, and optical and magnetic media.

[0109] Embodiments of the invention have been described through functional modules at times, which are defined by executable instructions recorded on computer readable media which cause a computer, microprocessors or chipsets to perform method steps when executed. The modules have been segregated by function for the sake of clarity. However, it should be understood that the modules need not correspond to discreet blocks of code and the described functions can be carried out by the execution of various code portions stored on various media and executed at various times.

[0110] It should be understood that processes and techniques described herein are not inherently related to any particular apparatus and may be implemented by any suitable combination of components. Further, various types of general purpose devices may be used in accordance with the teachings described herein. It may also prove advantageous to construct specialized apparatus to perform the method steps described herein. The invention has been described in relation to particular examples, which are intended in all respects to be illustrative rather than restrictive. Those skilled in the art will appreciate that many different combinations of hardware, software, and firmware will be suitable for practicing the present invention. Various aspects and/or components of the described embodiments may be used singly or in any combination. It is intended that the specification and examples be

considered as exemplary only, with a true scope and spirit of the invention being indicated by the claims.

1. A method comprising:

extracting an intent profile of a user based on activities executed and information shared on a social media site by the user, wherein extracting the intent profile comprises:

identifying structured user data from the social media site;

identifying user activities exclusively on the social media site;

identifying user activities involving the Internet;

identifying incoming likes, sharing, recommendation on the social media site;

identifying user searches;

determining a social graph and connections of the user; and

determining a weighted cluster and sub-graph of a global concept graph, wherein the global concept graph comprises nodes that are concepts, and edges that are relationships among such concepts; and

determining a user intent from the intent profile of the user.

2. The method of claim **1** further comprising

targeting an advertisement to the user based on the determined user intent.

3. The method of claim **2** further comprising:

generating a score for each vertical in a targeting vertical list.

4. The method of claim **3**, wherein the score comprises a time factor and an interest factor.

5. The method of claim **3**, wherein the score is generated based on user activities exclusively on the social media site, user activities involving the Internet, incoming likes, sharing, recommendation on the social media site, and user searches.

6. The method of claim **1** further comprising:

modifying a user experience of the user by organizing a contact list of the user and conversations and posts of the user based on their content and conceptual context.

7. The method of claim **6**, further comprising:

performing an aggregation scoring.

8. The method of claim **1**, wherein the structured data comprises an email identification, phone number, geo-location, friends and links.

9. The method of claim **1**, wherein the concepts comprise phrases that represent entities (e.g. people, companies, drugs, diets, films, shows, events etc.), domain-specific terms (e.g., sports and medical terminologies, specific treatments, procedures etc.), and common expressions that are used to convey information.

10. The method of claim **9**, wherein the relationships are identified by annotated edges among concepts.

11. The method of claim **9**, wherein the relationships comprise measures of closeness among the concepts, including at least one of co-occurrence statistics and explicit semantic relationships.

12. The method of claim **1**, further comprising tagging the user intent profiles with temporal data.

13. The method of claim **1**, wherein determining the weighted sub-graph is based on a number of the user's friends that share the same nodes or edges.

14. The method of claim **13**, further comprising tagging unstructured data using collective activities of users at the social media site.

15. The method of claim **13**, further comprising prioritizing unstructured data using collective activities of users at the social media site.

16. The method of claim **1**, further comprising post-processing and tagging the intent profile with weighted category scores defined over a structured taxonomy of interest.

17. The method of claim **1**, further comprising assigning each such advertiser a set of advertiser target profiles.

18. The method of claim **17**, wherein the advertiser target profile comprises one or more weighted and time-tagged sub-graphs of the global concept graph.

19. The method of claim **17**, wherein the advertiser target profile comprises a weighted list of categories picked from a structured taxonomy.

20. The method of claim **1**, further comprising determining for each user a weighted and prioritized list of advertisers that best match the user intent.

21. The method of claim **20**, wherein matching user intent to an advertiser comprises computing an overlap or distance between the user intent profile and the advertiser target profile by computing a measure of the distance between the two corresponding weighted sub-graphs in the global concept graph.

22. The method of claim **20**, wherein selecting a final set of advertisement units is completed by an optimization process that maximizes objective functions of interest, including revenue for the social media site, value and Return-On-Investment (ROI) for the advertisers, while considering the device and media that the user is on at the time of the impression.

23. The method of claim **6**, wherein modifying a user engagement comprises organizing friends or contact lists of a user into potentially overlapping groups by computing similarity between a user's profile and the profiles of those of his friends and contacts.

24. The method of claim **23**, wherein organizing friends or contact lists of a user into potentially overlapping groups comprise computing similarity between a user's profile and the profiles of those of the user's friends and contacts.

25. The method of claim **6**, wherein modifying a user engagement comprises organizing posts, comments and social interactions between a user and his friends based on an underlying context.

26. The method of claim **25**, wherein a linear list of posts on page of the social media site can be organized into categories by automatically classifying the posts by mapping the posts to categories in the global concept graph.

27. The method of claim **1**, further comprising post-processing and tagging the user profile with weighted category scores defined over a structured taxonomy of interest.

28. The method of claim **1**, further comprising determining a suggested search term based on the intent profile.

29. The method of claim **1**, further comprising:

grouping content based on communal user actions; and

dividing the grouped content into clusters.

30. The method of claim **29**, further comprising:

generating a collective content profile; and

tagging the corresponding users.

31. The method of claim **30**, further comprising prioritizing the grouped content based on page and domain statistics.

* * * * *