



US 20150095339A1

(19) **United States**

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

(10) **Pub. No.: US 2015/0095339 A1**

(43) **Pub. Date: Apr. 2, 2015**

(54) **IDENTIFYING MEMBERS OF A SMALL & MEDIUM BUSINESS SEGMENT**

(52) **U.S. Cl.**  
CPC ..... *G06F 17/3071* (2013.01); *G06Q 50/01* (2013.01)

(71) Applicant: **LinkedIn Corporation**, Mountain View, CA (US)

USPC ..... **707/740**

(72) Inventors: **Jiaqi Guo**, Mountain View, CA (US);  
**Baoshi Yan**, Belmont, CA (US); **Alexis B. Baird**, San Francisco, CA (US);  
**Daniel Tunkelang**, Mountain View, CA (US)

(57) **ABSTRACT**

A system, apparatus, computer-program product and method are provided for identifying members of a particular user segment. One particular user segment of interest includes members who work for employers having a number of employees within a predetermined range. Members of an online service provide data purporting to identify their employers and/or other personal or professional attributes. The data entries are normalized by standardizing terms, removing superfluous or unneeded terms, and/or performing other processing. The members are then clustered according to their normalized employer names, and members within clusters that have sizes within the range are added to the user segment. Invalid employer names (e.g., fictitious companies, non-existent entities) may be filtered out. Within a cluster, a professional social networking site (or other site) may be analyzed to determine if the clustered members have developed relationships; if not, the cluster may be cancelled.

(21) Appl. No.: **14/318,326**

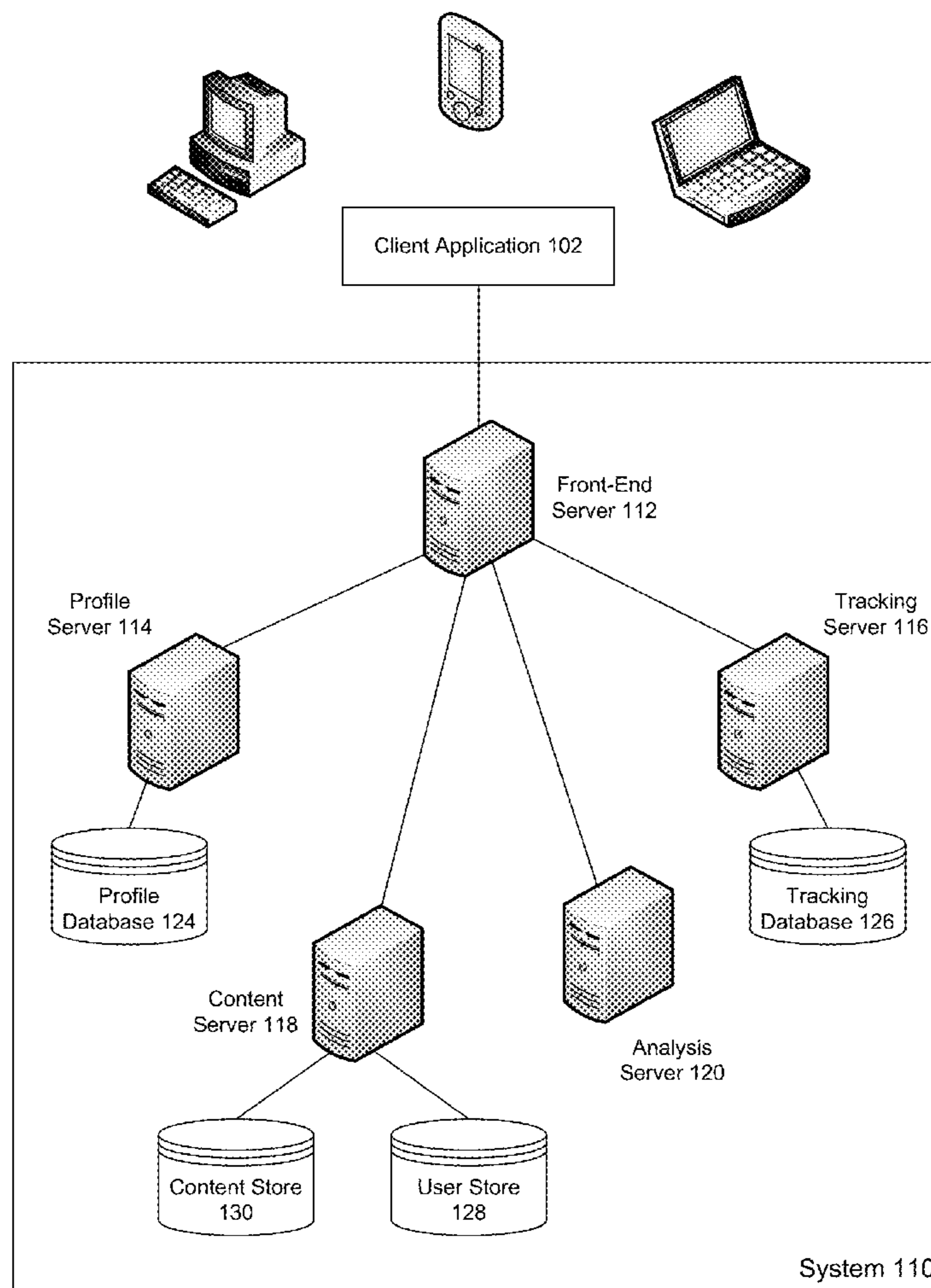
(22) Filed: **Jun. 27, 2014**

**Related U.S. Application Data**

(63) Continuation of application No. 14/044,218, filed on Oct. 2, 2013, now Pat. No. 8,831,969.

**Publication Classification**

(51) **Int. Cl.**  
*G06F 17/30* (2006.01)  
*G06Q 50/00* (2006.01)



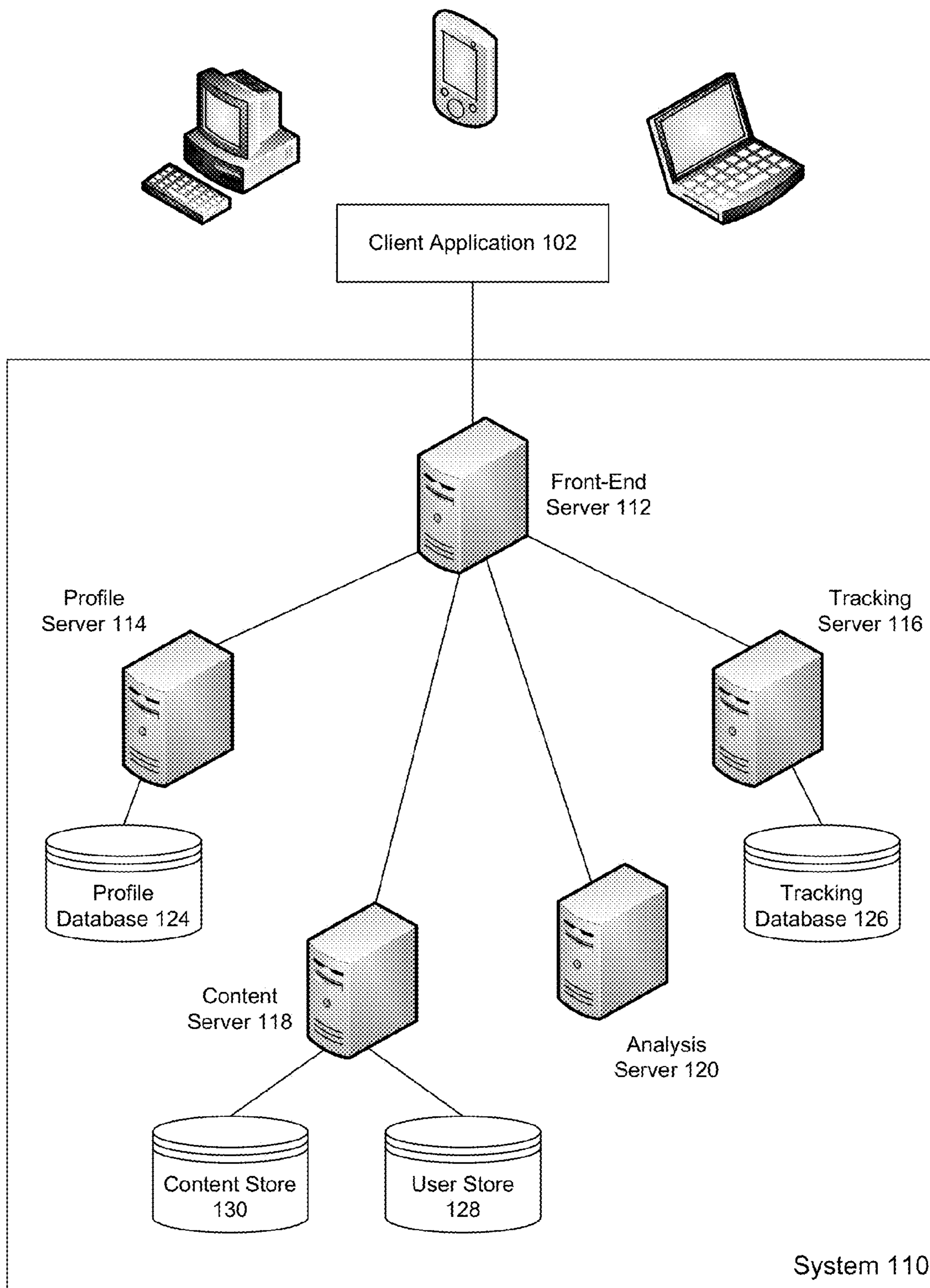


Fig. 1

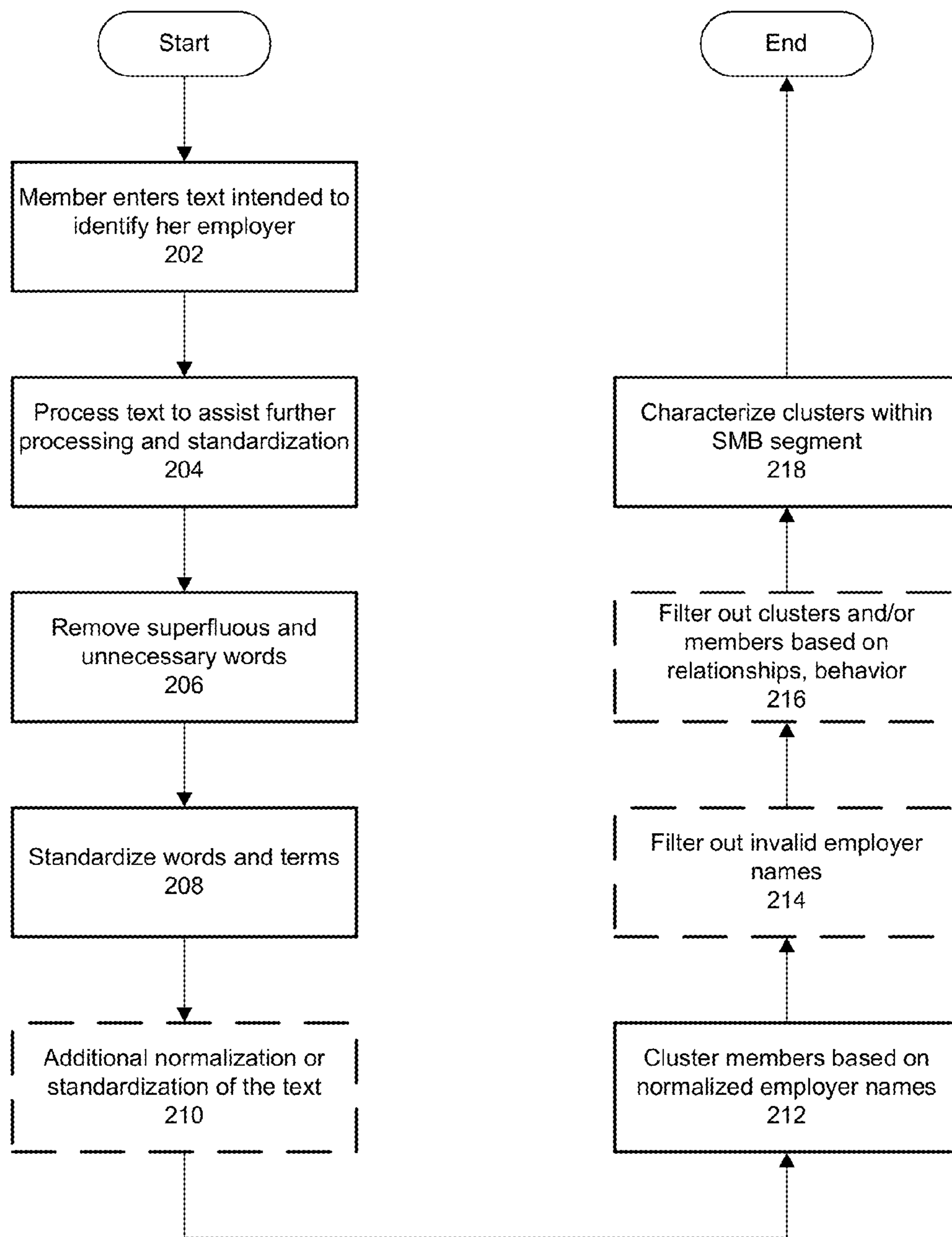


Fig. 2

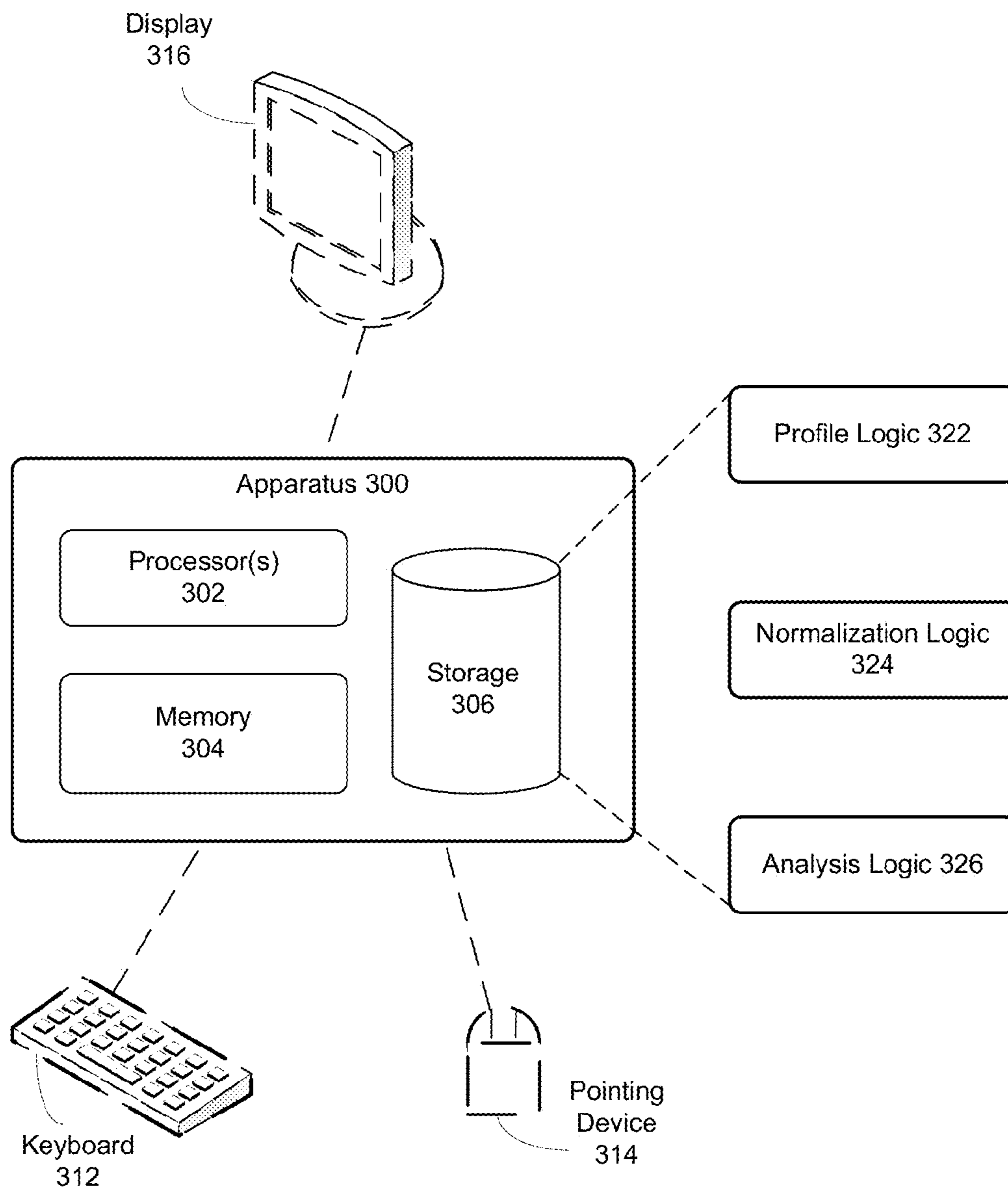


Fig. 3



## IDENTIFYING MEMBERS OF A SMALL & MEDIUM BUSINESS SEGMENT

### RELATED APPLICATION

[0001] This application is a continuation of, and claims priority under 35 U.S.C. §120 to, pending U.S. patent application Ser. No. 14/044,218, filed Oct. 2, 2013.

### BACKGROUND

[0002] A system, apparatus, method and computer-program product are provided for processing and analyzing profiles of users of a computer or online system or service, to determine whether they belong to a predefined category or segment of users.

[0003] Online systems and services maintain profiles of their users, which the users themselves may be able to configure or modify. Some users, however, do not correctly or completely configure their profiles, whether due to a lack of interest, lack of time and/or some other reason. For example, they may provide a partial answer to a requested field or data attribute, may provide a short-hand answer that is not readily understood, may misspell words or use unusual abbreviations when entering information, etc.

[0004] An online system or service may desire to make an offer to users that match a particular attribute or characteristic, but for those users who have not fully or correctly configured their profiles, or who have otherwise provided incorrect or incomplete information, the system will be unable to determine whether they qualify.

### DESCRIPTION OF THE FIGURES

[0005] FIG. 1 is a block diagram depicting a system in which users or members are categorized or placed into predefined user segments, in accordance with some embodiments.

[0006] FIG. 2 is a flow chart demonstrating a method of inferring membership in a particular user segment, in accordance with some embodiments.

[0007] FIG. 3 is a block diagram of an apparatus for segmenting or characterizing users or members of an online system or service, in accordance with some embodiments.

### DETAILED DESCRIPTION

[0008] Embodiments of a system, apparatus, method and computer-program product are provided for determining whether a user (e.g., a user of an online service) belongs to a particular category or segment of users, without any clear or explicit indication of such membership. In these embodiments, for example, different users that characterize themselves differently, such as by configuring a given user attribute(s) in different manners, are nonetheless determined to belong to the same category or segment of users.

[0009] FIG. 1 is a block diagram depicting a system in which user profiles are analyzed in order to categorize the corresponding users or determine whether they belong to a particular predefined segment, according to some embodiments.

[0010] System 110 hosts an online service or application for use by operators or users of client devices that execute a compatible client application 102. A client device may be stationary (e.g., a desktop computer, a workstation) or mobile (e.g., a smart phone, a tablet computer, a laptop computer) in nature. Client application 102 may be a browser program or

an application designed specifically to access the service(s) offered by system 110. Users of system 110 may be termed members because they may be required to register with the system in order to access all system services or receive all benefits offered by the system.

[0011] User connections are generally made through front-end server 112, which may comprise a portal, web server, application server, data server and/or other software and hardware modules. System 110 also includes profile server 114, tracking server 116, content server 118, analysis server 120, profile database 124, tracking database(s) 126, user data store 128 and content store 130. The system may include other components not depicted in FIG. 1.

[0012] In some specific embodiments, system 110 hosts a professional social networking service or site that allows and helps users to create, develop and maintain professional (and personal) relationships. As part of the service, system 110 serves content for presentation to users via the client application, which may include content generated or suggested by other users (e.g., text, images, video, audio), messages between users, offers from members and/or non-members, advertisements, announcements, job listings, status updates, and so on.

[0013] Profile server 114 maintains profiles, in profile database 124, of members of the service hosted by system 110. A member's profile may reflect any number of attributes or characteristics of the member, including personal (e.g., gender, age or age range, interests, hobbies), professional (e.g., employment status, job title, employer, skills, endorsements, professional awards), social (e.g., organizations the user is a member of, geographic area of residence, friends), educational (e.g., degree(s), university attended, other training), etc.

[0014] Tracking server 116 monitors and records activity of system 110 and/or members (e.g., in tracking database(s) 126). For example, whenever content is served from front-end server 112 (e.g., to a client device), the tracking server records what is served, to whom (e.g., which member), and when. Similarly, the tracking server also records member actions regarding advertisements and/or other content presented to the members (e.g., clicks, conversions, follow-on requests), to include identities of the member and the content acted upon, what action was taken, when, etc.

[0015] Content server 118 maintains one or more repositories of content items for serving to members (e.g., content store 130), an index of the content items, and user store 128. Content store 130 may primarily store advertisements for serving to members, including both revenue and non-revenue ads, but it may also store other content, such as images, documents, video, audio, announcements and so on, for use by various components of system 110. Content server 118 (or some other component of system 110) may include a recommendation module for recommending content to serve.

[0016] When content is stored in content store 130, it may be stored with attributes, indications, characteristics and/or other information describing one or more target audiences (e.g., recipients) of the content. For example, a provider of an advertisement may identify relevant attributes and desired values of target members, a provider of a job listing may identify attributes of members that should be informed of the job opening, an organization wishing to obtain more followers/subscribers/fans may identify the type(s) of members it would like to attract, and so on.



[0017] In particular, a content item's target audience may identify a particular segment of members or users that the item should be served to. A user segment may be defined in many different ways. For example, different segments may be defined for different values (or ranges of values) for different attributes of profiles of members of the service hosted by system 110. Thus, depending on a given member's data entered for a particular attribute, he or she may be placed in one or more user segments.

[0018] User store 128 stores, for each member of system 110, a record of content items served to the member, or served for presentation to the member, and when they were served. In particular, user store 128 may be configured to allow the content server and/or other components of system 110 to quickly determine whether a particular content item was previously presented to a particular member, how many times it was presented, when it was presented, how it was presented (e.g., how prominently or where it was presented within a web page or other page of content), and/or other information. Although some of this data may duplicate what is stored by tracking server 116, contents of user store 128 are rapidly accessible to one or more servers (especially content server 118), and may be used to help select a content item to serve in response to a current request.

[0019] Analysis server 120, discussed in more detail below, analyzes member profiles (in database 124) and/or other data (e.g., contents of user store 128 and/or tracking database 126) and attempts to characterize members in ways they have not explicitly characterized themselves, possibly using incomplete or unclear data.

[0020] System 110 may include other components not illustrated in FIG. 1. Also, in other embodiments, the functionality of the system may be distributed differently among the illustrated components, such as by merging or further dividing functions of one or more components, or may be distributed among a different collection of components. Yet further, while implemented as separate hardware components (e.g., computer servers) in FIG. 1, one or more of front-end server 112, profile server 114, tracking server 116, content server 118 and analysis server 120 may alternatively be implemented as separate software modules executing on one or more computer servers.

[0021] As indicated above, in some embodiments analysis server 120 is configured to identify members who can be justifiably placed into predefined user segments, with or without explicit knowledge that the categorization is apt. For example, it may be helpful to know which members of the site hosted by system 110 are employed by a small- or medium-sized business. Such members may be tagged accordingly, as belonging to a corresponding segment of the user population. This illustrative segment may be termed an SMB (Small and Medium Business) segment. A small or medium business may be defined in a particular manner by system 110 (e.g., as an organization having less than 200 employees), or a definition employed by a third party may be applied (such as a definition used by the Small Business Administration).

[0022] Some members may explicitly (and correctly) identify their employer. Further, these members may indicate the sizes of their employers, or their employers' sizes may be readily learned through other sources (e.g., websites of those employers, a web page on system 110 dedicated to the employer, a business directory). These members may be explicitly determined to belong to the SMB segment.

[0023] Other members, however, may misspell their employers' names, may use an abbreviation or alternate form of an employer's name, may provide only a partial name, may provide an incorrect name, etc. The system may attempt to infer whether these members belong to the SMB segment.

[0024] In other embodiments, analysis server 120 operates to identify members that belong to other segments, such as a segment encompassing employees of not-for-profit organizations, employees of educational institutions, employees within particular industries (e.g., finance, manufacturing), employees or members of specific organizations (e.g., the U.S. government, a state agency, a fraternal organization), etc.

[0025] FIG. 2 is a flow chart demonstrating a method of inferring membership in a particular member segment, according to some embodiments. Although the method is described as it may be applied to infer membership in an SMB segment, it may be modified for use with other member segments, as will be understood by one of ordinary skill in the art.

[0026] In operation 202, a member (or user) populates her user profile for an online system or service, and provides information that ostensibly identifies her employer. However, the information does not match any known employer, or at least does not match a standardized or normalized form of known employers' names, and may or may not match any other member's employer data.

[0027] The name or description entered by the member may be a partial name, an abbreviation or acronym, or may include a spelling error, a format error or some other error. In short, the data she provides does not identify her employer using a standard or normal name. For example, the member may have entered her preferred or pet name for her employer, rather than selecting from a list of known employers offered by the system or service (if such a list is offered). As another example, she may be the first employee of her employer to have registered with the system.

[0028] In some embodiments, the data entered by the member in operation 202 is compared to a list of known employers. If the data matches a known employer, the method may be aborted if the system already knows whether the employer is a small- or medium-sized business, in which case the system can explicitly determine whether the member belongs to the SMB segment.

[0029] In some embodiments, an employer is any organization that employs people, including companies, corporations, partnerships, government agencies and departments, not-for-profit entities, etc. A "known" employer may be any employer that has a page or presence on the system (e.g., an employer that is a member of the system's service(s)), that is identified in at least one user's profile (in a normalized/standardized format), or that is otherwise known to exist.

[0030] The purported employer name will now be normalized or standardized to allow the system to compare and/or aggregate it with data entered by other members.

[0031] In operation 204, some initial processing is applied to the employer name. For example, non-ASCII characters (e.g., å, ñ) may be changed to ASCII characters (e.g., a, n), parts of speech may be identified (e.g., nouns, adjectives), the entry may be parsed (e.g., if it contains a period or comma), extraneous white space (e.g., space characters) may be removed, etc.

[0032] In operation 206, select words are removed as necessary from the entry. Words that are removed may include 'stop' words—such as "a," "an," "and," and "the"—except,



perhaps, for those that appear as the first word in the employer description. For example, an employer such as “The Johns Hopkins University” may properly include a stop word, whereas other stop words found in the middle or at the end of an entry are likely to be superfluous. However, special cases may be identified that allow retention of a stop word that would otherwise be excised.

**[0033]** Other words that do little or nothing to disambiguate or distinguish between employers may be removed, such as Corporation, Limited, Company, Office, etc. (or abbreviations of these words). These words often appear as suffixes at the end of employers’ names.

**[0034]** Removal of stop words and superfluous terms (and/or abbreviations) may be context-sensitive so as not to remove distinctive uses of such words. For example, “Office” may be removed from an employer name such as “Acme Law Office,” but not from a name such as “United States Post Office.”

**[0035]** Before removing a word in operation **206**, the system may try to determine if the word is an acronym, a proper name or whether it otherwise adds meaning to the employer name. For example, in an employer name such as “AND Consulting,” the word “AND” should not be removed because it is a distinguishing feature of the name.

**[0036]** In operation **208**, words or terms may be standardized if appropriate. For example, abbreviations, acronyms or monikers such as “Ltd.,” “Corp.,” “Company,” “LLP,” “Inc.,” and so on, may be standardized into formats that do or do not include abbreviations (e.g., “Ltd.” versus “Limited”, “Corp.” versus “Corporation”).

**[0037]** In optional operation **210**, other normalization or standardization action may be taken, such as to capitalize (or un-capitalize) words, change non-standard spellings (e.g., words spelled differently in British English than in American English), translate from one language into another, add or remove punctuation (e.g., periods, commas), expand an abbreviation or acronym to its full text (or vice versa), etc.

**[0038]** In operation **212**, the system clusters members, including the member who provided data in operation **202**, based on their normalized employer names. For each unique normalized employer name, clustering may involve counting the number of members that submitted data entries that were normalized into that employer name. In some implementations, the system periodically or regularly performs the processing described herein to normalize unknown employer information of multiple members, and then aggregates any matches.

**[0039]** Thus, if different users entered “JHU,” “J.H.U.,” “The JHU” and/or other variations of the name “The Johns Hopkins University,” the system aggregates some or all of these, with the knowledge that those members were referring to the same employer when they provided their employer data. However, depending on the type and amount of processing above, multiple clusters for one employer may be formed. For example, for an organization whose name includes suffixes, prefixes, stop words, common words and/or words that are often abbreviated or represented by initials, a great variety of name forms may be possible and multiple clusters may result. These clusters may or may not be combined.

**[0040]** In optional operation **214**, clusters corresponding to invalid employer names are removed from consideration. Invalid names include names of employers that are non-existent, fictitious, erroneous or invented (e.g., “unemployed,” “Retired,” “work sucks”), those that consist of profane words

or are random sequences of characters that are unlikely to be valid employer names (e.g., “xyzyzy”), etc. Illustratively, a member may have entered an invalid name to avoid admitting that she is unemployed, to avoid revealing her employer’s name, or for some other purpose.

**[0041]** A human operator of the system may manually identify some invalid names (and/or valid names), and/or may train the system to recognize them. In some embodiments, operation **214** may be executed earlier (e.g., during or after normalization of a member’s data entry).

**[0042]** In optional operation **216**, members’ relationships or behavior are examined to determine whether members that have been clustered together are likely to actually work for the same employer. This operation helps the system avoid improperly characterizing members as belonging to the SMB segment.

**[0043]** In some implementations, in which the system hosts a professional social networking site, the system looks for links between clustered members—such as whether they are friends or associates, whether they share content or exchange messages, etc. Because employees of a single small business (or a medium-sized business) are likely to have at least some common professional relationships and/or person relationships, if few or none of the members grouped into one cluster have formed relationships or shared content, it is possible that they invented the same (or similar) employer name. In some implementations, if at least a threshold of members in the same cluster have not formed at least one social relationship with another member in that cluster (e.g., 10%, 25%, 40%), the cluster may be deemed invalid or some members (e.g., those that have not formed any relationships) may be dropped from the cluster.

**[0044]** In operation **218**, remaining clusters that have appropriate numbers of members (e.g., less than 200) are deemed to represent small- and medium-sized businesses. Profiles of these members may be notated appropriately, the members may receive offers, ads and/or other content directed to employees of small- and medium-sized businesses, etc.

**[0045]** The method illustrated in FIG. **2** and described above is but one method by which a system or service that hosts users or members is able to characterize or categorize those users by segment. In other embodiments, the same and/or different operations may be performed in a similar or different manner to obtain a similar result, as will be understood by one of ordinary skill in the art.

**[0046]** In particular, normalization or standardization of employer names entered by members may vary from one embodiment to another, individual normalization operations (e.g., operations **204** through **210** of FIG. **2**) may be performed in a different order, fewer (or more) such operations may be performed, etc.

**[0047]** The method of FIG. **2** may be repeated on a regular or irregular basis. A given member’s entry for her employer’s name may be analyzed multiple times, or may be analyzed only until that name has been normalized and matched with a known employer. As new employer names are learned through this process, they may be added to a list or lists of known employer names.

**[0048]** FIG. **3** is a block diagram of an apparatus for segmenting users or members of an online system or service, according to some embodiments.

**[0049]** Apparatus **300** of FIG. **3** comprises processor(s) **302**, memory **304** and storage **306**, which may comprise one



or more optical, solid-state and/or magnetic storage components. Apparatus **300** may be coupled (permanently or transiently) to keyboard **312**, pointing device **314**, and display **316**.

**[0050]** Storage **306** of the apparatus may store content for serving to/for members, member data (e.g., profiles), tracking data, and/or other information. Storage **306** also stores logic that may be loaded into memory **304** for execution by processor **302**. Such logic includes profile logic **322**, normalization logic **324** and analysis logic **326**. In other embodiments, any or all of these logic modules or other content may be combined or divided to aggregate or separate their functionality as desired.

**[0051]** Profile logic **322** comprises processor-executable instructions for assembling or populating member profiles. In some implementations, logic **322** solicits and receives from members information intended to identify their employers, other organizations they are affiliated with (e.g., their schools or universities), products or brands they enjoy, content they enjoy, identities of friends and associates, etc.

**[0052]** Normalization logic **324** comprises processor-executable instructions for normalizing or standardizing data obtained by profile logic **322**. As described above, for example, the normalization logic may modify a member's purported employer name to place it into a format with which it can be easily compared and clustered with names of other members' employers. Normalization may include virtually any type of natural language processing, editing, formatting and/or other manipulation.

**[0053]** Analysis logic **326** comprises processor-executable instructions for analyzing member information and characterizing or categorizing them into one or more member segments (e.g., the SMB segment discussed above). Logic **326** may therefore group members that reported the same or similar information for a particular field or part of their profile, filter out invalid data and appropriately categorize members according to the member segments they match. Analysis logic **326** may also operate to filter out invalid data (e.g., fictitious employer names), and may look for interrelationships among members believed to belong to the same category or segment.

**[0054]** In some embodiments of the invention, apparatus **300** performs most or all of the functions described in relation to system **110** of FIG. **1**, and such functions may be divided among the same collections of logic depicted in FIG. **3** and/or others. In some particular implementations, apparatus **300** may host multiple virtual computer servers performing the functions of one or more of the servers of system **110**.

**[0055]** An environment in which some embodiments of the invention are executed may incorporate a general-purpose computer or a special-purpose device such as a hand-held computer or communication device. Some details of such devices (e.g., processor, memory, data storage, display) may be omitted for the sake of clarity. A component such as a processor or memory to which one or more tasks or functions are attributed may be a general component temporarily configured to perform the specified task or function, or may be a specific component manufactured to perform the task or function. The term "processor" as used herein refers to one or more electronic circuits, devices, chips, processing cores and/or other components configured to process data and/or computer program code.

**[0056]** Data structures and program code described in this detailed description are typically stored on a non-transitory computer-readable storage medium, which may be any

device or medium that can store code and/or data for use by a computer system. Non-transitory computer-readable storage media include, but are not limited to, volatile memory, non-volatile memory, magnetic and optical storage devices such as disk drives, magnetic tape, CDs (compact discs) and DVDs (digital versatile discs or digital video discs), solid-state drives and/or other non-transitory computer-readable media now known or later developed.

**[0057]** Methods and processes described in the detailed description can be embodied as code and/or data, which can be stored in a non-transitory computer-readable storage medium as described above. When a processor or computer system reads and executes the code and manipulates the data stored on the medium, the processor or computer system performs the methods and processes embodied as code and data structures and stored within the medium.

**[0058]** The foregoing descriptions of embodiments of the invention have been presented for purposes of illustration and description only. They are not intended to be exhaustive or to limit the invention to the forms disclosed. Accordingly, many modifications and variations will be apparent to practitioners skilled in the art. The scope of the invention is defined by the appended claims, not the preceding disclosure.

What is claimed is:

1. A computer system-implemented method for identifying users associated with employers having a number of employees within a predetermined range, the method comprising:
  - receiving, from multiple users, employer information intended to identify the users' employers;
  - normalizing the employer information, using a computer processor, to yield normalized employer names, wherein said normalizing is context-sensitive so as not to remove a distinguishing feature of the normalized employer names;
  - using the computer processor to cluster the multiple users to produce a set of clusters based on the normalized employer names, wherein each cluster is associated with a different normalized employer name; and
  - for each cluster containing a number of users within the predetermined range, categorizing users within the cluster as members of a segment of users associated with employers having a number of employees within the predetermined range.
2. The method of claim **1**, further comprising:
  - analyzing relationships between users in a single cluster to determine if they are likely to be employed by the same employer.
3. The method of claim **2**, wherein analyzing the relationships comprises determining whether the users have formed relationships in a social network.
4. The method of claim **2**, wherein analyzing the relationships comprises determining whether the users have exchanged content.
5. The method of claim **1**, further comprising:
  - for each of one or more clusters, determining whether the cluster's associated normalized employer name is invalid.
6. The method of claim **5**, wherein determining whether the cluster's associated normalized employer name is invalid comprises comparing the associated normalized employer name to a list of known invalid names.
7. The method of claim **5**, wherein determining whether the cluster's associated normalized employer name is invalid



comprises analyzing a social networking site to determine if users within the cluster have formed social relationships.

**8.** The method of claim **1**, wherein normalizing the employer information comprises standardizing one or more terms of an employer name.

**9.** The method of claim **8**, wherein said normalizing the employer information further comprises removing one or more superfluous terms.

**10.** The method of claim **1**, wherein the predetermined range is between 1 and approximately 200.

**11.** The method of claim **1**, further comprising:  
analyzing, using the computer processor, social network relationships between users in a single cluster associated with a single employer name, to determine whether they are likely to be employed by the same employer; and  
determining whether at least a predetermined threshold number of users in the same cluster have formed at least one social relationship, in a social network, with another user in that cluster.

**12.** An apparatus for identifying users associated with employers having a number of employees within a predetermined range, the apparatus comprising:

one or more processors; and  
memory configured to store instructions that, when executed by the one or more processors, cause the apparatus to:  
receive from multiple users employer information intended to identify the users' employers;  
normalize the employer information to yield normalized employer names, wherein said normalizing is context-sensitive so as not to remove a distinguishing feature of the normalized employer names;  
cluster the multiple users to produce a set of clusters based on the normalized employer names, wherein each cluster is associated with a different normalized employer name; and  
for each cluster containing a number of users within the predetermined range, categorize users within the cluster as members of a segment of users associated with employers having a number of employees within the predetermined range.

**13.** The apparatus of claim **12**, wherein the memory is further configured to store instructions that, when executed by the one or more processors, cause the apparatus to:

analyze relationships between users in a single cluster to determine if they are likely to be employed by the same employer.

**14.** The apparatus of claim **13**, wherein analyzing the relationships comprises determining whether the users have formed relationships in a social network.

**15.** The apparatus of claim **12**, wherein the memory is further configured to store instructions that, when executed by the one or more processors, cause the apparatus to:

for each of one or more clusters, determine whether the cluster's associated normalized employer name is invalid.

**16.** The apparatus of claim **15**, wherein determining whether a cluster's associated normalized employer name is invalid comprises comparing the associated normalized employer name to a list of known invalid names.

**17.** The apparatus of claim **15**, wherein determining whether a cluster's associated normalized employer name is invalid comprises analyzing a social networking site to determine if users within the cluster have formed social relationships.

**18.** The apparatus of claim **12**, wherein normalizing the employer information comprises:

standardizing one or more terms of an employer name; and  
removing one or more superfluous terms.

**19.** The apparatus of claim **12**, wherein the predetermined range is between 1 and approximately 200.

**20.** The apparatus of claim **12**, wherein the memory is further configured to store instructions that, when executed by the one or more processors, cause the apparatus to:

analyze, using the computer processor, social network relationships between users in a single cluster associated with a single employer name, to determine whether they are likely to be employed by the same employer; and  
determine whether at least a predetermined threshold number of users in the same cluster have formed at least one social relationship, in a social network, with another user in that cluster.

\* \* \* \* \*