



(19) **United States**

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

(10) **Pub. No.: US 2018/0014038 A1**

(43) **Pub. Date: Jan. 11, 2018**

(54) **METHOD AND APPARATUS FOR  
SELECTING A NETWORK RESOURCE AS A  
SOURCE OF CONTENT FOR A  
RECOMMENDATION SYSTEM**

*G06F 9/50* (2006.01)

*H04N 21/80* (2011.01)

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

CPC ..... *H04N 21/235* (2013.01); *G06K 9/6256*

(2013.01); *H04L 47/80* (2013.01); *G06K*

*9/6257* (2013.01); *G06F 9/50* (2013.01); *H04N*

*21/80* (2013.01); *H04L 41/08* (2013.01)

(71) Applicant: **YANDEX EUROPE AG**, Luzern (CH)

(72) Inventors: **Victor Grigorievich LAMBURT**,  
Moscow (RU); **Igor Igorevich LIFAR**,  
Novorossiysk (RU)

(57) **ABSTRACT**

There are disclosed a method of and a system for selecting a network resource as a source of a content item, the content item to be analyzed by a recommendation system as part of a plurality of content items to generate a set of recommended content items as a recommendation for a given user of the recommendation system. The method comprises, for a network resource, receiving, by the server, a plurality of features associated with a network resource to be processed; generating given network resource profile for the network resource, the given network resource profile being based on the plurality of features; executing a machine learning algorithm in order to determine a source suitability parameter for the network resource, selecting at least one content item from the network resource if the source suitability parameter is determined to be above a pre-determined threshold.

(21) Appl. No.: **15/606,326**

(22) Filed: **May 26, 2017**

(30) **Foreign Application Priority Data**

Jul. 7, 2016 (RU) ..... 2016127446

**Publication Classification**

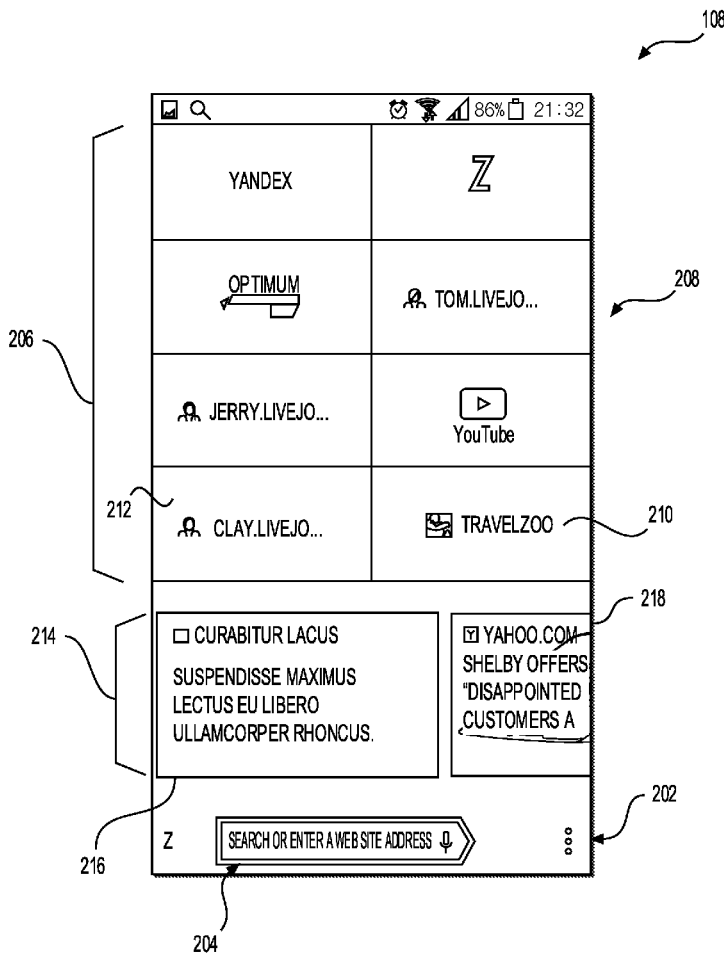
(51) **Int. Cl.**

*H04N 21/235* (2011.01)

*G06K 9/62* (2006.01)

*H04L 12/927* (2013.01)

*H04L 12/24* (2006.01)



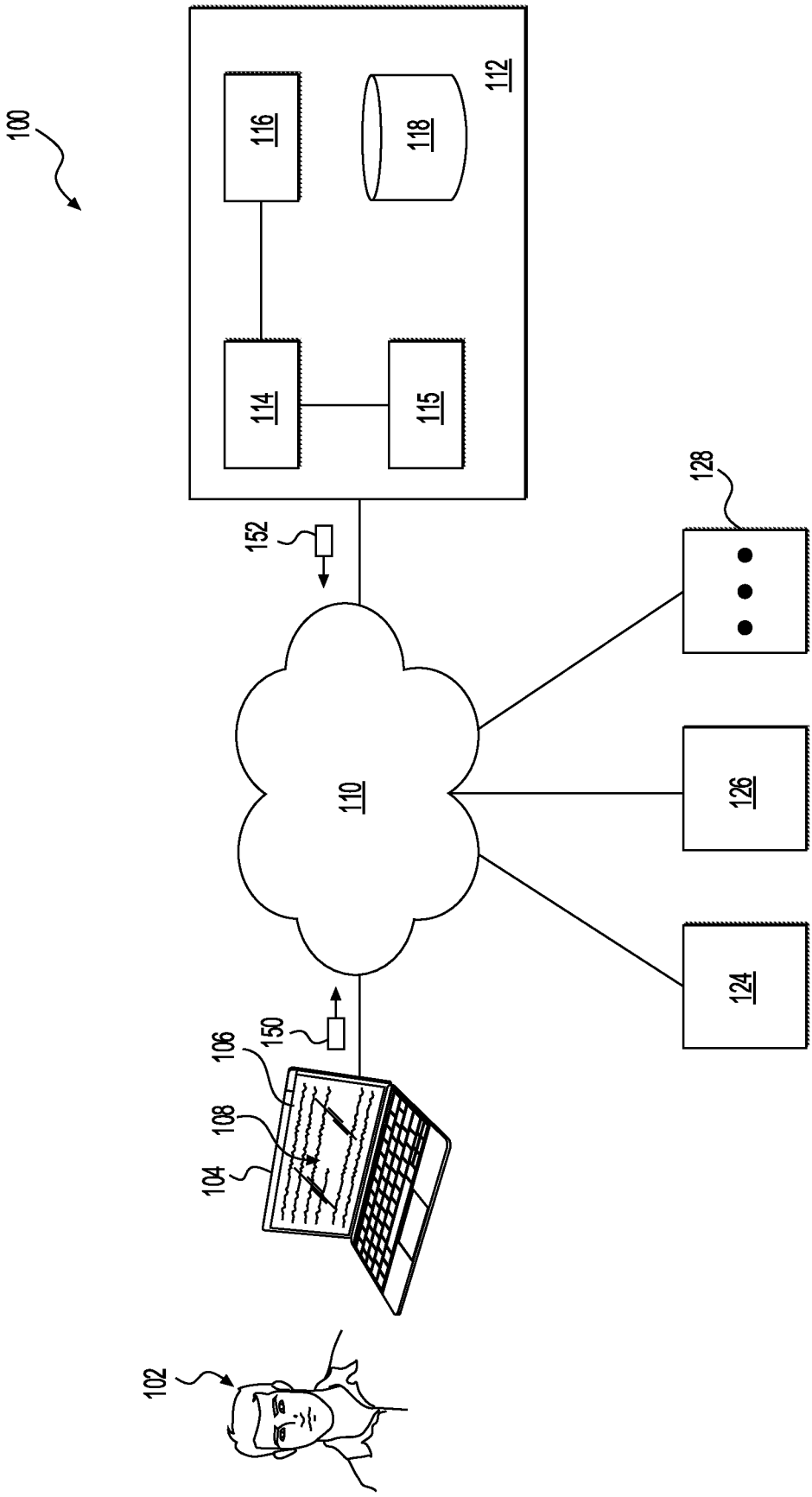


FIG. 1

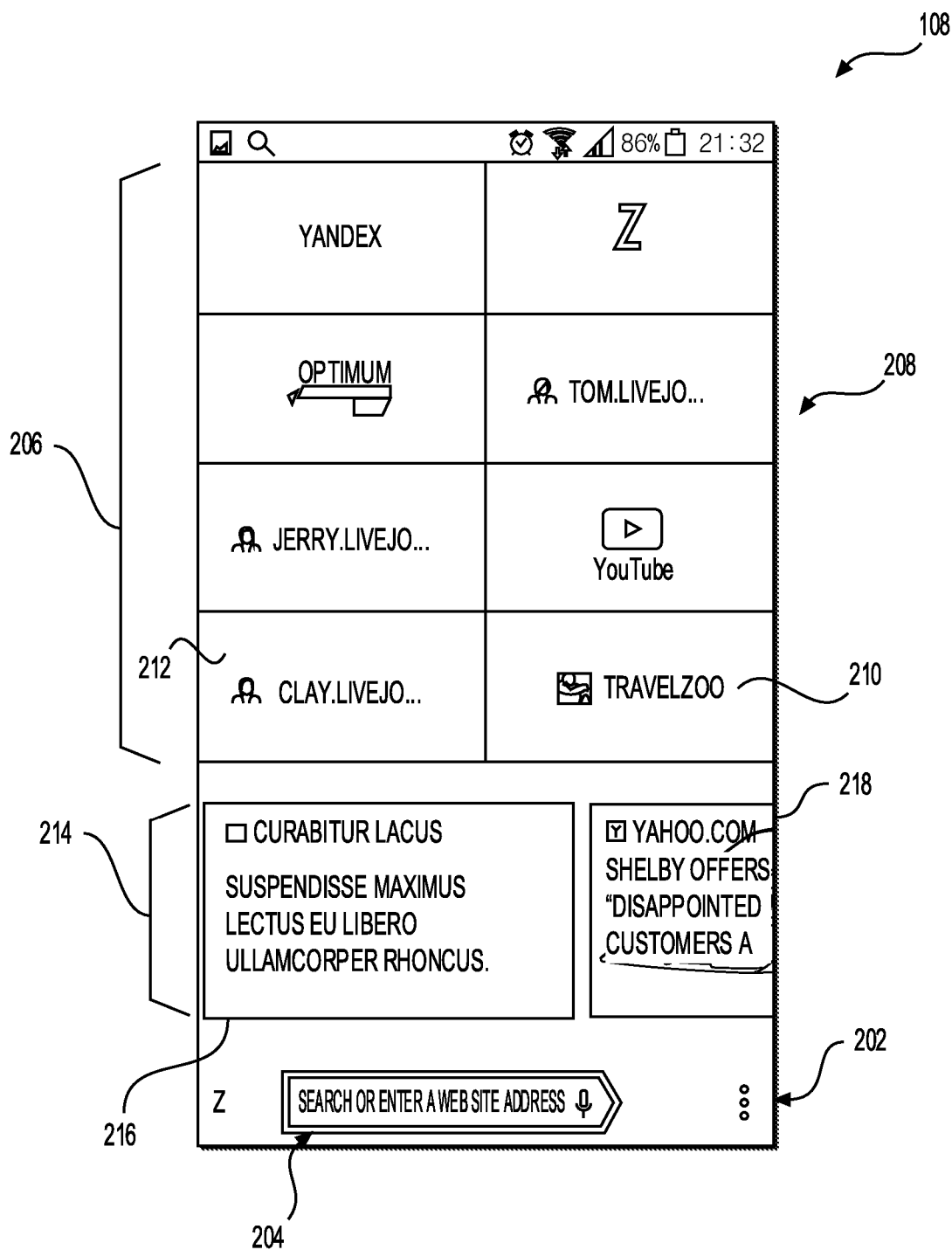
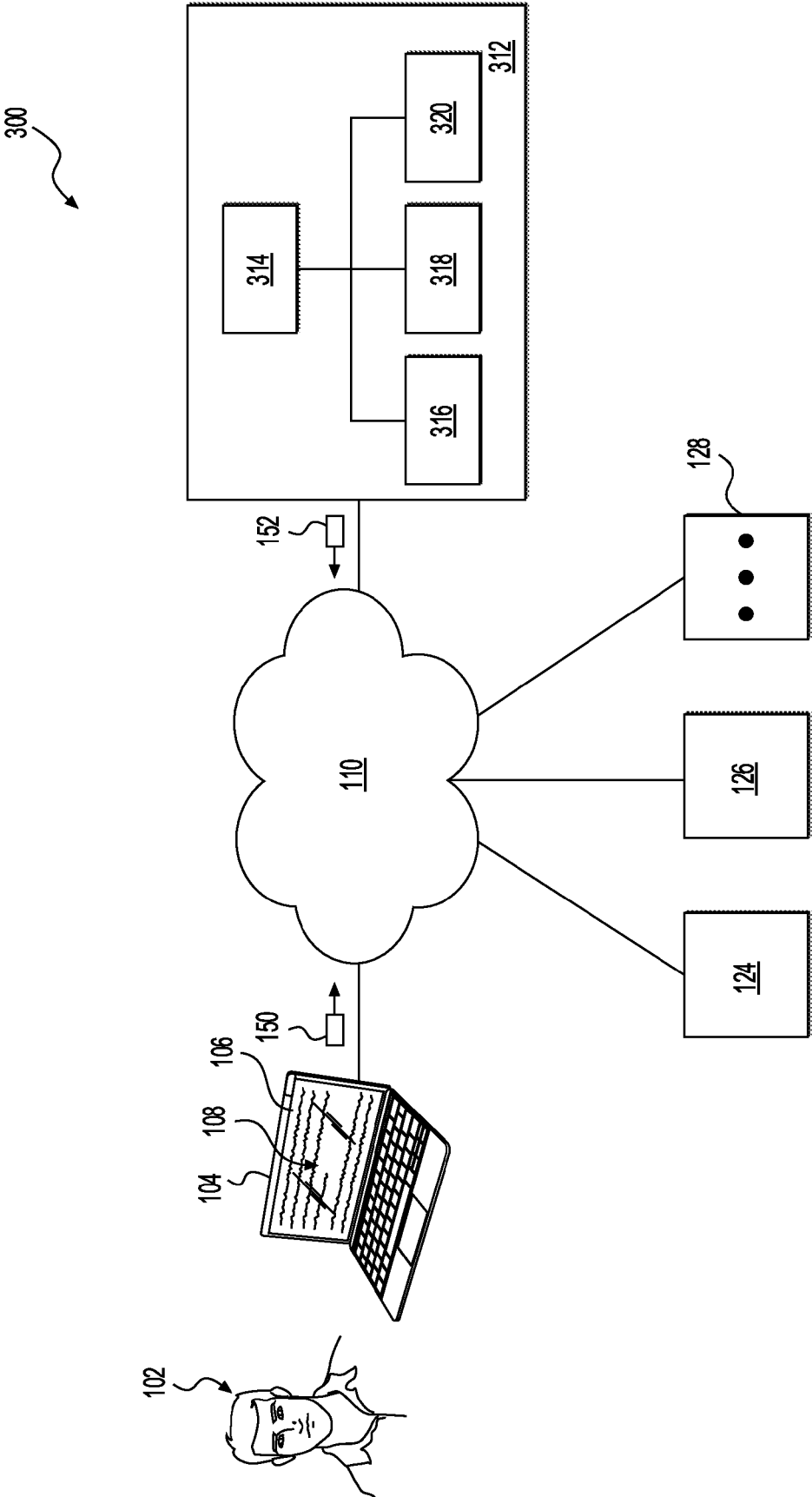
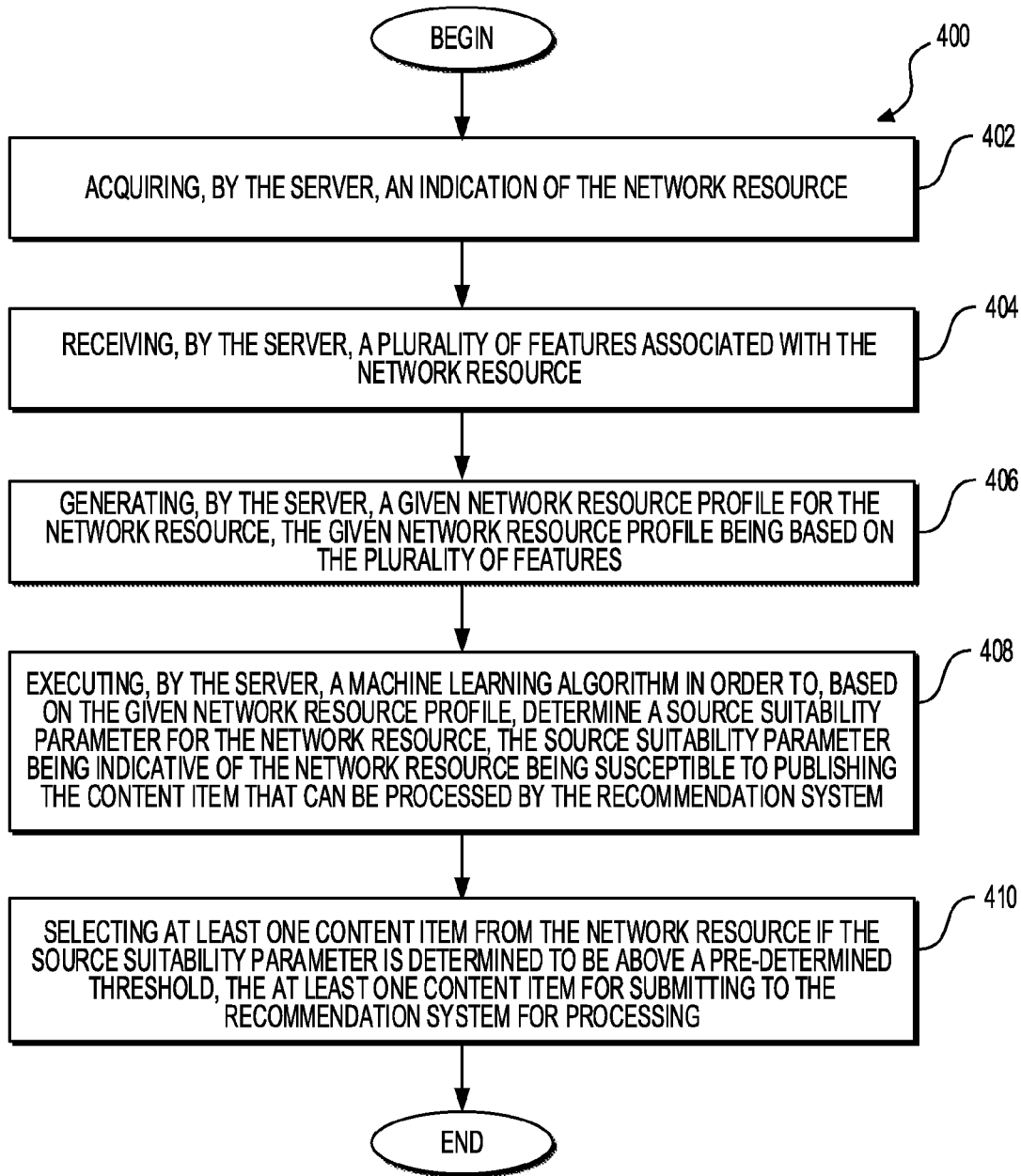


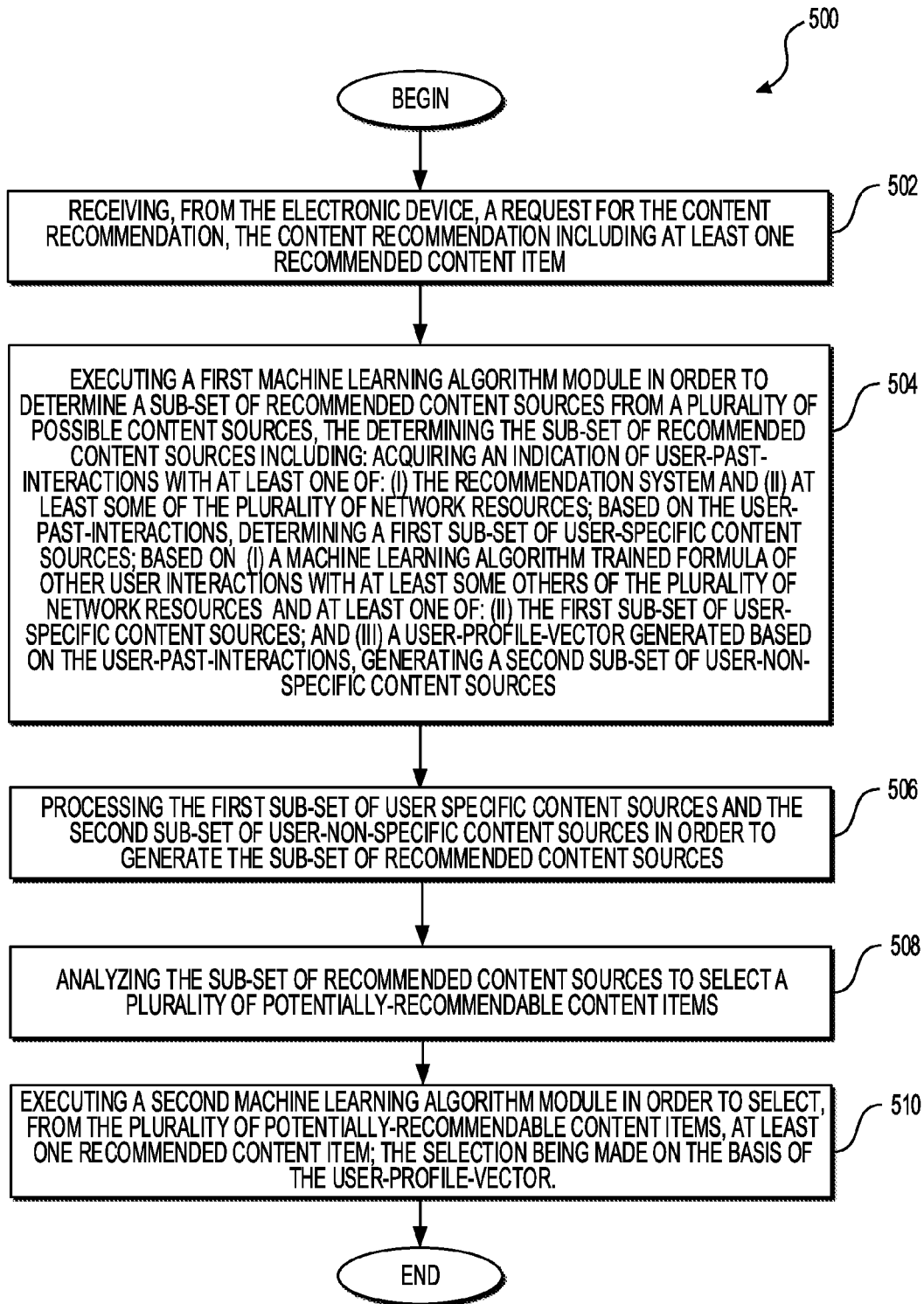
FIG. 2



**FIG. 3**



**FIG. 4**



**FIG. 5**

**METHOD AND APPARATUS FOR  
SELECTING A NETWORK RESOURCE AS A  
SOURCE OF CONTENT FOR A  
RECOMMENDATION SYSTEM**

**CROSS-REFERENCE**

**[0001]** The present application claims priority to Russian Patent Application No. 2016127446, filed Jul. 7, 2016, entitled “Method And Apparatus For Selecting A Network Resource As A Source Of Content For A Recommendation System” the entirety of which is incorporated herein.

**TECHNICAL FIELD**

**[0002]** The present technology relates to recommendation systems in general and specifically to a method and apparatus for selecting a network resource as a source of content for a recommendation system.

**BACKGROUND**

**[0003]** Various global or local communication networks (the Internet, the World Wide Web, local area networks and the like) offer a user a vast amount of information. The information includes a multitude of contextual topics, such as but not limited to, news and current affairs, maps, company information, financial information and resources, traffic information, games and entertainment related information. Users use a variety of client devices (desktop, laptop, notebook, smartphone, tablets and the like) to have access to rich content (like images, audio, video, animation, and other multimedia content from such networks).

**[0004]** The volume of available information through various Internet resources has grown exponentially in the past couple of years. Several solutions have been developed in order to allow a typical user to find the information that the user is looking for. One example of such a solution is a search engine. Examples of the search engines include GOOGLE™ search engine, YANDEX™ search engine, YAHOO!™ search engine and the like. The user can access the search engine interface and submit a search query associated with the information that the user is desirous of locating on the Internet. In response to the search query, the search engine provides a ranked list of search results. The ranked list of search results is generated based on various ranking algorithms employed by the particular search engine that is being used by the user performing the search. The overall goal of such ranking algorithms is to present the most relevant search results at the top of the ranked list, while less relevant search results would be positioned on less prominent positions of the ranked list of search results (with the least relevant search results being located towards the bottom of the ranked list of search results).

**[0005]** The search engines typically provide a good search tool for a search query that the user knows a priori that she/he wants to search. In other words, if the user is interested in obtaining information about the most popular destinations in Italy (i.e. a known search topic), the user could submit a search query: “The most popular destinations in Italy?” The search engine will then present a ranked list of Internet resources that are potentially relevant to the search query. The user can then browse the ranked list of search results in order to obtain information she/he is interested in as it related to places to visit in Italy. If the user, for whatever reason, is not satisfied with the uncovered

search results, the user can re-run the search, for example, with a more focused search query, such as “The most popular destinations in Italy in the summer?”, “The most popular destinations in the South of Italy?”, “The most popular destinations for a romantic getaway in Italy?”.

**[0006]** There is another approach that has been proposed for allowing the user to discover content and, more precisely, to allow for discovering and/or recommending content that the user may not be expressly interested in searching for. In a sense, such systems recommend content to the user without an express search request based on explicit or implicit interests of the user.

**[0007]** An example of such a system is a FLIPBOARD recommendation system, which system aggregates and recommends content from various social networks. The FLIPBOARD recommendation system presents the uncovered content in a “magazine style” format, where the user can “flip” through the pages with the recommended/aggregated content. The recommendation system collects content from social media and other websites, presents it in magazine format, and allows users to “flip” through their social-networking feeds and feeds from websites that have partnered with the company, effectively “recommending” content to the user even though the user may not have expressly expressed her/his desire in the particular content.

**[0008]** In order to generate the ranked search results in a search engine system or a list of recommended resources in a typical recommendation system, the respective system utilizes a machine learning algorithm to select search results and/or recommended content. There are several machine learning algorithms known in the art and utilized by search engines and/or recommendation systems. As is known, a typical machine learning algorithm is first “trained” using a training set (whether marked or unmarked) to generate a machine learning algorithm formula, which is then applied to determine at run time an output based on the system inputs.

**[0009]** In the typical recommendation system, the trained machine learning algorithm of the recommendation system selects a number of potential recommended items from a number of potential sources for the recommended items. Typically, the machine learning algorithm associated with such recommendation system, acquires (over time) knowledge about various potential sources of content items, past-user interactions with the various potential sources of content items (as well as generated recommendations) and uses this knowledge to generate a recommendation for a particular user of the recommendation system.

**SUMMARY**

**[0010]** It is an object of the present technology to ameliorate at least some of the inconveniences present in the prior art.

**[0011]** Embodiments of the present technology have been developed based on developers’ appreciation that the machine learning algorithms used for the selection of the recommended content are very computer processing resource consuming. However, the recommendation system needs to output content recommendations relatively fast, when a given user of the recommendation system requests the content recommendations. In other words, developers of the present technology have realized that it may be desirable to move certain tasks of the recommendation system from

run time to off line, i.e. to execute them before the given user of the recommendation system has requested the content recommendation.

**[0012]** Developers of the present technology have further appreciated that the volume of potential sources of content recommendations are large. Analyzing all potential sources of content recommendations can take a considerable time and require considerable computing resources. Thus, developers of the present technology have set out to develop a method and a system that pre-qualifies potential sources of content recommendations. The technical result of such pre-qualification is two fold. On the one hand, the pre-qualification of potential sources can be executed offline. And, secondly, content source profile data can be used when the recommendation system is used on a new territory, hence requiring pre-qualification of a large number of potential sources of content recommendations. Embodiments of the present technology are based on developers' premise that certain network resources are better suited to be sources of content recommendations. Without wishing to be bound to any specific theory, developers of the present technology have developed systems and methods that pre-qualify network resources as "potential content sources" (based on the network resource profile), the system utilizing a machine learning algorithm that places a larger weight on those network resources with a "fast paced" content, i.e. content that is frequently updated. These network resources can be (but are not so limited) news aggregators, media content aggregators and the like.

**[0013]** Other embodiments of the present technology aim at developing sources of content recommendations for the given user of the recommendation system. These embodiments analyze user-specific sources of content recommendations (those sources that the given user has interacted in the past) and enrich them with content sources that the given user may not have interacted with in the past. Thus, some embodiments of the present technology enable the recommendation system to mix content recommendation from known sources for the given user and unknown sources, thus enabling the recommendation system to "discover" new sources of recommended content for the given user.

**[0014]** According to a first broad aspect of the present technology, there is provided a method of selecting a network resource as a source of a content item, the content item to be analyzed by a recommendation system as part of a plurality of content items to generate a set of recommended content items as a recommendation for a given user of the recommendation system. The recommendation system is executed at a server coupled to a communication network. The method comprises: acquiring, by the server, an indication of the network resource; receiving, by the server, a plurality of features associated with the network resource; generating, by the server, a given network resource profile for the network resource, the given network resource profile being based on the plurality of features; executing, by the server, a machine learning algorithm in order to, based on the given network resource profile, determine a source suitability parameter for the network resource, the source suitability parameter being indicative of the network resource being susceptible to publishing the content item that can be processed by the recommendation system, selecting at least one content item from the network resource if the source suitability parameter is determined to be above a

pre-determined threshold, the at least one content item for submitting to the recommendation system for processing.

**[0015]** In some implementations of the method, the machine learning algorithm having been trained, prior to said acquiring, to determine the source suitability parameter based on training web site profiles generated with an associated plurality of training features at least some of which are of a same category as the plurality of features.

**[0016]** In some implementations of the method, plurality of features includes at least an absolute feature and a relative feature.

**[0017]** In some implementations of the method, the absolute feature and the relative feature are base features and wherein the method further comprises:

**[0018]** generating at least one derivative feature based on at least one of the absolute feature and the relative feature; and wherein

**[0019]** the given web site profile is further based on the at least one derivative feature.

**[0020]** In some implementations of the method, the plurality of features includes an indication of a proportion of visit sources to the network resource.

**[0021]** In some implementations of the method, the plurality of features includes at least one relative feature selected from:

**[0022]** hits per visit (HPV);

**[0023]** sessions per visit (SPV);

**[0024]** dated-pages per visit (DPV);

**[0025]** home page per visit (MPV);

**[0026]** time spent on the network resource (MTIME);

**[0027]** last domain per session (LPS);

**[0028]** first domain per session (FPS);

**[0029]** hits per session (HPS);

**[0030]** visits per user (VPU);

**[0031]** hits per user (HPU);

**[0032]** sessions per user (SPU);

**[0033]** dated-pages per hit (DPH);

**[0034]** home pages per hit (MPH); and

**[0035]** an indication of a visit source being one of:

**[0036]** direct access visit-source (TT\_GR);

**[0037]** search engine visit-source (TT\_SE);

**[0038]** social network visit-source (TT\_SN);

**[0039]** electronic message visit-source (TT\_IM);

**[0040]** URL-condenser visit-source (TT\_US); and

**[0041]** ad-platform visit-source (TT\_AD).

**[0042]** In some implementations of the method, the plurality of features includes at least one of: a number of users visiting the network resource per time interval, number of visits per time interval, number of transitions within various web pages within the network resource per time interval, and a number of unique sessions with the network resource per time interval.

**[0043]** In some implementations of the method, the plurality of features is a subset of all possible features, the method further comprising executing a feature selection to define the subset.

**[0044]** In some implementations of the method, the subset of all possible features consists of:

**[0045]** at least one of: a proportion of a landing page of the network resource visits within an initial transition to the network resource within a session and a proportion of the landing page visits during transitions within the network resource;



**[0046]** at least one of: a proportion of a web page associated with an URL containing a date being a first page visited within the initial transition to the network resource and a proportion of the web pages within the network resource visited during the session, the web pages being associated with an URL containing a date;

**[0047]** a proportionate profile of visit sources for the network resource, the visit sources being selected from: search engine/social networks/mail and messenger application/URL condensers/direct visits;

**[0048]** a proportion of session where the network resource was last in the session.

**[0049]** In some implementations of the method, the source suitability parameter is one of a suitable and non-suitable.

**[0050]** In some implementations of the method, the source suitability parameter is determined using at least one of:

**[0051]** a logistical regression analysis;

**[0052]** naive Bayes algorithm;

**[0053]** k-nearest neighbor algorithm;

**[0054]** random forest algorithm; and

**[0055]** hard voting algorithm.

**[0056]** In some implementations of the method, said selecting is executed in response to a request from the given user for a recommendation.

**[0057]** In accordance with another broad aspect of the present technology, there is provided a server. The server comprises:

**[0058]** a data storage medium;

**[0059]** a network interface configured for communication over a communication network;

**[0060]** a processor operationally coupled to the data storage medium and the network interface, the processor being configured to:

**[0061]** acquire an indication of the network resource, the network resource to be processed to determine its suitability as a source of a content item, the content item to be analyzed by a recommendation system as part of a plurality of content items to generate a set of recommended content items as a recommendation for a given user of the recommendation system;

**[0062]** receive a plurality of features associated with the network resource;

**[0063]** generate a given network resource profile for the network resource, the given network resource profile being based on the plurality of features;

**[0064]** execute a machine learning algorithm in order to, based on the given network resource profile, determine a source suitability parameter for the network resource, the source suitability parameter being indicative of the network resource being susceptible to publishing the content item that can be processed by the recommendation system,

**[0065]** select at least one content item from the network resource if the source suitability parameter is determined to be above a pre-determined threshold, the at least one content item for submitting to the recommendation system for processing.

**[0066]** In some implementations of the server, the machine learning algorithm has been trained, prior to said acquiring, to determine the source suitability parameter based on training web site profiles generated with an associated plurality of training features at least some of which are of a same category as the plurality of features.

**[0067]** In some implementations of the server, the plurality of features includes at least an absolute feature and a relative feature.

**[0068]** In some implementations of the server, the absolute feature and the relative feature are base features and wherein the processor is further configured to:

**[0069]** generate at least one derivative feature based on at least one of the absolute feature and the relative feature; and wherein

**[0070]** the given web site profile is further based on the at least one derivative feature.

**[0071]** In some implementations of the server, the plurality of features includes an indication of a proportion of visit sources to the network resource.

**[0072]** In some implementations of the server, the plurality of features includes at least one relative feature selected from:

**[0073]** hits per visit (HPV);

**[0074]** sessions per visit (SPV);

**[0075]** dated-pages per visit (DPV);

**[0076]** home page per visit (MPV);

**[0077]** time spent on the network resource (MTIME);

**[0078]** last domain per session (LPS);

**[0079]** first domain per session (FPS);

**[0080]** hits per session (HPS);

**[0081]** visits per user (VPU);

**[0082]** hits per user (HPU);

**[0083]** sessions per user (SPU);

**[0084]** dated-pages per hit (DPH);

**[0085]** home pages per hit (MPH); and

**[0086]** an indication of a visit source being one of:

**[0087]** direct access visit-source (TT\_GR);

**[0088]** search engine visit-source (TT\_SE);

**[0089]** social network visit-source (TT\_SN);

**[0090]** electronic message visit-source (TT\_IM);

**[0091]** URL-condenser visit-source (TT\_US); and

**[0092]** ad-platform visit-source (TT\_AD).

**[0093]** In some implementations of the server, the plurality of features includes at least one of: a number of users visiting the network resource per time interval, number of visits per time interval, number of transitions within various web pages within the network resource per time interval, and a number of unique sessions with the network resource per time interval.

**[0094]** In some implementations of the server, the plurality of features is a subset of all possible features, the method further comprising executing a feature selection to define the subset.

**[0095]** In some implementations of the server, the subset of all possible features consists of:

**[0096]** at least one of: a proportion of a landing page of the network resource visits within an initial transition to the network resource within a session and a proportion of the landing page visits during transitions within the network resource;

**[0097]** at least one of: a proportion of a web page associated with an URL containing a date being a first page visited within the initial transition to the network resource and a proportion of the web pages within the network resource visited during the session, the web pages being associated with an URL containing a date;

**[0098]** a proportionate profile of visit sources for the network resource, the visit sources being selected from:

search engine/social networks/mail and messenger application/URL condensers/direct visits;

**[0099]** a proportion of session where the network resource was last in the session.

**[0100]** In some implementations of the server, the source suitability parameter is one of a suitable and non-suitable.

**[0101]** In some implementations of the server, the source suitability parameter is determined using at least one of:

**[0102]** a logistical regression analysis;

**[0103]** naive Bayes algorithm;

**[0104]** k-nearest neighbor algorithm;

**[0105]** random forest algorithm; and

**[0106]** hard voting algorithm.

**[0107]** In some implementations of the server, the processor executes the selecting in response to a request from the given user for a recommendation.

**[0108]** In accordance with yet another broad aspect of the present technology, there is provided a computer-implemented method of generating a content recommendation for a user of an electronic device. The method is executable by a recommendation server accessible by the electronic device via a communication network, the content recommendation being associated with a content item available at one of a plurality of network resources accessible via the communication network. The method comprises: receiving, from the electronic device, a request for the content recommendation, the content recommendation including at least one recommended content item; executing a first machine learning algorithm module in order to determine a sub-set of recommended content sources from a plurality of possible content sources, the determining the sub-set of recommended content sources including: acquiring an indication of user-past-interactions with at least one of: (i) the recommendation system and (ii) at least some of the plurality of network resources; based on the user-past-interactions, determining a first sub-set of user-specific content sources; based on (i) a machine learning algorithm trained formula of other user interactions with at least some others of the plurality of network resources and at least one of: (ii) the first sub-set of user-specific content sources; and (iii) a user-profile-vector generated based on the user-past-interactions, generating a second sub-set of user-non-specific content sources; processing the first sub-set of user specific content sources and the second sub-set of user-non-specific content sources in order to generate the sub-set of recommended content sources; analyzing the sub-set of recommended content sources to select a plurality of potentially-recommendable content items; executing a second machine learning algorithm module in order to select, from the plurality of potentially-recommendable content items, at least one recommended content item; the selection being made on the basis of the user-profile-vector.

**[0109]** In some implementations of the method, the determining the first sub-set of user-specific content sources comprises applying an SVD algorithm.

**[0110]** In some implementations of the method, the determining the first sub-set of user-specific content sources is based on:

**[0111]** (i) the machine learning algorithm trained formula of other user interactions with at least some others of the plurality of network resources; and

**[0112]** (ii) the user-profile-vector generated based on the user-past-interactions.

**[0113]** In some implementations of the method, the method further comprises, prior to the receiving the request for the content recommendation, generating by the SVD algorithm a matrix of network resources events and users.

**[0114]** In some implementations of the method, the determining the first sub-set of user-specific content sources comprises applying a PMI algorithm.

**[0115]** In some implementations of the method, the determining the first sub-set of user-specific content sources is based on:

**[0116]** (i) the machine learning algorithm trained formula of other user interactions with at least some others of the plurality of network resources; and

**[0117]** (ii) the first sub-set of user-specific content sources.

**[0118]** In some implementations of the method, the selection from the sub-set of recommended content sources the plurality of potentially-recommendable content items comprises determining a pre-defined number of most recent content items.

**[0119]** In some implementations of the method, the selection from the sub-set of recommended content sources the plurality of potentially-recommendable content items comprises determining a pre-defined number of most popular content items.

**[0120]** In some implementations of the method, the selection from the sub-set of recommended content sources is specific to a geographic region where the request for content recommendation was executed from.

**[0121]** In some implementations of the method, the method further comprises filtering the selected potentially-recommendable content items.

**[0122]** In some implementations of the method, the filtering is based on a language setting of the electronic device and a language setting of the content item.

**[0123]** In accordance with another broad aspect of the present technology, there is provided a server. The server comprises: a data storage medium; a network interface configured for communication over a communication network; a processor operationally coupled to the data storage medium and the network interface, the processor configured to: receive, from an electronic device, a request for the content recommendation, the content recommendation including at least one recommended content item; the content recommendation being associated with a content item available at one of a plurality of network resources accessible via the communication network; execute a first machine learning algorithm module in order to determine a sub-set of recommended content sources from a plurality of possible content sources, the determining the sub-set of recommended content sources including: acquiring an indication of user-past-interactions with at least one of: (i) the recommendation system and (ii) at least some of the plurality of network resources; based on the user-past-interactions, determining a first sub-set of user-specific content sources; based on (i) a machine learning algorithm trained formula of other user interactions with at least some others of the plurality of network resources and at least one of: (ii) the first sub-set of user-specific content sources; and (iii) a user-profile-vector generated based on the user-past-interactions, generating a second sub-set of user-non-specific content sources; processing the first sub-set of user specific content sources and the second sub-set of user-non-specific content sources in order to generate the sub-set of recommended content sources; analyze the sub-set of recom-

mended content sources to select a plurality of potentially-recommendable content items; execute a second machine learning algorithm module in order to select, from the plurality of potentially-recommendable content items, at least one recommended content item; the selection being made on the basis of the user-profile-vector.

**[0124]** In some embodiments of the server, to determine the first sub-set of user-specific content sources, the processor is configured to apply an SVD algorithm.

**[0125]** In some embodiments of the server, the determining the first sub-set of user-specific content sources is based on:

**[0126]** (i) the machine learning algorithm trained formula of other user interactions with at least some others of the plurality of network resources; and

**[0127]** (ii) the user-profile-vector generated based on the user-past-interactions.

**[0128]** In some embodiments of the server, the server is further configured, prior to the receiving the request for the content recommendation, to generate, using the SVD algorithm, a matrix of network resources events and users.

**[0129]** In some embodiments of the server, to determine the first sub-set of user-specific content sources, the processor is configured to apply a PMI algorithm.

**[0130]** In some embodiments of the server, the determining the first sub-set of user-specific content sources is based on:

**[0131]** (i) the machine learning algorithm trained formula of other user interactions with at least some others of the plurality of network resources; and

**[0132]** (ii) the first sub-set of user-specific content sources.

**[0133]** In some embodiments of the server, the selection from the sub-set of recommended content sources the plurality of potentially-recommendable content items comprises determining a pre-defined number of most recent content items.

**[0134]** In some embodiments of the server, the selection from the sub-set of recommended content sources the plurality of potentially-recommendable content items comprises determining a pre-defined number of most popular content items.

**[0135]** In some embodiments of the server, the selection from the sub-set of recommended content sources is specific to a geographic region where the request for content recommendation was executed from.

**[0136]** In some embodiments of the server, the processor is further configured to filter the selected potentially-recommendable content items.

**[0137]** In some embodiments of the server, the filtering is based on a language setting of the electronic device and a language setting of the content item.

**[0138]** In the context of the present specification, a “server” is a computer program that is running on appropriate hardware and is capable of receiving requests (e.g., from client devices) over a network, and carrying out those requests, or causing those requests to be carried out. The hardware may be one physical computer or one physical computer system, but neither is required to be the case with respect to the present technology. In the present context, the use of the expression a “server” is not intended to mean that every task (e.g., received instructions or requests) or any particular task will have been received, carried out, or caused to be carried out, by the same server (i.e., the same software and/or hardware); it is intended to mean that any

number of software elements or hardware devices may be involved in receiving/sending, carrying out or causing to be carried out any task or request, or the consequences of any task or request; and all of this software and hardware may be one server or multiple servers, both of which are included within the expression “at least one server”.

**[0139]** In the context of the present specification, “client device” is any computer hardware that is capable of running software appropriate to the relevant task at hand. Thus, some (non-limiting) examples of client devices include personal computers (desktops, laptops, netbooks, etc.), smartphones, and tablets, as well as network equipment such as routers, switches, and gateways. It should be noted that a device acting as a client device in the present context is not precluded from acting as a server to other client devices. The use of the expression “a client device” does not preclude multiple client devices being used in receiving/sending, carrying out or causing to be carried out any task or request, or the consequences of any task or request, or steps of any method described herein.

**[0140]** In the context of the present specification, a “database” is any structured collection of data, irrespective of its particular structure, the database management software, or the computer hardware on which the data is stored, implemented or otherwise rendered available for use. A database may reside on the same hardware as the process that stores or makes use of the information stored in the database or it may reside on separate hardware, such as a dedicated server or plurality of servers.

**[0141]** In the context of the present specification, the expression “information” includes information of any nature or kind whatsoever capable of being stored in a database. Thus information includes, but is not limited to audiovisual works (images, movies, sound records, presentations etc.), data (location data, numerical data, etc.), text (opinions, comments, questions, messages, etc.), documents, spreadsheets, lists of words, etc.

**[0142]** In the context of the present specification, the expression “component” is meant to include software (appropriate to a particular hardware context) that is both necessary and sufficient to achieve the specific function(s) being referenced.

**[0143]** In the context of the present specification, the expression “computer usable information storage medium” is intended to include media of any nature and kind whatsoever, including RAM, ROM, disks (CD-ROMs, DVDs, floppy disks, hard drives, etc.), USB keys, solid state drives, tape drives, etc.

**[0144]** In the context of the present specification, the words “first”, “second”, “third”, etc. have been used as adjectives only for the purpose of allowing for distinction between the nouns that they modify from one another, and not for the purpose of describing any particular relationship between those nouns. Thus, for example, it should be understood that, the use of the terms “first server” and “third server” is not intended to imply any particular order, type, chronology, hierarchy or ranking (for example) of/between the server, nor is their use (by itself) intended imply that any “second server” must necessarily exist in any given situation. Further, as is discussed herein in other contexts, reference to a “first” element and a “second” element does not preclude the two elements from being the same actual real-world element. Thus, for example, in some instances, a “first” server and a “second” server may be the same

software and/or hardware, in other cases they may be different software and/or hardware.

**[0145]** Implementations of the present technology each have at least one of the above-mentioned object and/or aspects, but do not necessarily have all of them. It should be understood that some aspects of the present technology that have resulted from attempting to attain the above-mentioned object may not satisfy this object and/or may satisfy other objects not specifically recited herein.

**[0146]** Additional and/or alternative features, aspects and advantages of implementations of the present technology will become apparent from the following description, the accompanying drawings and the appended claims.

#### BRIEF DESCRIPTION OF THE DRAWINGS

**[0147]** For a better understanding of the present technology, as well as other aspects and further features thereof, reference is made to the following description which is to be used in conjunction with the accompanying drawings, where:

**[0148]** FIG. 1 depicts a system suitable for implementing non-limiting embodiments of the present technology.

**[0149]** FIG. 2 depicts a screen shot of a recommendation interface implemented in accordance with a non-limiting embodiment of the present technology, the recommendation interface being depicted as displayed on the screen of an electronic device of the system of FIG. 1, the electronic device being implemented as a smart phone.

**[0150]** FIG. 3 depicts another embodiment of a system, the system being implemented in accordance with another non-limiting embodiment of the present technology.

**[0151]** FIG. 4 depicts a block diagram of a method, the method executable in accordance with non-limiting embodiments of the present technology, the method executable within the system of FIG. 1.

**[0152]** FIG. 5 depicts a block diagram of a method, the method executable in accordance with another non-limiting embodiment of the present technology, the method executable within the system of FIG. 3.

#### DETAILED DESCRIPTION

**[0153]** Referring to FIG. 1, there is shown a schematic diagram of a system 100, the system 100 being suitable for implementing non-limiting embodiments of the present technology. It is to be expressly understood that the system 100 as depicted is merely an illustrative implementation of the present technology. Thus, the description thereof that follows is intended to be only a description of illustrative examples of the present technology. This description is not intended to define the scope or set forth the bounds of the present technology. In some cases, what are believed to be helpful examples of modifications to the system 100 may also be set forth below. This is done merely as an aid to understanding, and, again, not to define the scope or set forth the bounds of the present technology. These modifications are not an exhaustive list, and, as a person skilled in the art would understand, other modifications are likely possible. Further, where this has not been done (i.e., where no examples of modifications have been set forth), it should not be interpreted that no modifications are possible and/or that what is described is the sole manner of implementing that element of the present technology. As a person skilled in the art would understand, this is likely not the case. In addition

it is to be understood that the system 100 may provide in certain instances simple implementations of the present technology, and that where such is the case they have been presented in this manner as an aid to understanding. As persons skilled in the art would understand, various implementations of the present technology may be of a greater complexity.

**[0154]** Generally speaking, the system 100 is configured to provide content recommendations to a user 102 of the system 100. The user 102 may be a subscriber to a recommendation service provided by the system 100. However, the subscription does not need to be explicit or paid for. For example, the user 102 can become a subscriber by virtue of downloading a recommendation application from the system 100, by registering and provisioning a log-in/password combination, by registering and provisioning user preferences and the like. As such, any system variation configured to generate content recommendations for the given user can be adapted to execute embodiments of the present technology, once teachings presented herein are appreciated. Furthermore, the system 100 will be described using an example of the system 100 being a recommendation system (therefore, the system 100 can be referred to herein below as a “recommendation system 100” or a “prediction system 100” or a “training system 100”). However, embodiments of the present technology can be equally applied to other types of the systems 100, as will be described in greater detail herein below.

**[0155]** The system 100 comprises an electronic device 104, the electronic device 104 being associated with the user 102. As such, the electronic device 104 can sometimes be referred to as a “client device”, “end user device” or “client electronic device”. It should be noted that the fact that the electronic device 104 is associated with the user 102 does not need to suggest or imply any mode of operation—such as a need to log in, a need to be registered, or the like.

**[0156]** The implementation of the electronic device 104 is not particularly limited, but as an example, the electronic device 104 may be implemented as a personal computer (desktops, laptops, netbooks, etc.), a wireless communication device (such as a smartphone, a cell phone, a tablet and the like), as well as network equipment (such as routers, switches, and gateways). The electronic device 104 comprises hardware and/or software and/or firmware (or a combination thereof), as is known in the art, to execute a recommendation application 106. Generally speaking, the purpose of the recommendation application 106 is to enable the user to receive (or otherwise access) content recommendations provided by the system 100, as will be described in greater detail herein below.

**[0157]** How the recommendation application 106 is implemented is not particularly limited. One example of the recommendation application 106 may include a user accessing a web site associated with a recommendation service to access the recommendation application 106. For example, the recommendation application 106 can be accessed by typing in (or otherwise copy-pasting or selecting a link) an URL associated with the recommendation service. Alternatively, the recommendation application 106 can be an app downloaded from a so-called app store, such as APPSTORE™ or GOOGLEPLAY™ and installed/executed on the electronic device 104. It should be expressly understood that the recommendation application 106 can be accessed using any other suitable means.

[0158] Generally speaking, the recommendation application 106 comprises a recommendation interface 108, the recommendation interface 108 being displayed on a screen (not separately numbered) of the electronic device 104. With reference to FIG. 2, there is depicted a screen shot of the recommendation interface 108 implemented in accordance with a non-limiting embodiment of the present technology (the example of the recommendation interface 108 being depicted as displayed on the screen of the electronic device 104 being implemented as a smart phone).

[0159] In some embodiments of the present technology the recommendation interface 108 is presented when the user 102 of the electronic device 104 actuates (i.e. executes, run, background-run or the like) the recommendation application 106. Alternatively, the recommendation interface 108 can be presented when the user 102 opens a new browser window and/or activates a new tab in the browser application. For example, in some embodiments of the present technology, the recommendation interface 108 can act as a “home screen” in the browser application.

[0160] The recommendation interface 108 includes a search interface 202. The search interface 202 includes a search query interface 204. The search query interface 204 can be implemented as an “omnibox” which allows entry of a search query for executing a search or a network address (such as a Universal Remote Locator) for identifying a network resource (such as a web site) to be accessed. However, the search query interface 204 can be implemented as configured to receive one or both of: entry of the search query for executing the search or the network address (such as a Universal Remote Locator) for identifying the network resource (such as a web site) to be accessed.

[0161] The recommendation interface 108 further includes a links interface 206. The links interface 206 includes a plurality of tiles 208—of which eight are depicted in FIG. 2—only two of which are numbered in FIG. 2—a first tile 210 and a second tile 212.

[0162] Using the example of the first tile 210 and the second tile 212—each of the plurality of tiles 208 includes (or acts as) a link to either (i) a web site marked as “favorite” or otherwise marked by the user 102, (ii) a previously visited web site or (iii) the like. The plurality of tiles 208, in the depicted embodiment, is visually presented to the user 102 as square buttons with a logo and/or a name of the resource depicted therein, the logo and the name for enabling the user 102 to identify which resource the particular one of the plurality of tiles (not separately numbered) is linked to. However, it should be expressly understood that the visual representation of some or all of the plurality of tiles 208 can be different. As such, some or all of the plurality of tiles 208 can be implemented as differently shaped buttons, as hyperlinks presented in a list or the like.

[0163] As an example, the first tile 210 contains a link to a TRAVELZOO™ web site and the second tile 212 contains a link to a personal live journal web site. Needless to say, the number and content of the individual ones of the plurality of tiles 208 is not particularly limited.

[0164] For example, the number of the tiles within the plurality of tiles 208 can be pre-selected by the provider of the recommendation application 106. In some embodiments of the present technology, the number of tiles within the plurality of tiles 208 is pre-selected based on the size and/or resolution of the screen of the electronic device 104 executing the recommendation application 106. For example, a

first number of tiles can be pre-selected for the electronic device 104 executed as a smartphone, a second number of tiles can be pre-selected for the electronic device 104 executed as a tablet, and a third number of tiles can be pre-selected for the electronic device 104 executed as a laptop or desktop computer.

[0165] The recommendation interface 108 further includes a recommended content set 214. The recommended content set 214 includes one or more recommended content items, such as a first recommended content item 216 and a second recommended content item 218 (the second recommended content item 218 only partially visible in FIG. 2). Naturally, the recommended content set 214 can have more recommended content items. In the embodiment depicted in FIG. 2 and in those embodiments where more than one recommended content item are present, the user 102 can scroll through the recommended content set 214. The scrolling can be achieved by any suitable means. For example, the user 102 can scroll the content of the recommended content set 214 by means of actuating a mouse device (not depicted), a key board key (not depicted) or interacting with a touch sensitive screen (not depicted) of or associated with the electronic device 104.

[0166] Example provided in FIG. 2 is just one possible implementation of the recommendation interface 108. Another example of the implementation of the recommendation interface 108, as well as an explanation of how the user 102 can interact with the recommendation interface 108 is disclosed in a co-owned Russian Patent Application entitled A COMPUTER-IMPLEMENTED METHOD OF GENERATING A CONTENT RECOMMENDATION INTERFACE, filed on May 12, 2016 and bearing an application number 2016118519; content of which is incorporated by reference herein in its entirety.

[0167] How the content for the recommended content set 214 is generated will be described in greater detail herein below.

[0168] Returning to the description of FIG. 1, the electronic device 104 is communicatively coupled to a communication network 110 for accessing a recommendation server 112. In some non-limiting embodiments of the present technology, the communication network 110 can be implemented as the Internet. In other embodiments of the present technology, the communication network 110 can be implemented differently, such as any wide-area communication network, local-area communication network, a private communication network and the like. A communication link (not separately numbered) between the electronic device 104 and the communication network 110 is implemented will depend inter alia on how the electronic device 104 is implemented. Merely as an example and not as a limitation, in those embodiments of the present technology where the electronic device 104 is implemented as a wireless communication device (such as a smartphone), the communication link can be implemented as a wireless communication link (such as but not limited to, a 3G communication network link, a 4G communication network link, Wireless Fidelity, or WiFi® for short, Bluetooth® and the like). In those examples where the electronic device 104 is implemented as a notebook computer, the communication link can be either wireless (such as Wireless Fidelity, or WiFi® for short, Bluetooth® or the like) or wired (such as an Ethernet based connection).

[0169] The recommendation server 112 can be implemented as a conventional computer server. In an example of

an embodiment of the present technology, the recommendation server 112 can be implemented as a Dell™ PowerEdge™ Server running the Microsoft™ Windows Server™ operating system. Needless to say, the recommendation server 112 can be implemented in any other suitable hardware, software, and/or firmware, or a combination thereof. In the depicted non-limiting embodiments of the present technology, the recommendation server 112 is a single server. In alternative non-limiting embodiments of the present technology, the functionality of the recommendation server 112 may be distributed and may be implemented via multiple servers.

[0170] The recommendation server 112 comprises a processing module 114. The processing module 114 is coupled to or otherwise has access to a content source selection module 116. The content source selection module 116 has access to a data storage device 118. The processing module 114 is further coupled to a content recommendation selection module 115. Operation of the recommendation server 112 and its components will be described herein below in greater detail.

[0171] Also coupled to the communication network 110 are multiple network resources, including a first network resource 124, a second network resource 126 and a plurality of additional network resources 128. The first network resource 124, the second network resource 126 and the plurality of additional network resources 128 are all network resources accessible by the electronic device 104 (as well as other electronic devices potentially present in the system 100) via the communication network 110. Respective content of first network resource 124, the second network resource 126 and the plurality of additional network resources 128 is not particularly limited. A given one of the first network resource 124, the second network resource 126 and the plurality of additional network resources 128 can contain (or in other words, host) digital content such as: audio content for streaming, video content for streaming, news, blogs, information about various government institutions, information about points of interest, thematically clustered content (such as content relevant to those interested in kick-boxing), and the like.

[0172] The content is potentially “discoverable” to the electronic device 104 by various means. For example, the user 102 of the electronic device 104 can use a browser application (not depicted) and enter a Universal Resource Locator (URL) associated with the given one of the first network resource 124, the second network resource 126 and the plurality of additional network resources 128. Alternatively, the user 102 of the electronic device 104 can execute a search using a search engine (not depicted) to discover the content of one or more of the first network resource 124, the second network resource 126 and the plurality of additional network resources 128. As has been mentioned above, these are useful when the user 102 knows a priori which content the user 102 is interested in.

[0173] In accordance with embodiments of the present technology, the recommendation application 106 can recommend content items available from the given one of the first network resource 124, the second network resource 126 and the plurality of additional network resources 128 to the user 102, the content items that the user 102 may not a priori know about. The recommendation server 112 is configured to select content for the one or more recommendation items to be presented to the user 102 via the recommendation

application 106. More specifically, the processing module 114 is configured to (i) receive from the electronic device 104 a request for the content recommendation 150 and (ii) responsive to the request, generate a recommended content message 152 specifically customized for the user 102 associated with the electronic device 104. The processing module 114 can further coordinate execution of various routines described herein as performed by the given one of the first network resource 124, the second network resource 126 and the plurality of additional network resources 128 and the content source selection module 116, for example.

[0174] In some embodiments of the present technology, the request for the content recommendation 150 can be generated in response to the user 102 providing an explicit indication of the user desire to receive the content recommendation. For example, the recommendation interface 108 can provide a button (or another actuatable element) to enable the user 102 to indicate her/his desire to receive a new or an updated content recommendation. As a non-limiting example, the recommendation interface 108 can provide an actuatable button that reads “Request a content recommendation”. Within these embodiments, the request for the content recommendation 150 can be thought of as “an explicit request” in a sense of the user 102 expressly providing a request for the recommended content.

[0175] In other embodiments, the request for the content recommendation 150 can be generated in response to the user 102 providing an implicit indication of the user desire to receive the content recommendation. In some embodiments of the present technology, the request for the content recommendation 150 can be generated in response to the user 102 starting the recommendation application 106.

[0176] Alternatively, in those embodiments of the present technology where the recommendation application 106 is implemented as a browser (for example, a GOOGLE™ browser, a YANDEX™ browser, a YAHOO!™ browser or any other proprietary or commercially available browser application), the request for content recommendation 150 can be generated in response to the user 102 opening the browser application and can be generated, for example, without the user 102 executing any additional actions other than activating the browser application. As another example, the request for content recommendation 150 can be generated in response to the user 102 opening a new tab of the already-opened browser application and can be generated, for example, without the user 102 executing any additional actions other than activating the new browser tab. In other words, the request for the content recommendation 150 can be generated even without the user 102 knowing that the user 102 may be interested in obtaining a content recommendation.

[0177] As another example, the request for content recommendation 150 can be generated in response to the user 102 selecting a particular element of the browser application and can be generated, for example, without the user 102 executing any additional actions other than selecting/activating the particular element of the browser application.

[0178] Examples of the particular element of the browser application include but are not limited to:

[0179] An address line of the browser application bar

[0180] A search bar of the browser application and/or a search bar of a search engine web site accessed in the browser application

[0181] An omnibox (combined address and search bar of the browser application)

[0182] A favorites or recently visited network resources pane

[0183] Any other pre-determined area of the browser application interface or a network resource displayed in the browser application

[0184] In some embodiments of the present technology, each of the content recommendation selection module 115 and the content source selection module 116 can be configured to execute their respective one or more machine learning algorithms. In some embodiments of the present technology, one or more machine learning algorithms can be any suitable or semi-supervised supervised machine learning algorithm, such as but not limited to:

[0185] Artificial neural network

[0186] Bayesian statistics

[0187] Gaussian process regression

[0188] Decision trees

[0189] And the like

[0190] It is noted that the respective one or more machine algorithms executed by the content recommendation selection module 115 and the content source selection module 116 do not need to be the same relative to each other.

[0191] Even though the content recommendation selection module 115 and the content source selection module 116 have been described as separate entities each executing its respective one or more machine learning algorithms, in alternative embodiments of the present technology, the one or more machine learning algorithms respectively executed by the content recommendation selection module 115 and the content source selection module 116 can be executed by a single entity (such as the processing module 114, for example). Alternatively, the one or more machine learning algorithms respectively executed by the content recommendation selection module 115 and the content source selection module 116 can be distributed amongst more modules than the ones depicted in FIG. 1 and can be executed as part of multiple instances of the recommendation server 112.

[0192] Furthermore, each one of the content recommendation selection module 115 and the content source selection module 116 can execute additional functions (i.e. others than the respective one or more machine learning algorithms).

[0193] For example, the content source selection module 116 can execute a “crawling function”, which can be broadly described as a function for discovering new network resources and/or new content published by the first network resource 124, the second network resource and a number of additional network resources 128. To that end, the content source selection module 116 is configured to execute what is known a “crawler robot” that “visit” various network resources available/accessible via the communication network 110 (including the first network resource 124, the second network resource and a number of additional network resources 128) to discover and catalogue newly available resources (such as one of the first network resource 124, the second network resource and a number of additional network resources 128) or newly published content available from the first network resource 124, the second network resource and a number of additional network resources 128.

[0194] As part of the crawling function, the content source selection module 116 is configured to maintain (in the data storage device 118) information representative of the newly discovered network resources and/or newly discovered con-

tent available therefrom. In some embodiments, the content source selection module 116 can be configured to maintain an inverted index within the data storage device 118, but the content source selection module 116 can arrange the information representative of the newly discovered network resources and/or newly discovered content available therefrom in a data structure other than an inverted index.

[0195] The content source selection module 116 can further collect and maintain additional information about the available network resources and/or content available from those network resources such as one of the first network resource 124, the second network resource and a number of additional network resources 128). Additional information can be generally associated with a visit (i.e. a given user browsing activity of a particular network resource), a hit (i.e. the given user transition into the particular network resource, away from the particular network resource and between various web pages of the particular network resource) or a session (i.e. the particular user browsing session that can include visits to one or more network resources).

[0196] Some examples of the additional information collected about the available resources and/or content available from those resources include (but is not limited to):

[0197] hits per visit (HPV)—how many pages, within a given network resource, a given user has visited during browsing the given network resource

[0198] sessions per visit (SPV)—how many sessions contained visits to the particular network resource

[0199] dated-pages per visit (DPV)—how many pages of a given network resource with a date in the page URL, a given user has visited during a given visit to the given network resource

[0200] home page per visit (MPV)—how many times during a given visit a given user visited a home page of a given network resource

[0201] time spent on the network resource (MTIME)—how much time a given user has spent browsing a given network resource

[0202] last domain per session (LPS)—whether the given network resource was a last network resource in a given web browsing session of a given user

[0203] first domain per session (FPS)—whether the given network resource was a first network resource in a given web browsing session of a given user

[0204] hits per session (HPS)—how many pages of any network resources, a given user has visited during a given browsing sessions (i.e. how many transitions were made during the given session)

[0205] visits per user (VPU)—an average number of visits to the given network resource

[0206] hits per user (HPU)—an average number of transitions between various pages of the given network resource (or, alternatively, all network resources)

[0207] sessions per user (SPU)—an average number of sessions that included visits the given network resource

[0208] dated-pages per hit (DPH)—how many pages of a given network resource with a date in the page URL, a given user has visited during a given visit to the given network resource

[0209] home pages per hit (MPH)—how many times a given user has navigated to the home page of a given network resource during the visit to the given network resource

[0210] an indication of a source of a transition to the given network resource being one of:

- [0211] a. direct access visit-source (TT\_GR);
- [0212] b. search engine visit-source (TT\_SE);
- [0213] c. social network visit-source (TT\_SN);
- [0214] d. electronic message visit-source (TT\_IM);
- [0215] e. URL-condenser visit-source (TT\_US); and
- [0216] f. ad-platform visit-source (TT\_AD).

[0217] How the content source selection module 116 collects the additional information is not particularly limited. For example, some sources of the additional information include (but are not limited to): search engine logs, browser application logs, network resource tracking widgets, and the like.

[0218] For example, the content source selection module 116 can have access to a search engine log and obtain information of whether a given user has accessed a given network resource from a search engine. As another example, the content source selection module 116 may have access to the logs of the browser application executed on the electronic device 104 and receive an indication (which can be anonymized to protect a given user privacy) in regard to the given user 102: transition to the given network resource, transitions of the given user between various pages of the given network resource, information about the web pages of the given network resource visited (such as URLs, dates in the URLs and the like), time spent on the various web pages of the given network resource, user browsing sessions and the various network resources visited during the browsing session, and the like.

[0219] As another example, many of the available network resources install resource tracking widgets (such as, for example, Google™ Analytics, Yandex™ Metrika or the like). Naturally, the content source selection module 116 can obtain the additional information in a number of additional ways from a number of additional sources.

[0220] As such, in accordance with the non-limiting embodiments of the present technology, the recommendation server 112 is configured to execute a two-stepped approach to selection recommendations (i.e. recommended items) for inclusion into the recommended content message 152 in response to the request for the content recommendation 150.

[0221] The two-stepped approach is particularly useful (but is not limited) for those circumstances when the content recommendation service provided by the recommendation server 112 is offered in a new territory—such as a country, a region of a country, a new city or the like. A particular technical effect arising from the use of embodiments of the present technology, is the ability to “pre qualify” potential sources of content recommendations in an “offline” mode (“offline” referring to processing prior to receiving the request for the content recommendation 150).

[0222] Embodiments of the present technology are based on the premise that not all of the network resources (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128) contain potential content item suitable for processing and recommendation by the recommendation server 112. For example, some of the network resources (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128) may contain

content item that would not be of interest to the subscribers to the recommendation service provided by the recommendation server 112.

[0223] Embodiments of the present technology are further based on the premise that content of certain network resources (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128) is more suitable for processing and recommendation by the recommendation server 112. For example, if a given one of the network resources (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128) is an aggregator of news, it is more likely to contain content items suitable for processing and recommendation by the recommendation server 112. Such the network resource, is likely to have frequently updated content items (i.e. newly published news, articles, etc), as well as host content items that are of potential interest to a wide audience.

[0224] On the other hand, some other network resources (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128) contain content items that are less susceptible for processing and recommendation by the recommendation server 112. For example, a personal web site of a cat lover from Winnipeg, posting pictures of her three cats, is unlikely to provide potential content items for processing and recommendation by the recommendation server 112. This is due to the fact that the content of such the network resource is rarely updated and is associated with a relatively niche interest.

[0225] Generally speaking, the two-stepped approach is aimed at:

[0226] At a First Stage:

[0227] Select those ones of the plurality of network resources (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128) that contain potential content items for recommendation—more specifically, the recommendation server 112 determines a source suitability parameter for the network resource, the source suitability parameter being indicative of the network resource being susceptible to publishing the content item that can be processed by the recommendation server 112;

[0228] At a Second Stage:

[0229] From the so-selected network resources, the recommendation server 112 selects at least one content item, the at least one content item for submitting to the content recommendation selection module 115 for processing and selecting personalized content item recommendations for the user 102, in response to receiving the request for content recommendation 150.

[0230] Broadly speaking, in some embodiments of the present technology, the two-stepped approach can be implemented as follows:

[0231] Training of the Machine Learning Algorithm

[0232] Prior to execution of the two-stepped selection process, the recommendation server 112 (or, more specifically, the processing module 114) executes training of the machine learning algorithm that the content source selection module 116 uses to select content sources. Generally speaking, the machine learning algorithm is trained to determine the source suitability parameter of a given network resources (i.e. the first network resource 124, the second network



resource 126 and the plurality of additional network resources 128) based on a training set of network resources and associated labels assigned by assessors. The label can be, for example, “low sources suitability parameter”, “medium sources suitability parameter”, and “high sources suitability parameter”. Alternatively, the label can be, for example, “suitable” and “unsuitable”.

[0233] During training, the machine learning algorithm of the content source selection module 116 is presented with the training set of network resources and the associated labels in order to process various features associated with the web pages of the network resources in order to develop a given network resource profile and correlate the so-generated network resource profile to the associated labels in order to generate a “machine learning algorithm formula” that enables the machine learning algorithm of the content source selection module 116 to (i) be presented with an unknown network resource (such as a new of the first network resource 124, the second network resource 126 and the plurality of additional network resources 128), (ii) to analyze features of a given web page or the entire network resource (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128) to generate its respective network resource profile; and (iii) to determine, based on the so-generated network resource profile, a source suitability parameter associated therewith.

[0234] In some embodiments of the present technology, the plurality of features used by the machine learning algorithm includes at least an absolute feature (for example, a total number of visits to the network resource) and a relative feature (for example, an average number of visits in a given period of time or a number of hits per user). In some embodiments of the present technology, the machine learning algorithm used the absolute feature and the relative feature as base features and, additionally, generates at least one derivative feature based on at least one of the absolute feature and the relative feature to use for generating the network resource profile.

[0235] In various embodiments of the present technology, the derivative features are generated using one or more of: a logarithmic formula, a sigmoid function, and a root function. In some embodiments, the machine learning algorithm normalizes the features of the network resource. In some embodiments, the features are normalized to mean=0, stddev=1.

[0236] In accordance with some embodiments of the present technology, as part of training the machine learning algorithm of the content source selection module 116, the machine learning algorithm is configured to determine out of all possible features of the network resource a sub-set of features that better correlate to the source suitability parameter.

[0237] In a specific non-limiting embodiment of the present technology, out of all the possible features associated with the network resource, the machine learning algorithm of the content source selection module 116 can select a sub-set of features that consists of:

[0238] at least one of: a proportion of a landing page of the network resource visits within an initial transition to the network resource within a session and a proportion of the landing page visits during transitions within the network resource;

[0239] at least one of: a proportion of a web pages associated with an URL containing a date being a first page visited within the initial transition to the network resource and a proportion of the web pages within the network resource visited during the session, the web pages being associated with an URL containing a date;

[0240] a proportionate profile of visit sources for the network resource, the visit sources being selected from: search engine/social networks/mail and messenger application/URL condensers/direct visits;

[0241] a proportion of session where the network resource was last in the session.

[0242] Naturally, in alternative embodiments of the present technology, a different sub-set of entirety of features associated with the network resource can be used by the machine learning algorithm.

[0243] Examples of the features include but are not limited to:

[0244] hits per visit (HPV)—how many pages, within a given network resource, a given user has visited during browsing the given network resource

[0245] sessions per visit (SPV)—how many sessions contained visits to the particular network resource

[0246] dated-pages per visit (DPV)—how many pages of a given network resource with a date in the page URL, a given user has visited during a given visit to the given network resource

[0247] home page per visit (MPV)—how many times during a given visit a given user visited a home page of a given network resource

[0248] time spent on the network resource (MTIME)—how much time a given user has spent browsing a given network resource

[0249] last domain per session (LPS)—whether the given network resource was a last network resource in a given web browsing session of a given user

[0250] first domain per session (FPS)—whether the given network resource was a first network resource in a given web browsing session of a given user

[0251] hits per session (HPS)—how many pages of any network resources, a given user has visited during a given browsing sessions (i.e. how many transitions were made during the given session)

[0252] visits per user (VPU)—an average number of visits to the given network resource

[0253] hits per user (HPU)—an average number of transitions between various pages of the given network resource (or, alternatively, all network resources)

[0254] sessions per user (SPU)—an average number of sessions that included visits the given network resource

[0255] dated-pages per hit (DPH)—how many pages of a given network resource with a date in the page URL, a given user has visited during a given transition to the given network resource

[0256] home pages per hit (MPH)—how many times a given user has navigated to the home page of a given network resource during the visit to the given network resource

[0257] an indication of a source of a transition to the given network resource being one of:

[0258] direct access visit-source (TT\_GR);

[0259] search engine visit-source (TT\_SE);

[0260] social network visit-source (TT\_SN);

[0261] electronic message visit-source (TT\_IM);

**[0262]** URL-condenser visit-source (TT\_US); and

**[0263]** ad-platform visit-source (TT\_AD).

**[0264]** It should be noted that when the so-trained machine learning algorithm is trained and is used for determining a source suitability parameter of an unknown network resource, the machine learning algorithm uses at least some (or all) of the features that it has been trained on (or at least features of the same category as it has been trained upon).

**[0265]** In some embodiments of the present technology, the machine learning algorithm is trained to determine the source suitability parameter using a logistical regression analysis. In alternative non-limiting embodiments, the machine learning algorithm is trained to determine the source suitability parameter using one or more of: a logistical regression analysis; naive Bayes algorithm; k-nearest neighbor algorithm; random forest algorithm; and hard voting algorithm.

**[0266]** In some embodiments of the present technology, for a given network resource (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128), features thereof are grouped based on their geographical location. For example, if the given network resource is the online resource of “New York Times”, one feature can be based on network resource visitors from the United States, another feature based on network resource visitors from the United Kingdom.

**[0267]** First Stage—Content Recommendation Source Selection

**[0268]** During the first stage, the content source selection module 116 receives an indication of one or more network resources (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128). In some embodiments, the received indication is an indication of a new network resource that has not previously been crawled. In other embodiments, the received indication is for an updated network resource, i.e. the network resource that has been previously crawled but may have been updated with new content. In some embodiments of the present technology, the content source selection module 116 receives an indication of one or more network resources (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128) from the data storage device 118.

**[0269]** The content source selection module 116 then executes the trained machine learning algorithm to determine the source suitability parameter associated with the network resource in question (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128).

**[0270]** In some embodiments of the present technology, the content source selection module 116 generates the network resource profile for an entire network resource (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128) by analysing features of one or more web pages of the entire network resource (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128).

**[0271]** In other embodiments of the present technology, the content source selection module 116 generates a respective network resource profile for a sub-set of web pages of the network resource (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128) by analysing features of one or more

web pages of that sub-set of the network resource (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128).

**[0272]** In yet further embodiments, the content source selection module 116 can generate a respective network resource profile for each web page of the network resource (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128) by analysing features of each individual web page of the network resource (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128).

**[0273]** Then, based on the so-generated profile (whether for an individual web page, a sub-set of web pages within the given network resource, or the entire given network resource), the content source selection module 116 determines the content source suitability parameter. The content source selection module 116 generates the content source suitability parameter based on the machine learning algorithm formula generated during the training phase.

**[0274]** The content source selection module 116 then determines if the given network resource (or a sub-set of web pages or individual web pages thereof, as the case may be) are potential sources of content items for recommendation. In some embodiments of the present technology, the content source selection module 116 compares the so-determined source suitability parameter to a pre-determined threshold—in response to the source suitability parameter being above the pre-determined threshold, the content source selection module 116 determines that the given network resource (or a sub-set of web pages or individual web pages thereof, as the case may be) is a potential source of the recommendation items. It is noted that the pre-determined threshold can be determined by the machine learning algorithm as part of the training phase.

**[0275]** Second Stage—Content Recommendations Selection from the Selected Content Sources

**[0276]** Next, the content source selection module 116 selects, out of each of the network resources having been determined to be potential sources of content items, one or more content items for submitting to the recommendation engine (i.e. content recommendation selection module 115).

**[0277]** In some embodiments of the present technology, the selection of the content items from the so-selected content sources is executed “on-line”, i.e. in response to the receipt of the request for the content recommendation 150. In other embodiments, the selection of the content items from the so-selected content sources can be executed “off-line”, i.e. before the receipt of the request for the content recommendation 150 (for example, such selection can be executed on a regular basis, such as every day, every 7 days, every month or the like, based on how “perishable” content items are).

**[0278]** The selection of the content items from the so-selected content sources can be executed using one or more heuristics. For example, the content source selection module 116 selects a pre-defined number of content items (such as 2 content items, 5 content items, 10 content items, and the like) being associated with one or more: (i) most recent publication date, (ii) most visits during the past X number of days, (iii) being associated with a particular section of the given network resource (such as “most recent”, “hot”, etc).

[0279] Processing by the Recommendation System

[0280] Once the content source selection module 116 selects the content items candidates from one or more of the identified content sources, the so-selected candidates are submitted to the content recommendation selection module 115. The content recommendation selection module 115 executes its own machine learning algorithm that is trained and configured to select from the content items candidates submitted thereto, a sub-set of content items for the given user 102 in response to receipt of the request for the content recommendation 150.

[0281] How the content recommendation selection module 115 is implemented is not particularly limited. An example of the machine learning algorithm that can be used by the content recommendation selection module 115 is disclosed in a co-owned Russian Patent Application entitled: "METHOD AND APPARATUS FOR GENERATING A RECOMMENDED CONTENT LIST" and bearing an application number 2015136684 filed on Aug. 22, 2015; content of which is incorporated herein by reference in its entirety.

[0282] Once the content recommendation selection module 115 generates the sub-set of content items for the given user 102, the content recommendation selection module 115 generates a recommended content message 152 and transmits same to the electronic device 104. When the electronic device 104 receives the recommended content message 152, the recommendation application 106 processes content thereof and displays the recommended content items on the recommendation interface 108. How the recommendation interface 108 is implemented is not particularly limited—the recommendation interface 108 depicted in FIG. 2 can be used. As another example, the recommendation interface 108 can be implemented as disclosed in the co-owned Russian Patent Application entitled: "METHOD OF AND SYSTEM FOR INTERACTING WITH A CONTENT ELEMENT OF A CONTENT STREAM" and bearing an application number 2015141291 filed on Sep. 29, 2015; content of which is incorporated herein by reference in its entirety.

[0283] Given the architecture and examples provided herein above, it is possible to execute a method of selecting a network resource as a source of a content item, the content item to be analyzed by a recommendation system (i.e. the content recommendation selection module 115) as part of a plurality of content items to generate a set of recommended content items as a recommendation for the given user 102 of the system 100.

[0284] FIG. 4 depicts a block diagram of a method 400, the method 400 executable in accordance with non-limiting embodiments of the present technology. The method 400 can be executed on the recommendation server 112 and, more specifically, using the content source selection module 116. For the purposes of the description to be presented herein below, it is assumed that the machine learning algorithm to be executed by the content source selection module 116 has been trained to determine a source suitability parameter based on training web site profiles generated with an associated plurality of training features at least some of which are of a same category as the plurality of features to be used during execution of the method 200.

[0285] Step 402—acquiring, by the server, an indication of the network resource

[0286] The method 400 begins at step 402, where the content source selection module 116 receives an indication of the network resource.

[0287] As has been previously alluded to, the content source selection module 116 can execute a "crawling function", which can be broadly described as a function for discovering new network resources and/or new content published by the first network resource 124, the second network resource and a number of additional network resources 128. To that end, the content source selection module 116 is configured to execute what is known as a "crawler robot" to visit various network resources available/accessible via the communication network 110 (including the first network resource 124, the second network resource and a number of additional network resources 128) to discover and catalogue newly available resources (such as one of the first network resource 124, the second network resource and a number of additional network resources 128) or newly published content available from the first network resource 124, the second network resource and a number of additional network resources 128.

[0288] As part of the crawling function, the content source selection module 116 is configured to maintain (in the data storage device 118) information representative of the newly discovered network resources and/or newly discovered content available therefrom. In some embodiments, the content source selection module 116 can be configured to maintain an inverted index within the data storage device 118, but the content source selection module 116 can arrange the information representative of the newly discovered network resources and/or newly discovered content available therefrom in a data structure other than an inverted index.

[0289] As such, in some embodiments of the present technology, the content source selection module 116 can receive an indication of the network resource to be processed from the data storage device 118.

[0290] Step 404—receiving, by the server, a plurality of features associated with the network resource

[0291] At step 404, the content source selection module 116 receives a plurality of features associated with the network resource. For example, the content source selection module 116 can retrieve features associated with the network resource from the data storage device 118.

[0292] As will be recalled, the machine learning algorithm executed by the content source selection module 116 has been trained on certain features of the training set of network resources and, as such, the machine learning algorithm executed by the content source selection module 116 retrieves those features associated with the network resource that it has been trained to process.

[0293] In some embodiments of the method 400, the machine learning algorithm having been trained, prior to said acquiring, to determine the source suitability parameter based on training web site profiles generated with an associated plurality of training features at least some of which are of a same category as the plurality of features.

[0294] In some embodiments of the method 400, the plurality of features includes at least an absolute feature and a relative feature.

[0295] In some embodiments of the method 400 the absolute feature and the relative feature are base features and the method 400 further comprises: generating at least one derivative feature based on at least one of the absolute feature and the relative feature; and wherein the given web site profile is further based on the at least one derivative feature.

[0296] In some embodiments of the method 400, the plurality of features includes an indication of a proportion of visit sources to the network resource.

[0297] In some embodiments of the method 400, the plurality of features includes at least one relative feature selected from:

- [0298] hits per visit (HPV);
- [0299] sessions per visit (SPV);
- [0300] dated-pages per visit (DPV);
- [0301] home page per visit (MPV);
- [0302] time spent on the network resource (MTIME);
- [0303] last domain per session (LPS);
- [0304] first domain per session (FPS);
- [0305] hits per session (HPS);
- [0306] visits per user (VPU);
- [0307] hits per user (HPU);
- [0308] sessions per user (SPU);
- [0309] dated-pages per hit (DPH);
- [0310] home pages per hit (MPH); and
- [0311] an indication of a visit source being one of:
- [0312] direct access visit-source (TT\_GR);
- [0313] search engine visit-source (TT\_SE);
- [0314] social network visit-source (TT\_SN);
- [0315] electronic message visit-source (TT\_IM);
- [0316] URL-condenser visit-source (TT\_US); and
- [0317] ad-platform visit-source (TT\_AD).

[0318] In some embodiments of the method 400, the plurality of features includes at least one of: a number of users visiting the network resource per time interval, number of visits per time interval, number of transitions within various web pages within the network resource per time interval, and a number of unique sessions with the network resource per time interval.

[0319] In some embodiments of the method 400, the plurality of features is a subset of all possible features, the method further comprising executing a feature selection to define the subset.

[0320] In some embodiments of the method 400, the subset of all possible features consists of:

- [0321] at least one of: a proportion of a landing page of the network resource visits within an initial transition to the network resource within a session and a proportion of the landing page visits during transitions within the network resource;
- [0322] at least one of: a proportion of a web page associated with an URL containing a date being a first page visited within the initial transition to the network resource and a proportion of the web pages within the network resource visited during the session, the web pages being associated with an URL containing a date;
- [0323] a proportionate profile of visit sources for the network resource, the visit sources being selected from: search engine/social networks/mail and messenger application/URL condensers/direct visits;
- [0324] a proportion of session where the network resource was last in the session.

[0325] Step 406—generating, by the server, a given network resource profile for the network resource, the given network resource profile being based on the plurality of features

[0326] At step 406, the content source selection module 116 generates a given network resource profile for the network resource, the given network resource profile being based on the plurality of features.

[0327] In some embodiments of the present technology, the content source selection module 116 generates a vector representative of the given network resource profile based on some or all of:

- [0328] hits per visit (HPV);
- [0329] sessions per visit (SPV);
- [0330] dated-pages per visit (DPV);
- [0331] home page per visit (MPV);
- [0332] time spent on the network resource (MTIME);
- [0333] last domain per session (LPS);
- [0334] first domain per session (FPS);
- [0335] hits per session (HPS);
- [0336] visits per user (VPU);
- [0337] hits per user (HPU);
- [0338] sessions per user (SPU);
- [0339] dated-pages per hit (DPH);
- [0340] home pages per hit (MPH); and
- [0341] an indication of a visit source being one of:
- [0342] direct access visit-source (TT\_GR);
- [0343] search engine visit-source (TT\_SE);
- [0344] social network visit-source (TT\_SN);
- [0345] electronic message visit-source (TT\_IM);
- [0346] URL-condenser visit-source (TT\_US); and
- [0347] ad-platform visit-source (TT\_AD).

[0348] In some embodiments of the present technology, the content source selection module 116 executes a feature selection routine and, as such, the vector representative of the given network resource profile can be based on a pre-selected set of network resource features.

[0349] Step 408—executing, by the server, a machine learning algorithm in order to, based on the given network resource profile, determine a source suitability parameter for the network resource, the source suitability parameter being indicative of the network resource being susceptible to publishing the content item that can be processed by the recommendation system

[0350] At step 408, the content source selection module 116 executes a machine learning algorithm in order to, based on the given network resource profile, determine a source suitability parameter for the network resource, the source suitability parameter being indicative of the network resource being susceptible to publishing the content item that can be processed by the recommendation system.

[0351] In some embodiments of the method 400, the source suitability parameter is one of a suitable and non-suitable. The source suitability parameter can be determined using at least one of:

- [0352] a logistical regression analysis;
- [0353] naive Bayes algorithm;
- [0354] k-nearest neighbor algorithm;
- [0355] random forest algorithm; and
- [0356] hard voting algorithm.

[0357] Step 410—selecting at least one content item from the network resource if the source suitability parameter is determined to be above a pre-determined threshold, the at least one content item for submitting to the recommendation system for processing

[0358] At step 410, the content source selection module 116 selects at least one content item from the network resource if the source suitability parameter is determined to be above a pre-determined threshold, the at least one content item for submitting to the recommendation system for processing (i.e. the content recommendation selection module 115).

[0359] In some embodiments of the present technology, step 410 is implemented in response to a request from the given user 102 for a recommendation (i.e. in response to receiving the request for the content recommendation 150).

[0360] The method 400 can then terminate or return to step 402 and await for another one of the request for content recommendation 150.

[0361] With reference to FIG. 3, there is depicted another embodiment of a system 300, the system 300 being implemented in accordance with another non-limiting embodiment of the present technology. The system 300 includes certain components that are implemented substantially similar to those of the system 100. These components are denoted with the same reference numerals as components of the system 100. To that end, the system 300 comprises: the user 102, the electronic device 104 (with the recommendation application 106 executed thereon, the recommendation application 106 having the recommendation interface 108), and the communication network 110. Also coupled to the communication network 110 are the first network resource 124, the second network resource 126 and the plurality of additional network resources 128 that are all network resources accessible by the electronic device 104 (as well as other electronic devices potentially present in the system 300) via the communication network 110.

[0362] Respective content of the first network resource 124, the second network resource 126 and the plurality of additional network resources 128 is not particularly limited. As such a given one of the first network resource 124, the second network resource 126 and the plurality of additional network resources 128 can contain (or in other words, host) digital content such as: audio content for streaming, video content for streaming, news, blogs, information about various government institutions, information about points of interest, thematically clustered content (such as content relevant to those interested in kick-boxing), and the like.

[0363] In accordance with the non-limiting embodiment depicted in FIG. 3, the system 300 further comprises a recommendation server 312. The recommendation server 312 can be implemented as a conventional computer server. In an example of an embodiment of the present technology, the recommendation server 312 can be implemented as a Dell™ PowerEdge™ Server running the Microsoft™ Windows Server™ operating system. Needless to say, the recommendation server 312 can be implemented in any other suitable hardware, software, and/or firmware, or a combination thereof. In the depicted non-limiting embodiments of the present technology, the recommendation server 312 is a single server. In alternative non-limiting embodiments of the present technology, the functionality of the recommendation server 312 may be distributed and may be implemented via multiple servers.

[0364] The recommendation server 312 comprises a processing module 314. The processing module 314 is coupled to a content source determination module 316. The content source determination module 316 is configured to determine out of a plurality of potential content sources (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128) a sub-set of content sources.

[0365] In some embodiments of the present technology, the content source determination module 316 is configured to execute a first machine learning algorithm in order to determine the sub-set of recommended content sources from

the plurality of possible content sources (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128).

[0366] In some embodiments of the present technology, in order to determine the sub-set of recommended content sources, the content source determination module 316 is configured to:

[0367] acquire an indication of user-past-interactions with at least one of: (i) the system 100 and (ii) at least some of the plurality of network resources (such as the first network resource 124, the second network resource 126 and the plurality of additional network resources 128);

[0368] based on the user-past-interactions, determine a first sub-set of user-specific content sources (i.e. some of the first network resource 124, the second network resource 126 and the plurality of additional network resources 128);

[0369] based on (i) a machine learning algorithm trained formula of other user interactions with at least some others of the plurality of network resources (i.e. the first network resource 124, the second network resource 126 and the plurality of additional network resources 128) and at least one of: (ii) the first sub-set of user-specific content sources; and (iii) a user-profile-vector generated based on the user-past-interactions, to generate a second sub-set of user-non-specific content sources;

[0370] process the first sub-set of user specific content sources and the second sub-set of user-non-specific content sources in order to generate the sub-set of recommended content sources;

[0371] More specifically, the function of the content source determination module 316 can be implemented as follows, in at least some of the non-limiting embodiments of the present technology.

[0372] Generating a User-Profile-Vector

[0373] The content source determination module 316 can acquire an indication of user-past-interactions with at least one of: (i) the system 300 and (ii) at least some of the plurality of network resources (such as the first network resource 124, the second network resource 126 and the plurality of additional network resources 128).

[0374] In some embodiments of the present technology, the content source determination module 316 receives the indication of the user-past-interaction from the browser application executed on the electronic device 104. In other embodiments, the content source determination module 316 receives the indication of the user-past-interaction from the search engine logs. In yet other embodiments, the content source determination module 316 receives the indication of the user-past-interaction from the tracking widgets installed at the various network resources (such as the first network resource 124, the second network resource 126 and the plurality of additional network resources 128).

[0375] Determining a User-Specific Sub-Set of Content Sources

[0376] Based on the user-past-interactions, the content source determination module 316 can determine a first sub-set of user-specific content sources (i.e. some of the first network resource 124, the second network resource 126 and the plurality of additional network resources 128). More specifically, the content source determination module 316 determines one or more of the potentially content sources

(i.e. some of the first network resource **124**, the second network resource **126** and the plurality of additional network resources **128**) that the user **102** has interacted with in the past—for example, by means of browsing to or through one or more of the network resources, receiving and/or actuating a content recommendations from one or more of the network resources, etc. In other words, the first sub-set of user-specific content sources can be thought of “user-specific” or “expressly liked content sources” in the sense that the user **102** has specifically interacted with these network resources and can be deemed to like the content thereof due to the interaction therewith.

**[0377]** Enriching the User-Specific Sub-Set of Content Sources to Generate a User-Non-Specific Sub-Set of Content Sources

**[0378]** Next, the content source determination module **316** enriches the user-specific sub-set of content sources with content sources that the user **102** may not have previously seen. More specifically, based on (i) a machine learning algorithm trained formula of other user interactions with at least some others of the plurality of network resources (i.e. the first network resource **124**, the second network resource **126** and the plurality of additional network resources **128**) and at least one of: (ii) the first sub-set of user-specific content sources; and (iii) a user-profile-vector generated based on the user-past-interactions, to generate a second sub-set of user-non-specific content sources.

**[0379]** In some embodiments of the present technology, the machine learning algorithm used can be a Singular Value Decomposition (SVD) algorithm. Within these implementations of the present technology, the machine learning algorithm determines the user-non-specific sub-set of content sources based on a user-profile-vector generated based on the user-past-interactions. More specifically, the machine learning algorithm, implemented as the SVD algorithm, generates a matrix of (i) network resources events and (ii) users (based on information about past interactions of a plurality of users with a plurality of network resources, i.e. the first network resource **124**, the second network resource **126** and the plurality of additional network resources **128**).

**[0380]** The machine learning algorithm then applies the SVD algorithm to decompose the matrix into vectors of network resources and vectors of users. Then, the SVD algorithm discards the vectors of users and saves the vectors of network resources (for example stores them in a data storage device, which is not depicted in FIG. 3).

**[0381]** When the content source determination module **316** receives the user-profile-vector, the SVD algorithm multiplies the so-received user-profile-vector and the saved vectors of network resources to recreate the matrix of network resources with their associated ranks specific for the user **102** associated with the user-profile-vector. Based on the ranked network resources, the content source determination module **316** selects a pre-determined number of top-ranked network resources to generate the user-non-specific sub-set of content sources.

**[0382]** In some embodiments of the present technology, the machine learning algorithm used can be Pointwise Mutual Information (PMI) algorithm. Within these implementations of the present technology, the machine learning algorithm determines the user-non-specific sub-set of content sources based on the first sub-set of user-specific content sources. More specifically, the machine learning algorithm determines for each of the network resources within the first

sub-set of user-specific content sources and each of the crawled potential content sources: (i) how often the pair has occurred together (were browsed together, appeared together as part of search results, etc) during a particular interaction; and (ii) how often a respective one of the network resources was interacted with. The machine learning algorithm then calculates a proportion of the number of occurrences together (as per (i)) to a multiplication of the individual values of occurrences of the respective network resources (as per (ii)).

**[0383]** For each of the network resources in the first sub-set of user-specific content sources, the machine learning algorithm selects a pre-determined number of top-ranked resources determined by the PMI algorithm to generate the user-non-specific sub-set of content sources.

**[0384]** Processing User-Specific and User-Non-Specific Subsets of Content Sources to Generate the Sub-Set of Recommended Content Sources

**[0385]** The content source determination module **316** then processes the first sub-set of user specific content sources and the second sub-set of user-non-specific content sources in order to generate the sub-set of recommended content sources.

**[0386]** In some embodiments of the present technology, the content source determination module **316** selects a pre-determined number of the content sources from the user-non-specific sub-set of content sources. In other embodiments of the present technology, the content source determination module **316** selects a pre-determined number of the content sources from the user-non-specific sub-set of content sources and a pre-determined number of content sources from the first sub-set of user-specific content sources. The content source determination module **316** can be configured to determine the respective numbers of the pre-determined content sources, considering that the larger the number selected from user-non-specific sub-set of content sources, the more content items recommendations from “new” sources the user **102** is likely to get. Conversely, the larger the number selected from first sub-set of user-specific content sources, the more content items recommendations from “known” sources (and therefore, more likely to be acceptable) the user **102** is likely to get

**[0387]** The recommendation server **312** further comprises a potentially-recommendable content items selection module **318** for analyzing the sub-set of recommended content sources to select a plurality of potentially-recommendable content items.

**[0388]** The selection of the content items from the so-selected content sources can be executed using one or more heuristics. For example, the content source selection module **116** selects a pre-defined number of content items (such as 2 content items, 5 content items, 10 content items, and the like) being associated with one or more: (i) most recent publication date, (ii) most visits during the past X number of days, (iii) being associated with a particular section of the given network resource (such as “most recent”, “hot”, etc).

**[0389]** The recommendation server **312** further comprises a content recommendation selection module **320** for executing a second machine learning algorithm module in order to select, from the plurality of potentially-recommendable content items, at least one recommended content item; the selection being made on the basis of the user-profile-vector.

**[0390]** How the content recommendation selection module **320** is not particularly limited. An example of the

machine learning algorithm that can be used by the content recommendation selection module 320 is disclosed in a co-owned Russian Patent Application entitled: “METHOD AND APPARATUS FOR GENERATING A RECOMMENDED CONTENT LIST” and bearing an application number 2015136684 filed on Aug. 22, 2015; content of which is incorporated herein by reference in its entirety.

[0391] Once the content recommendation selection module 320 generates the sub-set of content items for the given user 102, the content recommendation selection module 320 generates a recommended content message 152 and transmits same to the electronic device 104. When the electronic device 104 receives the recommended content message 152, the recommendation application 106 processes content thereof and displays the recommended content items on the recommendation interface 108. How the recommendation interface 108 is implemented is not particularly limited—the recommendation interface 108 depicted in FIG. 2 can be used. As another example, the recommendation interface 108 can be implemented as disclosed in the co-owned Russian Patent Application entitled: “METHOD OF AND SYSTEM FOR INTERACTING WITH A CONTENT ELEMENT OF A CONTENT STREAM” and bearing an application number 2015141291 filed on Sep. 29, 2015; content of which is incorporated herein by reference in its entirety.

[0392] Given the architecture and examples provided herein above, it is possible to execute a method of generating a content recommendation for the user 170 of the electronic device 104. FIG. 5 depicts a block diagram of a method 500, the method 500 executable in accordance with non-limiting embodiments of the present technology. The method 500 can be executed on the recommendation server 312 and, more specifically, using the content source determination module 316 and the potentially-recommendable content items selection module 318. For the purposes of the description to be presented herein below, it is assumed that the machine learning algorithm to be executed by the content source selection module 116 has been trained to determine a source suitability parameter based on training web site profiles generated with an associated plurality of training features at least some of which are of a same category as the plurality of features to be used during execution of the method 200.

[0393] Step 502—receiving, from the electronic device, a request for the content recommendation, the content recommendation including at least one recommended content item

[0394] The method 500 starts at step 502, where the content source selection module 116 receives, from the electronic device 104, the request for the content recommendation 150, the content recommendation including at least one recommended content item.

[0395] Step 504—executing a first machine learning algorithm module in order to determine a sub-set of recommended content sources from a plurality of possible content sources, the determining the sub-set of recommended content sources including: acquiring an indication of user-past-interactions with at least one of: (i) the recommendation system and (ii) at least some of the plurality of network resources; based on the user-past-interactions, determining a first sub-set of user-specific content sources; based on (i) a machine learning algorithm trained formula of other user interactions with at least some others of the plurality of network resources and at least one of: (ii) the first sub-set of user-specific content sources; and (iii) a user-profile-vector

generated based on the user-past-interactions generating a second sub-set of user-non-specific content sources

[0396] At step 504, the content source selection module 116 executes a first machine learning algorithm module in order to determine a sub-set of recommended content sources from a plurality of possible content sources, the determining the sub-set of recommended content sources including:

[0397] acquiring an indication of user-past-interactions with at least one of: (i) the recommendation system and (ii) at least some of the plurality of network resources;

[0398] based on the user-past-interactions, determining a first sub-set of user-specific content sources;

[0399] based on

[0400] (i) a machine learning algorithm trained formula of other user interactions with at least some others of the plurality of network resources; and at least one of:

[0401] (ii) the first sub-set of user-specific content sources; and

[0402] (iii) a user-profile-vector generated based on the user-past-interactions

[0403] generating a second sub-set of user-non-specific content sources

[0404] In some embodiment of the method 500, the determining the first sub-set of user-specific content sources comprises applying an SVD algorithm. Within these embodiments, the determining the first sub-set of user-specific content sources is based on: (i) the machine learning algorithm trained formula of other user interactions with at least some others of the plurality of network resources; and (ii) the user-profile-vector generated based on the user-past-interactions.

[0405] Within these embodiments of the method 500, the method 500 further comprises, prior to the receiving the request for the content recommendation, generating by the SVD algorithm a matrix of network resources events and users.

[0406] In other embodiments of the method 500, the determining the first sub-set of user-specific content sources comprises applying a PMI algorithm. In these embodiments of the method 500, the determining the first sub-set of user-specific content sources is based on: (i) the machine learning algorithm trained formula of other user interactions with at least some others of the plurality of network resources; and (ii) the first sub-set of user-specific content sources.

[0407] Step 506—processing the first sub-set of user specific content sources and the second sub-set of user-non-specific content sources in order to generate the sub-set of recommended content sources

[0408] At step 506, the content source selection module 116 processes the first sub-set of user specific content sources and the second sub-set of user-non-specific content sources in order to generate the sub-set of recommended content sources.

[0409] Step 508—analyzing the sub-set of recommended content sources to select a plurality of potentially-recommendable content items

[0410] At step 508, the content source selection module 116 analyzes the sub-set of recommended content sources to select a plurality of potentially-recommendable content items

[0411] Step 510—executing a second machine learning algorithm module in order to select, from the plurality of potentially-recommendable content items, at least one recommended content item; the selection being made on the basis of the user-profile-vector

[0412] At step 510, the comprises a potentially-recommendable content items selection module 318 executes a second machine learning algorithm module in order to select, from the plurality of potentially-recommendable content items, at least one recommended content item; the selection being made on the basis of the user-profile-vector.

[0413] In some embodiments of the method 500, the selection from the sub-set of recommended content sources the plurality of potentially-recommendable content items comprises determining a pre-defined number of most recent content items.

[0414] In other embodiments of the method 500, the selection from the sub-set of recommended content sources the plurality of potentially-recommendable content items comprises determining a pre-defined number of most popular content items.

[0415] In other embodiments of the method 500, the selection from the sub-set of recommended content sources is specific to a geographic region where the request for content recommendation was executed from.

[0416] In other embodiments of the method 500, the method 500 further comprises filtering the selected potentially-recommendable content items. In other embodiments of the method 500, the filtering is based on a language setting of the electronic device and a language setting of the content item.

[0417] Some embodiments of the present technology can lead to a technical effect of arising from the use of embodiments of the present technology, is the ability to “pre qualify” potential sources of content recommendations in an “offline” mode (“offline” referring to processing prior to receiving the request for the content recommendation 150). In some embodiments of the present technology, a technical effect arises from an ability to pre qualify network resources as sources of recommended content items when entering a new territory.

[0418] It should be expressly understood that not all technical effects mentioned herein need to be enjoyed in each and every embodiment of the present technology. For example, embodiments of the present technology may be implemented without the user enjoying some of these technical effects, while other embodiments may be implemented with the user enjoying other technical effects or none at all.

[0419] Modifications and improvements to the above-described implementations of the present technology may become apparent to those skilled in the art. The foregoing description is intended to be exemplary rather than limiting. The scope of the present technology is therefore intended to be limited solely by the scope of the appended claims.

1. A method of selecting a network resource as a source of a content item, the content item to be analyzed by a recommendation system as part of a plurality of content items to generate a set of recommended content items as a recommendation for a given user of the recommendation system, the recommendation system being executed at a server coupled to a communication network, the method comprising:

acquiring, by the server, an indication of the network resource;

receiving, by the server, a plurality of features associated with the network resource;

generating, by the server, a given network resource profile for the network resource, the given network resource profile being based on the plurality of features;

executing, by the server, a machine learning algorithm in order to, based on the given network resource profile, determine a source suitability parameter for the network resource, the source suitability parameter being indicative of the network resource being susceptible to publishing the content item that can be processed by the recommendation system,

selecting at least one content item from the network resource if the source suitability parameter is determined to be above a pre-determined threshold, the at least one content item for submitting to the recommendation system for processing.

2. The method of claim 1, wherein:

the machine learning algorithm having been trained, prior to said acquiring, to determine the source suitability parameter based on training web site profiles generated with an associated plurality of training features at least some of which are of a same category as the plurality of features.

3. The method of claim 1, wherein the plurality of features includes at least an absolute feature and a relative feature.

4. The method of claim 3, wherein the absolute feature and the relative feature are base features and wherein the method further comprises:

generating at least one derivative feature based on at least one of the absolute feature and the relative feature; and wherein

the given web site profile is further based on the at least one derivative feature.

5. The method of claim 1, wherein the plurality of features includes an indication of a proportion of visit sources to the network resource.

6. The method of claim 1, wherein the plurality of features includes at least one relative feature selected from:

hits per visit (HPV);  
sessions per visit (SPV);  
dated-pages per visit (DPV);  
home page per visit (MPV);  
time spent on the network resource (MTIME);  
last domain per session (LPS);  
first domain per session (FPS);  
hits per session (HPS);  
visits per user (VPU);  
hits per user (HPU);  
sessions per user (SPU);  
dated-pages per hit (DPH);  
home pages per hit (MPH); and  
an indication of a visit source being one of:  
direct access visit-source (TT\_GR);  
search engine visit-source (TT\_SE);  
social network visit-source (TT\_SN);  
electronic message visit-source (TT\_IM);  
URL-condenser visit-source (TT\_US); and  
ad-platform visit-source (TT\_AD).

7. The method of claim 1, wherein the plurality of features includes at least one of: a number of users visiting the network resource per time interval, number of visits per time interval, number of transitions within various web pages



within the network resource per time interval, and a number of unique sessions with the network resource per time interval.

**8.** The method of claim **1**, wherein the plurality of features is a subset of all possible features, the method further comprising executing a feature selection to define the subset.

**9.** The method of claim **8**, wherein the subset of all possible features consists of:

at least one of: a proportion of a landing page of the network resource visits within an initial transition to the network resource within a session and a proportion of the landing page visits during transitions within the network resource;

at least one of: a proportion of a web page associated with an URL containing a date being a first page visited within the initial transition to the network resource and a proportion of the web pages within the network resource visited during the session, the web pages being associated with an URL containing a date;

a proportionate profile of visit sources for the network resource, the visit sources being selected from: search engine/social networks/mail and messenger application/URL condensers/direct visits;

a proportion of session where the network resource was last in the session.

**10.** The method of claim **1**, wherein the source suitability parameter is one of a suitable and non-suitable.

**11.** The method of claim **10**, wherein the source suitability parameter is determined using at least one of:

a logistical regression analysis;  
naive Bayes algorithm;  
k-nearest neighbor algorithm;  
random forest algorithm; and  
hard voting algorithm.

**12.** The method of claim **1**, wherein said selecting is executed in response to a request from the given user for a recommendation.

**13.** A server comprising:

a data storage medium;

a network interface configured for communication over a communication network;

a processor operationally coupled to the data storage medium and the network interface, the processor being configured to:

acquire an indication of the network resource, the network resource to be processed to determine its suitability as a source of a content item, the content item to be analyzed by a recommendation system as part of a plurality of content items to generate a set of recommended content items as a recommendation for a given user of the recommendation system;

receive a plurality of features associated with the network resource;

generate a given network resource profile for the network resource, the given network resource profile being based on the plurality of features;

execute a machine learning algorithm in order to, based on the given network resource profile, determine a source suitability parameter for the network resource, the source suitability parameter being indicative of the network resource being susceptible to publishing the content item that can be processed by the recommendation system,

select at least one content item from the network resource if the source suitability parameter is determined to be above a pre-determined threshold, the at least one content item for submitting to the recommendation system for processing.

**14.** The server of claim **13**, wherein:

the machine learning algorithm has been trained, prior to said acquiring, to determine the source suitability parameter based on training web site profiles generated with an associated plurality of training features at least some of which are of a same category as the plurality of features.

**15.** The server of claim **13**, wherein the plurality of features includes at least an absolute feature and a relative feature.

**16.** The server of claim **15**, wherein the absolute feature and the relative feature are base features and wherein the processor is further configured to:

generate at least one derivative feature based on at least one of the absolute feature and the relative feature; and  
wherein

the given web site profile is further based on the at least one derivative feature.

**17.** The server of claim **13**, wherein the plurality of features includes an indication of a proportion of visit sources to the network resource.

**18.** The server of claim **13**, wherein the plurality of features includes at least one relative feature selected from:

hits per visit (HPV);  
sessions per visit (SPV);  
dated-pages per visit (DPV);  
home page per visit (MPV);  
time spent on the network resource (MTIME);  
last domain per session (LPS);  
first domain per session (FPS);  
hits per session (HPS);  
visits per user (VPU);  
hits per user (HPU);  
sessions per user (SPU);  
dated-pages per hit (DPH);  
home pages per hit (MPH); and  
an indication of a visit source being one of:

direct access visit-source (TT\_GR);  
search engine visit-source (TT\_SE);  
social network visit-source (TT\_SN);  
electronic message visit-source (TT\_IM);  
URL-condenser visit-source (TT\_US); and  
ad-platform visit-source (TT\_AD).

**19.** The server of claim **13**, wherein the plurality of features includes at least one of: a number of users visiting the network resource per time interval, number of visits per time interval, number of transitions within various web pages within the network resource per time interval, and a number of unique sessions with the network resource per time interval.

**20.** The server of claim **13**, wherein the plurality of features is a subset of all possible features, the method further comprising executing a feature selection to define the subset; and wherein the subset of all possible features consists of:

at least one of: a proportion of a landing page of the network resource visits within an initial transition to the

- network resource within a session and a proportion of the landing page visits during transitions within the network resource;
- at least one of: a proportion of a web page associated with an URL containing a date being a first page visited within the initial transition to the network resource and a proportion of the web pages within the network resource visited during the session, the web pages being associated with an URL containing a date;
- a proportionate profile of visit sources for the network resource, the visit sources being selected from: search engine/social networks/mail and messenger application/URL condensers/direct visits;
- a proportion of session where the network resource was last in the session.

\* \* \* \* \*