Title: A CONTENT ITEM DISTRIBUTION SYSTEM AND METHOD OF DISTRIBUTION THEREOF

Abstract: A content item distribution system comprises a server determining content items to remote user devices. The server stores a hierarchical user cluster structure wherein user clusters at a higher hierarchical level comprise one or more user clusters of a next lower hierarchical level. An initialisation processor selects a first user cluster at a first hierarchical level for a content item. A sample processor recommends the content item to a sample group of user devices of the first user cluster and receives user rating feedback from these. A propagation processor determines if the user rating feedback meets a popularity criterion for the first content item and if so recommends the content item to user devices of a second user cluster which is at a higher hierarchical level than the first hierarchical level and comprises the first user cluster group and at least one other user cluster of the first hierarchical level.
Published:
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments
— with international search report
A CONTENT ITEM DISTRIBUTION SYSTEM AND METHOD OF DISTRIBUTION THEREFOR

Field of the invention

The invention relates to a content item distribution system and method of distribution therefor and in particular, but not exclusively, to a recommendation system for recommending content items such as articles, video clips, music etc.

Background of the Invention

In recent years, the availability and provision of information and entertainment content has increased substantially. For example, the number of online news and entertainment articles available to the average user has grown considerably with the increased popularity of the Internet. Consequently, users are increasingly provided with a plethora of different types of content from different sources. In order to identify and select the desired content, the user must typically process large amounts of information which can be very cumbersome and impractical.

Accordingly, significant resources have been invested in research into techniques and algorithms that may provide an improved user experience and assist a user in identifying and selecting content, personalising services etc. Specifically, recommendation systems have been developed which can automatically search through vast amounts of content and select specific content that is recommended to the individual user.

In order to enhance the user experience, it is advantageous to personalise the recommendations for the individual user as much as possible. In this context, a recommendation can be considered to consist in predicting how much a user may like a particular content item and recommending it if it is considered of sufficient interest.

People increasingly use a wide range of electronic devices for different purposes and with different capabilities (e.g. cell phones, music players, set-top boxes, personal computers, etc.) All these devices are running more and more complex applications with many applications using some element of personalisation to provide a better or simplified user experience.
Thus, user communication systems and data networks are increasingly used for distribution of content items. However, such data networks are typically designed and developed for point to point communication between two or more parties and have not been specifically designed for distribution and broadcasting of content. Accordingly, a problem in such systems is how to provide efficient and reliable distribution of content. For example, in order to reduce the communication resource usage, it is desirable that the content items are only distributed to users which are likely to have an interest in the content item. At the same time, it is desirable that the content is distributed such that it reaches all users that have a potential interest.

Such problems may be addressed by the use of recommendation systems wherein a central server determines which users are likely to be interested in specific content thereby allowing the server to target the distribution of content targeted to the identified users.

However, such conventional systems tend to have a number of disadvantages. For example, conventional systems tend to be inflexible and require that a given content item must be distributed to a large number of users with a relatively low preference in order to ensure that the content is sufficiently distributed to all users that may consume the content. Furthermore, conventional systems tend to treat all content in the same way resulting in all content having being distributed equally broadly. Furthermore, in order to vary the degree of distribution, manual intervention is required to e.g. change the preference threshold used to select which users the content is distributed to. Such an approach is cumbersome and impractical for even a moderate number of content items.

Furthermore, the distribution of content in conventional systems is based on stored user preference data which represent an average and historic preference that may not accurately reflect the user’s current instantaneous preference. For example, if an unexpected but significant event occurs, the preferences of the users may change quickly and thus the current preferences may deviate from the stored user preferences. Furthermore, this preference change may be relatively short lived and therefore should not result in a modification to the stored long term user preference profile.
As a specific example, the Internet may support news item distribution systems wherein news items are distributed to users of user devices coupled to the Internet. Such a distribution system may use a centralised server comprising user preference profiles for a large number of users to select which articles are sent to which users. This may substantially reduce the number of items that must be sent to the individual user thereby substantially reducing the communication resource usage and providing an improved user experience as the user is not required to filter a large number of items to select those of particular interest. In order to achieve this, the central server correlates each content item with the user’s preferences and transmits content items only to those users for which a close match is achieved.

However, in order to achieve an efficient system with low communication resource usage and with a low requirement for manual item filtering by the user, the preference threshold must be set relatively high. However, some news items may be of more general interest to most users even if the articles do not closely match the specific user preferences stored for the users. For example, news items relating to important and recent events may be of interest to most users even if the topic of the news item is not historically of interest to the users. Such, articles tend not to be distributed widely in conventional systems unless a manual intervention is carried out to expand the distribution.

Thus, on one hand, it is preferable that users only receive content items closely matched to their specific interests. However, it is also important that users are provided with more general "headline" news items (i.e. news items of interest to many users despite not being closely correlated with their specific long term historical preferences). For example, many users would want to read about a current significant natural disaster, such as a large earthquake, although they may not previously have expressed an interest in natural disasters or consumed similar articles. In conventional system it is a problem to achieve a wide distribution of such content without high communication resource usage or manual intervention.

Also, conventional collaborative filtering systems allow recommendation of items based on user ratings of other similar users. However, the requirement to achieve accurate recommendations in such systems typically means that popular items
will propagate very slowly from the initial seed users to the wider population. On the other hand, performing an initial sampling of user opinions across a larger user population, whose content preferences are not a close match for the given content item may result in user irritation with the system and would waste system resources such as network bandwidth.

Hence, an improved content distribution would be advantageous and in particular a system allowing reduced communication resource usage, sufficient distribution of content, reduced manual intervention, improved flexibility, improved adaptability, facilitated operation, improved performance and improved user experience would be advantageous.

Summary of the Invention

Accordingly, the Invention seeks to preferably mitigate, alleviate or eliminate one or more of the above mentioned disadvantages singly or in any combination.

According to a first aspect of the invention there is provided a content item distribution system comprising a central recommendation server recommending content items to a plurality of remote user devices, the central recommendation server comprising: means for providing a hierarchical user cluster structure of user clusters of the remote user devices, the user clusters at a higher hierarchical level comprising one or more user clusters of a next lower hierarchical level; cluster determining means for determining a first user cluster for a first content item, the first user cluster being at a first hierarchical level; sample means for recommending the first content item to a first sample group of remote user devices of the first user cluster; means for receiving user rating feedback from the first sample group; propagation means for determining if the user rating feedback meets a popularity criterion for the first content item and if so recommending the content item to remote user devices of a second user cluster, the second user cluster being a user cluster of a higher hierarchical level of the hierarchical user cluster structure than the first hierarchical level and comprising the first user cluster and at least one other user cluster of the first hierarchical level.

The invention may provide an improved content item distribution system. The invention may allow reduced communication resource usage in a data network or
communication system supporting the distribution of content items. Reduced communication usage may be achieved while ensuring sufficient distribution of content items. The degree of distribution may automatically adapt to the current preferences of users in the system. The system may allow facilitated operation and/or reduce or eliminate the need for manual intervention. In particular, the system may in many embodiments provide an automated propagation of a content item that can not only reflect specific historic user preference but can also adapt to dynamic changes (e.g. sudden and/or short term changes) in the preferences of the users. The degree of propagation may automatically adapt to the current characteristics of the users and/or of the content item. An improved user experience may be achieved and in particular a user may be presented with content items which reflect the user's preferences as well as general content items of interest to many users.

The first user cluster may be an initial user cluster in which the first content item is initially recommended to remote user devices or may be a user cluster reached by previous iterations. The popularity criterion may be any criterion applied to the user ratings. In particular, it may comprise a requirement that an accumulated and/or averaged rating value for received user rating feedback exceeds a threshold. Recommendation of the first content item to a remote user device may specifically comprise transmitting the first content item to the remote user device and/or may e.g. comprise transmitting an indication of the first content item to the remote user device.

According to another aspect of the invention there is provided a central recommendation server for a content item distribution system, the central recommendation server comprising: means for providing a hierarchical user cluster structure of user clusters of the remote user devices, the user clusters at a higher hierarchical level comprising one or more user clusters of a next lower hierarchical level; cluster determining means for determining a first user cluster for a first content item, the first user cluster being at a first hierarchical level; sample means for recommending the first content item to a first sample group of remote user devices of the first user cluster; means for receiving user rating feedback from the first sample group; and propagation means for determining if the user rating feedback meets a popularity criterion for the first content item and if so recommending the content item
to remote user devices of a second user cluster, the second user cluster being a user cluster of a higher hierarchical level of the hierarchical user cluster structure than the first hierarchical level and comprising the first user cluster and at least one other user cluster of the first hierarchical level.

According to another aspect of the invention there is provided a method of content item distribution for a content item distribution system comprising a central recommendation server recommending content items to a plurality of remote user devices, the method comprising the central recommendation server performing the steps of: providing a hierarchical user cluster structure of user clusters of the remote user devices, the user clusters at a higher hierarchical level comprising one or more user clusters of a next lower hierarchical level; determining a first user cluster for a first content item, the first user cluster being at a first hierarchical level; recommending the first content item to a first sample group of remote user devices of the first user cluster; receiving user rating feedback from the first sample group; and determining if the user rating feedback meets a popularity criterion for the first content item and if so recommending the content item to remote user devices of a second user cluster, the second user cluster being a user cluster of a higher hierarchical level of the hierarchical user cluster structure than the first hierarchical level and comprising the first user cluster and at least one other user cluster of the first hierarchical level.

These and other aspects, features and advantages of the invention will be apparent from and elucidated with reference to the embodiment(s) described hereinafter.
**Brief Description of the Drawings**

Embodiments of the invention will be described, by way of example only, with reference to the drawings, in which

FIG. 1 illustrates an example of a distributed content item recommendation system in accordance with some embodiments of the invention;

FIG. 2 illustrates an example of a recommendation server in accordance with some embodiments of the invention;

FIG. 3 illustrates an example of a method of content distribution in accordance with some embodiments of the invention;

FIG. 4 illustrates an example of a user preference taxonomy for content items; and

FIG. 5 illustrates an example of a propagation of a popular news item through a hierarchical user cluster structure in accordance with some embodiments of the invention.
Detailed Description of Some Embodiments of the Invention

The following description focuses on embodiments of the invention applicable to a news article distribution system using the Internet. However, it will be appreciated that the invention is not limited to this application but may be applied to many other distribution systems and communication networks/system including for example a distribution system for music or video clips.

FIG. 1 illustrates an example of a content item recommendation system supported by the Internet. The system comprises a central recommendation server 101 which supports a plurality of remote user devices 103 coupled to the central recommendation server 101 via a communication system/network which in the specific example is the Internet 105. The remote user devices 103 may for example include personal consumer devices such as cell phones, personal digital assistants, set-top boxes, personal computers, etc.

FIG. 2 illustrates elements of the central recommendation server 101 of FIG. 1. The central recommendation server 101 is capable of distributing news items (such as text documents or video clips) to a large number of users supported by the user devices 103. Thus, each of the user devices 103 may comprise functionality for receiving and presenting the received news items. Furthermore, the user devices 103 comprise functionality for generating user rating feedback for the received news items. For example, a user of a user device 103 may manually enter a rating for a given news item. As another example, the user rating may be generated automatically, e.g. as a measure of the proportion of e.g. a news article which has been read.

The user devices are arranged to transmit user rating feedback back to the central recommendation server 101 and in response the central recommendation server 101 can generate and maintain a user preference profile for each supported user. As another example, a user may manually provide a rating/preference value for different news categories and topics and this may be transmitted to the recommendation server 101 to be used in a user preference profile for the user.

The recommendation server 101 comprises functionality for distributing news items in the network and specifically comprises functionality for distributing the items
in accordance with the user preferences while at the same time allowing a broad and automatic dissemination of headline news items which may be of interest to many users despite not matching their specific preference profile.

The central recommendation server comprises a cluster processor 201 which is operable to generate a hierarchical user cluster structure of user clusters of the remote user devices. In the hierarchical user cluster structure each hierarchical level comprises the remote user devices divided into a number of clusters referred to as user clusters. The user clusters are at a given level formed by combining user clusters from the next lower level, known as agglomerative clustering. Thus, user clusters at a given hierarchical level can be formed by merging user clusters at the lower level or equivalently can be created by further dividing user clusters of the next higher level into smaller user clusters, known as divisive clustering.

It will be appreciated that the number of user clusters of a lower level that are combined into a user cluster at a higher level may depend on the preferences and characteristics of the individual embodiment and that user clusters of a given level does not need to comprise the same number of user clusters from the lower level. As a specific example, if each user cluster is formed by merging two user clusters of the next lower level a hierarchal structure corresponding to a binary bifurcation tree is obtained.

The cluster processor 201 is coupled to a cluster store 203 in which the hierarchical user cluster structure is stored.

The cluster store 203 is furthermore coupled to an initialisation processor 205 which is operable to initiate a distribution process for a first content item. In particular, when a new news item is received for distribution, the initialisation processor 205 determines a first user cluster at the lowest hierarchal level of the hierarchical user cluster structure. The initial cluster is determined as the cluster for which the user preference profiles most closely match the characteristics of the news item. Specifically, the item characteristics may be matched to the centroid of the user preferences for the given cluster.
The initialisation processor 205 is coupled to a sample processor 207 which is operable to select a subset of the user devices of the first user cluster to be included in a sample group. This sample processor 207 is coupled to a network interface 209 which interfaces the recommendation server 101 to the Internet 105. The sample processor 207 proceeds to recommend the first content item to the sample group. The recommendation may for example be performed by transmitting the news item itself to the user devices or may e.g. be achieved by transmitting an identification of the news item allowing the individual user devices 103 to retrieve the news item from a different source e.g. by accessing a specific web site address of the Internet.

The user devices 103 then generate user ratings for the news item and transmit these back to the recommendation server 101. For example, in a simple embodiment, whenever the news item is selected by a user of a specific user device 103 the user device 103 transmits an indication of the selection to the recommendation server 101. Thus, the user rating feedback from the sample group can simply consist in indications of whether the news item was selected by the users. In other embodiments, more complex user rating feedback may be used. E.g. the users may provide a manual and detailed preference indication for the news item.

The user ratings are received by the network interface 209 and fed to a user rating processor 211 coupled to the network interface 209 and the sample processor 207.

The user rating processor 211 is furthermore coupled to a propagation processor 213 which evaluates the user rating feedback to see if it meets a popularity criterion for the news item. It will be appreciated that any suitable criterion can be used. As a simple example, the criterion may be considered to be met if a sufficient proportion of users in the sample group have selected the news item within a given time interval. As another example, the criterion may be considered to be met if a sufficient proportion of users in the sample group have rated the news item higher than a given threshold.

In the specific example, if the popularity criterion is not met the distribution of the news item terminates and no further distribution is performed. In some
embodiments, the distribution may be continued for first user cluster using collaborative filtering.

However, if the criterion is met the central recommendation server 101 proceeds to recommend the content item to remote user devices of a second user cluster where the second user cluster is at a higher hierarchical level of the hierarchical user cluster structure than the hierarchical level of the initial user cluster. The second user cluster specifically comprises the initial user cluster as well as at least one other user cluster.

Specifically, the propagation processor 213 is coupled to the sample processor 207 which is provided with an indication that the popularity criterion has been met. In response, the sample processor 207 identifies the next higher level cluster comprising the initial user cluster. It then selects a second sample group of user devices in this broader user cluster. The new site is then recommended to the user devices of the broader user cluster and the described process is repeated for the second user cluster.

In addition, the propagation processor 213 is coupled to a collaborative filtering processor 215 which can recommend the first news item to remote user devices of the first user cluster which were not included in the first sample group. Thus, the collaborative filtering processor 215 can expand the recommendations to user devices 103 within the first cluster which were not included in the original sample group thereby distributing the news item to all interested users in the first cluster.

The distribution of the news item in the first cluster is based on a collaborative filtering process using the received user rating feedback and a comparison of the user preferences for the individual user and the user preferences for the users providing the feedback. The collaborative filtering process will be described in more detail later.

Thus, the recommendation server 101 comprises functionality for automatically propagating a content item within a network of user devices. The content item is initially distributed to users which are highly likely to be interested in the news item. The distribution is then automatically increased to user devices with less predicted interest in the content item. The distribution is automatically increased
in ever expanding circles until the interest for the content item falls below a given level. Thus, the expansion of the distribution is determined by the current user ratings for the specific content item thereby providing a distribution which matches the current preferences in the system. This may allow headline news items which tend to be of high interest even if not matching the users’ long term preference profiles to be broadly distributed in the system. At the same time, the initialisation of the process is controlled by the stored user preference profiles thereby preventing that non-headline news items (i.e. news items with interest only to limited groups of users) are broadly distributed. Thus, the system may automatically provide a broad propagation of general interest content items and a more narrow propagation of specific interest content items.

FIG. 3 illustrates an example of a method of content distribution in accordance with some embodiments of the invention. The method is applicable to the recommendation server 101 of FIG. 1 and 2 and will be described with reference thereto.

The method initiates in step 301 wherein the cluster processor 201 generates the hierarchical user cluster structure. The hierarchical user cluster structure is generated in response to user preference profiles associated with the remote user devices. Specifically, each user may manually create a user preference profile and transmit it to the recommendation server 101.

It will be appreciated that many different techniques for creating and maintaining user preference profiles are known to the skilled person and that any suitable method may be used without detracting from the invention.

In the specific example, user preferences are represented as a vector (with binary or positive real value entries), with each term representing the strength of interest in a particular topic. In order to simplify the manual entry of user preferences, topic categories are represented as a taxonomy e.g. as indicated by FIG. 4. A preference for a non-leaf node topic implies a preference for the descendents. The preference vector represents the user interests in the leaf nodes of the topic taxonomy.
Additionally user preferences can be augmented by preferences for a fixed set of predefined keywords (e.g. names of sports teams or entertainers).

In step 301, the cluster processor 201 thus uses these user preference vectors to create a hierarchical multi-level clustering of users. Users can specifically be clustered according to their explicit and/or implicit topic preferences where explicit preferences refer to user preferences that are entered manually and implicit preferences are derived from the history of previously consumed news items.

The clustering is performed such that similar user preference profiles are clustered together. Specifically, standard clustering techniques, such as k-Means, are applied to cluster the preference vectors.

In the specific example of FIG. 3, the hierarchical user cluster structure is generated by agglomerative clustering of the user preferences. Initially, all user preference profiles/vectors are clustered into N clusters using a K-means clustering process. Thus, a first set of user clusters are generated at the lowest hierarchical level of the hierarchical user cluster structure.

Secondly, a common user preference vector is then calculated for each of the N clusters. For example, the common user preference vector may be a centroid vector of the user preference vectors of the cluster members. In the case where the clusters are generated by k-means clustering, such a vector is generated as a by-product of the clustering algorithm.

Next, the common user preference vectors for the N generated user clusters are clustered into M (M<N) clusters using a K-means clustering process. Thus, this step generates M clusters where each cluster comprises one or more of the N clusters. The M clusters accordingly form the second hierarchical level.

This clustering process may be iterated until a sufficient number of hierarchical levels have been generated for the hierarchical user cluster structure. Each cluster of a given hierarchical level thus comprises one or more user clusters from the next lower hierarchical level with at least some and typically all of them comprising a plurality of clusters from the lower level. Thus, equivalently, clusters at
a lower hierarchical level are subsets of the corresponding clusters at a higher hierarchical level.

As the cluster sizes increase for higher hierarchical levels, the common user preferences (centroids) for a cluster at a lower level will tend to be more specific than common user preferences for user clusters of higher hierarchical levels. Specifically, the correlation or similarity between users tends to decrease for increasing sizes of the clusters.

Step 301 is followed by step 303 wherein the initialisation processor 205 determines an initial user cluster for a first news item which is to be distributed in the system. It will be appreciated that in some scenarios or embodiments, more than one cluster may be selected and the content item may be sent to sample groups of more than one cluster.

Specifically, the initialisation processor 205 can compare characteristics of the news item to user cluster preference profiles of the lowest hierarchical level of the hierarchical user cluster structure. For each cluster, a preference value is determined which indicates the predicted interest of users in the user cluster for the news item. The initial cluster (or clusters) for the distribution of the news item is then selected as the user cluster having the highest preference value.

In more detail, an initial user cluster is selected as the cluster having user profiles that best match metadata characterising the news item. An efficient method for matching the topic preferences is to compute a centroid for each user cluster and measure the relevance of the news items to the cluster based on a comparison of news item metadata and centroid characteristics. The cluster centroids may in the specific example be calculated as part of the clustering process. Matching of keyword preferences can be carried out by counting the frequency or relative frequency of keywords of all user profiles in the cluster, and comparing this to a keyword list in the news item metadata.

In some embodiments, the initialisation processor 205 may always select the initial user cluster as a user cluster of the lowest hierarchical level. However, in other embodiments the initialisation processor 205 may select the initial hierarchical level
in response to a user preference indication for user clusters of different hierarchical levels.

For example, if the preference value for the news item and a centroid of the user cluster at a higher hierarchical level is above a threshold, this may indicate that the news item is predicted to be of high interest to all users in that user cluster. Accordingly, the recommendation server 101 may introduce the news item at a higher hierarchical level and thus may introduce the news item to a larger group of users. This may allow faster propagation of content items likely to be of specific interest to a large number of users.

Step 303 is followed by step 305 when the sample processor 207 recommends the news item to a sample group of the selected cluster. The sample group selection may be carried out randomly and/or standard statistical sampling techniques can be applied.

The news item can then be delivered to the selected user devices as priority items (e.g. they can be included in a list of recommended news items sent to the user devices 103 of the sample group).

Step 305 is followed by step 307 wherein the user rating processor 211 collects user rating feedback for the news item from the user devices 103.

Thus, the user ratings for the news item are collated at the recommendation server 101. The user ratings can e.g. be in the form of:

- Explicit ratings requested from or given by the user.
- Implicit ratings obtained by monitoring links in the content that the user clicks on.
- Implicit ratings obtained from time viewing the items and the proportion of the news item which has been accessed by a user.
User studies have shown that the implicit indicators listed above are highly correlated with user preferences for news items. Explicit ratings can be provided by requesting user feedback as items are consumed.

Step 307 is followed by step 309 wherein the collaborative filtering processor 215 proceeds to recommend the news item to user devices of the first user cluster which were not included in the sample group. In this example, user-to-user based collaborative filtering is used to distribute the news item to the remaining user devices of the first cluster.

The basic object of collaborative filtering is to provide predictions or recommendations based on the opinions of other like-minded users. A collaborative filtering algorithm generally relies on user preference information and uses a similarity analysis algorithm to identify similar users (user-to-user collaborative filtering) or content items (item-to-item collaborative filtering). The similarity analysis is then used to generate groups of users/ content items with ratings similar to the user/content item for which the recommendation is generated. In this way a recommendation can be generated based on the combined preferences from the subgroup of users/items.

In the specific example, user-to-user collaborative filtering is used to determine which user devices the news item is distributed to.

As a simple example, the user preference of a first user device belonging to the first cluster but not to the sample group may be compared to the user preferences of all the users of the sample group. The user device of the sample group which has the closest matching user preference is then identified and the user ratings for the news item provided by that user device may be considered to also apply to the first user device. Thus, if the matching sample group user has rated the news item sufficiently high, the recommendation server will recommend the news item to the first user device and otherwise it will not recommend it to the first user device.

Hence, if a matching user of the sample group has rated the news item highly, this will indicate that the user of the first user device also has a high preference for the second content item whereas a low rating may indicate that the user of the first user
device has a low preference for the news item. In the former case, the second content item is likely to be recommended to the user whereas it is likely to be discarded in the latter case.

It will be appreciated that different and more complex correlation algorithms based on user-to-user correlation data will be known to the skilled person and may be used without detracting from the invention. In such more complex algorithms the user ratings of the sample group are used as seeds to initialise the collaborative filtering.

Step 309 is followed by step 311 wherein the propagation processor 213 determines whether the user rating feedback meets a popularity criterion for the first content item.

The criterion may for example require that an average user rating provided by the sample group exceeds a predetermined threshold which may e.g. be determined based on a training algorithm.

The evaluation of the ratings can be performed based on a time limit and/or a requirement that a minimum proportion of the sampled users provide ratings.

For example, the criterion can require that a preference criterion must be met within a time constraint. E.g. if the average user rating has not risen above the predetermined threshold within a given time from the initial recommendation to the sample group, the popularity criterion is considered not to be met.

Also, the criterion may require that a number of received user ratings for the news item exceeds a threshold. If an insufficient number of user ratings have been received, the user feedback is considered to be insufficiently reliable to have indicated a strong preference for the news item.

If the popularity criterion is not met, the method continues in step 313 wherein the process terminates and thus the news item only propagates to user devices of user clusters of the hierarchical level at which the process terminates. In some embodiments, the collaborative filtering and subsequent recommendation of the news item within the current cluster may be continued following step 311.
It will be appreciated that in the example of FIG. 3, the news item is distributed to all users of the first cluster which are predicted to have an interest in the item regardless of whether the user rating from the sample group indicated a high preference for the item or not. In other words, the collaborative filtering is performed regardless of whether the user rating feedback meets the popularity criterion.

In other embodiments, the distribution to the whole of the user cluster may only be performed if the sample group provides user ratings that meet a popularity criterion (which may be the same criterion that is used to determine whether the item should be propagated to a cluster of a higher hierarchical level).

If the popularity criterion is met, the method continues in step 315 wherein the next higher hierarchical level is selected. The user cluster at this hierarchical level which comprises the first user cluster is selected and the method returns to step 305. The process is then iterated for the broader and more general cluster of the higher hierarchical level. Thus, if the news item received sufficient interest within the first user cluster the distribution of the news item is extended to a second broader user cluster at a higher hierarchical level. Accordingly, distribution of the news item is expanded to a larger user group.

Specifically, the sample processor 207 proceeds to select a sample of user devices from the second user cluster. The user devices are selected from the user devices of the second user cluster that were not included in the first user cluster, i.e. they are selected from the new additional user devices considered to have potential interest in the news item.

The news item is then recommended to the user devices of this second sample group and user rating feedback from this group is collected. The recommendation server then proceeds to use collaborative filtering to distribute the news item further in the second cluster. It furthermore evaluates if the user rating feedback meets a popularity criterion. If not, the distribution process terminates and otherwise the next higher hierarchical level is selected. The process is then repeated for a user cluster of this higher level.
Thus, the central recommendation server is arranged to iterate the process of recommending the first content item to user clusters at increasingly higher hierarchical levels until a popularity criterion for a given hierarchical level is not met. Thus, a news item will automatically propagate through the system in increasing circles until the interest falls below a given level. This approach may allow news items of specific interest to be distributed only to the user likely to have the specific interest whereas news items of more general interest will be distributed to larger groups of users even if the current user profile for these users do not indicate a preference for the news item. Thus, the system provides an efficient, automated and dynamic process for propagating a content item to an appropriate degree. For example, the system may allow a wide dissemination of content items of general current interest while ensuring that content items of specific interest are less widely dissemination. Thus, the system may reduce the communication resource required for dissemination of content items while ensuring sufficient distribution of content items of general interest. The system may further reduce the inconvenience of the user having to manually filter and select specific news items.

In some embodiments, the popularity criterion can depend on the hierarchical level for which it is applied. For example, the threshold that must be exceeded by the average user rating from the sample group may decrease for higher levels of the hierarchy reflecting that the preference is likely to decrease as the distribution is increased to user clusters having lower average preferences for the specific characteristics of the news item.

FIG. 5 illustrates an example of the propagation of popular news item through a hierarchical user cluster structure. The clustering of users is represented from left to right in increasing size, reflecting the iterations of the process of FIG. 3. In the example, a popular news items is propagated throughout all user clusters of a hierarchical user cluster structure comprising five hierarchical levels. Thus, the news item will be available to be propagated using collaborative filtering for all users in the given population. The stars indicate respectively sample users who receive the given content item as a sample, users who belong to a cluster which has rated the news item
above the threshold, users who have received the content item based on recommendation of other users via collaborative filtering and not by sampling.

After identifying the most appropriate cluster(s), a sample set of users is selected to receive the current article in the first hierarchical level 501. If the users' ratings are higher than the threshold the sampling process is continued to the next level 503 with a different set of users from the ones selected for the previous level 501. This process is repeated for the higher levels 505.

In some embodiments, multiple sampling and thresholding stages are adopted within each level. For example, the news item is distributed to a small sampling group of a user cluster. If the user ratings of this group meet a popularity criterion the sampling group is enlarged and the user ratings for the entire sampling group are used to determine whether the news item should propagate to the next higher level. This may ensure that the sampling process is terminated early for news items of little or no interest to the users while ensuring a sufficient reliability for the determination of whether to move to the next hierarchical level. It will be appreciated that more than two iterations can be implemented.

It will be appreciated that the above description for clarity has described embodiments of the invention with reference to different functional units and processors. However, it will be apparent that any suitable distribution of functionality between different functional units or processors may be used without detracting from the invention. For example, functionality illustrated to be performed by separate processors or controllers may be performed by the same processor or controllers. Hence, references to specific functional units are only to be seen as references to suitable means for providing the described functionality rather than indicative of a strict logical or physical structure or organization.

The invention can be implemented in any suitable form including hardware, software, firmware or any combination of these. The invention may optionally be implemented at least partly as computer software running on one or more data processors and/or digital signal processors. The elements and components of an embodiment of the invention may be physically, functionally and logically
implemented in any suitable way. Indeed the functionality may be implemented in a
single unit, in a plurality of units or as part of other functional units. As such, the
invention may be implemented in a single unit or may be physically and functionally
distributed between different units and processors.

Although the present invention has been described in connection with some embodiments, it is not intended to be limited to the specific form set forth herein. Rather, the scope of the present invention is limited only by the accompanying claims. Additionally, although a feature may appear to be described in connection with particular embodiments, one skilled in the art would recognize that various features of the described embodiments may be combined in accordance with the invention. In the claims, the term comprising does not exclude the presence of other elements or steps.

Furthermore, although individually listed, a plurality of means, elements or method steps may be implemented by e.g. a single unit or processor. Additionally, although individual features may be included in different claims, these may possibly be advantageously combined, and the inclusion in different claims does not imply that a combination of features is not feasible and/or advantageous. Also the inclusion of a feature in one category of claims does not imply a limitation to this category but rather indicates that the feature is equally applicable to other claim categories as appropriate. Furthermore, the order of features in the claims does not imply any specific order in which the features must be worked and in particular the order of individual steps in a method claim does not imply that the steps must be performed in this order. Rather, the steps may be performed in any suitable order.
CLAIMS

1. A content item distribution system comprising a central recommendation server recommending content items to a plurality of remote user devices, the central recommendation server comprising:

   means for providing a hierarchical user cluster structure of user clusters of the remote user devices, the user clusters at a higher hierarchical level comprising one or more user clusters of a next lower hierarchical level;

   cluster determining means for determining a first user cluster for a first content item, the first user cluster being at a first hierarchical level;

   sample means for recommending the first content item to a first sample group of remote user devices of the first user cluster;

   means for receiving user rating feedback from the first sample group;

   propagation means for determining if the user rating feedback meets a popularity criterion for the first content item and if so recommending the content item to remote user devices of a second user cluster, the second user cluster being a user cluster of a higher hierarchical level of the hierarchical user cluster structure than the first hierarchical level and comprising the first user cluster and at least one other user cluster of the first hierarchical level.

2. The content distribution system of claim 1, wherein the central recommendation server further comprises recommending means for recommending the first user content item to remote user devices of the first user cluster not comprised in the first sample group.

3. The content distribution system of claim 2, wherein the recommending means is arranged to recommend the first content item in response to a collaborative filtering based on the user rating feedback.

4. The content distribution system of claim 3, wherein the collaborative filtering is a user-to-user collaborative filtering.
5. The content distribution system of claim 2, wherein the recommending means is arranged to recommend the first content item to a first remote user device in response to a comparison of a user preference profile for the first remote user device and at least one user preference profile associated with the user rating feedback.

6. The content distribution system of claim 1, wherein the propagation means is arranged to distribute the first content item to a second sample group of remote user devices of the second user cluster, the remote user devices of the second sample group not being comprised in the first user cluster.

7. The content distribution system of claim 6, wherein the central recommendation server further comprises:

   means for receiving second user rating feedback from the second sample group; and

   means for determining if the second user rating feedback meets a second popularity criterion for the first content item and if so recommending the first content item to remote user devices of a third user cluster, the third user cluster being a user cluster of a higher hierarchical level of the hierarchical user cluster structure than the second user cluster and comprising the second user cluster and at least one other user cluster of the hierarchical level of the second user cluster.

8. The content distribution system of claim 6, wherein the central recommendation server is arranged to select remote user devices for the second sample group from remote user devices belonging to the second user cluster but not the first user cluster.

9. The content distribution system of claim 6, wherein the central recommendation server further comprises means for recommending the first user content item to remote user devices of the second cluster not comprised in the first cluster or the second sample group.

10. The content distribution system of claim 1, wherein the central recommendation server comprises user cluster means for generating the hierarchical user cluster structure in response to user preference profiles associated with the remote user devices.
11. The content distribution system of claim 10, wherein the user cluster means comprises:
   means for performing a first user clustering in response to user preference profiles for users of the remote user devices to generate a first set of user clusters at a first hierarchical level;
   means for determining a common user preference profile for each user cluster of the first set of user clusters;
   means for performing a second user clustering in response to the common user preference profiles to generate a second set of user clusters at a second hierarchical level, the second hierarchical level being a next higher hierarchical level for the first hierarchical level.

12. The content distribution system of claim 1, wherein the cluster determining means is arranged to determine a preference value for each cluster of the first hierarchical level in response to a comparison between a cluster user preference profile for each user cluster and the first content item; and to select the first user cluster as a user cluster of the first hierarchical level having the highest preference value.

13. The content distribution system of claim 1, wherein the popularity criterion comprises a requirement that a preference criterion must be met within a time constraint.

14. The content distribution system of claim 1, wherein the popularity criterion comprises a requirement that a number of received user ratings for the first content item must exceed a threshold.

15. The content distribution system of claim 1, wherein user preferences for user clusters of lower hierarchical levels are more specific than user preferences for user clusters of higher hierarchical levels.

16. The content distribution system of claim 1, wherein the central recommendation server comprises means for selecting an initial hierarchical level for an initial recommendation of the first content item in response to a user preference indication for user clusters of different hierarchical levels.
17. The content distribution system of claim 1, wherein the central recommendation server is arranged to iterate the process of recommending the first content item to user clusters at increasingly higher hierarchical levels until a popularity criterion for a hierarchical level is not met.

18. The content distribution system of claim 17, wherein the popularity criterion depends on the hierarchical level.
19. A central recommendation server for a content item distribution system, the central recommendation server comprising:

   means for providing a hierarchical user cluster structure of user clusters of the remote user devices, the user clusters at a higher hierarchical level comprising one or more user clusters of a next lower hierarchical level;

   cluster determining means for determining a first user cluster for a first content item, the first user cluster being at a first hierarchical level;

   sample means for recommending the first content item to a first sample group of remote user devices of the first user cluster;

   means for receiving user rating feedback from the first sample group; and

   propagation means for determining if the user rating feedback meets a popularity criterion for the first content item and if so recommending the content item to remote user devices of a second user cluster, the second user cluster being a user cluster of a higher hierarchical level of the hierarchical user cluster structure than the first hierarchical level and comprising the first user cluster and at least one other user cluster of the first hierarchical level.
20. A method of content item distribution for a content item distribution system comprising a central recommendation server recommending content items to a plurality of remote user devices, the method comprising the central recommendation server performing the steps of:

- providing a hierarchical user cluster structure of user clusters of the remote user devices, the user clusters at a higher hierarchical level comprising one or more user clusters of a next lower hierarchical level;

- determining a first user cluster for a first content item, the first user cluster being at a first hierarchical level;

- recommending the first content item to a first sample group of remote user devices of the first user cluster;

- receiving user rating feedback from the first sample group; and

- determining if the user rating feedback meets a popularity criterion for the first content item and if so recommending the content item to remote user devices of a second user cluster, the second user cluster being a user cluster of a higher hierarchical level of the hierarchical user cluster structure than the first hierarchical level and comprising the first user cluster and at least one other user cluster of the first hierarchical level.
FIG. 2
CREATE HIERARCHICAL USER CLUSTER STRUCTURE
SELECT LOWEST LEVEL USER CLUSTER
SEND CONTENT ITEM TO SAMPLE USERS IN CURRENT CLUSTER
COLLECT USER RATINGS
PERFORM COLLABORATIVE FILTERING
DO RATINGS MEET CRITERION?
STOP

SELECT CLUSTER IN NEXT HIGHER LEVEL
FIG. 4
A. CLASSIFICATION OF SUBJECT MATTER

INV. G06Q30/00

According to International Patent Classification (IPC) or to both national classification and IPC.

B. RELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>The technical aspects identified in the present application (Art. 15 PCT) are considered part of common general knowledge. Due to their notoriety no documentary evidence is found to be required. For further details see the accompanying Opinion and the reference below. XP002456414</td>
<td>1-20</td>
</tr>
</tbody>
</table>

D. See patent family annex.

Date of the actual completion of the international search
9 September 2008

Date of mailing of the international search report
16/09/2008

Name and mailing address of the ISA/ European Patent Office, P.O. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016

Authorized officer
CTrstet, Andrei