

(21) Application No: 0624552.6  
(22) Date of Filing: 09.12.2006

(71) Applicant(s):  
**Motorola Inc**  
(Incorporated in USA - Delaware)  
1301 East Algonquin Road, Schaumburg,  
Illinois 60196, United States of America

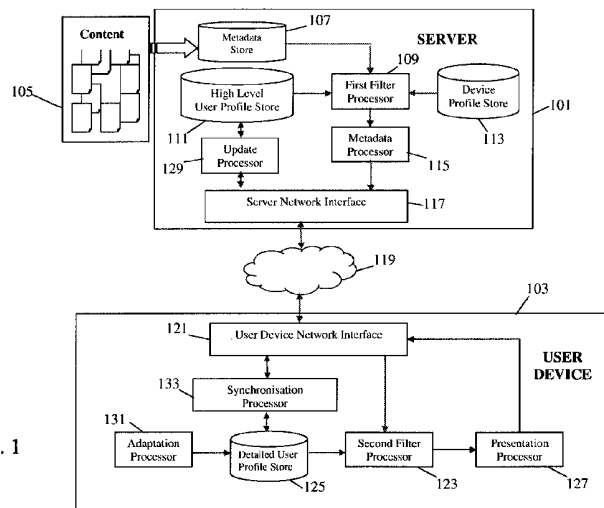
(72) Inventor(s):  
**Simon Waddington**  
**Ben Bratu**  
**Ioannis Kompatsiaris**  
**Evangelia Nidelkou**  
**Maria Papadogiorgaki**  
**Vasileios Papastathis**

(74) Agent and/or Address for Service:  
**Eltima Consulting**  
Grove House, Lutyens Close,  
Chineham Court, BASINGSTOKE,  
Hampshire, RG24 8EG, United Kingdom

(51) INT CL:  
**G06F 17/30** (2006.01) **G06Q 30/00** (2006.01)  
(56) Documents Cited:  
**EP 1484692 A1** **WO 1999/028837 A1**  
(58) Field of Search:  
INT CL **G06Q**  
Other: **Online: EPODOC, FULLTEXT, WPI**

(54) Abstract Title: **A content recommendation system**

(57) A content recommendation system comprises a server 101 and a user device 103. The server 101 comprises a first filter processor 109 filtering a plurality of content items to generate a first subset of content items. The filtering comprises selecting content items in response to characterising data for the content items and data of a first user profile for a first user. A metadata processor 115 then generates subset characterising data for the first subset which is transmitted to the user device 103. This comprises a second filter processor 123 which filters the first subset to generate a second subset of content items. The filtering comprises selecting a content item from the first subset in response to the subset characterising data and a second user profile for the first user. The second user profile is more detailed than the first user profile. Improved privacy and security may be achieved as the user profile is distributed over different devices.



**A CONTENT RECOMMENDATION SYSTEM AND A METHOD OF OPERATION  
THEREFOR**

**5 Field of the invention**

The invention relates to a content recommendation system and a method of operation therefor and in particular, but not exclusively, to a system for recommendation of text  
10 documents.

**Background of the Invention**

15 In recent years, the availability and provision of multimedia and entertainment content has increased substantially. For example, the number of available television and radio channels has grown considerably and the popularity of the Internet has provided new content  
20 distribution means. 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  
25 impractical.

Also the availability of electronic text documents has increased explosively with many user being provided with daily access to numerous text files such as emails, web  
30 pages, downloadable documents etc.

Similarly, an increasing number of services and applications with many different options and customisation features are becoming available to the user.

- 5 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.
- 10 In order to enhance the user experience, it is advantageous to personalise the recommendations to the individual user as much as is possible. In this context, a recommendation consists in predicting how much a user may like a particular content item and recommending it if it is considered of
- 15 sufficient interest. The process of generating recommendations requires that user preferences have been captured so that they can be used as input data by a prediction algorithm.
- 20 For example, recommendation systems for evaluating, categorising and recommending text documents are receiving significant interest. Such systems may for example retrieve large number of online text documents and web pages and compare them to a user's recorded preferences in order to
- 25 generate recommendations for text documents of particular interest to the user.

- Also, people increasingly use a wide range of electronic devices for different purposes and with different
- 30 capabilities (e.g. cell phone, PDA, MP3 players, set-top boxes, personal computers, etc.). It is accordingly becoming increasingly important to provide personalised user

experiences for e.g. portable devices which typically have relatively low capabilities in terms of communication resources, computational resources etc.

- 5 Proposals have been made for centrally generating specific recommendations and transmitting these to (portable) user devices for presentation to the user. However, using such a server based approach comprises a number of disadvantages.
- 10 For example, the approach requires that a detailed user profile for each user is stored and maintained centrally. This is difficult to achieve and typically requires large degrees of communication between user device and server in order to continually maintain and adapt the user profile to
- 15 the user's preferences and behaviour. Also, a centralised approach tends to require complex and expensive servers. Furthermore, re-use of preference data or recommendations between different user applications is difficult as different servers typically are used for different purposes
- 20 and content items. In addition, a centralised storage of complex and detailed user profiles provides a less secure system with an increased privacy risk for the individual user.
- 25 Accordingly, it has been proposed to implement the recommendation and content filtering in the individual user device. However, this also has a number of associated disadvantages.
- 30 For example, the available storage and computational resource is typically severely limited resulting in a restriction in the complexity of the recommendation

algorithm and user profile data leading to a reduced quality of the recommendations. Furthermore, in order to provide the user device with information of all available content items, a large amount of data will need to be transmitted to each  
5 user device thereby increasing communication resource usage, delay and costs. Indeed as ever more content is made available for consumption on e.g. mobile handsets, a handset driven solution is not feasible as the volume of available content typically far exceeds what reasonably can be  
10 transmitted to and processed by the handset. The additional processing also increases power consumption which is critical for battery driven devices.

Hence, an improved recommendation system would be  
15 advantageous and in particular a system allowing increased flexibility, facilitated operation, improved performance, reduced device resource requirements, reduced communication resource requirements, facilitated implementation and/or improved recommendations would be advantageous.

20

### **Summary of the Invention**

25 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  
30 provided a content recommendation system, comprising a recommendation server comprising: means for providing characterising data for a plurality of content items, first

filtering means for filtering the plurality of content items to generate a first subset of content items, the filtering comprising selecting content items from the plurality of content items for the first subset in response to the

5 characterising data and data of a first user profile for a first user, means for generating subset characterising data for the first subset of content items from the characterising data, and a transmitter for transmitting the subset characterising data to a user device; and

10 the user device comprising: a receiver for receiving the subset characterising data, and second filtering means for filtering the first subset of content items to generate a second subset of content items, the filtering comprising selecting at least one content item from the first subset

15 for the second subset in response to the subset characterising data and data of a second user profile for the first user, the second user profile being more detailed than the first user profile.

20 The invention may allow an improved recommendation system. In particular, a highly accurate recommendation performance can be achieved based on specific and detailed information localised at the individual user device while at the same time maintaining low resource requirements. For example, a

25 low computational and/or communication resource requirement can be achieved. The user device may apply a highly accurate user profile to generate accurate recommendations but need only evaluate a small subset of the content items which specifically may be selected as content items which are

30 highly likely to be of interest to the user. Also, the complexity of the recommendation server may be reduced in many embodiments. In addition, user profile management,

adaptation and/or updating may be substantially simplified as only more general information need to be stored centrally. The approach may facilitate re-use of user preference information by different user applications of the  
5 user device. Improved privacy and security may be achieved as the user profile is distributed over different devices.

The recommendation server may serve a plurality or multiplicity of user devices. The data of the first and/or  
10 second user profile may comprise community data. Thus the data may include user preference data relating to a group of users as well as or rather than an individual user. In some embodiments, the filtering of content items into subsets may be iterated a number of times by the recommendation server,  
15 one or more user devices or both.

The first user profile may be a high level user profile and the second user profile may be a low level user profile. The second user profile is more detailed than the first user  
20 profile so that it may provide differentiation between content items for which the first user profile cannot provide differentiation. The second user profile may be more detailed than the first user profile in that it comprises separate data for content items for which the first user  
25 preference comprises only common data.

According to an optional feature of the invention, the first user profile comprises a first set of categories for content items and the second user profile comprises a division of at  
30 least one of the first set of categories into further subcategories.

This may allow improved and/or simplified operation, improved recommendation performance and/or facilitated development and/or implementation. The second user profile may be more detailed than the first user profile by  
5 comprising characterising data which belongs to different subcategories in the second user profile but which belongs to the same category in the first user profile.

According to a second aspect of the invention, there is  
10 provided: a method of operation for a content recommendation system including a recommendation server and a user device; the method comprising: the recommendation server performing the steps of: providing characterising data for a plurality of content items, filtering the plurality of content items  
15 to generate a first subset of content items, the filtering comprising selecting content items from the plurality of content items for the first subset in response to the characterising data and data of a first user profile for a first user, generating subset characterising data for the  
20 first subset of content items from the characterising data, and transmitting the subset characterising data to a user device; and the user device performing the steps of: receiving the subset characterising data, and filtering the first subset of content items to generate a second subset of  
25 content items, the filtering comprising selecting at least one content item from the first subset for the second subset in response to the subset characterising data and data of a second user profile for the first user, the second user profile being more detailed than the first user profile.

30



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

5

### **Brief Description of the Drawings**

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

10

FIG. 1 is an illustration of an example of a recommendation system in accordance with some embodiments of the invention; and

15 FIG. 2 is an illustration of an example of a method of operation for a content recommendation system in accordance with some embodiments of the invention.

### **20 Detailed Description of Some Embodiments of the Invention**

The following description focuses on embodiments of the invention applicable to a recommendation system for recommending text documents such as online text files.

25 However, it will be appreciated that the invention is not limited to this application but may be applied to many other scenarios and applications. In particular, the described principles may be applied to other multimedia content including text, video and audio. For example, non-textual  
30 content can be provided with textual annotation or other data characterising the content items.

The described recommendation system utilises a distributed client-server architecture for personalised filtering of content.

5 In the system, high-level filtering of available content is performed by the server followed by more accurate matching of detailed user preferences by a client application of a user device such as a portable device, mobile phone or set top box. Specifically, the system employs a user profile for  
10 a user which is distributed between the server and the user device client. The high-level user preferences are stored in a high-level user profile on the server and the low-level preferences are stored in a detailed user profile in the user device.

15

The described embodiments specifically focus on the use of NLP (Natural Language Processing) and text classification methods which perform high-level classification of content items according to a given taxonomy of topic categories. A  
20 given content item can be classified into one or more topic categories. These topic categories are matched with the high-level user preferences by the server. The high-level filtering step can also include selection of items according to a user's preferred content providers, user device  
25 characteristics, communication characteristics etc.

The output of the first filtering step is a short-list of recommended contents items for the user. The distribution of items in the topic categories is determined by the user's  
30 high-level preferences and the relevance of items to each topic category. The number of items remaining after the first filtering step may be considerably larger than the

amount of content the user can consume. The total size of the filtered content set can be limited by the network bandwidth and user device processing capabilities. The user device can then compare the received set of content items to  
5 the locally stored detailed user profile to determine ratings for each content item. The highest rated item(s) may then be recommended to the user.

FIG. 1 is an illustration of an example of a recommendation  
10 system in accordance with some embodiments of the invention. FIG. 2 is an illustration of an example of a method of operation for a content recommendation system in accordance with some embodiments of the invention. FIG. 2 further illustrates data elements used by the method. The method of  
15 FIG. 2 may specifically be used by the system of FIG. 1 and will be described with reference to this scenario.

The recommendation system uses a two-step content filtering process wherein a server 101 performs an initial filtering  
20 based on a user's high-level content preferences (specifically preferences for topic categories). Data for the resulting content items are transmitted to a user device 103 which performs a second stage of low-level filtering of the pre-filtered content based on more detailed user  
25 preferences. The resulting content item(s) may then be retrieved and presented to a user.

The method initiates in step 201 wherein the server 101 retrieves content items from a content store 105 and  
30 generates characterising data for the content items. Specifically, the server 101 generates metadata for each content item which is stored in a metadata store 107.

It will be appreciated that in other embodiments, the content items may e.g. be stored in one or more locations which are external to the server 101; and that the server  
5 101 for example receives metadata for the content items from an external source rather than generate the metadata itself.

In the specific example, the content items are text documents which are analysed to generate metadata describing  
10 the contents of the text document in some way. The text document may for example be a text file or an internet web page (e.g. an HTML document) and may e.g. be newspaper articles, books or any other text.

15 Specifically, the server 101 generates a document term vector for each document. The document term vector comprises a set of terms/keywords found in the text document plus a weight indicating how relevant each of the terms is to the text document.

20

Specifically, the server 101 can detect all nouns in the text document, determine the stem of the noun and generate a set of terms comprising the stems of the nouns which are present in the document. The server 101 then determines  
25 relevance indications for each term in the set and generates a document term vector comprising the identified terms and the associated relevance indications.

Thus, when a document is processed, it is converted to a  
30 vector space model where each element corresponds to a pair of parameters  $(i, w_i)$ , where  $i$  is a unique term identified in

the document and  $w_1$  is its relevance indication for the current document.

The relevance indication can be determined as a numeric  
5 value representing the frequency of its appearance in the given document. For example, the server 101 can determine how many nouns are present in the text and can determine the relevance indication for a noun as the number of times this noun is used divided by the total number of nouns.

10

Alternatively or additionally, the relevance of the terms can also be dependent on the document structure or style. E.g. a term's frequency can be increased by a multiplication factor if the term is found in the title, headings, summary,  
15 and conclusions etc or if it is clearly written in a different style.

The server may alternatively or additionally determine a topic category for the text documents e.g. by comparing  
20 terms of the document with terms known to be associated with specific topics.

The metadata store 107 is coupled to a first filter processor 109 which executes step 203 wherein a plurality of  
25 content items are filtered to generate a first subset (henceforth referred to as the filtered subset) of content items. Specifically, the first filter processor 109 evaluates the characterising data (the metadata) for the content items stored in the metadata store 107 with  
30 reference to a user profile in order to generate a subset comprising of a suitable number of content items.

Accordingly, the first filter processor 109 is coupled to a high level user profile 111. The first filter processor 109 retrieves the user profile for the appropriate user. It then proceeds to compare the metadata for each content item with  
5 the user profile and to select a number of content items to include in the filtered subset based on this comparison.

In the example, the user profile comprises a number of categories with a user preference value assigned to each  
10 category. The user preference values may for example be manually entered by a user using a suitable interface (e.g. a website provided by the server 101).

The first filter processor 109 may match the metadata for a  
15 content item with one of the categories of the user profile. The preference value of the matching category may then be assigned to the content item. The first filter processor 109 may then proceed to select the content items for the filtered subset depending on the assigned preference value.

20 For example, the first filter processor 109 may select a given number,  $N$ , of content items as the content items which have the highest assigned preference values. Specifically, the first filter processor 109 may select  $N$  content items by  
25 selecting content items from each topic category such that the number of content items from each category is in proportion to the user preferences for that category. Within each category, a simple method of selecting can be employed such as selecting the most recent items, the best matching  
30 items, random items etc. As another example, the first filter processor 109 may select a given number of content

items for each category having a preference value above the threshold.

In the case where some content items are classified to  
5 belong to multiple categories, the preference value may be selected to correspond to an average of the individual topic-specific preference values.

The first filter processor 109 may specifically identify a  
10 matching category for a first content item by comparing the metadata for that content item with similar metadata stored in the user profile for each category. For example, the keywords/terms extracted from the first content item may be compared to a set of keywords/terms stored for each category  
15 of the user profile. The matching category may be selected as the category of the user profile having the most keywords in common with the first content item. The user preference value stored for the matching category can then be retrieved and assigned to the first content item. More specifically,  
20 the first filter processor 109 can match the user's weight topic preferences to the topic weights of each document in the content store 105.

As an example, a similarity measure may be determined for  
25 topic categories of the user profile using the document term vector for the content item and the corresponding topic term vectors stored for each topic category.

As a detailed example, the high level user profile may  
30 comprise a number of topic categories where each topic category is represented by characterising data in the form of a topic term vector of terms and their relevance values

for the given topic. The relevance value of a term can be calculated based on the frequency of the term occurring in the document of the topic.

- 5 The determination of topic term vectors for topic categories can specifically be based on training text documents for the different text categories. The documents used for training are typically organized in collections with one collection for each specified topic. There are different corpora of
- 10 pre-classified documents available for training such as the Reuters Corpus  
(<http://about.reuters.com/researchandstandards/corpus/>).

The training system initially detects terms to be included

15 in the topic term vector. Specifically, distinct nouns used in the training documents are detected and arranged to determine the stem of the nouns. Furthermore, it is determined how many times the term is used in the training document and therefrom a suitable relevance value is found.

20 Thus, an initial term vector is generated comprising all unique terms encountered in the collection of training text documents for the category.

As a specific example, each term can be associated with a

25 relevance value calculated by a frequency- inverse document frequency (*tf\_idf*) function given by:

$$w_{i,k} = TF(w_i, k) \times IDF(w_i)$$

30 where  $w_{i,k}$  represents the weight of the term  $i$  in the particular topic category  $k$ . There are different functions



available to compute the term frequency (*tf*) and the inverse document frequency (*idf*), two of them are the following:

$$TF(w_i, k) = \frac{tf_{ik}}{\max\_tf_k}$$

$$IDF(w_i) = \log\left(\frac{N}{n(i)}\right)$$

where  $tf_{ik}$  is the number of occurrences of term  $i$  in topic  $k$ ;  $\max\_tf_k$  is the maximum value scored by a term in topic  $k$ ;  $N$  is the total number of topics; and  $n(i)$  is the number of  
 10 topics in which the term  $i$  is present.

In order to determine *tf\_idf*, a table containing the frequencies of all unique terms found in the given collection of documents is first generated for each topic  
 15 category. Based on the values in the table, the term frequency is then calculated. Finally, using all the generated tables, the inverse document frequency is calculated.

20 In some embodiments, further operations may be performed, such as eliminating the 0-value terms, normalizing the vectors and performing a dimensional reduction of the vectors.

25 When a text document is assigned to a matching category, the first filter processor 109 may first generate the document term vector containing the pairs  $(i, w_i)$  where  $i$  is a unique term identified in the document and  $w_i$  is its relevance for the current document as previously described. Specifically,

the weight of a term can represent the frequency of its occurrence in the given document.

In the specific example, the categorisation or  
 5 classification is based on the similarity measurement between the two vectors. The match indication can specifically be calculated based on a cosine similarity formula, and can be given by:

$$10 \quad \text{sim}(d, T_k) = \frac{\sum_i (w_{i,d} * w_{i,k})}{\sqrt{\sum_i w_{i,d}^2} \sqrt{\sum_i w_{i,k}^2}}$$

where  $w_{i,d}$  are the relevance indications for the text document and  $w_{i,k}$  are the relevance values of the topic term vector for category k.

15

The first filter processor 109 can then select the matching category as the category for which the similarity measure is the highest. It can then proceed to allocate the user preference value of that category to the document.

20

In the specific example, the first filter processor 109 is furthermore coupled to a device profile store 113 and may be arranged to further generate the first subset of content items in response to a device profile for the user device  
 25 103. The device profile comprises data representing the capabilities of the user device.

In the specific example, the device profile for the user device 103 is used by the first filter processor 109 to  
 30 determine the specific characteristics of the user device 103 (e.g. processor speed, memory capacity etc). Based on

the device profile, the first filter processor 109 determines the number of pre-filtered content items that should be included in the filtered subset and thus should be delivered to the client. For example, the number N of  
5 content items to be included in the filtered subset may be selected such that the processing and memory capabilities of the user device 103 are not exceeded.

The device profiles may for example be obtained using pre-  
10 determined look-up tables. Examples of device profiles are the following:

- The OMA UAProf uses the W3C Composite Capability/Preference Profile (CC/PP) model to define a  
15 framework for describing and transmitting information about the client and the used network  
<http://www.openmobilealliance.org/tech/affiliates/wap/wapindex.html>.
- MPEG-21 has a complete specification for device  
20 profiling. This information is represented using XML.  
<http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm>.

25

In some embodiments, the first filter processor 109 may also be arranged to generate the filtered subset of content items in response to a characteristic of a communication link between the server 101 and the user device 103.

30

For example if the communication link between the server 101 and user device 103 has a bandwidth/capacity below a given

threshold, a lower number of content items may be included in the prefiltered subset than if the bandwidth/capacity is above this threshold. Thus, the first filter processor 109 may select the number, N, of content item indications to  
5 transmit to the user device 103 based on a communication characteristic such as e.g. a cost, delay, maximum data rate etc. of the communication link supporting the exchange.

The characteristic may be a specific characteristic for the  
10 connection used to transmit the relevant data or may simply be a characteristic of the communication system or network used for this communication. For example, a higher number of content items may be used if the user device 103 is served by a GPRS network than if it is served by a 3<sup>rd</sup> generation  
15 cellular communication system or wireless local area network.

Step 203 is followed by step 205 wherein a metadata processor 115 coupled to the first filter processor 109  
20 generates subset characterising data for the filtered subset of content items from the characterising data. The subset characterising data may specifically be generated by selecting a subset of the metadata which relates to the content items that are included in the filtered subset.  
25 Thus, in a simple embodiment, the metadata processor 115 may simply select the stored metadata of the filtered content items.

Step 205 is followed by step 207 wherein the resulting  
30 subset characterising data/metadata is transmitted to the user device 103. Step 207 is followed by step 209 wherein the user device 103 receives the subset metadata.

Specifically, the server 101 comprises a server network interface 117 which is coupled to the metadata processor 115 and which receives the subset metadata. The server network interface 117 couples the server 101 to an external network 5 119. The user device 103 comprises a user device network interface 121 which couples the user device 103 to the external network 119. Thus, the server 101 can transmit data to the user device 103 via the external network 119.

10 It will be appreciated that any suitable network, communication system or communication means can be used to transmit data from the server 101 to the user device 103 depending on the preferences and requirements of the individual embodiment. For example, the user device 103 may 15 be a cellular mobile phone and the network 119 may be a cellular communication system such as GSM or UMTS. As another example, the user device 103 may be a Personal Digital Assistant (PDA) which is coupled to the server 101 via a WiFi™ network. As another example, the user device 103 20 may be a set-top box connected to the server 101 via a cable.

Step 209 is followed by step 211 wherein a second filter processor 123 coupled to the user device network interface 25 121 receives the subset metadata and proceeds to perform a further filtering of the filtered subset of content items to generate a second subset of content items. The second filter processor 123 is coupled to a detailed user profile store 125.

30

Specifically, the second filter processor 123 evaluates the subset metadata for the filtered subset of content items

with reference to a low level, more detailed, more accurate user profile in order to generate a subset comprising a suitable number of content items. The second filter processor 123 retrieves the user profile for the appropriate  
5 user. It then proceeds to compare the metadata for each content item of the filtered subset with the user profile and to select a number of content items.

The detailed user profile is more detailed than the high-  
10 level user profile. Specifically, the detailed user profile contains user preference information that allows a finer differentiation of the content items than the high-level user profile. For example, based on the high-level user profile it may be determinable that a number of documents  
15 belong to the same category. However, based on the detailed user profile it is possible to further allocate these content items into different subcategories such that it is possible to differentiate between the content items.

20 In the example, the detailed user profile comprises a number of categories corresponding to the categories of the high-level user profile of the server 101. However, in addition, one or more (and possible all) of the categories are further divided into subcategories. Thus, the detailed user profile  
25 may contain more topic categories than the high-level user profile.

In some embodiments, the detailed user profile may e.g. use the same topic term vectors as the high-level user profile  
30 but may in addition comprise intra-topic data related to intra-topic differentiators. This intra-topic data may be used by the second filter processor 123 to differentiate

between different user preference values for each topic. Thus, in some embodiments each subcategory may have an individual term topic vector which can be used by the second filter processor 123 whereas in other exemplary embodiments  
5 the subcategories of a specific category may use the same term topic vector but be differentiated by additional data. Furthermore, in some embodiments the metadata sent to the user device may be chosen by the server 101 as intra-topic differentiators.

10

It will be appreciated that in some embodiments the subcategories may not merely be further divisions of the categories of the high-level user profile but may be narrower categories having different boundaries and topic  
15 associations than the categories of the high-level user profile.

It will also be appreciated, that the detailed user profile may be generated in a similar way to the high-level user  
20 profile and that it specifically may be generated based on a training process. Specifically, the detailed user profile may be generated either by learning of keywords in the low-level metadata generated at the server 101 and corresponding to items the user has read, or by explicit input by the user  
25 of keywords.

Furthermore, the detailed user profile may be continuously updated and adapted depending on the usage and preferences of a user of the user device 103. For example, the user of  
30 the user device 103 may manually input preference values for subcategories and indeed may in some embodiments assist in defining these subcategories.

The second filter processor 123 may operate using a similar or potentially identical algorithm to that of the first filter processor 109 but may use the more detailed user profile information. Thus, specifically, the second filter processor 123 may receive document term vectors for each of the text documents of the filtered subset and may apply the similarity measure also used by the first filter processor 109 to identify a matching subcategory. The preference value of that subcategory may then be assigned to the content item. The second filter processor 123 may then proceed to select the content items for the second subset in response to the assigned preference value.

It will be appreciated that most of the previous comments and descriptions of the first filter processor 109 applies equally well to the second filter processor 123.

The second filter processor 123 is coupled to a presentation processor 127. The presentation processor 127 may present an identification of at least one content item of the second subset of content items to a user. Specifically, the presentation processor 127 can extract the title and author of the selected text document(s) from the metadata of that document and can present this information on a display of the user device.

In some embodiments, the user device 103 may be arranged not only to present the user with an indication of the recommended content item(s) but also to present the content item(s) itself.



In the specific example, step 211 is followed by step 213 wherein the presentation processor 127 retrieves at least one content item of the second subset from a remote server in response to a user selection of the at least one content  
5 item. For example, the second filter processor 123 may generate a list of a few recommended content items. This list may be presented on a user display of the user device 103. In response, the user may select one of the recommended content items e.g. by entering an appropriate number on a  
10 keypad of the user device 103. In response the presentation processor 127 generates a request message identifying the selected content item and transmit this message to the server 101. In response, the server extracts the identified content item from the content item store 105 and transmits  
15 this back to the user device 103. The presentation processor 127 then proceeds to execute step 215 wherein the text document is presented on the display of the user device 103.

It will be appreciated that in the example, the selected  
20 content item is retrieved from the server 101 but in other embodiments, the content item may be retrieved from other external servers.

In the described example, the metadata processor 115 simply  
25 selected the metadata to transmit as the metadata of the content items of the filtered subset. However, in some embodiments the metadata processor 115 may be arranged to further reduce and/or modify the metadata to be transmitted. Specifically, for a given content item a subset of the  
30 provided metadata can be selected depending on the matching category identified by the first filter processor 109.

The aim of the data reduction is to reduce the amount of metadata that needs to be communicated while still providing the user device 103 with metadata that allows filtering of items within a given topic category. The reduction of the  
5 data may be based on the assumption that after different documents are classified into a given topic, the differentiation between them can be made using a subset of the metadata. Thus, the server 101 may retain only those terms in the metadata that are relevant for intra-topic  
10 classification.

For example, if the subcategories used in the detailed user profile corresponds to a direct division of the categories of the high-level user profile, the subset metadata  
15 transmitted to the user device 103 may comprise an indication of the matching category. Accordingly, the second filter processor 123 need not identify the main matching category as it can directly select the category identified by the first filter processor 109. Thus, the second filter  
20 processor only needs to identify which subcategory within the main matching category the individual content item belongs to.

In some embodiments, the metadata processor 115 is arranged  
25 to exclude first metadata from the transmitted subset metadata in response to a determination that the first metadata is not associated with the matching category. The first metadata is not associated with the matching category if it is not related to and/or descriptive for that  
30 category. For example, metadata may not be associated with the matching category if it is not part of the metadata stored for that category. Specifically, any terms which are

not included in the topic term vector for the matching category can be deleted from the document term vector for that content item.

5 Thus, a document's metadata representation can be reduced only to those terms that are also present in the corresponding topic term vector. This reduction is made according to the fact that this new metadata will be used to identify the particular sub-area of the given topic in which  
10 a user is interested. In this case, if a terms is not relevant for the topic category in which the document was classified it is unlikely to be relevant for a particular sub-area of that topic.

15 In some embodiments, the metadata processor 115 is arranged to exclude first metadata from the subset metadata for a content item in response to a determination that the first characterising data does not provide intra-category differentiation. Intra-category differentiation may for  
20 example require that the corresponding data is different for at least two of the subcategories for the matching category. Thus, any data which does not allow the selection of a subcategory within the matching category may be deleted from the transmitted metadata. In particular, terms of the  
25 document term vector which are identical for all subcategories of the matching category can be deleted.

Thus, a document's metadata representation can be reduced by deletion of all the terms that are not relevant for an  
30 intra-topic classification. This reduction is made according to the fact that the topic vector representation constructed during the offline learning contains terms that are relevant

for making a differentiation relative to other topics but may have a low differentiation value for any intra-topic classification. For example, if an article was classified into the topic Tennis in order to determine the sub-area of the Wimbledon event from other tennis events, terms like *tennis, set, game* will have less relevance than *grass, July, slam* and of course *Wimbledon*.

As a result of the data reduction process, the metadata representation of the documents sent to the user's device may contain only those terms that are present in both the document and the intra-topic classification vector of the corresponding topic. This may substantially reduce the amount of data to be transmitted thereby reducing the delay and resource needed for the communication. For example, weights can be computed for each term to reflect both the semantic relevance to the article and the relevance of the term for intra-topic classification.

In many embodiments both the detailed user profile and the high-level user profile are continually updated and adapted to the user.

Specifically, in some embodiments the user of the user device 103 may modify data of the high-level user profile by manually entering data into the user device 103. For example, the presentation processor 127 may receive user preference data in the form of a user manually entering a preference value for a given topic category. This may be fed to the user device network interface 121 which transmits the user preference data to the server 101. The server 101 can comprise an update processor 129 coupled to the server

network interface 117 and the high-level user profile store 111. The update processor 129 receives the user preference data and updates the high-level user profile accordingly. Specifically, if the user preference data is a specific user  
5 preference value for a given topic category, the update processor 129 allocates this preference value to the specified category. It will be appreciated that this approach may be used both for changing user preference values as well as for providing initial preference values  
10 for the categories.

It will also be appreciated that in many embodiments the user profiles may be updated based on the behaviour of the user. For example, the user device 103 can comprise an  
15 adaptation processor 131 which is coupled to the detailed user profile store 125 and which continuously monitors the content items that are selected by the user. The adaptation processor 131 may then proceed to change the detailed user profile depending on the user selection. For example, the  
20 adaptation processor 131 may increase the preference value for the matching subcategory of a selected content item. Thus, the more a specific content item is selected the higher will the preference value be for the subcategory to which the content item belongs. It will be appreciated that  
25 the adaptation process 139 may modify other aspects of the detailed user profile such as the terms comprised in the topic term vectors or the division into subcategories.

In some embodiments the user device 103 and the server 101  
30 may comprise functionality for updating one user profile in response to the other user profile. Specifically, in the example of FIG.2, step 217 is performed wherein the high-

level preferences of the high-level user profile are modified in response to changes in the detailed user profile.

5 In the example, the user device 103 comprises a synchronisation processor 133 coupled to the detailed user profile store 125 and the user device network interface 121. The synchronisation processor 133 continuously monitors the detailed user profile to determine whether any changes  
10 occur.

If the synchronisation processor 133 detects that a change has occurred (or that a sufficient amount of changes have occurred according to a suitable criterion), it proceeds to  
15 generate a user profile modification indication for this change. For example, if the user preference values for one of the subcategories of a specific category have changed in the detailed user profile, the synchronisation processor 133 can generate a user profile modification indication in the  
20 form of data specifying the new preference values for the subcategories.

The user profile modification indication is transmitted to the server 101 where it is fed to the update processor 129.  
25 The update processor 129 then proceeds to modify the first user profile in response to the user profile modification indication. For example the update processor 129 may determine a new user preference value for a category of the high-level user profile by adding or subtracting the changes  
30 in the user preference value for the individual subcategories of that category.

A more detailed example of the adaptation of the user profiles with specific reference to text documents will be described in the following. In the example, the user device 103 collects implicit and explicit user profile data and  
5 updates the low-level user preferences in the detailed user profile. The high-level user preferences are then determined for the high level user profile such that the high-level and detailed user profiles remain synchronised.

10 In the example, an initial detailed user profile consists in a reduced vector representation of all the topics a user has selected. (The vectors can additionally be aggregated with explicit keywords that the user wishes to provide for each selected topic).

15

A low-level learning process is used to modify the detailed user profile using the metadata associated with the items selected for the corresponding user. This process has three distinct goals:

20

- To modify the relevance of the common terms based on the obtained feedback.
- To insert new terms into the vectors based on the feedback.
- 25 • To eliminate old terms from the vectors based on a time depreciation algorithm.

The weights of the existing terms in the detailed user profile are updated according to the formula given below.

30 The factors in this formula are related both to the content of a text document (i.e. the extracted terms) and to the overall user behaviour towards the personalisation system:

- The weight of the appropriate topic category in the high level user profile.
- 5 • The cosine similarity measure between the item and the user's detailed user profile.
- The amount of time the user spent reading the document.
- The length of the document.
- The average number of documents that the user reads per day.
- 10 • The number of documents interesting for the user and which contains the term.
- A beta constant value that is used to differentiate between the changing rate of the weight if the update is performed in relation to an interesting or a non-interesting document.
- 15

The specific mathematical formula used for the update rule for the weight of each term that exists in the detailed user profile is:

20

$$w_{new} = w_{old} \pm w_T * Sim(I,U) * e^{-\beta * t * y} * \log \frac{time}{log length}$$

where:

- $w_{old}$ : The current weight to be updated
- +/-: Demonstrates positive/negative feedback
- 25 •  $w_T$ : The weight (in the high level user profile) of the topic category to which the item has been classified
- $Sim(I,U)$ : The cosine similarity measure between the content item (I) and the user's detailed user profile
- time: The time spent reading the content item



- length: The length of the content item (in chars or bytes)
  - $\beta$ : A constant differentiating between the positive/negative feedback
- 5
- x: The mean number of documents the user reads per day
  - y: The number of the selected documents where the term exists

Individual terms of the topic term vectors may thus be added  
10 or deleted. Specifically, when a user selects a text  
document which contains new terms, each of these terms is  
placed in a subordinate waiting stack. Then, each time the  
user selects a document that contains a specific term, the  
usage history of the term changes (i.e. there is an increase  
15 in the value for the term). The metric that determines the  
insertion of a new term into the detailed user profile is  
whether the term usage history exceeds a certain threshold.

When a term is inserted into the detailed user profile, its  
20 initial weight is related to the weight of the topic in the  
high level user profile to which the most recently read  
document has been classified. The default values for the  
initial entry into the system are similar to those used  
during the initialization of the detailed user profile.

25

The criterion for the removal of terms is also the metric  
relating to the usage history. More specifically, a number  
of terms having a lower usage history can be deleted from  
the detailed user profile when new terms are added.

30

Updates to the high level user profile can be determined by a number of methods. For example, the system may monitor the total implicit feedback (selections and viewing times) of items in each topic category and adjust the preference values of the high level user profile accordingly.

Alternatively or additionally, preference values may be determined based on the preference values of each topic subcategory in the detailed user profile.

10 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  
15 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  
20 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  
25 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  
30 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.

5 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  
10 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  
15 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  
20 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  
25 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  
30 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 recommendation system, comprising  
5 a recommendation server comprising:  
means for providing characterising data for a plurality  
of content items,  
first filtering means for filtering the plurality of  
content items to generate a first subset of content  
10 items, the filtering comprising selecting content  
items from the plurality of content items for the  
first subset in response to the characterising data  
and data of a first user profile for a first user,  
means for generating subset characterising data for the  
15 first subset of content items from the  
characterising data, and  
a transmitter for transmitting the subset  
characterising data to a user device; and  
the user device comprising:  
20 a receiver for receiving the subset characterising  
data, and  
second filtering means for filtering the first subset  
of content items to generate a second subset of  
content items, the filtering comprising selecting at  
25 least one content item from the first subset for the  
second subset in response to the subset  
characterising data and data of a second user  
profile for the first user, the second user profile  
being more detailed than the first user profile.  
30
2. The content recommendation system of claim 1 wherein  
the first filtering means is further arranged to generate

the first subset of content items in response to a device profile for the user device, the device profile being indicative of at least one capability of the user device.

5 3. The content recommendation system of claim 1 wherein the first filtering means is further arranged to generate the first subset of content items in response to a characteristic of a communication link between the transmitter and the receiver.

10

4. The content recommendation system of claim 1 wherein the first user profile comprises a first set of categories for content items and the second user profile comprises a division of at least one of the first set of categories into  
15 further subcategories.

5. The content recommendation system of claim 4 wherein at least one of the first and second means for filtering is arranged to determine user preference values for content  
20 items in response to the characterising data and the data of the user profile, and to generate a subset of content items by selecting a number of content items having the highest user preference value.

25 6. The content recommendation system of claim 5 wherein the first filtering means comprises:  
means for determining a matching category for a first content item in response to a comparison of characterising data for the first content item and characterising data for  
30 the first set of categories; and

means for determining a user preference value for the first content item in response to a user preference value for the matching category.

5 7. The content recommendation system of claim 5 wherein the second filtering means comprises:

means for determining a matching subcategory for a first content item in response to a comparison of subset characterising data for the first content item and

10 characterising data for the first set of subcategories; and

means for determining user preference value for the first content item in response to a user preference value for the matching subcategory.

15 8. The content recommendation system of claim 4 wherein the transmitting means is arranged to generate the subset characterising data for a first content item of the first subset by selecting a subset of the characterising data for the first content item in response to a characteristic of a  
20 matching category for the first content item.

9. The content recommendation system of claim 8 wherein the transmitting means is arranged to exclude first characterising data from the subset characterising data for  
25 the first content item in response to a determination that the first characterising data is not associated with the matching category.

10. The content recommendation system of claim 8 wherein  
30 the transmitting means is arranged to exclude first characterising data from the subset characterising data for the first content item in response to a determination that

the first characterising data does not provide intra-category differentiation.

11. The content recommendation system of claim 1 wherein  
5 the user device further comprises:

means for receiving a user selection of at least one selected content item of the second subset of content items; and

10 means for modifying the second user profile in response to the characterising data of the selected content item.

12. The content recommendation system of claim 1 wherein the user device comprises means for transmitting a user profile modification indication to the recommendation  
15 server, the user profile modification indication being indicative of a change to the second user profile; and

the recommendation server further comprises:

means for receiving the user profile modification indication, and

20 modifying means for modifying the first user profile in response to the user profile modification indication.

13. The content recommendation system of claim 12 wherein  
25 the user profile modification indication comprises an indication of a user preference value for subcategories of a first category of the first user profile; and the modifying means is arranged to determine a user preference in response to the user preference values for the subcategories.

30

14. The content recommendation system of claim 1 wherein the user device further comprises means for retrieving at

least one content item of the second subset from a remote server in response to a user selection of the at least one content item.

5 15. The content recommendation system of claim 1 wherein the user device further comprises:

means for receiving user preference data from a user of the user device;

means for transmitting the user preference data to the  
10 recommendation server; and

the recommendation server further comprises means for modifying the first user profile in response to the user preference data.

15 16. The content recommendation system of claim 1 wherein the content items are text documents and the characterising data comprises keywords for the text documents.

17. The content recommendation system of claim 1 wherein  
20 the user device comprises means for presenting an identification of at least one content item of the second subset to a user.

18. A method of operation for a content recommendation  
25 system including a recommendation server and a user device; the method comprising:

the recommendation server performing the steps of:

providing characterising data for a plurality of  
content items,

30 filtering the plurality of content items to generate a first subset of content items, the filtering comprising selecting content items from the



- plurality of content items for the first subset in response to the characterising data and data of a first user profile for a first user,  
generating subset characterising data for the first  
5 subset of content items from the characterising data, and  
transmitting the subset characterising data to a user device; and  
the user device performing the steps of:
- 10 receiving the subset characterising data, and  
filtering the first subset of content items to generate a second subset of content items, the filtering comprising selecting at least one content item from  
15 the first subset for the second subset in response to the subset characterising data and data of a second user profile for the first user, the second user profile being more detailed than the first user profile.
- 20 19. A computer program product enabling the carrying out of a method according to claim 18.

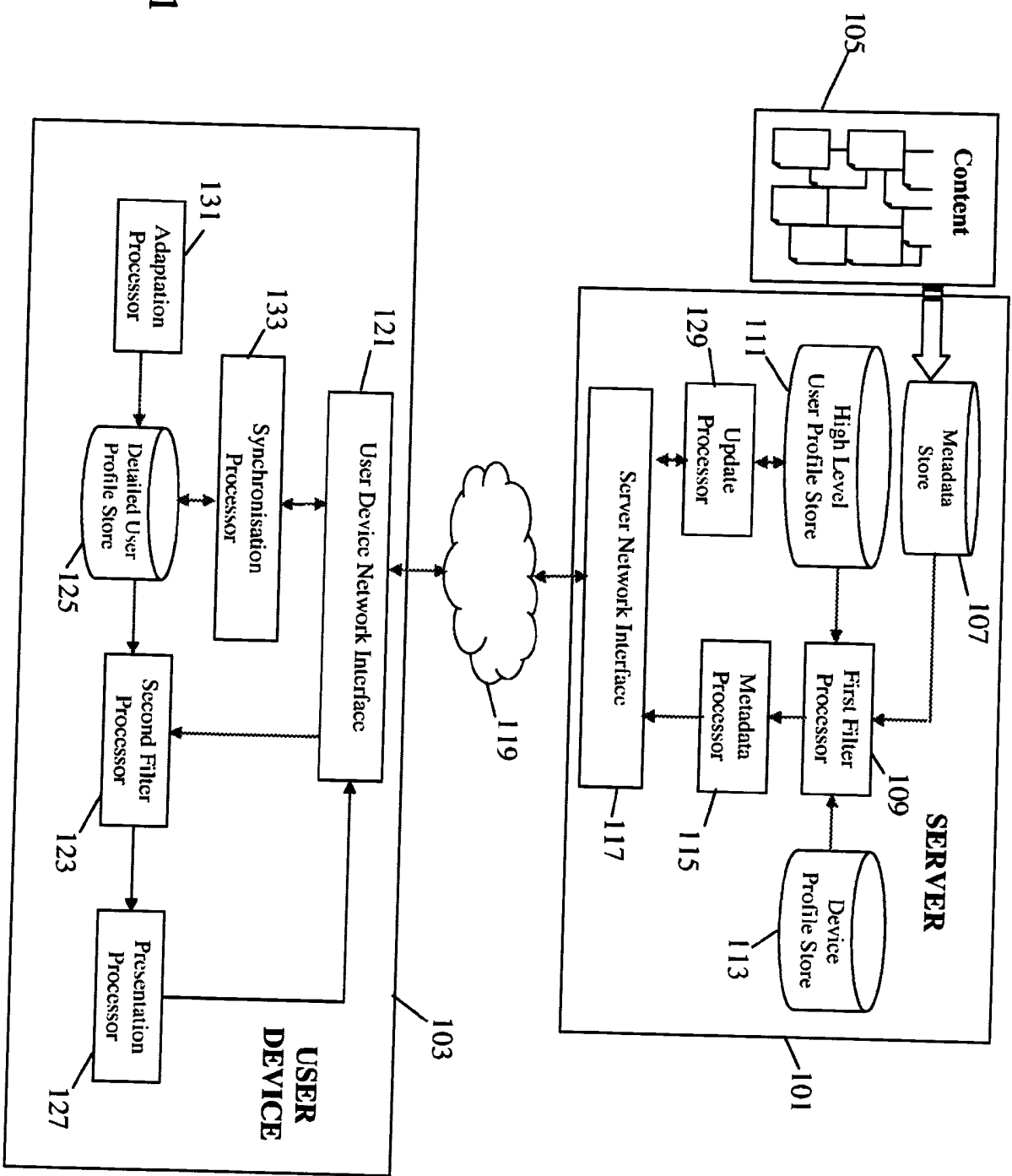


FIG. 1

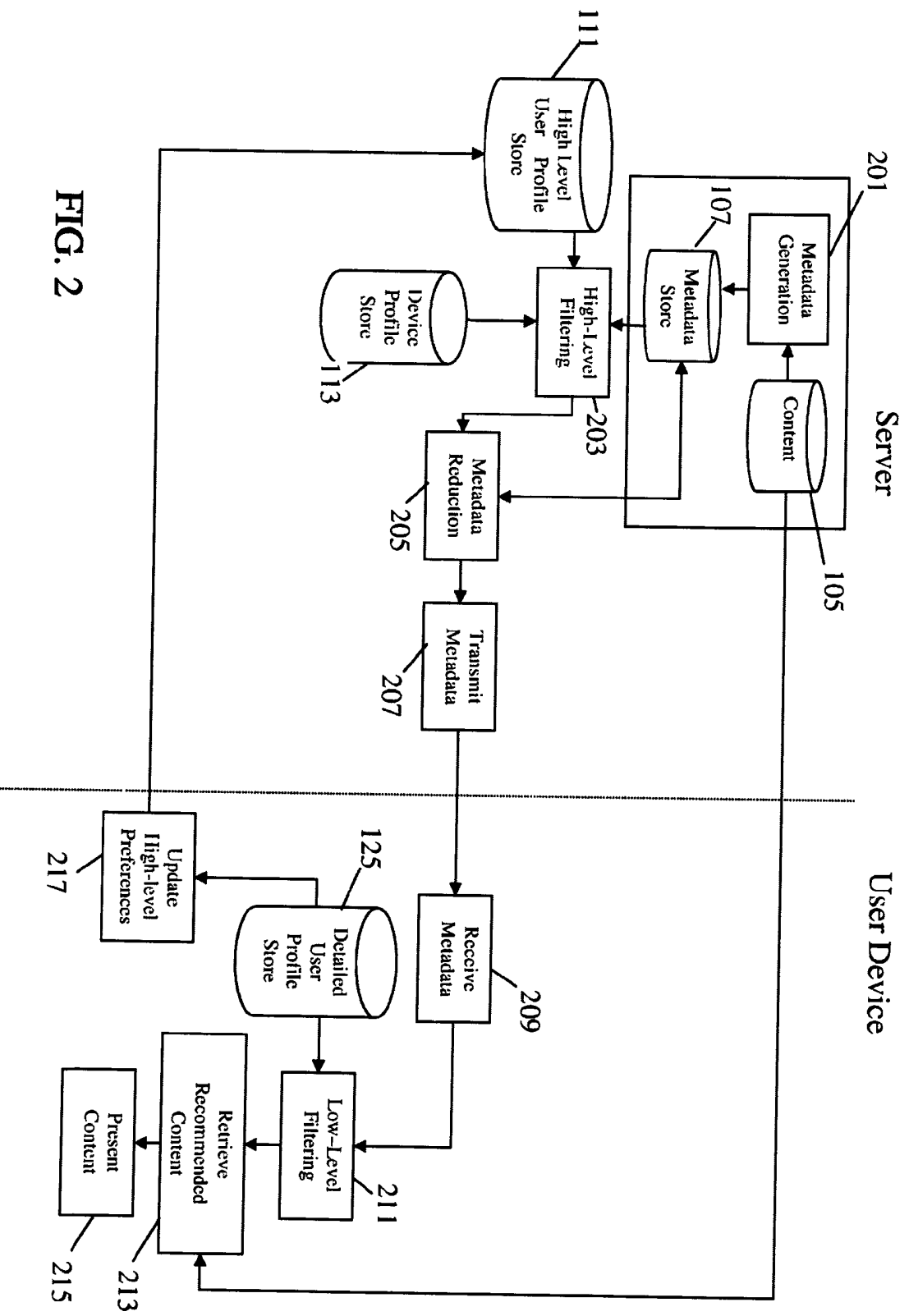


FIG. 2



For Innovation

41

**Application No:** GB0624552.6

**Examiner:** Kalim Yasseen

**Claims searched:** 1-19

**Date of search:** 29 March 2007

## Patents Act 1977: Search Report under Section 17

### Documents considered to be relevant:

Category	Relevant to claims	Identity of document and passage or figure of particular relevance
X,Y	X:1, 4, 6, 16-19, Y: 11, 15	WO99/28837 A1 (AVEO) page 2 line 19 to page 3 line 8, page 4 line 14 to page 5 line 10
Y	11, 15	EP1484692 A1 (SONY NETSERVICES) see whole document especially abstract

### Categories:

X	Document indicating lack of novelty or inventive step	A	Document indicating technological background and/or state of the art.
Y	Document indicating lack of inventive step if combined with one or more other documents of same category.	P	Document published on or after the declared priority date but before the filing date of this invention.
&	Member of the same patent family	E	Patent document published on or after, but with priority date earlier than, the filing date of this application.

### Field of Search:

Search of GB, EP, WO & US patent documents classified in the following areas of the UKC<sup>X</sup> :

Worldwide search of patent documents classified in the following areas of the IPC

G06Q

The following online and other databases have been used in the preparation of this search report

Online: EPODOC, FULLTEXT, WPI