ASSOCIATIVE CONTENT RETRIEVAL

Abstract: A retrieval system is provided. First description data including dimension data for a first identified content item is extracted. This process may be repeated for additional available identified content items. Candidate description data is extracted. Then, a set of vector values for each candidate content item may be generated, each vector value representing a degree of similarity between the dimension data for a dimension, for example, metadata, usage history, genre, content type, of the first description data and the corresponding dimension data of the candidate description data. A similar candidate content item from the candidate content items may be selected based on the degrees of similarity represented by the generated set of vector values, and is provided.
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.
ASSOCIATIVE CONTENT RETRIEVAL

The present invention relates to the field of content retrieval, management and presentation.

The storage capacity of storage devices and databases, including hard drives on personal computers and on other types of storage media has been rapidly increasing in recent years. It has been estimated that storage capacity doubles approximately every 12 months, while network bandwidth also has been increasing very rapidly. As a result, storage devices store a greater amount of content to which user access needs to be facilitated. Content that is not indexed or organized in a manner transparent to the user may be "lost" as far the needs of the user are concerned and be unlikely to be retrieved. On the other hand, a user can be overloaded with content stored on a storage device or database, and may not be able to retrieve content that is available on a network, such as the internet, unless the content is somehow managed or organized to provide convenient access for the user.

Various schemes for data retrieval exist. Platt, U.S. Patent Publication No. 2003/0221541 discloses an automatic playlist generator, in which several seed songs, including "undesirable seed" songs are used to generate songs on a playlist. Cluts, U.S. Patent No. 5,616,876 discloses selecting additional songs that are like a first set of songs, based on "style labels" for each song previously written by an editor. However, none of these references disclose providing content to a user of a content type different from the content type of the user-designated identified content item.

It is also possible of course for a user to retrieve content items, however attempting to locate similar items can be a time-consuming and onerous job, particularly if the content type of desirable items is not known or specified by the user. Further, as content items continue to accumulate in a storage device or database controlled by the user, the job of retrieving content items becomes ever more difficult.

Provided are a method, system, device, engine, apparatus, and computer-readable media that embodies or carries out the functions of a retrieval system. First description data including dimension data for a first user-selected content item is extracted. Then, candidate description data including corresponding dimension data for candidate content items is extracted, each candidate content item being of a content type different from the content type of the user-selected content item. A first set of vector values for each
candidate content item may be generated, each vector value representing a degree of
similarity between the dimension data for a dimension of the first description data and the
 corres panding dimension data of the candidate description data. A candidate content item
from the candidate content items can then be selected based on the degrees of similarity
represented by the generated first set of vector values. The selected candidate content item
or items are then provided by the retrieval system, such as via a user interface.

A dimension of the dimension data represents a content type of the item, a content
style for the item, a genre of the item, item metadata, usage history of the item, a performer
performing in the item, a director associated with the item, a creator associated with the
item, or rendering requirements for the item. As used herein, the metadata may include a
time of creation of the item, a place of creation of the item, a time of acquisition of the
item, and/or a place of acquisition of the item.

The candidate content item may be selected only if a total degree of similarity
represented by the first set of vector values surpasses a minimum threshold.

The candidate content item with the highest total degree of similarity as represented
by the first set of vector values may be selected.

Additional content items may be identified. Description data including the
dimensions data for a second identified content item grouped with the first identified
content item is extracted. The candidate content item is then selected based also on a
second set of vector values representing degrees of similarity between the dimension data
for the second identified content item and the dimension data of the similar candidate
content item. Accordingly, the candidate content item may be selected such that the first
set of vector values and the second set of vector values is averaged, weighted averaged, or
added.

A commonality vector may also be chosen for weighting results. A commonality
vector, a vector that represents a dimension for which dimension data of the first identified
content item is closest to the second identified content item is selected, and in selecting the
candidate content item a value of the commonality vector may be weighted more than
remaining vector values of the first set of vector values and the second set of vector values.

A virtual content item may be constructed. Description data including dimension
data for a first and a second user-selected content item are extracted. Candidate description
data including corresponding dimension data for candidate content items are extracted,
each candidate content item being of a content type different from the content type of the
user-selected content item. Then, a virtual item may be constructed by averaging or
weighted averaging a virtual item set of vector values, each vector value of the virtual item
set of vector values representing a degree of similarity between a dimension of the
dimension data of the first description data and a corresponding dimension of the
dimension data of the second description data. A set of vector values for each candidate
content item can be generated, each vector value representing a degree of similarity
between the dimension data for a dimension of the virtual content item and corresponding
dimension data for the candidate content item. A candidate content item from the candidate
content items may thus selected by computing as a testing value one of an average, a
weighted average, and a sum for each set of vector values of the candidate content items,
and determining as the selected candidate content item the candidate content item whose
testing value surpasses a threshold. The selected candidate content item or items are
provided.

Figure 1 is a schematic view all of a retrieval system according to an embodiment
of the present invention.

Figures 2A-2C are flowcharts operations of a system according to the present
invention.

Figure 3 shows a data chart of vector value alignment according to an embodiment
of the present invention.

The following discussion and the foregoing figures describe embodiments of
Applicant’s invention as best understood presently by the inventors however, it will be
appreciated that numerous modifications of the invention are possible and that the
invention may be embodied in other forms and practiced in other ways without departing
from the spirit of the invention. Further, features of embodiments described may be
omitted, combined selectively or as a whole with other embodiments, or used to replace
features of other embodiments, or parts thereof, without departing from the spirit of the
invention. The figures and the detailed description are therefore to be considered as an
illustrative explanation of aspects of the invention, but should not be construed to limit the
scope of the invention.

As shown in Figure 1, the retrieval system 1-1 includes several modules, which will
be described below. Modules of the retrieval system 1-1, or portions thereof, and/or the
retrieval system as a whole, may be comprised of hardware, software, firmware, or a combination of the foregoing, however some modules may be comprised of hardware for example, while other modules may be comprised of software, firmware or a combination thereof.

It is to be understood that modules of the retrieval system need not all be located or integrated with the same device. A distributed architecture is also contemplated for the retrieval system, which may “piggy-back” off of suitable modules provided by existing devices.

The following description will refer to a retrieval system 1-1 that is physically integrated with or connected to a database 1-2 via a wired or wireless connection thereto. The database 1-2 may be embodied on a storage device such as on a hard drive of a personal computer, a personal video recorder, an entertainment system, an electronic organizer, a personal handheld device, a Jaz drive, or may be embodied as a commercial storage facility, such as a disk drive. It will be understood that the database 1-2 may include several storage devices that are connected, such that organization or grouping of content items on two or more of such devices is possible. It will further be understood that the database may be understood to include one or more storage media, such as disks, including CDs, DVDs, zip disks, floppy disks, data cartridges, or the like, which can be loaded onto and retrieved by the database 1-2. However, it will be understood that the retrieval system 1-1 is also capable of retrieving content via a network 1-9, such as a LAN, WAN, the Internet, or the like.

As shown in Figure 1, the retrieval system 1-1 includes a description data extractor 1-11, which is a module that collects certain types of data from a content item. The content item may be a video, or a video clip, a movie, a photo, a text file, music data, an audio file, or other type of multimedia data, a JPEG file, or XML data. For example, the video may be a home video shot on a digital video recorder, the movie may be commercially distributed film data, such as a film encoded as MPEG (including MPEG-2, MPEG-3, or the like), the photo may be a digital photograph data, or series of photographs or a photograph album, the text file may be a word processor produced file, a spreadsheet, or a computer code file, the music data may be an MP3 file or the like, and so forth.

The description data extracted by the description data extractor 1-11 includes information about the content item. Such description data describe the dimensions of the
content item. Such dimensions may include any one or more of the following:

the content type, including the medium, such as the video, audio, photo, text file, etc.

cetera;

the content style or genre, such as holiday movie, personal landscape photography,
jazz music or the like;

metadata for the item, such as time and/or location of the creation of the item, time
and/or place of acquisition of the item;

usage history of the item, such as the last/first/penultimate etc. time and/or location
and/or context of playback and/or editing, a time period of most usage (for example, the
item is mostly used at night, or on Monday afternoons, or 6-8 AM or the like), a time of
acquisition of the item, a place of creation of the item, a place of acquisition of the item, a
place of last usage, and a place of most usage (for example, the item is mostly used in the
living room, or in the user’s home, or the like); such usage history data is sometimes
known as metadata, and conversely, types of metadata are sometimes referred to as usage
history data; and

an actor, director, creator, artist, performer, photographer or the like associated with
the content item.

It will be understood that such description data about the item may be located and
extracted in a variety of ways, including from the item, from an index or database
management file, or from an outside source such as from the World Wide Web connected
to the retrieval system 1-1 via a wired or a wireless connection to the Internet 1-9.

The identified content item may be identified in one of several ways. A user may
designate the item based on which other items, sometimes referred to as “candidate content
items” are to be retrieved. Alternatively, a content item newly added or created may
automatically be designated as an identified content item based on which other items are to
be retrieved.

Based on these compiled dimensions of the description data extracted by
description data extractor 1-11, content item identifier 1-12 identifies candidate content
items in the database, over the network connection or from other sources that are similar
with respect to these dimensions of their description data to the first identified content
item. Vector constructor 1-13 then creates a first set of vector values by assigning vector
values to each of a number of vector as follows: each vector corresponds to a dimension,
and a value for the vector reflects a degree of similarity or matching of a dimension of the first identified content item with the candidate content item.

For example, a vector that corresponds to the dimension of the content item termed style or genre would get a high value if both the identified content item and the candidate content item are of the same genre, such as “Spanish holiday.” A vector value of 1 or 0 may indicate little or no correlation or matching for the particular dimension between the first identified content item and the candidate content item, while a vector value of 9 or 10 may indicate a high degree of similarity or match. For example, when both content items have a genre of “Spanish holiday” then for the vector corresponding to the genre dimension, a 9 or 10 value would be assigned. Alternatively, instead of using a scale of 1 to 10, vector values may merely represent a “strong”, “normal”, or “weak” match for the dimension. It will be understood that numerous other schemes for such vector values may be used without departing from the spirit of the present invention. An average or a sum of such a set of vector values for a pair of content items would then be calculated as an overall degree of similarity between the two content items.

If a second identified content item is available, then a second set of vector values may be similarly constructed by vector constructor 1-13 based on description date extracted by description data extractor 1-11 for the second content item, such that this second set represents a degree of similarity between corresponding dimensions of this second identified content item and a candidate content item. There may be additional available identified content items. Thus, this process of description data extraction and vector value set generation may be repeated for any number of available identified content items 1-N, N being a positive integer greater than 1. Then, the candidate content item selection is performed based on all such generated vector value sets, or their average.

If more than one identified content items are available, then a commonality vector generator/threshold setter 1-14 may select one or more vectors for which the vector values of the first set and the second set are consistently high. Such vector values may then be weighted more than values for the other vectors in the average or sum of the set of vector values representing the overall degree of similarity between the two items. In this way, a dimension which is representative of the first and second identified content item, or which tends to capture the similarity between the first and second identified content item and is therefore characteristic of the group would be weighted more then other vector values.
Although shown as part of a single module 1-14, separate modules, a commonality vector generator module and a threshold setter module may be constructed as part of the retrieval system 1-1, or such modules may be incorporated into other modules.

Virtual item constructor 1-15 will be described below in the context of a discussion of an operation of embodiment of the present invention.

Content item selector 1-16 selects the candidate content item or items to be provided to user. This module may also handle other tasks necessary for the operation of the retrieval system, such as overall control and coordination of the modules of the retrieval system 1-1.

Retrieval result output 1-17 interfaces with other devices and communication with the outside, including interfacing with a user (not shown). In particular retrieval result output 1-17 signals about the user interface of content items retrieved by the retrieval system 1-1. User interface 1-3 may be a separate device or may be integrated with another device or system, such as a personal computer or a personal video recorder, or one or more of the storage and other devices enumerated above.

An operation of an embodiment of the present invention will now be described with reference to Figures 1-3. A first content item is identified, as described above, by a user via user interface 1-3 shown in Figure 1, or automatically by the system, for example by a detection of a newly added content item or an isolated content item in database 1-2.

Description data extractor 1-11 of retrieval system 1-1 extracts first description data for the first content item identified, as stated at S1 of Figure 2A. Figure 3 shows a box labeled 6-11 referencing identified content item 1. At S2, dimension data for each of the dimensions for the first identified content item are compiled. It will be understood, that depending on the needs of the user, some or all of the above-identified dimensions may be more relevant, while others may be completely irrelevant and unused by an retrieval system according to the present invention. Also other dimensions not explicitly recited here may be particularly relevant and used by the retrieval system 1-1.

If an additional second identified content item, shown in Figure 3 as 6-12, is available or has been identified, then steps S3 and S4 are performed: at S3 description data for the identified content item is extracted, and at S4, dimension data for each of the dimensions for the second identified content item are compiled. As shown in Figure 3 a number of content items may be identified as bases of content retrieval. Figure 3 shows
first identified content item, 6-11, second identified content item, 6-12, and identified content item N, 6-14. Therefore, this process would be repeated for each of the first – N identified content items.

Content item identifier 1-12 of Figure 1 identifies candidate content items in the database 1-2, over a network or elsewhere, while description data extractor 1-11 at S5 (Figure 2A) extracts description data for each of the candidate content items and, at S6, compiles the dimension data for each of the content items. The process of extracting the corresponding description data of a second candidate content item (represented in box 6-22), if found, is performed at S7, and the compilation of the dimension data for the second candidate content item is then performed at S8.

According to an aspect of the present invention, At S9, depending on the system settings or depending on the user’s setting or current command, it may be decided that a virtual item is to be constructed as a basis for determining the similarity of candidate content items, in which case processing will proceed as shown in Fig. 2C. Otherwise, processing would proceed as shown in Figure 2B.

Based on the similarity or match of each dimension of each identified content item with the corresponding dimension of each candidate content item, a vector value is constructed by a vector constructor 1-13 as shown in S11 of Figure 2B. Figure 3 shows a table 6-1 with a set of vectors 6-3 with values that reflect the degree of similarity for corresponding dimensions of first identified content item 6-11 with the first candidate content item 6-21. Similarly, a set of vector values 6-4 reflects the similarity of the dimensions of first identified content item, 6-11, with second candidate content item, 6-22. With respect to second identified content item, 6-12, the set of vector values 6-5 reflects the degrees of similarity for corresponding dimensions with first candidate content item 6-21, while the set of vector values 6-6 reflects the degree of similarity between dimensions of second identified content item, 6-12, with candidate content item 6-22.

Each set of vector values also may include an average vector value determined at S12, based on computation of the arithmetic mean, mode, median or sum of the vector values of this set, that reflects the average similarity for the pair of content items. Thus, for instance, vector values 6-3 of Figure 3, may include a first vector value, a second vector value, and h-th vector value, and an average value for the set.

Further identified content items may also be available, and the process of extracting
the dimension data and finding a set of vector values based on the similarity with corresponding dimensions of candidate content items would continue. Box 1-14 of Figure 3 shows identified content item N.

Also, further candidate content items may be found, and for each one, sets of vector values could be calculated for each identified content item. Box 6-23 references such a candidate content item M.

According to an embodiment of the present invention, at S13, a commonality vector value set is determined based on the similarity of dimensions between identified content items. Thus, dimensions that are most similar are identified, and representative vectors can be weighted more than the other vectors, or can be used exclusively. In this way, a dimension which is representative of the first and second (and additional) identified content items, and which therefore tends to capture the similarity between the identified content items and is therefore characteristic of the group being formed would be weighted more than other vector values, or would be used exclusively to determine similar candidate content items.

At S14, a further set of vector values 6-8 may be computed that reflect the overall similarity for each of the dimensions for each candidate content item, by averaging or adding corresponding vector values of the candidate content item 6-21. Thus, for instance by adding or averaging corresponding vector values for each set of vector values for that candidate content item (for the column 6-2), an overall degree of similarity with the identified content items for the dimension is attained for the first candidate content item. Further, all of the vector values of the set 6-8 may be added or averaged to obtain an total similarity value for that candidate content item.

It will be understood, that average as used herein may include an arithmetic mean, a mode, a median or some such other statistical function suitably selected to provide a composite view of the selected values. Further, a simple sum of the values may be used as well as some such statistical function. Depending on the type of content item, and depending on the database and the needs of the user, certain dimensions all of the content item may be more important than others, and for this reason it may be helpful to weight vectors corresponding to certain dimensions more than others. The degree to which such factors are weighted would depend on the application and the needs of the user.

Once the vector values of the overall similarity set 6-8 are generated, a minimal
similarity threshold may be used to eliminate non-similar candidate content items, as shown at S15 of Figure 2B.

Further, it is also contemplated that different thresholds may be employed for the various vectors, depending on the needs of the user and the application. Accordingly, candidate content items for which the vector values meet or surpass the threshold value are grouped with the identified content items by group organizer 1-17, while other candidate content items are rejected. Alternatively, the most similar candidate content item, or predetermined number of the most similar candidate content items may be selected for grouping with the identified content items, while the remainder of the candidate content items may be rejected.

According to an aspect of the present invention, the content item retrieved is of a content type different from the content type of the user-selected content item. For example, if the user-selected content item is of the type music file, or MP3, then the retrieved content item may be of the content type photograph data. In this way, for example, pictures of a certain genre may be retrieved to match user-selected music of the same genre.

This (or these) selected candidate content item(s) are provided to the user or to the user interface 1-3 at S16. A signal may be provided directly to the database 1-2 to cause retrieval of the selected candidate item to the database or to the user interface 1-3. A notification may be provided to user interface 1-3 to notify a user (not shown) of a retrievable content item. The notification may consist of an identification of the content item to be retrieved, a description of the content item, a URL or a link to the content item, a retrieval of the entire content item or a portion thereof, or a combination of the foregoing. At S17, processing terminates.

Figure 2C shows a further process according to an aspect of the present invention, using a virtual content item. At S21, virtual item constructor 1-15 analyzes the dimensions of the identified content items based on which a grouping is sought. At S22, a representative content item for all of the identified content items, called a virtual content item 6-15 is then constructed based on the average or weighted average dimensions of the identified content items. For example, if all of the identified content items are of the genre “Spanish holiday,” then the virtual content item would also have as its genre “Spanish holiday.” Then, at S23 sets of vector values 6-7 are generated based on the similarity with
the dimensions of this virtual content item with the candidate content items. At S24, the threshold is applied in selecting similar candidate content items are selected, or the highest scoring candidate content item or items are selected.

Based on the candidate content items selected as similar using thresholding, or based on the predetermined number of the most similar candidate content items that are selected, at S25, as discussed, notification signal is provided by retrieval result output 1-17. At S26, processing terminates.

For example, suppose a user is compiling digital data representing photographs of a recent holiday in Spain in a database and would like to retrieve other content items with a Spanish theme available in the database, in another connected storage medium, or available over the Internet. The user may select the three photos as identified content item 1, identified content item 2, and identified content item 3, respectively, via user interface 1-3. The retrieval system would then retrieve a data file representing Spanish music found as the selected candidate content item. The user may not have remembered the existence of the Spanish music, or where to look for it in the database 1-2, and indeed the data file may have been added by another user with access to the database 1-2, or may have been retrieved by the retrieval system 1-1 from another storage device or from the world wide web. In any event, the user would now be notified of the retrieved content item and/or the retrieved content item would be associated with the user-selected content items. The user would then be able to accompany the viewing of the Spanish holiday photographs with Spanish music.

Embodiments of the present invention provided in the foregoing written description are intended merely as illustrative examples. It will be understood however, that the scope of the invention is provided in the claims.
CLAIMS

1. A content retrieval method comprising:
   extracting (S1) first description data including dimension data for a first
   user-selected content item;
   extracting (S5) candidate description data including corresponding
   dimension data for candidate content items, each candidate content item being of a content
   type different from the content type of the user-selected content item;
   generating (S11) a first set of vector values for each candidate content item,
   each vector value representing a degree of similarity between the dimension data for a
   dimension of the first description data and the corresponding dimension data of the
   candidate description data;
   selecting (S15) a candidate content item from the candidate content items
   based on the degrees of similarity represented by the generated first set of vector values;
   and
   providing (S16) the selected candidate content item.

2. The method of claim 1, wherein a dimension of the dimension data
   represents one of a content type of the item, a content style for the item, a genre of the
   item, usage history of the item, a performer performing in the item, a director associated
   with the item, a creator associated with the item, rendering requirements for the item, and
   any metadata for the item.

3. The method of claim 2, wherein the metadata represents one of a time of
   creation of the item, a time of last usage, a time period of most usage, a time of acquisition
   of the item, a place of creation of the item, a place of acquisition of the item, a place of last
   usage, and a place of most usage.

4. The method of claim 1, wherein the candidate content item is selected
   only if a total degree of similarity represented by the first set of vector values surpasses a
   minimum threshold.
5. The method of claim 1, wherein the candidate content item with the highest total degree of similarity as represented by the first set of vector values is selected.

6. The method of claim 1, further comprising:
   extracting (S3) description data including the dimensions data for an N-th identified content item grouped with the first identified content item, N being any positive integer greater than 1; and
   automatically selecting (S15) the candidate content item based also on an N-th set of vector values representing degrees of similarity between the dimension data for the N-th identified content item and the dimension data of the similar candidate content item.

7. The method of claim 6, wherein the candidate content item is selected such that the first set of vector values and the N-th set of vector values is one of averaged, weighted averaged, and added.

8. The method of claim 6, comprising selecting, as a commonality vector, a vector that represents a dimension for which dimension data of the first identified content item is closest to the N-th identified content item, and in selecting the candidate content item weighting a value of the commonality vector more than remaining vector values of the first set of vector values and the N-th set of vector values.

9. A content retrieval method comprising:
   extracting (S1) first description data including dimension data for a first user-selected content item;
   extracting (S3) N-th description data including dimension data for an N-th user-selected content item, N being a positive integer greater than 1;
   extracting (S5) candidate description data including corresponding dimension data for candidate content items, each candidate content item being of a content type different from the content type of the user-selected content item;
   constructing (S22) a virtual item by one of averaging and weighted averaging a virtual item set of vector values, each vector value of the virtual item set of
vector values representing a degree of similarity between a dimension of the dimension
data of the first description data and a corresponding dimension of the dimension data of
the N-th description data;

generating (S23) a set of vector values for each candidate content item, each
vector value representing a degree of similarity between the dimension data for a
dimension of the virtual content item and corresponding dimension data for the candidate
content item;

selecting (S24) a candidate content item from the candidate content items by
computing as a testing value one of an average, a weighted average, and a sum for each set
of vector values of the candidate content items, and determining as the selected candidate
content item the candidate content item whose testing value surpasses a threshold; and

providing (S25) the selected candidate content item.

10. A content retrieval system comprising:

a description data extractor (1-11) configured to extract first description data
including dimension data for a first user-selected content item;

said description data extractor (1-11) further configured to extract candidate
description data including corresponding dimension data for candidate content items, each
candidate content item being of a content type different from the content type of the user-
selected content item;

a vector generator (1-13) configured to generate a first set of vector values
for each candidate content item, each vector value representing a degree of similarity
between the dimension data for a dimension of the first description data and the
corresponding dimension data of the candidate description data;

a content item selector (1-16) configured to select a candidate content item
from the candidate content items based on the degrees of similarity represented by the
generated first set of vector values; and

a retrieval results output (1-17) configured to provide the selected candidate
content item.

11. The system of claim 10, wherein a dimension of the dimension data
represents one of a content type of the item, a content style for the item, a genre of the
item, usage history of the item, a performer performing in the item, a director associated with the item, a creator associated with the item, rendering requirements for the item, and any metadata for the item.

12. The system of claim 11, wherein the metadata represents one of a time of creation of the item, time of last usage, time period of most usage, a time of acquisition of the item, a place of creation of the item, a place of acquisition of the item, a place of last usage, and a place of most usage.

13. The system of claim 10, wherein said content item selector (1-16) is configured to select the candidate content item only if a total degree of similarity represented by the first set of vector values surpasses a minimum threshold.

14. The system of claim 10, wherein said content item selector (1-16) is configured to select the candidate content item with the highest total degree of similarity as represented by the first set of vector values.

15. The system of claim 10, wherein said description data extractor (1-11) is configured to extract description data including the dimensions data for an N-th identified content item grouped with the first identified content item, N being a positive integer greater than 1, and

said content item selector (1-16) is configured to select automatically the candidate content item based also on an N-th set of vector values representing degrees of similarity between the dimension data for the M-th identified content item and the dimension data of the similar candidate content item.

16. The system of claim 15, wherein said content item selector (1-16) is configured to select the candidate content item such that the first set of vector values and the N-th set of vector values is one of averaged, weighted averaged, and added.

17. The system of claim 15, further comprising a commonality vector generator/threshold setter (1-14) configured to select, as a commonality vector, a vector
that represents a dimension for which dimension data of the first identified content item is closest to the N-th identified content item,

wherein said content item selector (1-16) is configured to select the candidate content item based on weighting a value of the commonality vector more than remaining vector values of the first set of vector values and the N-th set of vector values.
FIG. 2B
**INTERNATIONAL SEARCH REPORT**

**International application No**

PCT/IB2005/053986

---

**A. CLASSIFICATION OF SUBJECT MATTER**

G06F 17/30

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

---

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

G05F G06F

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

---

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

EPO-Internal, WIPO Data, INSPEC

---

**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>X</td>
<td>US 6 173 275 B1 (CAID WILLIAM R ET AL) 9 January 2001 (2001-01-09) abstract column 2, line 10 - line 34 column 4, line 25 - column 5, line 27; figures 1,2 column 12, line 50 - line 65; figure 13 column 14, line 29 - line 49</td>
<td>1-17</td>
</tr>
</tbody>
</table>

---

Further documents are listed in the continuation of Box C.

---

See patent family annex.

---

* Special categories of cited documents:

  * "A" document defining the general state of the art which is not considered to be of particular relevance
  * "E" earlier document but published on or after the international filing date
  * "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
  * "C" document referring to an oral disclosure, use, exhibition or other means
  * "P" document published prior to the international filing date but later than the priority date claimed

---

"*" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered involved in an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"S" document member of the same patent family

---

Date of the actual completion of the international search 4 April 2006

Date of mailing of the international search report 12/04/2006

---

Names and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 Hl Rijswijk
Tel. (+31-70) 340-3040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016

Authorized officer Barieux, M
<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>A</td>
<td>WO 03/009173 A (SAP AKTIENGESELLSCHAFT; SCHWEDES, HOLGER)</td>
<td>1, 9, 10</td>
</tr>
<tr>
<td></td>
<td>30 January 2003 (2003-01-30) abstract</td>
<td></td>
</tr>
<tr>
<td></td>
<td>paragraph [0016]; figure 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>paragraph [0021]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>paragraph [0026] - paragraph [0028]; figure 6</td>
<td></td>
</tr>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>US 6173275 B1</td>
<td>09-01-2001</td>
<td>NONE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 0146858 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2002178158 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 1410265 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2003018617 A1</td>
</tr>
</tbody>
</table>