2015 IEEE International Symposium on Multimedia

ExploratoryVideoSearch: A Music Video Search System Based on Coordinate Terms and Diversification

Kosetsu Tsukuda
National Institute of Advanced Industrial Science and Technology (AIST), Japan
k.tsukuda@aist.go.jp

Masataka Goto
National Institute of Advanced Industrial Science and Technology (AIST), Japan
m.goto@aist.go.jp

***Abstract*—Many people search and watch music videos on video sharing websites. Although a vast variety of videos are uploaded, current search systems on video sharing websites allow users to search a limited range of music videos for an input query. Especially when a user does not have enough knowledge of a query, the problem gets worse because the user cannot customize the range by changing the query or adding some keywords to the original query. In this paper, we propose a music video search system, called ExploratoryVideoSearch, that is coordinate term aware and diversity aware. Our system focuses on artist name queries to search videos on YouTube and has two novel functions: (1) given an artist name query, the system shows a search result for the artist as well as those for its coordinate terms; and (2) the system diversifies search results for the query and its coordinate terms, and allows users to interactively change the diversity level. Coordinate terms are obtained by utilizing the Million Song Dataset and Wikipedia, while search results are diversified based on tags attached to YouTube music videos. ExploratoryVideoSearch enables users to search a wide variety of music videos without requiring deep knowledge about a query.**

***Keywords*-music video search; coordinate term; diversification**

I. Introduction

It has become common in recent years to watch music videos on video sharing websites such as YouTube [1]. Formerly, most of music videos were uploaded by professional creators; but now ordinary people can also easily upload music videos in which they dance to or cover an artist's music [1]. As a result, there are various kinds of music videos on video sharing websites. However, despite the vast variety of uploaded videos, the current user interface for video searching allows users to browse only a limited range of videos, especially when they do not have enough knowledge about the query. Let us consider two examples.

EXAMPLE 1. A user becomes interested in Celine Dion and inputs her name as a query on YouTube. The user watches videos related to Celine Dion and now wants to know other artists like her. If the user has enough knowledge, the user can input queries such as Mariah Carey or Beyonce and then watch their videos. If the user does not, however, it is not always easy to find such keywords, and thus, the user misses a chance to watch their videos.

[1]http://www.youtube.com/

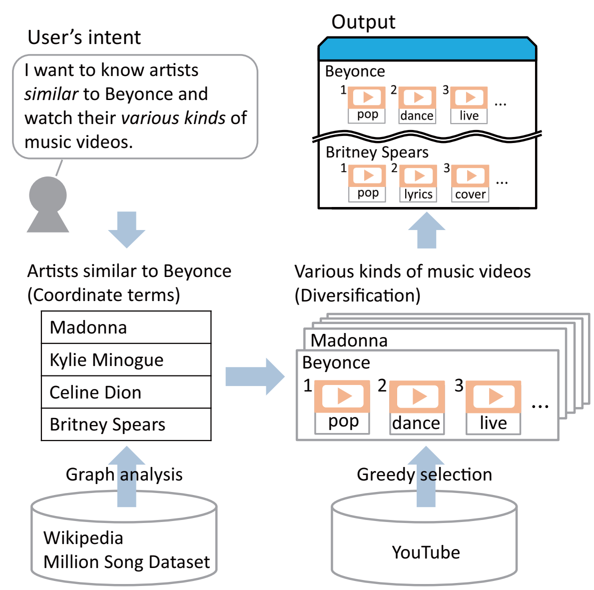

Figure 1. An overview of ExploratoryVideoSearch. Given user's intent as shown in the figure, the system first obtains artists similar to Beyonce (they are called coordinate terms) from Wikipedia and the Million Song Dataset by using graph analysis. Then the system searches for their various kinds of videos using YouTube data based on a greedy selection algorithm (diversification). Finally, the system outputs aggregated search results.

EXAMPLE 2. A user becomes interested in Michael Jackson and searches for videos by his name on YouTube. Almost all the top-ranked videos are original Michael Jackson songs and the user watches some of them and leaves the site. If the user has enough knowledge about Michael Jackson, the user can watch videos in which Michael Jackson is interviewed or a user dances to or covers his song by adding keywords such as "interview" or "cover." However, if the user does not have enough knowledge, it is difficult to think of such keywords. The user may be able to find such videos by scrolling through the search results, but most users are likely to look at only the top of the search results [2]. Thus, the user misses a chance to watch such videos.

In light of the above observations, we propose a new video search system called ExploratoryVideoSearch, which is *coordinate term aware* and *diversity aware*. Figure 1 shows an overview of ExploratoryVideoSearch. ExploratoryVideoSearch enables users to search a wide variety of videos without requiring a deep knowledge about a

978-1-5090-0379-2/15 $31.00 © 2015 IEEE
DOI 10.1109/ISM.2015.99

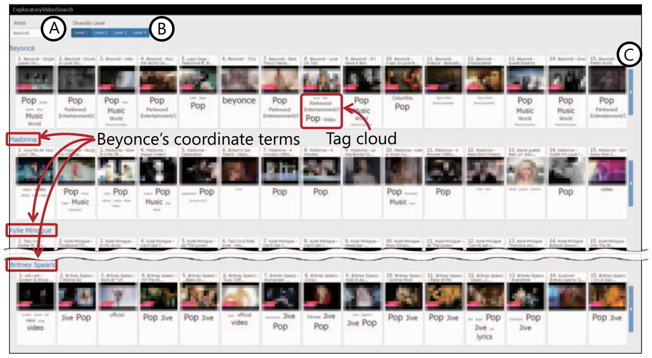

Figure 2. Search results for "Beyonce" with diversity level 1 (non-diversified search results).

query. More specifically, the main advantages of ExploratoryVideoSearch are twofold: (1) a user can find information about artists who are similar to an input query artist and browse their video search results at the same time; and (2) a user can view a wide variety of artist's videos that are related to not only the artist's popular subtopics but also the artist's minor or rare subtopics.

II. Functions

This section describes the functions of ExploratoryVideoSearch.

When a user thinks "I want to know artists similar to Beyonce and watch their various kinds of music videos," the user inputs an artist name "Beyonce" in the textbox (Ⓐ in Figure 2) and chooses the diversity level of the video search results by clicking a button (Ⓑ in Figure 2). Here, the diversity level ranges from one to four. The higher the diversity level is, the more diversified search results the user can see. In Figure 2, the user selects level 1, and then, ExploratoryVideoSearch generates search results for each of the input artist and its top four coordinate terms: Madonna, Kylie Minogue, Celine Dion, and Britney Spears (note that level 1 generates non-diversified search results). This enables the user to browse videos of artists that are similar to the input artist. Although ExploratoryVideoSearch shows only the top 15 search results for each artist, a user can examine lower ranked videos by clicking the scroll button (Ⓒ in Figure 2) and browsing a maximum of 200 videos for each artist. It also allows the user to change the diversity level by clicking another diversity level button.

In addition, ExploratoryVideoSearch displays a tag cloud under each video thumbnail so that the user can see tags that are related to each video. This tag cloud shows tags that appear equal to or more than five times in the artist's search results. The tags are displayed in random order and their font size represents its appearance frequency: 31pt, 24pt, 17pt, and 10pt for the frequencies of 50 $\geq$, 25 $\geq$, 15 $\geq$, and 15 $<$. Even if a user does not know much about an artist, they can understand the popular and rare tags for the artist from the tag cloud, and watch the corresponding videos.

ExploratoryVideoSearch also allows a user to click on an artist's name to see its coordinate terms and their search results. In addition, when a user clicks a tag in a tag cloud, the system adds the tag to the original query and its coordinate terms, and shows their search results. For example, if a user clicks "dance" in the search result of "Beyonce," the system shows search results not only for "Beyonce dance" but also for "Madonna dance," "Kylie Minogue dance," etc. The user can also change diversity level for the results. These functions enable the user to expand the user's interest from the initial query artist.

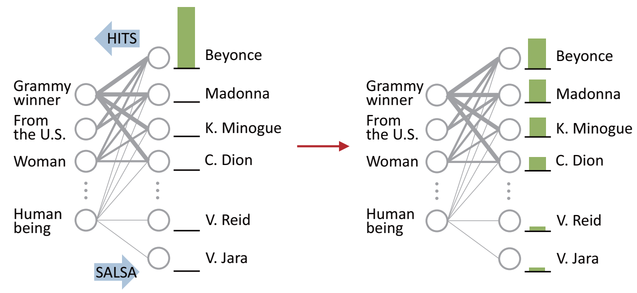

Figure 3. An example of calculating coordinate term scores for "Beyonce."

III. Algorithm

A. Coordinate Terms Collection

In this study, we collect the coordinate terms for a given query using the method proposed by Tsukuda *et al.* [3]. We first describe the dataset, and then introduce the algorithm in the following paragraphs.

This algorithm requires a dictionary containing hypernym-hyponym pairs. In this paper, we developed a dictionary containing only artists' names as its hypernyms. The first step to develop the dictionary obtains artists' names from the Million Song Dataset [4]. We use 7,576 artists who have Japanese Wikipedia articles from among the 44,745 artists in the dataset so that we can make effective use of the tool used in the second step. The second step obtains the hypernyms of each artist. We use an open source "hypernym/hyponym extraction tool" to obtain these hypernyms [5]. This tool uses machine learning to extract the hypernym-hyponym pairs from Japanese Wikipedia. We applied this tool to the 2014-11-04 version of the Wiki-JP dump data, and this gave us 71,780 hypernym-hyponym pairs for the 7,576 artists. For instance, "Beyonce" has 34 hypernyms such as "artist" and "winner of Grammy Awards."

Given a query q, we create a bipartite graph $G = (\{q\} \cup C_q \cup H_q, E)$, where C_q and H_q are the set of coordinate terms and the set of hypernyms of q, respectively, and E is a set of edges between H_q and $\{q\} \cup C_q$. An edge exists between $h_i \in H_q$ and $c_j \in \{q\} \cup C_q$ when h_i is a hypernym of c_j. Figure 3 shows the bipartite graph when q is "Beyonce." In the figure, the edge width and the height of a rectangle represent the edge weight and the coordinate term score, respectively. Only "Beyonce" has an initial value before applying the algorithm as shown in the left graph. By applying the method proposed by Tsukuda *et al.* [3], which calculates the appropriateness of each coordinate term in C_q using the HITS [6] and SALSA [7] algorithms, the scores of vertices that are connected to many vertices with bold edges

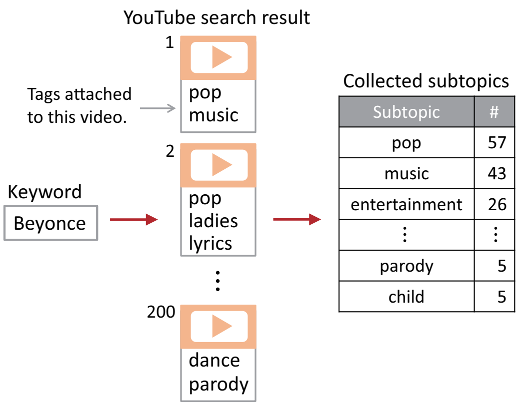

Figure 4. Subtopics mining from a YouTube search result.

become high as shown in the right graph. More details can be found in Tsukuda *et al.* [3].

B. Subtopic Mining

We obtain the subtopics for q in order to diversify the video search results retrieved using a keyword q. Dou *et al.* treated related and suggested queries scraped from the Web search engines as subtopics, while we treat a set of tags attached to the videos retrieved using a keyword as subtopics. We first generate the top N non-diversified search results for q using YouTube API [2] to obtain a set of tags. The videos are ranked in descending order by their view count. Then, we pool the tags whose appearance frequencies are equal to or higher than θ, and treat them as subtopics. In this paper, we use $N = 200$ and $\theta = 5$. Figure 4 shows the process example for the keyword "Beyonce."

C. Search Result Diversification

We utilize the diversification framework proposed by Dou *et al.* [8]. In the following paragraphs, we first intuitively explain their algorithm and then describe the detail.

In their algorithm, videos that are relevant to more important subtopics and ranked higher in the original search result (OSR) are ranked higher than those which are relevant to less important subtopics and ranked lower in OSR. The algorithm also tries to cover as many subtopics as possible in the diversified search result (DSR), by minimizing the redundant retrieved videos for each subtopic. Figure 5 shows an example of search result diversification. For simplicity, we consider top five videos in OSR and five subtopics in this example. The importance score of each subtopic is given in the table. First, the first ranked music video in OSR is also ranked first in DSR because it covers important subtopics. Next, the second ranked video in OSR has a subtopic "pop," but it is already included in DSR. Although the third ranked video in OSR also has the subtopic "pop," the video has a new subtopic "lyrics." Therefore, the third ranked video in OSR is ranked second in DSR. Similarly, the fifth ranked video in OSR is ranked third in DSR because it has two new

[2]https://developers.google.com/youtube/2.0/developers_guide_protocol_audience?hl=en

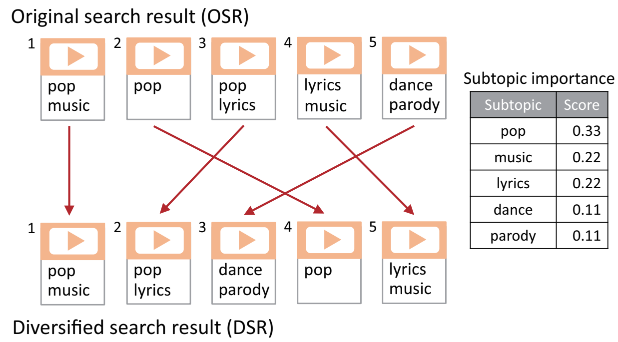

Figure 5. An example of search result diversification.

subtopics ("dance" and "parody"), even though its original rank is low. Then the second ranked video in OSR is ranked fourth because its original rank is high. Finally, the fourth ranked video in OSR is ranked fifth in DSR due to its redundant subtopics and the low rank in OSR.

We slightly modified the algorithm by Dou *et al.* [8] to diversify a music video search result based on the view counts and attached tags of the videos. We describe the modified algorithm below. Let C denote the set of subtopics and c be a member of C. Using the YouTube API, we first generate the top N non-diversified search results for a keyword q. Here, N videos are ranked in descending order of their view count. R denotes the results. Let $rank(q, v)$ denote the rank of video v in R. According to Dou *et al.*, the relevance score of video v with respect to keyword q is given by $rel(q, v) = \frac{1}{\sqrt{rank(q,v)}}$. When we compute $rel(c, v)$, which is the relevance score of v with respect to a subtopic c, we first obtain the videos that are tagged with c, and rank them in descending order by their view count. We denote the ranked list as R_c. Then, $rel(c, v)$ is $\frac{1}{\sqrt{rank(c,v)}}$ if $v \in R_c$, and 0 otherwise.

Dou *et al.* [8] uses a greedy algorithm that iteratively selects documents and generates a diversified ranking list. Let S_n denote the top n videos selected so far. The $n+1$-th video is given by

$$v_{n+1} = \underset{v \in R \setminus S_n}{\operatorname{argmax}}[\rho \cdot rel(q, v) + (1 - \rho) \cdot \Phi(v, S_n, C)]. \quad (1)$$

In ExploratoryVideoSearch, the diversity levels 1, 2, 3, and 4 correspond to ρ =1, 0.5, 0.1, and 0.001 in this equation. $\Phi(v, S_n, C)$ represents a subtopic richness score of v given the set S_n of videos already selected. In their algorithm, it is required to compute the importance of subtopic c, denoted by w_c. In this paper, w_c is given by $w_c = \frac{freq(c,R)}{\sum_{c_i \in C} freq(c_i,R)}$, where $freq(c, R)$ is the number of videos that are tagged with c in R. More details can be found in Dou *et al.* [8].

IV. Experiment

We discuss the actual performance of our system in this section. Figure 2 and Figure 6 are the original non-diversified search results with a diversity level of 1 and the diversified search results with a diversity level of 3 for "Beyonce." In terms of the coordinate terms, the results are reasonable:

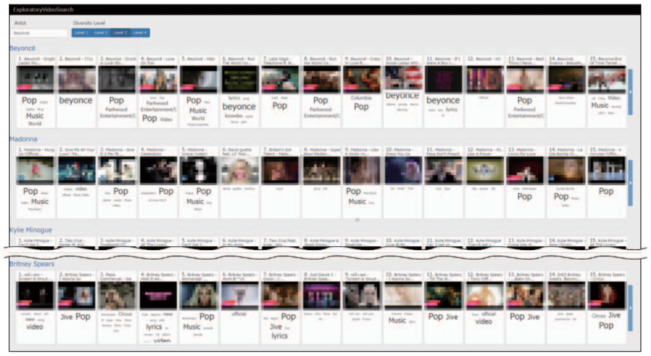

Figure 6. Search results for "Beyonce" with diversity level 3.

Madonna, Kylie Minogue, Celine Dion, and Britney Spears are among the top four coordinate terms for "Beyonce." Kylie Minogue, for example, shares nine hypernyms such as "artist," "winner of Grammy Awards," and "artist influenced by Donna Summer" with "Beyonce," and that resulted in the high coordinate score. By examining the results, a user can gain more information on artists like "Beyonce" and browse their video search results at the same time.

As for the video search results with diversity level of 3, ExploratoryVideoSearch generates well diversified search results that cover not only popular subtopics but also rare subtopics. For example, videos in the top 15 search results for "Beyonce" in Figure 2 are related to a subtopic "pop." Although the top 15 search results for "Beyonce" in Figure 6 allocate more space for videos that are related to "pop" because it is an important subtopic for "Beyonce," a video that focuses on "lyrics" is ranked at six, and a video related to "parody" and "dance" is ranked at 10: the former and latter are originally at ranks of 85 and 55, but are promoted by diversification. Similarly, in the search results for "Britney Spears" with diversity level of 3, a user can find a parody video ranked at 5 and a video in which computer graphic characters are dancing to her music at 8: they were promoted from ranks of 25 and 42 in the original search results. Figure 6 also covers more tags than Figure 2. For example, the top 15 search results for "Beyonce" in Figure 2 include 19 unique tags, while those in Figure 6 include 37 unique tags. Similarly, in Figure 2, there are 20 unique tags for "Madonna," 39 for "Kylie Minogue," 34 for "Celine Dion," and 19 for "Britney Spears," while in Figure 6, there are 29 for "Madonna," 50 for "Kylie Minogue," 46 for "Celine Dion," and 45 for "Britney Spears."

Finally, we evaluated how many subtopics were covered in the search results. 7,426 artists who had 50 $\geq$ retrieved YouTube videos were selected from 7,576 artists in our dataset for the evaluation. We counted the number of unique tags covered in the top k search results with diversity level l for each artist. Then the number of unique tags was averaged over the 7,426 artists. k was set to 5, 10, 15, and 20; l was set to 1, 2, 3, and 4. Figure 7 shows the results. Error bars indicate the standard error of the mean. It can be observed that the number of unique tags increases with the increase of diversity level at any rank. Even in the top five search

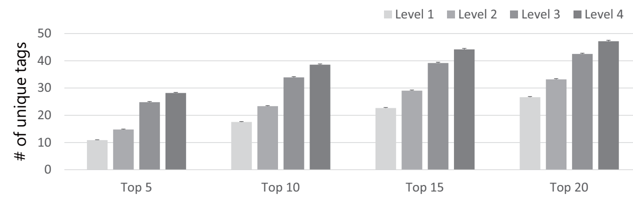

Figure 7. The average number of unique tags covered in the top 5, 10, 15, and 20 retrieved music videos with diversity level 1, 2, 3, and 4.

results, diversification has a positive effect: 10.88 tags are covered with diversity level 1, while 28.17 tags are covered with diversity level 4. These results indicate that users can watch wide variety of music videos for many queries in terms of covered subtopics by changing the diversity level.

V. Conclusion

In this paper we presented a novel music video search system called ExploratoryVideoSearch that enables users to search for music videos on YouTube in an exploratory way. Given an artist name query, ExploratoryVideoSearch generates diversified search results for the query and those for its coordinate terms. It also allowed users to interactively change the diversity level.

Our future work includes a user study that requires users to perform video search tasks and quantitatively discuss the usefulness of ExploratoryVideoSearch. Another future work is the exploration of more sophisticated diversification methods. Currently, we use only the tags attached by uploaders, but we could also use comments posted on videos by their viewers. Comment data could enable us to diversify the search results from the viewpoint of the viewers. Appropriate keywords must be extracted from the comments in order to realize this and a new diversification method needs to be proposed that balances between tags generated by the uploaders and comments generated by the viewers.

Acknowledgements

This work was supported in part by CREST, JST.

References

[1] M. Hamasaki, H. Takeda, and T. Nishimura, "Network analysis of massively collaborative creation of multimedia contents: Case study of hatsune miku videos on nico nico douga," in *UXTV*, 2008, pp. 165–168.

[2] T. Joachims *et al.*, "Accurately interpreting clickthrough data as implicit feedback," in *SIGIR*, 2005, pp. 154–161.

[3] K. Tsukuda, H. Ohshima, and K. Tanaka, "Ranking of coordinate terms and hypernyms using a hypernym-hyponym dictionary," in *WI-IAT*, 2014, pp. 15–21.

[4] T. Bertin-Mahieux *et al.*, "The million song dataset," in *ISMIR*, 2011, pp. 591–596.

[5] I. Yamada *et al.*, "Generating information-rich taxonomy from wikipedia," in *IUCS*, 2010, pp. 96–103.

[6] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *J. ACM*, vol. 46, pp. 604–632, 1999.

[7] R. Lempel and S. Moran, "SALSA: the stochastic approach for link-structure analysis," *ACM Trans. Inf. Syst.*, vol. 19, pp. 131–160, 2001.

[8] Z. Dou *et al.*, "Multi-dimensional search result diversification," in *WSDM*, 2011, pp. 475–484.