12-2011

Galaxy browser: Exploratory search of web videos

Lei PANG

Song TAN

Hung-Khoon TAN

Chong-wah NGO
Singapore Management University, cwngo@smu.edu.sg

Galaxy Browser: Exploratory Search of Web Videos

Lei Pang§ Song Tan§ Hung-Khoon Tan† Chong-Wah Ngo§
leipang3@student.cityu.edu.hk songtan@student.cityu.edu.hk
thkhoon@utar.edu.my cwngo@cs.cityu.edu.hk

Department of Computer Science§
City University of Hong Kong
Kowloon, Hong Kong

Faculty of Information & Communication Technology†
University Tunku Abdul Rahman
Perak, Malaysia

ABSTRACT

Most search engines return a ranked list of items in response to a query. The list however tells very little about the relationship among items. For videos especially, users often read to spend significant amount of time to navigate the search result. Exploratory search presents a new paradigm for browsing where the browser takes up the role of information exploring and presents a well-organized browsing structure for users to navigate. The proposed interface Galaxy Browser adopts the recent advances in near-duplicate detection and then synchronizes the detected near-duplicate information with comprehensive background knowledge derived from online external resources. The result is a topic structure on which users can easily browse and explore.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Retrieval models

General Terms

Algorithms Performance Experimentation

Keywords

Video browsing, information visualization, summarization

1. INTRODUCTION

Existing video search engines typically provide a *flat* list in response to a user query on a particular topic-of-interest. They are designed mainly to cater for *targeted search* where users have a certain degree of idea and knowledge of the topic-of-interest. However, more often than not, this is not the case. A user may not be familiar with the topic and indeed does not have a clear search goal. Instead, the search goal is changed and refined when more information is revealed. As indicated in a recent survey [1], 37% of the users were found to browse casually without any specific video in mind. Therefore, a more appealing paradigm for browse-and-search is by providing functionalities that expose different patches of visual information for efficient navigation, namely *exploratory search*. This demo will show our search engine which organizes and visualizes videos in a manner that enables users to grasp the gist of a search topic and interact with various information facets mined from videos and external resources.

To facilitate exploratory search, our system consolidates multi-faceted information to facilitate the switch of search goal and navigation path. First, the videos are linked based on overlapping content and this allows users to view the redundancy between the videos. Furthermore, when partially overlapping videos are linked, users may gain insight into how a topic evolves in content over time. Second,reliable milestone events are extracted. They constitute the high-level semantic organization of the topic which gives users a bird-eye's view of how critical a particular video and event is with respect to the whole topic. Basically, the system encompasses the *what*, *when* and *how* aspects of the topic and thus forms a very informative basis for exploratory search. Guided by the resultant structure of the topic, users can browse for videos intelligently without having to go through blindly all the videos in a flat list.

2. TECHNOLOGY

The metadata associated with web videos are often sparse and not informative which is difficult for users to find important videos. Our system addresses this challenge by exploiting visual redundancy among videos and authoring the search results with rich media information. The first part is achieved by hyperlinking hundreds to thousands of videos under a search topic to form and visualize different islands of visual content. The second part is to augment the islands with events extracted from multiple knowledge sources through media synchronization.

Given a set of somewhat correlated but diverse videos, a scene that are broadly found across different videos and in addition uploaded by different users, is usually related to an important event based on the fact that it triggers significant user interest. The set of videos associated to a common scene through near-duplicate links forms a *visual snippet*. To mine visual snippets, we employ the near-duplicate algorithm in [3] for its capability to match two partially near-duplicate videos. A video is represented using sparse keyframe representation and each keyframe is represented using Bag-Of-Words (BoW). Near-duplicate (ND) keyframe pairs are then extracted by BoW matching and Hamming Embedding [2] verification. The ND pairs are then aligned using temporal network [3] by imposing temporal constraints. The links detected by partial near-duplicate detection is used to hyperlink the videos, forming constellations of videos or visual snippets to present a galaxy view of the topic.

To infuse the visual snippets with semantic connotations,

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.
ACM 978-1-4503-0616-4/11/11.

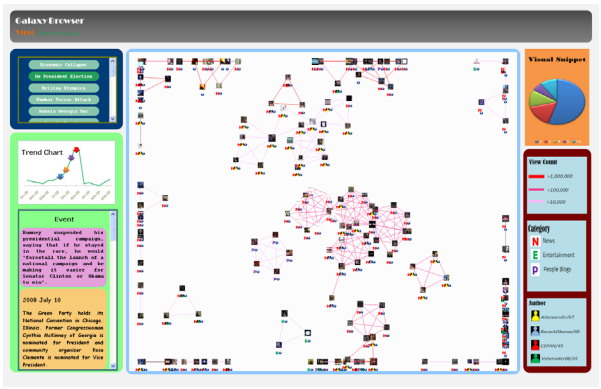

Figure 1: The main interface of the Galaxy Browser.

we rely on external resources to supplement the visual information. Directing the search using the original queries to *Wikipedia*, *Google News* and *Google Trend* retrieves useful patches of information related to the topic. Wikipedia articles provides a very concise and complete summary of the event while online news articles further offer the details of these events if they are present. Google Trend provides user browsing history where spikes in the query frequency hints the important time when milestone events occur. Hence, one major component of the system is to synchronize the visual snippets with these auxiliary modalities with respect to time. The similarity is based on textual information resulting in a heterogeneous network. Performing time-constraint alignment [4] on the network produces the milestone events, each defined by a visual snippet, atomic event and news article. With the milestone events, the Galaxy Browser can organize the video in a manner that is conducive towards exploratory video browsing.

3. BROWSING

Figure 1 shows the main interface of Galaxy Browser which comprises 3 panels. The left panel is the *key-info panel* which summarizes the topic structure mined by the system. Given a topic of interest selected from top left panel, it presents to the users the set of all milestone events in the *event list* sub-panel in the bottom of the panel. Another important feature is the trend chart grabbed from Google Trend which informs the user of the fluctuations in interest among the online community on the topic. Thus, user can acquire a certain degree of understanding on *what* happens through the milestone events, *when* it happens and *how crucial* these events are through the trend chard. In practice, the panel acts as a filtering tool. Users can rapidly grab all videos associated to a specific event (spatial filtering) in the event list or a hot time (temporal filtering) from the trend chart and the associated visual snippet will be highlighted in the middle panel, i.e., the *Galaxy Window*.

The *Galaxy Window* illustrates the set of all visual snippets which depicts how the videos are linked to each other topically. For example, a fully connected structure indicates high redundancy among the videos. Conversely. a sparsely connected graph expose the evolving nature of the particular event. This allows users to select and go through a bare minimum number of videos and yet acquire the most information within the shortest time span. More importantly, each snippet is annotated with the description from the associated Wikipedia atomic event. This gives users an in-depth textual prelude of the snippet content. An interested user can choose to venture deeper into the topic by exploring the details of the visual snippet or read the associated news article.

The right panel is the *info panel* which lists the details of a selected visual snippet or video. In this panel, detailed information for the visual snippets is given, including *category*, *author* and *view counts* of the videos. These auxiliary data is displayed in the Galaxy window as well. Indeed, though not directly describing videos' content, they tell a great deal about a snippet. For example, a snippet with mostly entertainment category videos or with very low view count is most likely unimportant. On the other hand, when videos in a snippet are uploaded by the same user with back-to-back partial links, the individual videos are likely separate parts derived from the same video and are equally important.

Another interesting feature is the *snippet window* which is activated when a user clicks a video in a certain snippet. The videos belonging to the snippet are visualized as a series of thumbnail keyframes aligned horizontally which allows user to have a first-hand preview of the video content itself without having to play them. The overlapping near-duplicate segments will also be highlighted and users will know directly how these videos are related. The users can choose a specific video for viewing.

In summary, Galaxy Browser is a new generation browser system which adopts the exploratory search paradigm. Different facets of information offered by search results are visualized by mining visual snippets. Search exploration is further facilitated by augmenting the snippets with knowledge such as the time duration of the topic, milestone events, the hot degree or intensity of each event, which provide navigational guidance. These features result in an engine that goes beyond traditional fact-finding browsers such as storyboard and video skimming for more interactive visual information seeking. Interested readers can find our demo video at http://vireo.cs.cityu.edu.hk/demovideo.wmv.

ACKNOWLEDGEMENT

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (CityU 119610).

4. REFERENCES

[1] R. D.Oliveira, M. Cherubinia, and N. Oliver. Understanding near-duplicate videos: A human-centric approach. In *ACM Multimedia*, 2009.

[2] H. Jégou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *European Conference on Computer Vision*, 2008.

[3] H. K. Tan, C. W. Ngo, R. C. Hong, and T. S. Chua. Scalable detection of partial near-duplicate videos by visual-temporal consistency. In *ACM Multimedia*, 2009.

[4] S. Tan, H. K. Tan, and C. W. Ngo. Topical summarization of web videos by visual-text time-dependent alignment. In *ACM Multimedia*, 2010.