COVER SHEET

This is the author-version of article published as:

Wang, Lei and Tjondronegoro, Dian (2006) Mining and Visualizing Multimedia Dataset on Mobile Devices in a Topic-oriented Manner. In Kotsis, G and Taniar, D and Pardede, E and Ibrahim, I.K., Eds. *Proceedings The fourth international conference on advances in mobile computing,* **pages pp. 131-140, Yogyakarta, Indonesia.**

Mining and Visualizing Multimedia Dataset on Mobile Devices in a Topic-oriented Manner

Lei Wang[1]), Dian Tjondrongoro[1])

Abstract

With significant enhancement of telecom bandwidth along with mobile devices getting cheaper and cheaper, consuming multimedia video clips "on the move" is no longer a hype. A great amount of telecom operators are now porting traditional TV service to PDA, cell phone and other mobile devices. However, a tailored solution to help mobile device user consuming multimedia items is yet available. In this paper, we propose a novel solution to organize multimedia dataset in a topic-oriented manner and visualize it into a graphic-rich map format. The main benefit is that mobile user can quickly locate the interesting topics intuitively to request the streaming of intended clips only. The topic-oriented navigation can further help users mining related multimedia data by using the automatically-constructed relationships between topics. A web-based multimedia data retrieval application has been implemented for PDA to demonstrate the feasibility and robustness of the proposed solution. Five sample video clips from four different domains (sports, news, talkshow and advertisement) have been included in the application.

1. Introduction

In the last five years, mobile devices have been undergoing a booming prosperity in our every day life. More and more handheld devices with video playback function are available in the market with a reasonable price. The advent of third generation communication networks further enables telecom operators to provide much better mobile multimedia service, such as smoother streaming time and higher quality of video resolution. Both client and server ends have been significantly improved to promote the era of consuming multimedia resources on the move.

There are several telecom operators that have already implemented mobile multimedia services in real life. The Finnish Broadcasting Company [7] have joined forces for a commercial pilot of mobile broadcasting services that will enable users to follow *on-the-air* TV and radio programs on their mobile devices. In UK, O_2 networks [8] provide the service which enable users to select their favorite program using an on-screen service guide and search for specific items. These two solutions are typical and sharing the same feature which is to stream live on-the-air TV as it is without much of users' interaction; just like the traditional TV services.

[1] School of Information Systems, Queensland University of Technology, Brisbane, Australia

Traditional TV viewing method was originally designed for static viewing from our lounge. Applying this viewing scheme to mobile services without any customization is far from perfect. On one hand, mobile device storage and battery life along with communication expense do not allow users to stream or download the whole video item (e.g. a 90 minutes soccer game or 30 minutes news). Moreover, [9] reports that most users just spend 15 or 20 minutes on mobile TV mostly during waiting time . On the other hand, the provided contents to mobile users are completely subject to content provider's will, meaning users are passively receiving the content "pushed" by provider rather than "pull" the interesting content with their own interest. In this respect, private and personal nature of mobile device is ignored. Several networks have conducted a survey on users' watching behavior [1] which reflects that most users are mainly using the mobile multimedia service to "snack on", as opposed to watching full program. Instead of replacing existing TV service, mobile services should be complementary [10] and offers more interactive means for users to watch their chosen content. As such, an effective retrieval mechanism should be tailored for mobile device users in order to help them quickly locate topic of interest and the related contents.

Researchers have already realized the importance of amplifying users' cognition of the media item's structure so that users can navigate through the content and locate interesting topic in a shorter time. Currently, browsing is one of the major solutions for navigating through multimedia contents. For example, a video can be broken down into segments and organized in a tree view [11] [18]. Some key frames from highlight events can be extracted and presented in a timeline format [11] [20]. The limitation of such browsing mechanism is that it is mostly based on temporal sequence, thus does not easily support semantic based search [22]. This can potentially stop users mining on the dataset. Most current video browsing tools are too complex, especially for mobile device user who can only afford few clicks due the limitation of input method and screen size. Video browsing mechanism can only help user discover a small portion of multimedia data and have no possibility to explore large, complex dataset. To remedy this deficiency, researchers [15] [16] have discussed information visualization concept for data mining. The basic idea of visual data exploration is to present data in a graphic format, thus user can have insight into the data, draw conclusions and directly involved themselves in the data mining process [16]. Following this paradigm, Bernd [13] pointed out the design issues when applying this concept to mobile devices. However, a concrete solution for multimedia dataset visualization on mobile device is yet available.

Based on this review, the paper will advance the state of arts by applying new approaches in the following aspects: a) ***Topicalization***: Reorganize large and complex multimedia dataset into a common topics-based layer. Thus, viewers can follow this topic-oriented manner to navigate through audiovisual content and directly involved themselves in the data mining process. b) ***Visualization***: Visualize the common topic-based layer into a graphic-rich format and illustrate relationships between topics. This makes video browsing becomes more intuitive and productive.

2. System overview

The system architecture is presented in Fig 1, which comprises two core components being *topicalization processor* and *visualization processor*. The system work flow can be formulated in the following steps:

- Step 1: Raw video clips are described in various content description metadata languages, for instance MPEG 7 [13] or SEO (Segment-Event-Object based indexing scheme).
- Step 2: Topicalization processor converts multimedia description metadata into topic-based knowledge model.
- Step 3: Visualization processor converts topic-based knowledge model into graphic navigation User Interface (UI) according to user's request.
- Step 4: users can request to stream down the intended clips based on the URL included in graphic navigation UI.

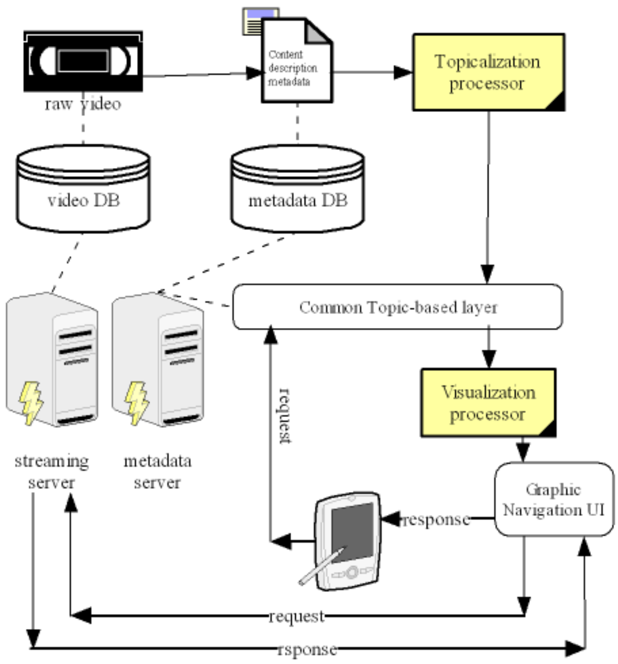


Fig 1. System architecture

3. Topicalization: Common Topic-based Layer

The common topic-based layer is to build up a universal medium on top of various formats of multimedia data and metadata. Building this layer can benefit from the following aspects: a) Unify various description metadata into one single format, thus multimedia content described in different languages can be integrated seamlessly b) Temporal sequence can be topicalized into abstract object (e.g. a goal event from frame 20 to frame 98 could be a topic) c) Topics can be cross-related so that user can be involved in data mining when exploring large, complex dataset (e.g. a goal event is related to a soccer player, the soccer player is related to a news interview). In this section, we will present the solution that formulates this topic-based layer.

3.1 Topic-based Knowledge Modeling: Topic Maps

On top of the semantic web concept, an upper layer is needed to unify various description syntaxes so that knowledge can be exchanged. Knowledge model determines how the resources are organized and presented to the user. As for now, there are two candidates for knowledge modeling being Topic Maps [2] and RDF [3]. Topic maps are an ISO standard for the representation and interchange of structured knowledge models. It is able to be utilized as an algorithm to organize information in a way that is optimized for navigation [12]. Topic maps are good at bridging topics with links such as glossaries, cross-references, thesauri and catalogs. It can also merge structured or unstructured information. The Resource Description Framework (RDF) is developed under W3C, it is an infrastructure that enables the encoding, exchange, and reuse of structured metadata.

Topic maps and RDF are sharing a very similar concept (if not the same) which is to model complex metadata and ontology as multi-dimensional graphs using XML as the interchange format [19]. However, there are some subtle differences between these models which lead to the choice according to the application context. Based on Garshol's findings [14], the major difference between these two models is how they represent association between two "things". RDF is a bit redundant and ambiguous while Topic Maps is structured more complex than RDF statements. Furthermore, Topic Maps also wrap more information in the form of the role types. As our application context focuses on locating, collating and cross-relating audiovisual contents, we identified that Topic Map is more appropriate for multimedia knowledge modeling. Another reason we adopt Topic Maps as the knowledge model is that it organizes information in a way that can be optimized for navigation. Moreover, Topic Maps can be used to break down the multimedia items, cluster resources by topics and link related topics together. After visualization of this XML-based Topic Maps into a MetaMap, users with vague exploration goal will have some clues, therefore able to navigate, browse and explore the multimedia collections intuitively.

3.2 Automatically Construct Topic-based Layer: SEO2TopicMaps Topicalization Processor

Manual construction of the topic-based layer is not feasible due to the large amount of metadata attached with multimedia items. Thus, a proper metadata model is the corner stone to success as it can feed efficient topicalization algorithm. This sub-section will present SEO modeling scheme and its conversion to topic maps (i.e. SEO2TopicMaps) algorithm for mass generation of the common topic-based layer.

SEO (Segment-Event-Object), A domain-neutral, semantic-rich modeling scheme

To have a comprehensive common topic-based layer, the modeling scheme should provide enough structural and semantic information otherwise knowledge will not be fully included and exchanged. Based on our previous work [21], we propose SEO (Segment-Event-Object) modeling scheme as the description scheme for audiovisual content. SEO highly abstracts a common ground from all kinds of audiovisual content. It identified that all audiovisual contents are structured by segment, event and objects that appears in. Segments can contain low-level (e.g. color, texture) and mid-level features (e.g. slow motion, replay, face), while event is a specialized segment which contains generic, specific and further tactical semantics of a certain domain. An object could be any entity appears in the audiovisual content which will be linked to the relevant segment or event. Using this well generalized structure, SEO can be applied to different domains as they all share the same structure. However, different domains keep their own semantic rules. For example, an object David Beckham, who is interviewed in a news program, can be annotated as an interviewee, whereas when he appears in a soccer match should be annotated as a soccerPlayer. SEO facilitates accurate description for segment, event and object depending on its domain context. The following is the SEO example to annotate an object considering the domain context it appears in:

Table 1. Object David Beckham appears in a news program

```
<SEO:person type="interviewee">
  <SEO:objectAlias>p_David_Beckham</SEO:objectAlias>
<SEO:objectId>P1</SEO:objectId>
<SEO:name>David Beckham</SEO:name>
<SEO:definition>interviewee</SEO:definition>
</SEO:person>
```

Both instances are sharing same name and referencing to the same object P1 (David Beckham as a person)

Table 2. Object David Beckham appears in a soccer game

```
<SEO:person type="soccerPlayer">
<SEO:objectAlias>p_David_Beckham</SEO:objectAlias>
<SEO:objectId>P1</SEO:objectId>
<SEO:name>David Beckham</SEO:name>
<SEO:definition>soccer player</SEO:definition>
</SEO:person>
```

Given this domain neutral description scheme, it is possible to automatically transform the structural data from various domains into a common knowledge model using one algorithm (processor) as they share common conventions of syntax, and structure.

SEO2TopicMaps Topicalization Processor

SEO2TopicMaps (read as SEO to Topic Maps) Topicalization Processor is the tool which utilizes XQuery to transform structural data SEO into Topic Maps. The algorithm constructs the Topic Maps based on SEO data instance by developing topics from segments, events and objects and building associations between them. The following algorithm topicalizes objects to Topic Maps format based on SEO instance in table 2 (in the previous page).

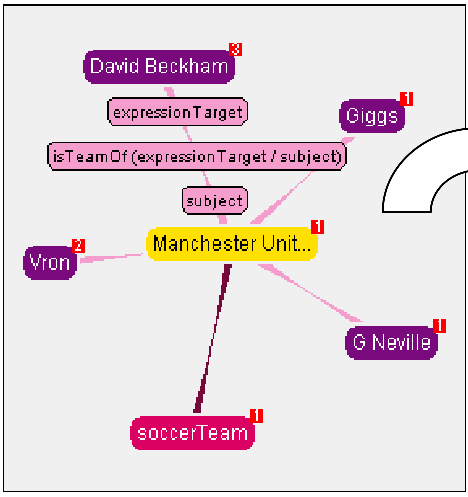
```
for $i in $vidLib/SEO:semanticObjectCollection/*
let $objectId:=$i/SEO:objectId  (:get object id = P1:)
let $objectAlias:=$i/SEO:objectAlias  (:get object AliasName = P_David_Beckham :)
let $type:=string($i/@type) (:get object Type = soccerPlayer:)
  return
    <topic id="{$objectId}">
      <instanceOf>
        <topicRef xlink:href="#{$type}"/>
       </instanceOf>
        <baseName>
          <baseNameString>{$objectAlias}</baseNameString>
        </baseName>
      ……
    </topic>
```

4. Visualization: Graphical Navigation for Multimedia Contents Exploration and Mining

In this section, we present novel strategies for multimedia data access, presentation and mining on mobile devices by visualizing XML- based Topic Maps into graphically rich format. Information visualization has been a hot topic in the data mining community as it involves human in the data exploration process by exploiting the flexibility, creativity and general knowledge of human [16]. It provides the possibility that inexperienced users can explore complex dataset with only mouse manipulation. As for mobile users, most of them are inexperienced user with a vague exploration goal and limited input ability when consuming multimedia contents. Therefore, we identified that information visualization can be well utilized on mobile devices.

Currently, there are already some middleware tools for visualizing Topic Maps. For instance, *Ontopia* [4] provides a complete set of tools for building, maintaining and deploying Topic Map based applications. Among the tools, *Ontopia Vizigator* uses java technology to provide a graphical navigation of Topic Maps, offering a visual alternative to text-only browsing. Fig 2A, B, C are rendered by Vizigator using the Topic Map generated by the SEO2TopicMaps algorithm (in the Topicalization Processor). As shown in these figures, *Manchester united* soccer club is the initial topic of a user's interest, Fig 2A shows the semantic object structure of *Manchester united.* User can keep on exploring *David Beckham* in Fig 2B and would like to know about *David Beckham is a player of foul.* Finally Fig2C denotes that this foul clip is in *play2* of *track2* which also contains one *excitement* clip and one *text* screen.

The above example demonstrates how user can explore and mine topics within one multimedia item. As discussed in the previous section, Topic Maps are able to build a universal medium on top of various multimedia data, thus knowledge can be cross-related and exchanged. The following diagram (Fig 3) depicts the Topic Map merged by two separate Topic Maps being a soccer game and sports news on the same day. Compared to Fig2B, we found that the topic *David Beckham is the news of the day* is also linked to the topic *David Beckham*.



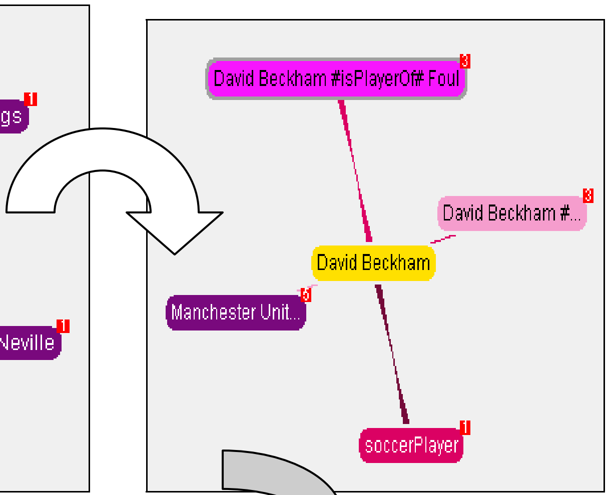


Fig 2A Initial topic "Manchester United" **Fig 2B Explore on topic "David Beckham"**

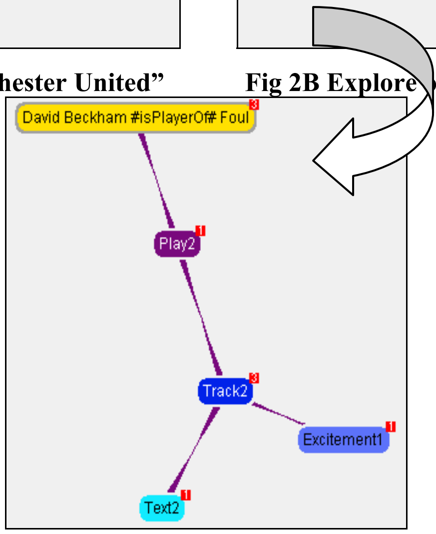


Fig 2C. Explore on topic "David Beckham is a player of foul"

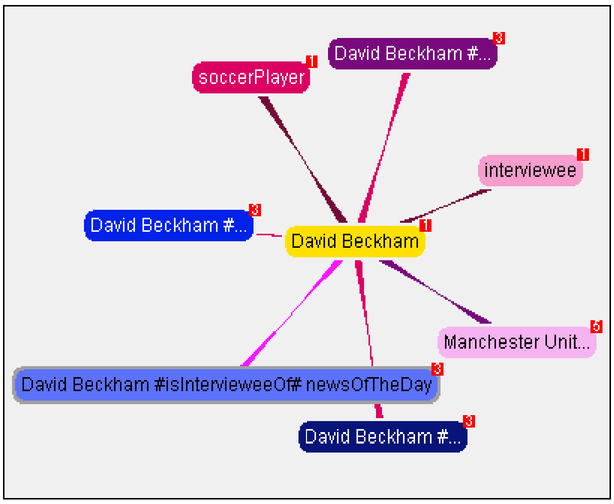


Fig 3. Topic Maps merged with soccer game and sports news

These two examples demonstrate that users can explore complex multimedia metadata more intuitively and productively with a graphical navigation interface. Furthermore, automatic data mining is achieved during exploration process with users' intention.

5. Experiment of the proposed solution

To demonstrate the feasibility and robustness of our proposed solution, we implemented a "concept-proof" web application which includes five sample video clips in four different domains being news, soccer game, talkshow and advertisement. Among them, four are described using the SEO scheme while one is described in MPEG-7 format. SEO instance is converted to Topic Maps format by SEO2TopMaps topicalization processor algorithm on the fly and stored in eXist [5] native XML database with *xtm4xmldb* [6] plug-in. MPEG-7 description is translated to Topic Maps semi-automatically. A snapshot of the dataset used in our web application is visualized in Fig 4.

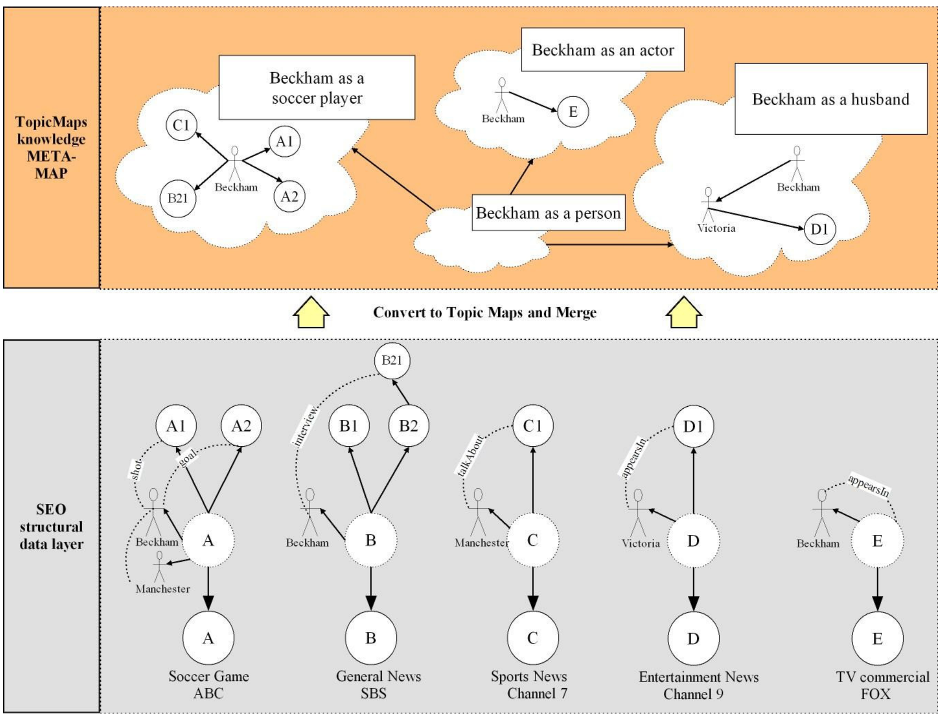


Fig 4. Top Maps Knowledge exchange

In Fig 4, David Beckham is the main topic which appears in several media items across different content providers. After the Topic Maps which represents different sources merged, the *MetaMap* automatically clustered his occurrences in different occasions by the role he played (e.g. soccer player or actor). In addition, *David Beckham* can be cross-related to other external resources via intermediate topics such as *Manchester United* or *Victoria Beckham*; viewers can keep on mining the intermediate topics if they found it interesting. The merging process is not in an ad-hoc manner as viewers should be able to control the granularity of merge vertically and horizontally in order to control the scope of result.

The following screenshots illustrate the workflow of the proposed application:

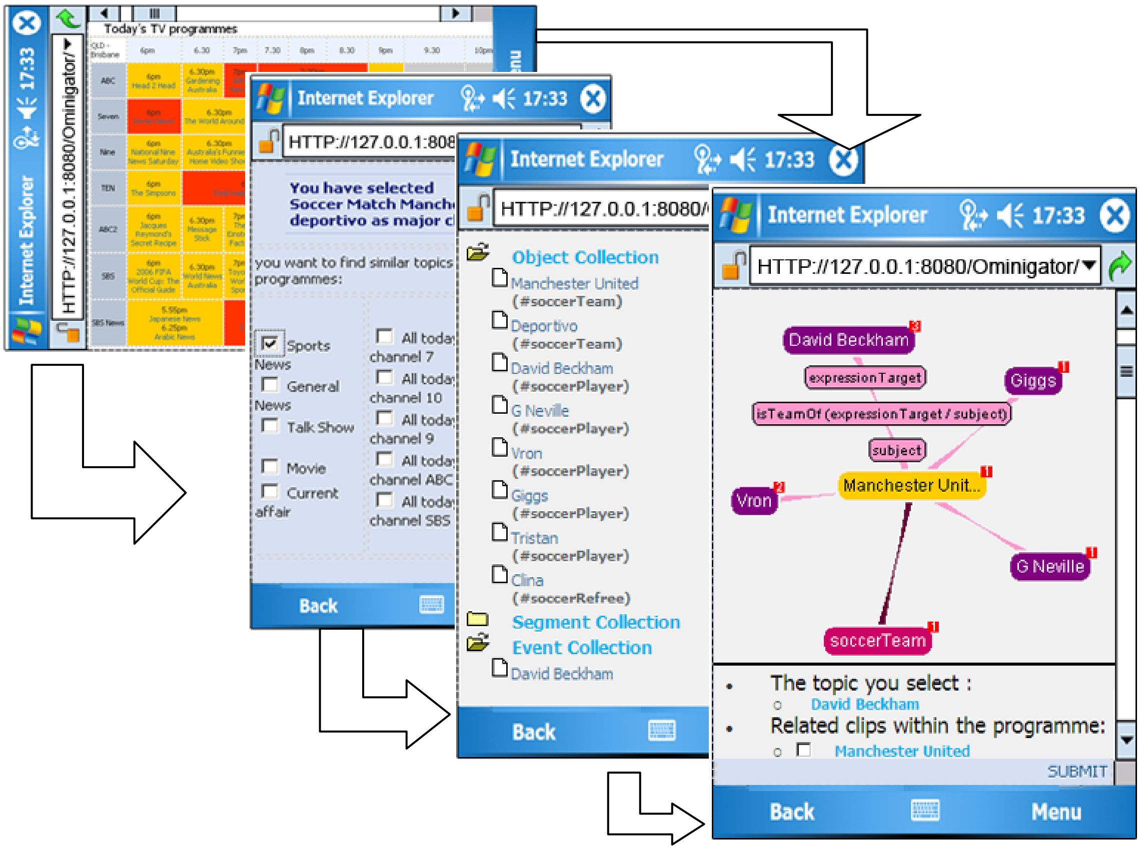


Fig 5. Screenshots of Application Work Flow

In Fig 5, the top screen is an EPG (Electronic Program Guide) which will guide user to select today's program of interest. After that, user can navigate within the selected item or go to the second screen. The second screen enables user to select other programs to merge with the program selected in the previous screen. For example, users may want to select a soccer game as the major clip and select sports news as complementary clip, thus he may view news about this match or a certain player's interview in the news if there are topics linked. By clicking submit, third screen will list all the available topics. Finally, last screen will demonstrate the selected topic's relationship with other topics both in visual and text form.

6. Conclusions

In this paper, we have presented a users-oriented solution for efficient multimedia retrieval on mobile devices which is adaptive for mobile devices and mobile consumers' behavior. Given these two contexts, we identified a new method which is based on Topic Maps to organize and present multimedia data and metadata. We have used Topic Maps as the knowledge model to reorganize complex multimedia dataset into a common topics-based layer. Therefore, viewers can follow topic-oriented objects to navigate through audiovisual content. During our experiment, we have used Ontopia Vizigator to visualize the XML-based Topic Maps files to enable user exploring and mining the interesting topics more intuitively and productively. We also presented SEO description scheme and SEO2TopicMaps algorithm to support mass generation of Topic Map files.

7. Reference

[1].Orange 3G TV Results. < http://www.3g.co.uk/PR/Sept2005/1943.htm >
[2].Topic Maps. < http://www.topicmaps.org/>
[3].RDF. < http://www.w3.org/rdf/>
[4].Ontopia. < http://www.ontopia.net/ >
[5].eXist. < http://exist.sourceforge.net/>
[6].xtm4xmldb. < http://sourceforge.net/projects/xtm4xmldb/>
[7].Finnish Broadcasting Company < http://www.finnishmobiletv.com>
[8].O2 < http://www.o2.com/about/tv_to_mobile_trial.asp>
[9].Mobile TV: About Mobile TV <http://focus.ti.com/general/docs/wtbu/wtbugencontent.tsp?templateId=6123&navigationId=12499&contentId=4445 >
[10].3 Signs UK Mobile TV Deal < http://www.cellular-news.com/story/18707.php >
[11].Belle L. Tseng, C.-Y.L., and John R. Smith. Video Summarization and Personalization for Pervasive Mobile Devices. in SPIE Electronic Imaging 2002 - Storage andRetrieval for Media Databases. January 2002. San Jose.
[12].Biezunski, M., Topic Maps at a glance.
[13].Brend Karstens, M.K., Heidrun Schumann. Visualization of Complex Structures on Mobile Handhelds. in International Workshop on MobileComputing. 2003.
[14]. Garshol, L.M., Living with topic maps and RDF. 2003.
[15]. Herman, I., Graph visualization and navigation in information visualization: Asurvey. Visualization and Computer Graphics, IEEE Transactions on, 2000. Volume: 6(Issue: 1): p. 24-43.
[16]. Keim, D.A., Information visualization and visual data mining. Visualization and Computer Graphics, IEEE Transactions on. Volume: 8(Issue: 1): p. 1-8.
[17]. Martinez, J., Overview of the MPEG-7 Standard (version 5.0).
[18]. Maryam Kamvar, P.C., Lynn Wilcox,Sandeep Casi,Surapong Lertsithichai. MiniMedia surfer: browsing video segments on small displays. in Conference on Human Factors in Computing Systems. 2004. Vienna, Austria.
[19]. Pepper, S. The TAO of Topic Maps: Finding the Way in the Age of Information. in in Proceedings of XML Europe 2000. 2000. GCA.
[20]. Qingshan Liu, Z.H., Cunxun Zang ,Xiaofeng Tong. Providing on-demand sports video to mobile devices. in International Multimedia Conference. 2005. Hilton, Singapore.
[21]. Tjondronegoro, D., Content-based Video Indexing for Sports Applications using Intergrated Multi-Modal Approach, *PHD thesis* Deakin University,Australia.
[22]. Zhang, W.Y.M.a.H. An indexing andbrowsing system for home video. in X European SignalProcessing conference. September2000. Finland.