Automatic Generation of Social Event Storyboard from Image Click-Through Data

Desam. Mangi Reddy[1], D.Rammohanreddy[2]
[1]*PursuingM.Tech (software engineering), Newton's Institute of Engineering College, Alugurajupally, macherla, Guntur dist, AP, India*
[2]*Associate Professor, Newton's Institute of Engineering College, Alugurajupally, macherla, Guntur dist, AP, India*

***Abstract*- Recent studies have shown that a noticeable percentage of web search traffic is about social events. While traditional websites can only show human-edited events, in this paper we present a novel system to automatically detect events from search log data and generate storyboards where the events are arranged chronologically. We chose image search log as the resource for event mining, as search logs can directly reflect people's interests. To discover events from log data, we present a Smooth Nonnegative Matrix Factorization framework (SNMF) which combines the information of query semantics, temporal correlations, search logs and time continuity. Moreover, we consider the time factor an important element since different events will develop in different time tendencies. In addition, to provide a media-rich and visually appealing storyboard, each event is associated with a set of representative photos arranged along a timeline. These relevant photos are automatically selected from image search results by analyzing image content features. We use celebrities as our test domain, which takes a large percentage of image search traffics. Experiments consisting of web search traffic on 200 celebrities, for a period of six months, show very encouraging results compared with handcrafted editorial storyboards**

I. INTRODUCTION

The events are detected from search log data and generate story boards where events are arranged along a time line. It is found that search log data is a good data resource for event detection because: (1) search logs cover a wide variety of real world events (2) search log directly reflect user's interests (3) search log respond to real time events. To discover events from log data, an approach called Smooth Non-negative Matrix Factorization (SNMF) framework is used. There are two basic ideas for SNMF: (1) It promotes event queries (2) It differs events from popular queries. SNMF guarantee weights for each topic to be nonnegative and considers time factor for event development. To make event detection easier, relevant images are attached for each event. There are two phases for the proposed approach: Event detection by SNMF and Event photo selection. In event detection, initially events are searched from log data. Then it discovers groups of queries that have high frequency which is known as topic factorization. Next topics with similar behaviors are merged together along a timeline which is called topic fusion. Event ranking happens in which topics like social events are highlighted. After ranking top topics are called social events and non top topics are called profile topics. In event photo selection, both the social events and profile topics are sent to search engines like Google or Bing. The search engines generate two sets of image thumbnails which contains relevant images to social events. Image similarity measures occur in which similarity between events and images are measured. Image ranking is done which is sorting of images in the social event image set. Finally all social events together with their images constructa storyboard.

II. LITERATURE SURVEY

2.1 Introduction

Literature Survey is the most important step in software program development manner. Before growing the tool it's miles necessary to decide the time thing, financial system n employer strength. Once these things are glad, ten subsequent steps are to determine which working device and language may be used for growing the tool and language can be used for developing the tool. Once The

programmers begin constructing the tool the programmers want lot of outside guide. This aid can be acquired from senior programmers, from e-book or from web sites. Before constructing the gadget the above attention r taken under consideration for developing the proposed device.

Literature survey Is the documentation of a comprehensive assessment of the published and unpublished work from secondary resources statistics inside the regions of particular interest to the researcher. The library is a wealthy storage base for secondary facts and researchers used to spend numerous weeks and sometimes months going through books, journals, newspapers, magazines, convention lawsuits, doctoral dissertations, grasp's theses, authorities guides and financial reports to locate statistics on their research topic .researcher.

2.2 Existing System

EXISTING SYSTEM:

- The most related research topics to this paper are event/topic detection from Web. There have been quite a few works that examine related directions. The most typical data sources for event/topic mining are news articles and weblogs. Various statistical methods have been proposed to group documents sharing the same stories. Temporal analysis has also been involved to recover the development trend of an event.
- The representative work for event/topic detection is the DARPA-sponsored research program called TDT (topic detection and tracking), which focus on discovering events from streams of news documents. With the development of Web 2.0, weblogs have become another data source for event detection. Some of these research efforts develop new statistical methods and some others focused on recovering the temporal structure of events.

DISADVANTAGES OF EXISTING SYSTEM:

- First, the coverage of human centre domains is small. Typically, one website only focuses on celebrities in one or two domains (most of them are entertainment and sports), and to the best of our knowledge, there are no general services yet for tracing celebrities over various domains.
- Second, these existing services are not scalable. Even for specific domains, only a few top stars are covered1, as the editing effort to cover more celebrities is not financially viable.
- Third, reported event news may be biased by editors' interests.
- Discovering events from a search log is not a trivial task.
- Existing work on log event mining mostly focus on merging similar queries into groups, and investigating whether these groups are related to semantic events like "Japan Earthquake" or "American Idol". Basically, their goals are to distinguish salient topics from noisy queries. Directly applying their approaches will fail as the discovered topics are more likely related to vast and common topics, which may be familiar to most users.

PROPOSED SYSTEM:

- In this paper, we aim to build a scalable and unbiased solution to automatically detect social events especially related to celebrities along a timeline. This could be an attractive supplement to enrich the existing event description in search result pages.
- In this paper, we will focus on those events happening at a certain time favored by users as our celebrity-related social events.

ADVANTAGES OF PROPOSED SYSTEM:

- To provide a comprehensive and vivid storyboard, in this paper, we also introduce an automatic way to attach a set of relevant photos to each piece of event news.
- We propose a novel framework to detect interesting events by mining users' search log data. The framework consists of two components, i.e., Smooth Non-Negative Matrix Factorization event detection and representative event related image photo selection
- We have conducted comprehensive evaluations on large scale real-world click through data to validate the effectiveness.

HARDWARE REQUIREMENTS:

- System : Pentium Dual Core.
- Hard Disk : 120 GB.
- Monitor : 15'' LED
- Input Devices : Keyboard, Mouse
- Ram : 1 GB

SOFTWARE REQUIREMENTS:
- Operating System : Windows 7.
- Coding Language : JAVA/J2EE
- Tool : Netbeans 7.2.1
- Database : Oracle

III. TESTING

Testing is a process of executing a program with the intent of finding an error. Testing is a crucial Detail of software fine assurance and gives ultimate evaluate of specification, layout and coding. System trying out is an crucial segment. Testing represents an thrilling anomaly for the software .

5.1 Testing Objectives:
- Testing is a process of executing a program with the intent of finding an error.
- A good test case is one that has a high probability of finding an as yet discovered error.
- A successful test is one that uncovers an as yet undiscovered error.

Testing Principles:
- All tests should be traceable to customer requirements.
- Tests should be planned long before testing begins.
- Testing should begin 'in the small" an progress toward testing "in the large"
- Exhaustive testing is not possible.
- To be most effective and independent third party should conduct testing.

Unit Testing:
Unit testing is done on individual modules as they are completed and become executable. It is confined only to the designer's requirements. Each module can be tested using the following

Two strategies:

Black Box Testing:
In this strategy some test cases are generated as input conditions that fully execute all functional requirements for the program. This testing has been uses to find errors in the following categories:
- Incorrect or missing functions.
- Interface errors
- Errors in data structure or external database access

White Box Testing:
In this the test cases are generated on the logic of each module by drawing flow graphs of that module and logical decisions are tested on all the cases. It has been uses to generate the test cases in the following cases:
- Guarantee that all independent paths have been executed.
- Execute all logical decisions on their true and false sides.
- Execute all loops at their boundaries and within

Integrating Testing: Integration testing ensures that software and subsystems work together as a whole. It tests the interface of all the modules to make sure that the modules behave properly when integrated together.

System Testing:
Involves in-house testing of the entire system before delivery to the user. Its aim is to satisfy the user the system meets all requirements of the client's specifications.

Acceptance Testing:
It is a pre-delivery test in which entire system is tested at client's site on real world data to find errors

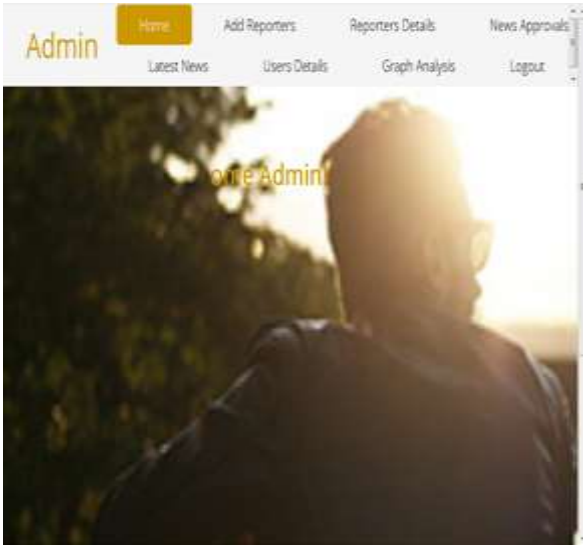
Validation:
The system has been tested and implemented successfully and thus ensured that all the requirements as listed in the software requirements specification are completely fulfilled In case of erroneous input corresponding error messages are displayed.

IV.OUTPUT SCREEN

Home Page:

Login page:

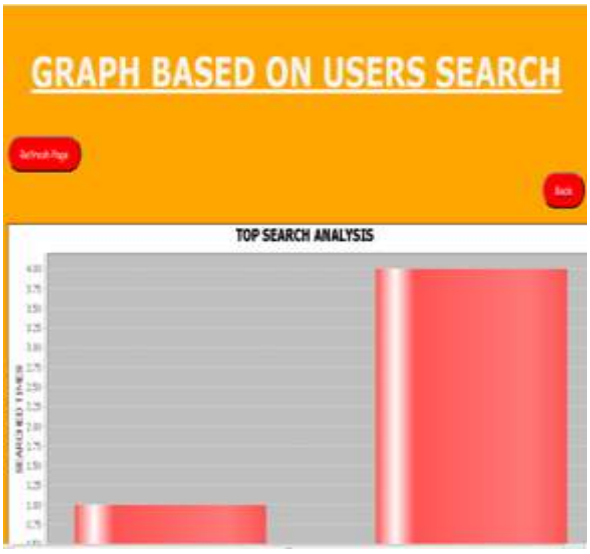
Admin home:



Add reporters:

Latest news:

Graph:



User search:

User Registration:

V. CONCLUSION

This survey has been performed for identifying the various event detection methods which are useful for event mining. It was found that search logs are a good data source for generating an efficient storyboard. SNMF together with time information is emerging as one of the better event detection methods. Moreover it highlights the benefits of mapping events to images along a timeline so as to generate automatically a storyboard. Some advantages of this approach are: there is a large coverage of domains e.g. Entertainment, sports etc., it was found more scalable i.e. it covers large number of topics and it is not at all biased by any editor's interest. Some of the applications of this approach are: monitors social events, creates storyboard and useful for content based news headings.

REFERENCES

[1] J.Allan,J.G.Carbonell,G.Doddington,J.Yamron and Y.Yang.Topic detection and tracking pilot study final report.1998.

[2] J.Weng and B.-S.Lee.Event detection in twitter,ICWSM,11:401-408,2011.

[3] D.M.Blei.Introduction to probabilistic topic models.Comm.ACM,55(4):77-84,2012.

[4] H.L.Chieu and Y.K.Lee.Query based event extraction along a timeline.In proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval,pages 425-432.ACM,2004.

[5] Y.Yang,T.Pierce & J.Carbonell.A study of retrospective and online event detection. In proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval,pages 28-36.ACM,1998.

[6] JunXu,TaoMei,Seniormember, IEEE,Rui Cai, Member,IEEE,Houqiang Li,Senior Member,IEEE and Yong Rui,Fellow,IEEE.Automatic Generation of Social Event Storyboard from Image Clickthrough Data, DECEMBER 2015