

Planned Events Across Social Media Sites using Association Rule Mining Based on Autocorrelation

Prajakta K. Sarolkar, Meghna Nagori

Abstract- User-contributed Web data contains rich and diverse information about a variety of events in the physical world, such as shows, festivals, conferences and more. This information ranges from known event features (e.g., title, time, location) posted on event aggregation platforms (e.g. Event Brite, Face book events) to discussions and reactions related to events shared on different social media sites (e.g., Twitter, YouTube, Flickr). In this paper, we propose the challenge of automatically identifying user-contributed content for events that are planned and, therefore, known in advance, across different social media sites. We mine event aggregation platforms to extract event features, which are often noisy or missing. We use these features to develop query formulation strategies for retrieving content associated with an event on different social media sites.

Keywords- This information ranges from known event features (e.g., title, time, location) posted on event aggregation platforms

I. INTRODUCTION

Event-based information sharing and seeking are common user interaction scenarios on the Web today. The bulk of information from events is contributed by individuals through social media channels: on photo and video-sharing sites (e.g., Flickr, YouTube), as well as on social networking sites (e.g., Face book, Twitter). This event-related information can appear in many forms, including status updates in anticipation of an event, photos and videos captured before, during, and after the event, and messages containing post- event reflections. Importantly, for known and upcoming events (e.g., concerts, parades, and conferences) revealing, structured information (e.g., title, description, time, and location) is often explicitly available on user-contributed event aggregation platforms (e.g., Last.fm events, Event Brite, Face book events). In this paper, we explore approaches for identifying diverse social media content for planned events.

Automatically identifying social media content associated with known events is a challenging problem due to the heterogeneous and noisy nature of the data. These properties of the data present a double challenge in our setting, where both the known event information and its associated social media content tend to exhibit missing or ambiguous information, and often include short, ungrammatical textual features.

Manuscript received June, 2013.

Prajakta K. Sarolkar, Dept. Of Computer Science & Engg. Government College of Engineering. Aurangabad (M.S.) India.

Meghna Nagori, Dept. Of Computer Science & Engg. Government College of Engineering. Aurangabad (M.S.) India.

II. RELATED WORK

We describe related work in three areas: quality content extraction in social media, event identification in textual news, and event identification in social media. Existing approaches to find and organize social media content associated with known events are limited in the amount and types of event content that they can handle.

Most related research relies on known event content in the form of manually selected terms (e.g., “earthquake,” “shaking” for an earthquake) to describe the event. These terms are used to identify social media documents, with the assumption that documents containing these select terms will also contain information about the event.

Unfortunately, manually selecting terms for any possible planned event is not a scalable approach. Improving on this point, a recent effort used graphical models to label artist and venue terms in Twitter messages, identifying a set of related Twitter messages for concert events. While this work goes a step further in automating the process of associating events with social media documents, it is still tailored to a particular type of event (i.e., concerts) and restricted to a subset of the associated social media documents (i.e., documents containing venue and artist terms). Importantly, these related efforts focus on identifying site-specific event content, often tailoring their approaches to a particular site and its properties. We experiment with formulating queries for each social media site individually, and also explore ways to use retrieved content from one site to improve the retrieval process on another site. Our contributions are as follows:

➤ We pose the problem of identifying social media content for known event features as a query generation and retrieval task.

We develop precision-oriented query generation strategies using known event features.

III. DATA FLOW DIAGRAM

It is a simple graphical formation that is used to represent a system in terms of the input data to the system, various processing carried out using these data, and then comes the output, output data is generated by the system. Figure 1 shows the data flow of the system. First step is the login step. For every login there is authorized login and the user login. Secondly, we can view the details of user. Afterwards user can create schedules which are further stored in the database, which can be used while searching schedules accordingly. This process is followed by twitter, orkut, facebook as well as social site but in social site it additionally includes databases of above mentioned three social sites for searching all schedules.

Planned Events Across Social Media Sites using Association Rule Mining Based on Autocorrelation

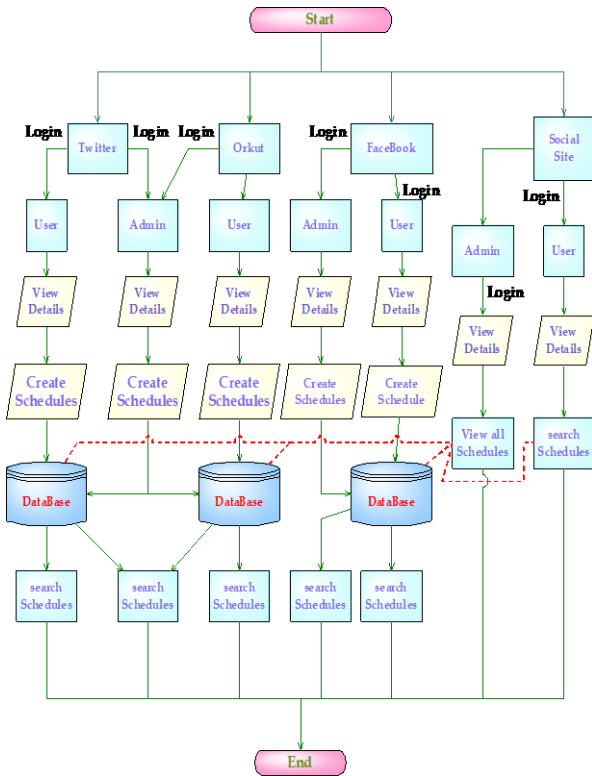


Fig1.DataFlow Diagram.

Use case diagram:

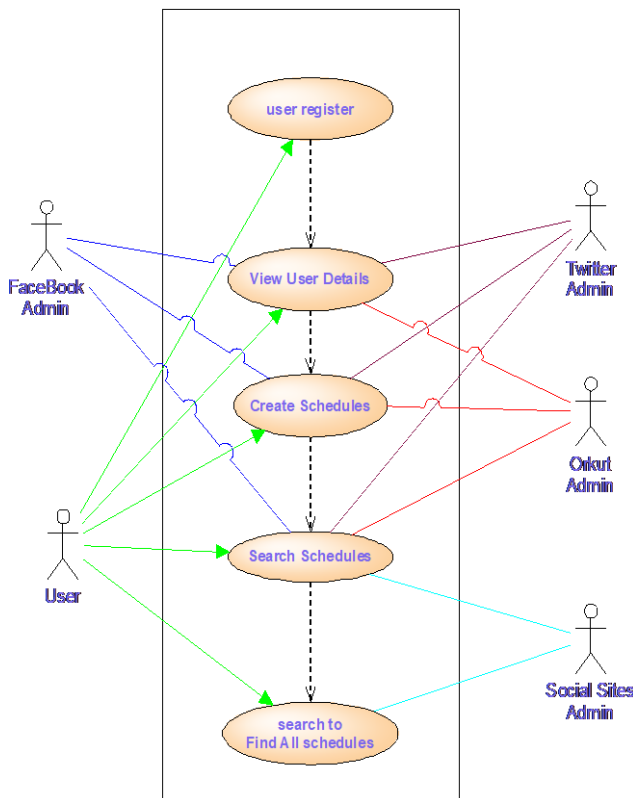


Fig2.Use-Case Diagram

Fig 2 . Shows the use-case diagram the which shows the steps evolved by user as well as by admin.The above fig describes the details done by user as well as by administration which are user registration, view user details, createshedules, searchshchedules, search to find all shedules

for user and viewing user details, creating shedules, search shedules for more popularity for admin respectively

Activity diagram:

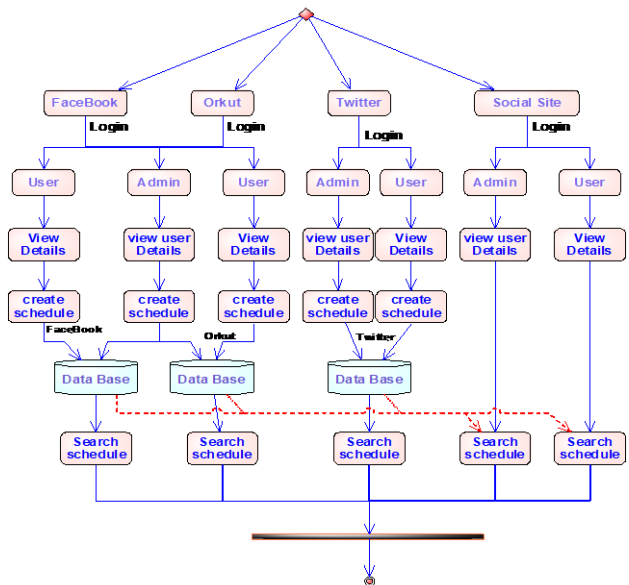


Fig.3 Activity Diagram

Sequence diagram:

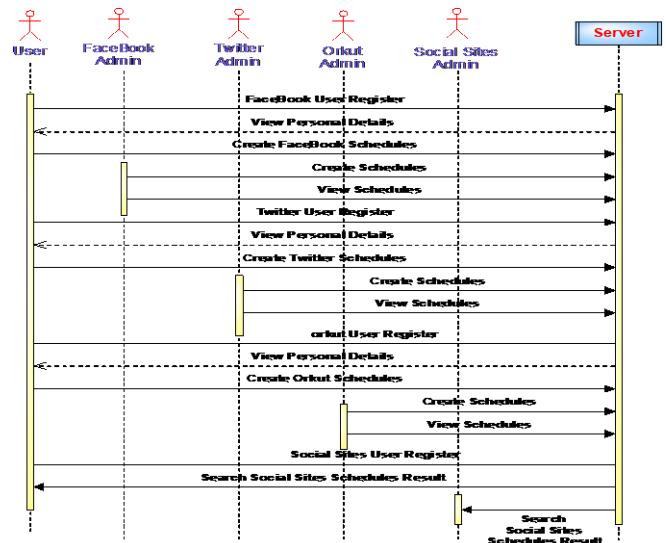


Fig.4 Sequence Diagram

Above diagram shows sequence processing of the events while user login.

IV. CONCLUSION

In this paper, we presented a query-oriented solution for retrieving social media documents for planned events across different social media sites. This work provides an essential step in the process of organizing social media documents for events, towards improved browsing and search for event media. Using a combination of precision-oriented and recall-oriented query generation techniques, we showed how to automatically and effectively associate social media documents with planned events from various sources.



Importantly, we demonstrated how social media documents from one social media site can be used to enhance document retrieval on another social media site, thus contributing to the diversity of information that we can collect for planned events.

REFERENCES

1. E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding high-quality content in social media. In Proceedings of the First ACM International Conference on Web Search and Data Mining (WSDM '08), 2008.
2. J. Allan, editor. Topic Detection and Tracking: Event-based Information Organization. Kluwer Academic Publisher, 2002.
3. H. Becker, F. Chen, D. Iter, M. Naaman, and L. Gravano. Automatic identification and presentation of Twitter content for planned events. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM'11), 2011.
4. H. Becker, M. Naaman, and L. Gravano. Learning similarity metrics for event identification in social media. In Proceedings of the Third ACM International Conference on Web Search and Data Mining (WSDM '10), 2010.
5. H. Becker, M. Naaman, and L. Gravano. Beyond trending topics: Real-world event identification on Twitter. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM '11), 2011.
6. H. Becker, M. Naaman, and L. Gravano. Selecting quality Twitter content for events. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM '11), 2011.
7. E. Benson, A. Haghighi, and R. Barzilay. Event discovery in social media feeds. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT '11), 2011.
8. R. Crane and D. Sornette. Robust dynamic classes revealed by measuring the response function of a social system. Proceedings of the National Academy of Sciences, 105(41):15649–15653, 2008.
9. W. B. Croft, D. Metzler, and T. Strohman. Search Engines: Information Retrieval in Practice. Addison-Wesley Publishing Company, 2009.
10. W. Dakka and P. G. Ipeirotis. Automatic extraction of useful facet hierarchies from text databases. In Proceedings of the IEEE 24th International Conference on Data Engineering (ICDE '08), 2008.
11. N. Diakopoulos, M. Naaman, and F. Kivran-Swaine. Diamonds in the rough: Social media visual analytics for journalistic inquiry. In Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST '10), 2010.
12. Events, 2002. In Stanford Encyclopedia of Philosophy. Retrieved June 2nd, 2010 from <http://plato.stanford.edu/entries/events/>.
13. V. Hatzivassiloglou, L. Gravano, and A. Maganti. An investigation of linguistic features and clustering algorithms for topical document clustering. In Proceedings of the 23rd ACM International Conference on Research and Development in Information Retrieval (SIGIR '00), 2000.