

Usability Design and Testing of an Interface for Search and Retrieval of Social Web Data

Dimitris Spiliotopoulos¹, Ruben Bouwmeester², Georgios Kouroupetroglou³,
Pepi Stavropoulou³, and Dimitrios Tsonos³

¹ Innovation Lab, Athens Technology Centre, Greece
d.spiliotopoulos@atc.gr

² New Media, Innovation, Deutsche Welle, Bonn, Germany
ruben.bouwmeester@dw.de

³ National and Kapodistrian University of Athens,
Dept. Informatics and Telecommunications, Greece
{pepis, koupe, dtsonos}@di.uoa.gr

Abstract. The vast amount of data on the web has been extensively harvested for many years for the purpose of digital archiving. In the recent years, however, the social networks contain the sources of most of the debating between the people. Recent approaches include social web information to the archived content for various reasons. This work reports on the usability design and evaluation of a search and retrieval user interface that was designed to retrieve web objects along with semantic information analyzed for the social web. The main task of the interface was to combine the social information with the standard archived content in meaningful and usable ways.

Keywords: Search and retrieval user interfaces, social network information, usability.

1 Introduction

The Social Media provide a rich source of, for the most part, user generated content that exhibits specific properties that are very useful for the web users. Events are reported as they happen by users that are witnessing them on sites [1]. Users voice their opinions on important aspects of life such as politics, economy, war, etc. Other people endorse those opinions by stating facts and discussing topics by referring to the important events and entities that play crucial roles to the topics discussed. Importance of events can be traced by analyzing the user involvement in the social media and diversity of opinions on them. The ARCOMEM project¹, among others, provides and integrating system for crawling and analyzing the social web in order to use the derived semantic information for the archiving process. That leads to the question: how can, then, the archived web objects be retrieved along with semantic information that will be used to provide the users with a complete picture of the events that occurred and the peoples' opinions on them at that time?

¹ ARCOMEM: Archive Communities Memories, www.arcomem.eu, FP7-ICT-270239.

Recent research stresses the importance of social media and the fact that it is imperative that they are accessible and usable by everybody [2]. From the other end, power users such as journalists and political analysts collect, aggregate and analyze the social web data in order to identify the society feedback on the events that are important [3,4]. That information is used by the journalists to report on the trends and stories [5]. It is also used by political analysts and parliamentary assistants to collect and appraise the societal opinion on specific political issues [6].

The archived social content that the ARCOMEM system collects, analyzes and preserves contains both web archives and social analysis results [7]. The challenge for archivists is identifying the important content that should be preserved for the future and how the social web analysis can be leveraged to succeed in this task [8]. Similarly the challenge for the HCI designer is providing the means for archivists and end users, such as journalists and political analysts, to access the archived content.

The motivation was to design a user interface that serves as the means to accessing the vast amount of information which can be categorized according to type (multimodal: text, video, sound, picture), role players (agents, users, opinion leaders) and actual semantic meta-data (entities, opinions on entities, etc.). The complexity of the diverse semantic data (huge lists of roles, entities, linked topics and events, opinions and trends, etc.) had to be controlled so that the users would be able to retrieve data that were both correctly filtered to their search parameters but also usable. The technical soundness depends on the indexing and retrieving of the correct information quickly and accurately and can be tested systemically. In terms of usability, the users were involved in all iterations of the development lifecycle in order to ensure that the requirements were refined and the results were adequately evaluated.

This work reports on the usability design and testing of a social web search and retrieval interface for archived documents. The next sections describe the requirements analysis, the usability testing methods and their results.

2 Requirements Analysis

Concrete requirements for reporting the peoples' opinions on important events and associating them with candidate content pages for have been laid out. These stress the important role of the social media for such task. Harvesting the opinions of the people and enabling the retrieval of such social information for the future generations will provide a unique eye in the history and preservation of events. It also validates the selection of the content appropriate for archiving based on the impact of the recorded information that was reported in it.

The requirements for the search and retrieval of semantically analyzed archived content were considered on the domain, content use and interaction levels. Certain events and the opinions on the social web may be short-lived (e.g. a political speech) while others may span several months, even years (e.g. EU economic crisis).

Related research denotes that for searching purposes the user's semantic intention can be identified and described by concrete high level semantic actions [9]. An example of semantic intention is "I want to find all web articles that discuss the EU plan for economy growth and are mentioned positively by the people". In order to accommodate such query, the analysis of the text contained in the web resources

should include EU politicians that talk about the economic growth and also be able to retrieve enough social media resources, such as from Twitter, that bear positive sentiment on the above events and entities [10].

Taking the above into consideration, a main functionality for our initial design approach is the continuity of the data provision as well as ensuring that the search queries return correct results from the web archive. Non functional requirements analysis was based on a pilot experiment that studied the user perception toward the social media derived information in order to understand the intentions of the user semantic search parameters [11].

3 Usability Testing Methodology

Based on high fidelity mockups with sample data, generic results regarding fusing and visualizing social content with semantically driven context-sensitive information were derived. The semantic meta-data were integrated to the design of the interaction flow, so that users may facet and filter their search by using several levels of semantic content as well as prominent traditional information such as social network source information. The aim of the experiments at this stage was to derive an automated testing specification based on user feedback on the usability issues. Twenty users with social media analysis and archivist background were asked to evaluate the UI using both ad-hoc and systematic testing. The human interpretation of the requirements was validated through the semantic search results and the scenarios (two transcribed and one free form) have been used to compute the expected results and filtering options from the users. At the same time, response feedback and user selection were logged by the system in order to formulate the base search model parameters. Information availability varied considerably throughout the tests. Semantic data from the social web included several layers of semantics such as:

- ranked user roles
- importance of users and opinion leaders
- user contribution levels (involvements and social media source)
- entities occurrence
- opinions (positive, negative, neutral) about entities and events
- trending (opinion and occurrence over time)

The results were analyzed and a specification model was created in order to optimize the information flow (type and volume) as semantic search-dependent rules. The users were monitored for the whole duration of their interaction with the system. Their semantic search intentions were recorded using the think aloud method and their activity was both logged from the system and verbally justified during the scenarios.

4 Results

The users completed the scenarios pretty quickly and without any generic issues. They verified findings from earlier evaluation cycle that the actual opinion marker for

each search result, as shown at the top right position of each item in Figure 1, bore more importance than the title text for initial selection.

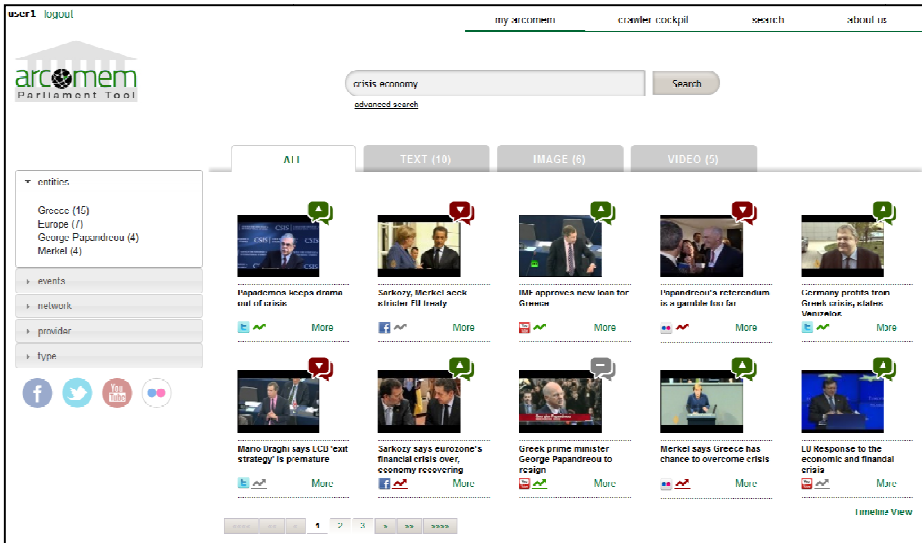


Fig. 1. Semantic search results within the political domain

Many users made the qualitative suggestion that having information for both positive and negative opinion statistics for the results would suffice. That should replace the information on trending and social media source that was used. The actual time that the users spend on each type of activity was computed and normalized (Fig. 2). It is obvious that interaction with the raw content, such as list of entities, events and opinions on them was the most important activity.

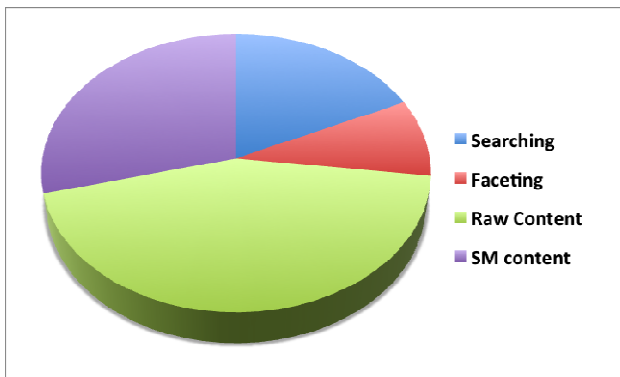


Fig. 2. Relevant time spent on activity type

The social media information was reported as interesting and the results in Figure 2 were deemed appropriate considering the much larger body of raw data compared to the social media data available for the examined evaluation cycle. Semantic analysis results have been used to enrich the archived content with semantic descriptions and tags that include entities (persons, locations, dates, etc.), events, topics, opinions on them, trending, cultural dynamics and more. Each web document may include several pieces of semantic information that may or may not be useful to the end user. That information was fully used by the evaluators during the interaction. At the end of all scenarios they were asked to provide feedback on a Likert scale 1-7 regarding their evaluation of the search functionality of the interface as well as the importance of the content provision according to the type itself (Fig. 3).

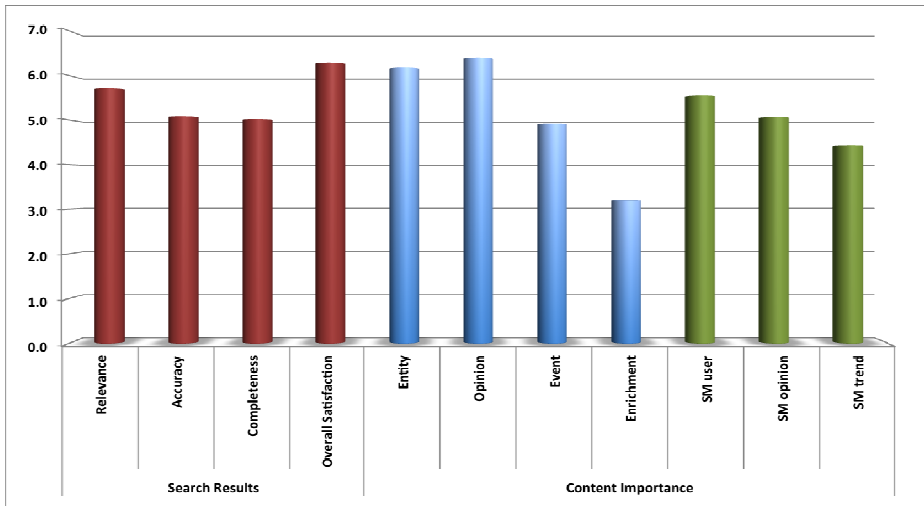


Fig. 3. User subjective feedback on search functionality and archived content provision

The user evaluation has indicated that the core semantic analysis can provide important information on the archived documents and that the search and retrieval process can be based on that. Other types of analyses, such as entity enrichment may be used but bear low level of importance. The importance of the semantic information is almost as important for the social media content (shown in blue) as is for the raw content (in green).

5 Conclusion

This work reports on the usability considerations and testing of a specialized user interface for socially-aware semantic search and retrieval of web documents. The user perception of importance of social media types as well as the analysis of semantic search patterns of users was evaluated and has lead to a specification for information selection and provision based on the importance for the search and retrieval of the archives as well as the appreciation of the actual types of semantic data.

Acknowledgements. The work described here was partially supported by the EU ICT research project ARCOMEM: Archive Communities Memories, www.arcomem.eu, FP7-ICT-270239.

References

1. Vaataja, H., Vainio, T., Sirkkunen, E., Salo, K.: Crowdsourced News Reporting: Supporting News Content Creation with Mobile Phones. In: Proc. 13th Int. Conf. on Human Computer Interaction with Mobile Devices and Services, pp. 435–444. ACM (2011)
2. Dale, O., Drivenes, T., Tollefsen, M., Reinertsen, A.: User Testing of Social Media – Methodological Considerations. In: Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W. (eds.) ICCHP 2012, Part I. LNCS, vol. 7382, pp. 573–580. Springer, Heidelberg (2012)
3. Spiliotopoulos, D., Cabulea, C., Frey, D., Lasfargues, F., Peters, W., Risse, T.: Journalists in Times of Social Media: An Application Centric Social Web Preservation Approach. In: World Wide Web Journal, Special Issue on Social Media Preservation and Applications, Springer, Heidelberg (2013)
4. Diakopoulos, N., De Choudhury, M., Naaman, M.: Finding and Assessing Social Media Information Sources in the Context of Journalism. In: Proc. 2012 ACM Annual Conf. Human Factors in Computing Systems, CHI-2012 (2012)
5. Hermida, A.: Twittering the News: The Emergence of Ambient Journalism. *Journalism Practice* 4(3), 297–308 (2010)
6. Spiliotopoulos, D., Schefbeck, G., Koryzis, D.: Obtaining Societal Feedback on Legislative Issues through Content Extraction from the Social Web. In: Proc. 16th Int. Symp. Legal Informatics, Salzburg, Austria, pp. 21–23 (February 2013)
7. Risse, T., Dietze, S., Maynard, D., Tahmasebi, N., Peters, W.: Using Events for Content Appraisal and Selection in Web Archives. In: Proc. Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web (DERIVE 2011), in Conjunction with the 10th Int. Semantic Web Conference 2011 (ISWC 2011), Bonn, Germany, October 23 (2011)
8. Schefbeck, G., Spiliotopoulos, D., Risse, T.: The Recent Challenge in Web Archiving: Archiving the Social Web. In: International Council on Archives Congress, Brisbane, Australia, 20–24 August (2012)
9. Chen, Z., Lin, F., Liu, H., Ma, W.-Y., Wenyan, L.: User Intention Modeling in Web Applications Using Data Mining. In: World Wide Web: Internet and Web Information Systems, 5th edn., pp. 181–191. Kluwer (2002)
10. Maynard, D., Funk, A.: Automatic detection of political opinions in tweets. In: García-Castro, R., Fensel, D., Antoniou, G. (eds.) ESWC 2011. LNCS, vol. 7117, pp. 88–99. Springer, Heidelberg (2012)
11. Spiliotopoulos, D., Tzoannos, E., Stavropoulou, P., Kouroupetroglou, G., Pino, A.: Designing User Interfaces for Social Media Driven Digital Preservation and Information Retrieval. In: Miesenberger, K., Karshmer, A., Penaz, P., Zagler, W. (eds.) ICCHP 2012, Part I. LNCS, vol. 7382, pp. 581–584. Springer, Heidelberg (2012)