BVG Trust

J. Sci. Trans. Environ. Technov. 2018, 11(3) : 149-154

Scientific Transactions in Environment and Technovation

Big Data Deliverables through Cross-Platform Interest-Based content discovery

https://doi.org/10.56343/STET.116.011.003.007
http://stetjournals.com

V.Geetha*, R.Anitha and M.V. Srinath

* Department of computer science, S.T.E.T Women's college,Sundarakottai,Mannargudi,Tamil Nadu.
Department of computer science, S.T.E.T Women's college, Sundarakottai,Mannargudi,Tamil Nadu.
Department of Computer Applications, S.T.E.T Women's college, Sundarakottai,Mannargudi,Tamil Nadu.

Abstract

Big data is a broad term for data sets so large or complex that traditional data processing applications are inadequate. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying and information privacy. The goal of this study was twofold: first we address the theoretical dilemmas of a cross-platform user experience; second, we implemented an Android-based mobile application and designed a cloud architecture to account for theoretical parameters of Big Data User-centric approach and interactivity. To address cross-platform Big Data challenges, we relied on cloud computing to perform computationally intensive operations such as searching, data mining, and data processing at large scale. on and content filtering across multiple radio content streams. The streams consisted of tags from radio stations' programming and social media content through a discovery process. User interaction was geared to enable preferred topic filtering, flexible shifting participation roles, notifications, and navigation through external data sources. We tested our application on a list of popular radio stations and their social media content streams (including Facebook, Twitter, Google+) to generate a Big Data scenario.

Key words: big data, interest-based content discovery, user-centric approach, cross-platform media, cloud computing.

Received : June 2017 Revised and Accepted : February 2018

INTRODUCTION

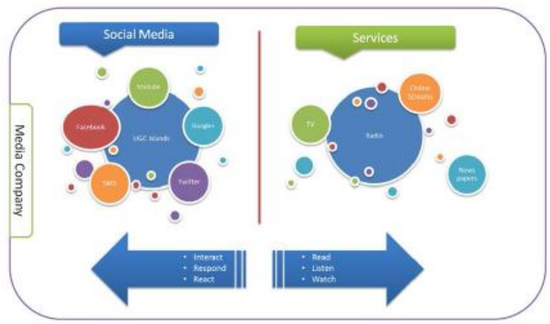
Big Data technology (Beyer *et al*., 2007) is a new technology that aims to efficiently obtain value from very big volumes of a wide variety of data, by enabling high velocity capture, process, store, discovery and / or analysis, while ensuring their veracity by an automatic quality control in order to obtain a big value and make decision. Big Data is described by what is often represented as a multi-V model. In multi-V model, volume, velocity and variety are the items most commonly recognized. Variety means the various data types, velocity represents the rate at which the data is produced, captured, processed and analyzed, and volume defines the amount of data from difference data sources and data produced from the system itself. Veracity means how much the data can be trusted given the reliability of its source (Ellison, 2007). A myriad of social media platforms emerged in recent years with services geared towards users through adds-on such as mobile texting, Facebook (Beyer *et al*., 2007) with increasing popularity of Twitter, Google+ and WhatsApp, especially in entertainment contexts. These social media platforms, predominantly consisting of social networking sites (SNSs), heavily rely on individual users for content creation, in contrast to professionally produced content. SNSs' success hinged on constant user involvement and participation (Ellison, 2007). With forty-one percent of the US population finding photos and videos online, interest-based content discovery became the driving force for new content generation and redistribution (Lee and Brenner, 2012). As social media permeates all spheres of our lives and these applications generate considerable percentage of Internet traffic, content streams remain fragmented thus limiting to discover interest-based relevant content to their users. We considered interest as an individual experience, continuously stimulated by relevant content discovery. Single-platform access inevitably leaves a proportion of interest-based content underexposed. Single-platform SNSs, even historical ones, varied technologically and scope-wise, ranging from user demographics, geographical attributes, or mere maintenance of pre-existing relationships. Since 2003 specialized social networking sites became mainstream, focusing on specific interests such as traveling, activism, religion, photo sharing, music listening, and video sharing to mention a few (Ellison, 2007). SNSs evolved and broadened their scope over time. For example, access to Facebook is open to

*Corresponding Author :
email: kkmannaig@gmail.com

everyone, even if initially it was restricted to college student networks. Some of these limitations were addressed to overcome limited content access, platform interoperability issues, and lack of relevant content segmentation across multiple platforms. Attempts to facilitate interest-based content access thus started to be modeled within a single platform. Some of the techniques included "like" feature on Facebook; Twitter content following and filtering were implemented by using "hashtags." Regardless of these attempts, interest-based content still can be searched solely within a single platform rather than across multiple platforms, not even considering user interaction with other users or content across through various platforms. Several social media cross-platform applications pioneered to account for single-platform content access limitations. Interest-based content redistribution was facilitated by "share" function easier content access to multiple platforms was provided through a open identity (Recordon and Reed , 2006) to account for increased content variety content aggregation tools we developed to combine functionalities from multiple external sources (Twitter,2013;Yoono,2013). In addition, for users, it takes time, effort, and cognitive capacity to follow multiple platforms with equal dedication Bucy, 2004. However, all these cross-platform applications continue to bear limitations. Although the "share" function allowed content to be broadcast or duplicated across various platforms, the downside of such approach was that the user could engage in one-to-many content distribution, but remained limited to receive contents from each separated platform individually. Open identity (Foundation, 2013) facilitated access to content by allowing users to sign in to multiple websites with a single identity (ID). Such an open ID remained limited to a targeted platform rather than to multiple parallel platforms. Content aggregation platforms in turn provided users with larger amounts of content access, yet did not support interaction and content discovery through other user experiences.

Related Works

To account for an enhanced user experience, we employed the Big Data User-centric model presented in (Zelenkauskaite and Simoes ,2013) foregrounded in a Big Data paradigm. Big Data paradigm explicates benefits and challenges of increasing quantities of data sources and serves as an analytical and conceptual framework through five parameters. The parameters include volume, variety, velocity, veracity, and value (Laney,2001;TDWI,2012). Volume involves large amounts of data that has to be manipulated. Variety refers to filtering of the uncorrelated and unstructured data sources as well as high scalability of the data (Cuzzocrea *et al*., 2011). Velocity refers to a gained advantage to process large and dynamic amounts of data in a timely manner. Veracity quantifies the trust associated with inhomogeneous data, which is often gathered from unverified sources. The Big Data User-centric model is particularly concerned with value creation for users (Zelenkauskaite and Simoes ,2013). We defined value as a meaningful information extraction through data analysis going beyond the plurality of its definitions such as "the monetary value of something," "the market price," or "something intrinsically valuable or desirable" Value (2013). Therefore the value extraction in a Big Data User-centric model for the purposes of this research dealt with interest-based relevance extracted from multiple streams and content navigation and interaction, focusing on dimensions of Interactivity defined in the following section.



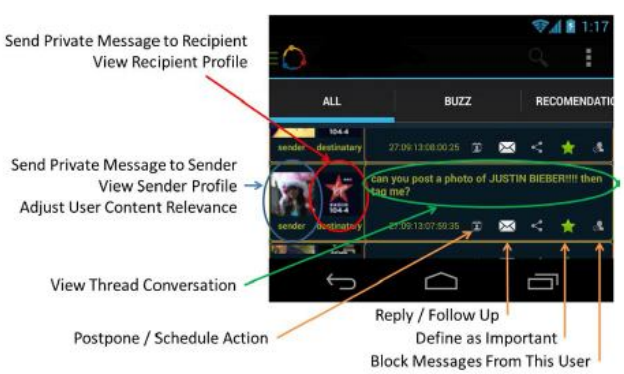
The above figure shows the roles and experiences in interactive services.

Interactive Interest-Based Environments

Departing from the Big Data User-centric model for social media platforms, we conceptualized value as a meaningful interest-based content discovery through interaction with other users, content, and the system. Interactivity refers to the "quality or condition of interaction" (Bucy, 2004). Interactivity conditions consists of non-interactive state, when new messages are not related to previous messages; reactive, when new messages are related only to a specific, often times prompted message; and interactive, when new messages can be related to a number of previous messages having a specific relationship between them (Rafaeli, 1988). Interactive interest-based environment thus aimed to increase degrees of adaptability, user agency, choices, and flexibility leading to some of the positive outcomes of Interactivity such as user satisfaction (Sundar and Marathe, 2010). In a specific mass media context, so far, Interactivity was included through social media outlets. Figure 1 exemplifies how traditional media – such as TV, radio, and newspapers – integrated interactive applications through social media to enhance user experience. Regardless of the promise of interactivity in traditional media settings, it was found to be limited. Limits were attributed to structural constraints inherent to a rigid hierarchical architecture, limited interaction (Van Dijck, 2009) and control over the content. Organizations still maintain control over 80% of the content, even if half of it is produced by users (Gantz and Reinsel, 2012). Even if users attempted to maximize content push by repeatedly posting content, user role still remains marginal (Zelenkauskaite Herring *et al*., 2008).

Challenges Of Interactive Interest-Based Content Discovery

Interest-based content discovery in a cross-platform context not only enhanced user interaction, but it also raised challenges to the Big Data repositories. One of the challenges was ascribed to a potential cognitive overload (Bucy, 2004). The issue of cognitive overload in mediated settings was discussed since mid-nineties (TDWI ,2012;Zelenkauskaite and Simoes, 2013). However, the increasing amounts of users and their produced content made the issue of cognitive overload even more pronounced. It was found that when information supply exceeded individual processing capacity, the overload led to stress and too segmented engagement (Mcmillan, 2002). Wilson (Downes and Mcmillan, 2000) identified seven techniques to ease the information overload associated with interactive Big Data models: information retrieval, aiming at finding information pertinent to a given subject through the use of keywords (Refaeli and Sudweeks, 1997) information filtering, relying on filtering techniques to highlight relevance from a continuous flow of information ; rank filtering, providing omission techniques to identify relevant items, using predefined factors such as the number of recommendations, user acceptance and popularity within the community; brute-force interaction, defining techniques that enable immediate and effortless initiation of interaction; content approximation, to help users selecting the most important and relevant items by providing users with a brief preview of a given item extracted from each of the properties; contextualization, introducing techniques to organize the information, i.e. to highlight its significance; and information stack, combining the aforementioned techniques with actions to postpone and redefine the priority of an item. The following figure shows the User-generated content feed and messenger in a seamless interface.



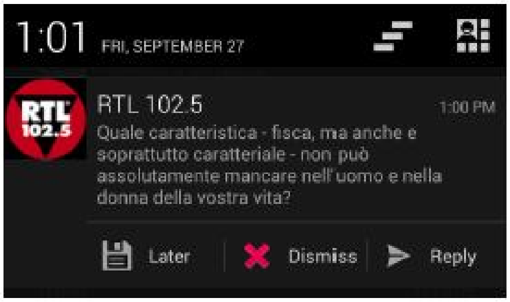
Mobile-Oriented Framework

Considering challenges of Big Data and cross-platform user interaction, we developed a mobile-oriented framework entitled SAMU. Its name was inspired by the word's dual meaning, translated as "of the strings or braids" and "I dig" (Wilson, 1996) which entails the thread multiplicity and a persistent search, associated with new and relevant interest-based content and users. This framework not only facilitates the use and understanding of different sources of information, but also locates, navigates, customizes, and interacts with flexible and fluid user roles in an interest-based context. Moreover, it capitalizes on cloud computing as a processing infrastructure, which aids users to discover and manage social media content through services that perform analytics at large scale. This mobile application accounts for six dimensions of perceived interactivity introduced previously in Section II, as well as the following interactivity traditions: human-to-human interactivity and human-to-system interactivity.

Human-To-System Interaction

Building on techniques to tackle information overload, SAMU delivers three essential mechanisms for accessing and managing social media content.

Subscription management: defines a set of features necessary to subscribe (or unsubscribe) to specific topics and user-related activities. Subscription topics are subdivided into two categories: topics expressed through well-defined user search rules (e.g. expression matching), which can also employ the use of tags – tag subscription; and topics emerging from the content analysis performed by cloud computing services, to discover new interest-aware topics that have good coverage and acceptance. We named this type of results as a "group" of content. We created a unique set of functionalities that were geared towards a seamless access to non-interoperable data and an environment to reduce information overload for users to make their reading-data discovery more effective.Content management: delivers a set of features to manage subscriptions' content, e.g. features to automatically bring to the top the latest and the most relevant items by clearly marking those that have been read. Three main components are activity buzz, content filtering, and analytics-oriented content management. The following figure shows Enhancing interaction using



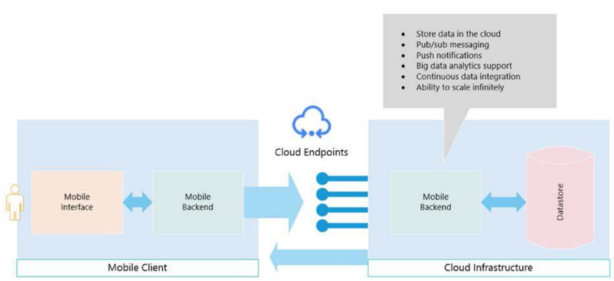
interaction- oriented notifications notifications Activity buzz. Primes all items classified as important. The algorithm used to select relevant items is based on several factors such as the most read items, created by "opinion makers" and user acceptance. It also provides a short glance of recent social media content. This feature aims to create a sense of community, encourage discussing commonly read content, and to support serendipity.

Content filtering through this feature users can define priority schemes for processing a subset of information while ignoring its complementary set, that is, to prioritize or exclude content using a set of conditional rules. Additionally, information concerning the usage of this feature is gathered and processed by the analytics-oriented content management module, to obtain further insights.

Content interactivity: Several mechanisms were implemented to provide higher levels of user-content interactivity. From the main screen, users can access all the content – i.e. user profiles and thread messages with a one-click mechanism. They can reply and / or identify the source, medium, destination, and message.

Human-To-Human Interaction

Apart from the spatial navigation, which is intrinsically associated with the features described in the previous section, the application also introduced novel mechanisms to foster temporal dimension i.e. real-time interaction across various platforms. On one hand, it focused on a notification system's ability to provide invitation opportunities to novice users to engage in a sustained interaction with and through the interface. It also served to inform experienced users – in real-time – about ongoing activities that might be of interest (Figure 3). Notification interaction alerted users about the latest updates, yet did not require them to take an immediate action; freeing users from the constant urge to pay attention to content streams. The following figure shows the Mobile-oriented Service Architecture



Cloud computing as a Big Data analytics solution

Although SAMU provided multiple opportunities to prioritize and filter information through content analysis, such techniques remain insufficient to overcome information overload. On one hand, the quality and quantity of results depend on how users configure their systems. For example, the use of erroneous content filters greatly contributes to the unsatisfactory user experience and low content quality (Edmunds and Morrisn ,2000). On the other hand, extracting and analyzing all the social media content is unfeasible, not only because of mobile devices' low processing capabilities – that would lead to incomplete outputs or responsiveness – but also because of the unrealistic amount of network connections required to access necessary data. Moreover, a restricted number of requests permitted within a certain time window by data providers – mainly social networking sites – introduces even greater complexity. Hence, SAMU required a backbone architecture to continuously extract social media content from Big Data repositories and to handle all the management and processing tasks, to overcome mobile devices' bandwidth and processing limitations.

Use-Case: A Seamless Radio Context

In traditional mass media, users are constrained to access radio content through the hierarchical architecture that is based on a one-to-many model, which focuses on an individual radio's content. In contrast to the hierarchical one-radio access architecture, we propose a user-centric architecture that departs from the interest-based content rather than a specific radio. This model is geared to enable a fluid content access to various radio stations and contents associated with it, to foster content discovery and content management. Access to specific radio programs can be exemplified by the following scheme: radio-programs / DJ-songs / artists / albums.

In addition to music content, the primary radio stations' content, users could access radio stations' social media content streams as well. Finally, all radio stations' contents would generate a unified Big Data repository form which individually tailored interests could be extracted. To run initial tests on a cross-radio Big Data repository, with the developed application KASU we analyzed the top ten radios from the "Top

Facebook Radio Pages" list. The sample of radios analyzed included RTL 102.5, Mosaique FM, Radio Sawa, Alwakeelnews, Jawhara FM, HIT Radio, Mazaj FM, Play 99.6, Virgin Dubai, and Hala FM Radio. Social media platforms consisted of Facebook, Twitter, and Google+. Table I breaks down the distribution of 12,312 messages produced during a period of one week by social media platforms (i.e. users) and radios.

Table.1. User activity on top Radios' Social Media platforms

	Facebook		Google+		Twitter	
	msgs	users	msgs	users	msgs	users
RTL 102.5	2363	1887	1	1	239	58
Mosaique FM	2255	1658	10	1	188	7
Radio Sawa	360	250	0	1	76	2
Alwakeelnews	570	252	0	1	23	1
Jawhara FM	689	479	20	1	133	1
HIT Radio	2392	1976	3	2	182	3
Mazaj FM	666	440	20	1	117	3
Play 99.6	296	239	0	1	109	2
Virgin Radio Dubai	466	395	62	30	127	5
Hala FM Radio	4369	137	0	1	126	1

Table I shows that majority of content traffic on these radios were generated by Facebook users (given that we have selected the top radios based on Facebook).

CONCLUSION

In this paper, we proposed a model that encompassed the interconnectedness between services and social media platforms. We considered Big Data as ways to highlight user value and bridge user needs through social media and professional content by enabling users to model interest-based relevant streams and create their own individualized social radio experience. In practical terms, we developed a mobile-oriented framework built on top of social media streams and professional radio content. This android application aimed to facilitate user immersion and an interactive experience through non-hierarchical customization and enhanced individual choice in a continuously dynamic unstructured content environment. Due to bandwidth and mobile processing limitations, we also developed a service architecture based on the concept of cloud computing. Cloud computing served as an essential element to ensure the access and processing of inhomogeneous data (SNSs).

REFERENCES

Beyer, Y., Enli, G.S., Maaso, A. and Ytreberg, E. 2007. Small talk makes a big difference: recent developments in interactive, SMS-based television, Television & New Media, P.213–234. https://doi.org/10.1177/1527476407301642

Bucy, E.P.2004.Interactivity in society: Locating an elusive concept The information society, vol. 20, no. 5, P. 373–383. https://doi.org/10.1080/01972240490508063

Cuzzocrea, A., Song, I.-Y. and Davis, K.C. 2011. Analytics over large-scale multidimensional data: the big data revolution! in Proceedings of the ACM 14th international workshop on Data Warehousing and OLAP, ser. DOLAP '11. New York, NY, USA: ACM,P.101–104. [Online]. Available: http://doi.acm.org/10.1145/2064676.2064695 https://doi.org/10.1145/2064676.2064695

Downes, E.J. and McMillan, S.J. 2000. Defining interactivity a qualitative identification of key dimensions," New Media & Society, vol. 2, no. 2, P. 157–179. https://doi.org/10.1177/14614440022225751

Edmunds A. and Morris A.,2000. The problem of information overload in business organizations: a review of the literature. International journal of information management, vol. 20, no. 1,P.17–28. https://doi.org/10.1016/S0268-4012(99)00051-1

Ellison N.B. 2007. Social network sites: Definition, history, and scholarship," Journal of Computer-Mediated Communication,P.210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Foundation, O. Sep 13,2013. OpenId foundation website. [Online]. Available: http://openid.net/

Gantz, J. and Reinsel, D. 2012. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east.Technical report, IDC, Tech. Rep.

Laney, D. 2001. 3-d data management: Controlling data volume, velocity and variety.META Group. http://blogs.gartner.com/doug-aney/files/2012/01/ad949-3DData-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Lee, K.P. and Brenner, J. Sep 13,2012. Photos and videos as social currency online. [Online]. Available: http://pewinternet.org/Reports/2012/Online-Pictures/MainFindings.aspx?view=al

McMillan, S.J. 2002. A four-part model of cyber-interactivity some cyberplaces are more interactive than others," New Media & Society, vol. 4, no. 2, P. 271–291. https://doi.org/10.1177/14614440222226370

Rafaeli,S.1988. Interactivity: From new media to communication. Sage annual review of communication research: Advancing communication science, vol. 16,P.110–134.

Rafaeli, S. and Sudweeks, F. 1997. Networked interactivity. Journal of Computer-Mediated Communication, vol. 2, no. 4, P.0–0. https://doi.org/10.1111/j.1083-6101.1997.tb00201.x

Recordon, D. and Reed, D. 2006. Opened 2.0: a platform for user-centric identity management, in Proceedings of the second ACM workshop on Digital identity management, ser. DIM '06. New York, NY, USA: ACM,P.11–16. [Online]. Available: http://doi.acm.org/10.1145/1179529.1179532 https://doi.org/10.1145/1179529.1179532

Sundar, S.S. and Marathe, S.S.2010. Personalization versus customization: The importance of agency, privacy, and power usage," Human Communication Research, vol. 36, no. 3, P.298–322. https://doi.org/10.1111/j.1468-2958.2010.01377.x

TDWI, 2012. Big data – why the 3vs just don't make sense. [Online]. Available: http://tdwi.org/Articles/2012/07/24/Big-Data-4thV.aspx?Page=1

Twitter, 2013. An app that brings more flexibility and insight to power users. [Online]. Available:http://tweetdeck.com/

Van Dijck, J. 2009. Do Users like you? theorizing agency in user-generated content.Media, Culture, and Society, vol. 31, no. 1, P. 41.
https://doi.org/10.1177/0163443708098245

Wilson, P.1996. Interdisciplinary research and information overload.Library Trends, vol. 45, no. 2, P.192–203.

Yoono, 2013. Your social networks united. [Online]. Available: http://www.yoono.com/

Zelenkauskaite Herring S. 2008. Television-mediated conversation: Coherence in Italian ITV SMS chat in Hawaii International Conference on System Sciences.
https://doi.org/10.1109/HICSS.2008.425

Zelenkauskaite, A. and Simoes, B. 2013. The big data user-centric model in The 2013 World Congress in Computer Science, Computer Engineering, and Applied Computing, ACM. CSREA Press.