Volume 4 Issue 3 January 2017 ISSN: 2321 – 788X

BIG DATA AND LIBRARIES: PROPHECY OF VANNEVER BUSH

Dr. P. Ramkumar
Librarian, Srimath Sivagnana Balaya Swamigal Tamil Arts and Science College, Mailam-604304

Kannan
University Assistant Librarian, Thiruvalluvar University, Vellore-632115

Abstract

The concept of Big Data was proposed as memex device and set of functions in 1945 by Vannever Bush who is remembered in MIT (USA) as an academic celbrity. Big Data by being customer centric is an innovation for business enterprises which aim to market products and services to customers while maximizing the profit of the organizations. Libraries are non-profit oriented service organizations with striking similarities to business transactions. Metadata and Library User Weblog/usage mining are potential areas for exploration to qualify library's participation in the Big Data trend. The prophecy of Vannever Bush has come true though the library model, he used to explain his memex concept is yet to see the light of the day. Library community has to come alive with plans and programmes.

Keywords*: Big Data; Memex; Web Usage Mining; Knowledge Discovery in libraries.*

Introduction

Universe of knowledge is a continuum. Reflecting this dictum, Sir Isaac Newton expressed that, "If I can see further than anyone else, it is only because I am standing on the shoulders of giants." Those giants for Isaac Newton were scientists like. Copernicus, Galileo, and Kepler. The inference from Newton's expression is that, without the findings and research of these earlier mathematical geniuses to follow, Newton may not have ever promulgated the famous three laws of motion. Such a tradition has become an in built mechanism even today with the citation culture. Scientists build on the theories of their predecessors to answer the questions from among an unlimited number that continues to mystify humanity.

Prophecy Validated

Big data, as it is known at present, was envisioned already by a giant on whose shoulder many giants are standing today. Yes, it was Vannever Bush (1945) who authentically prophecised his vision elaborately in his publication entitled, *As we may think*. The year 1995 marks the 50th anniversary of Vannevar Bush's landmark paper, "As We May Think," published first in the *Atlantic Monthly* and subsequently in Life magazine. In honor of Dr. Bush's vision there was a research symposium held at MIT, his academic home, on October 12 & 13, 1995. The man deserves all the celebrations. He had proposed a device called memex in his publication. According to Bush (1945), "A memex is a device in which an individual stores all his books, records, and communications, and which is

mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory." The definition may sound simple making one to think of a pen/flash drive today.

But, when one may go through his subsequent publication entitled, *Memex revisited* (Bush, 1965), published 20 years after his first one shall find the document encompassing all the aspects of big data and the networked environment. In his *Memex revisited*, Bush writes in the context of involving Library that, "It is thus fairly clear that there is no serious problem today in assembling, editing, and correcting the record, or in compressing it into as small a volume as we may need for memex. If we wish it, a whole private library could be reduced to the volume of a matchbox; similarly, a library of a million volumes could be compressed into one end of a desk. If the human race has produced, since the invention of movable type, a total record in the form of magazines, newspapers, books, tracts, advertising blurbs, correspondence, having a volume corresponding to a billion books, the whole affair, assembled and compressed, could be lugged off in a wheelbarrow." His prophecy regarding big data in the context of libraries of the present day are in the making.

Fifty years after the publication of the proposed concept, memex, an exclusive website with the title 'The Memex and Beyond' has been created by the MIT (USA) as a major research, educational, and collaborative web site integrating the historical record of and current research in hypermedia. The name honors the 1945 publication of Vannevar Bush's article "*As We May Think*" in which he proposed a hypertext engine called the Memex, and the web site is an outgrowth of the 1995 Brown/MIT Bush Symposium honoring the 50th anniversary of its publication. A specific point noted here is that, Bush was a student of the MIT, USA.

Even in 1965 when computers and their applications were in the making, Bush could open his far sighted view thus: It is thus fairly clear that there is no serious problem today in assembling, editing, and correcting the record, or in compressing it into as small a volume as we may need for memex. If we wish it, a whole private library could be reduced to the volume of a matchbox; similarly, a library of a million volumes could be compressed into one end of a desk. If the human race has produced, since the invention of movable type, a total record in the form of magazines, newspapers, books, tracts, advertising blurbs, correspondence, having a volume corresponding to a billion books, the whole affair, assembled and compressed, could be lugged off in a wheelbarrow.

Licklider: Yet Another milestone

Communication in general and networking is the point of focus today. Many documents while discussing the contrinutions of Bush, necessarily talk also about Licklider.

Among the significant contributions to the development of network communications, those of Licklider played pivotal roles. It is to be noted that Licklider was a member of the ARPANET Team in USA. Licklider (1960) published his work entitled Man-

Computer symbiosis followed by Licklider's (1962) the first ever recorded description of the social interactions that could be enabled through networking. He had discussed his "Galactic Network" concept. Surprisingly, Licklider (1965) worked on a sponsored project involving computers and libraries and brought out his book entitled *Libraries of the future*. He used Vannevar Bush's idea of an automated library system and expanded it to describe how computers could be used to make library resources available to multiple remote users through a single database.

He envisioned a globally interconnected set of computers through which everyone could quickly access data and programs from any site. In spirit, the concept was very much like the Internet of today. No wonder, the concept of networking kindled thoughts in Licklider a foremost consideration to combine libraries with social networking. In 1965 he published his first book, Libraries of the Future.

According to Leiner (2009) The Internet represents one of the most successful examples of the benefits of sustained investment and commitment to research and development of information infrastructure. Beginning with the early research in packet switching, the government, industry and academia have been partners in evolving and deploying this exciting new technology.

The Third Platform

Libraries are yet to enter the third platform of the big data stream. Becoming a data-driven or software-defined enterprise is not an easy task. It is a fact that it took long, say a few decades, to convert the document oriented, manually operated traditional library to adapt to the online world, and yet the number of struggling libraries left out is larger than one may imagine. Today, the online world is on steroids. Experts describe this as the era of the Third Platform which is built on a foundation of cloud, mobile, social, and Big Data technologies.

According to a white paper sponsored by the IDC (2013), "Major new sources of competitive advantage are being built by creatively leveraging cloud, mobile, social, and Big Data technologies. Enterprises and organizations alike are grappling with how to address the impact of these technologies not only on their decision-making processes, operations, product rollouts, and promotions but also, most importantly, on how they engage with their customers. To respond to these dynamics, enterprises are undertaking digital transformation initiatives. This is already happening today and will accelerate over the next three to five years. Your future and the future of your organization depend on your ability to master these 3rd Platform technologies and use them to your advantage."

Libraries to Get involved in the Third Platform

The issues, challenges and opportunities of the digital universe are technical, but they are first and foremost organizational. Here are three steps all libraries must take without delay if they want to thrive in the new era:

1. Create a C-level position in charge of developing new digital/data library service opportunities: Whether it is a new position (e.g., Chief Digital Librarian), or an enhanced existing one (e.g., redesignating and upgrading the Librarian's current responsibilities), this new executive role is in charge of identifying and pursuing new service streams based on both internal and/or external data and its analysis.
2. Develop and continuously revise an executive-team understanding of the new digital/data landscape for every individual library: How are digital librarians going to cooperate with other libraries and their professional community to anticipate and thwart any possible digital disruption. What are the short- and long-term steps they should must take to ensure a smooth and timely digital/data transformation?
3. Design and execute a plan for accelerated investments in library digital/data involving Library technologies and skills: Re-allocate resources across services based on digital/data transformation priorities, invest in promising data collection and analysis areas, and identify the gaps in talent and skills required for success in the era of the Third Platform.

Increased investment in human capital and the new skills required today is the first order of libraries as a business transaction centre/ organization. Another crucial action to take is launching an institution-wide initiative to determine where to focus the response to the opportunities presented by the Third Platform. What kind of new products and services based on data and analytics can a library develop and render? What are the areas of the library 'collection-web usage-user' in which the Internet of Things will have the biggest impact? These and similar questions must be answered and acted upon, then reviewed and the answers revisited in short periodic intervals as new applications and new sources of data become available.

Web Usage Mining in Libraries

Big data is customer centric. The potential areas for big data analysis are the metadata regarding the collection on the web and library web user behavior.

In the words of Sanjay Kumar Malik et al., (2010) Web Usage Mining can be used to make search relevant by determining frequent access behavior for users, needed links can be identified to improve the overall performance of future accesses. Web Usage mining has been defined as the application of data mining techniques to discover usage patterns from Web data in order to understand and better serve the needs of Web based applications. Web usage mining consists of three phases, namely preprocessing, pattern discovery, and pattern analysis. Web Usage Mining may be applied to data such as contained in logs files. A log file contains information related to the user queries on a website. Web usage mining may be used to improve the website structure or giving recommendations to visitors.

According to Han et al. (2004), "The principal objective of web usage mining is to discover and retrieve useful and interesting patterns from a large dataset. In web usage

mining, the dataset is the huge web data. Web data contains different kinds of information, including, web structure data, web log data, and user profiles data. Web mining is the application of data mining techniques to extract knowledge from web data, where at least one of structure or usage data is used in the mining process. Web usage mining has various application areas such as web pre-fetching, link prediction, site reorganization and web personalization. Most important phases of web usage mining is discovering useful patterns from web log data by using pattern discovery techniques such as Apriori, Frequency Pattern -Growth algorithm."

Each of the mouse click in a click stream of a user session on the net brought under web usage mining and data discovery contributes to Big Data unfurling user behavior profile so as to enable the library render tailored services to every user in an individual (may be personalized) context. This could help library achieve winning every user into the purview of library services either within the portals or the website of the library.

Conclusion

The requirement of the Third Platform is not a "strategic plan," but a continuously revised action plan. Several issues will have to be addressed to capture the full potential of big data. Policies related to privacy, security, intellectual property, and even liability will need to be addressed in a big data world. Organizations need not only to put the right talent and technology in place but also structure workflows and incentives to optimize the use of big data. Access to data is critical—companies will increasingly need to integrate information from multiple data sources, often from third parties, and the incentives have to be in place to enable this.

Experts opine that, 'the right skills, a laser-sharp focus, and a passion for continuous learning and adaptation are a must for library professionals to thrive in the digital/big data universe of tomorrow.' The big data universe will continue to expand at a rapid pace emptying into deluge with already saturated infrastructures and the people that use as well as those manage them. The Internet of Things will add new levels of complexity — and opportunity — on top of what Big Data has offered during just the last few years. The bright stars of the big data universe of libraries to come will ride the data deluge to new heights of research productivity and ultimately the human welfare.

The prophecy of Vannever Bush has come true though the library model, he used to explain his memex concept is yet to see the light of the day. Library community has to come alive with plans and programmes. The library professional community have an extra-ordinary responsibility to carry the services based on Big data analytics. The feel, functionality and performance of applications have a direct effect on library users, their preference related to relevant data and information to individual's context. The user experience is the currency of the library service application economy.

It should be opt to conclude with a quote from Vannever Bush (Memex revisited), "Presumably man's spirit should be elevated if he can better review his shady past and analyze more completely and objectively his present problems. He has built a civilization so complex that he needs to mechanize his record more fully if he is to push his experiment to its logical conclusion and not merely become bogged down part way there by overtaxing his limited memory. His excursion may be more enjoyable if he can reacquire the privilege of forgetting the manifold things he does not need to have immediately at hand, with some assurance that he can find them again if they prove important." One can visualize that Bush considered libraries as the good samaritan carrying knowledge in the form of data along with research publications to the academic community anywhere, anytime and any device.
Bush, Vannevar., (1945), As We May Think. The Atlantic Monthly, July 1945.

References

1. Hawkings, Stephen., (2003) Standing on the Giant shoulders. (First Printing Edition), New York: Running Press. 1286p.
2. Bush, Vannever., (1945), As we may think. The Atlantic monthly, July 1945. http://web.mit.edu/STS.035/www/PDFs/think.pdf Retrieved on 07-01-2015.
3. Nyce, James M., and Paul, Kahn., (1991), From Memex to Hypertext: Vannevar Bush and the Mind's Machine. London: Academic Press Limited.
4. Licklider, J.C.R. (1960). Man-Computer Symbiosis, IRE Transactions on Human Factors in Electronics, volume HFE-1, pages 4-11, March 1960.
5. Licklider, J.C.R., (1965), Libraries of the future. Cambridge in Mass.: M.I.T. Press.
6. Leiner, Barry M., (2009), A brief history of the Inernet. *ACM SIGCOMM Computer Communication Review* 22, 39(5), October 2009.
7. Gens, Frank., (2013) The 3rd Platform: Enabling Digital Transformation. http://www.tcs.com/SiteCollectionDocuments/White-Papers/3rd-Platform-Enabling-Digital-Transformation.pdf Retrieved on 06-01-2015.
8. Sanjay Kumar Malik, Nupur Prakash, and S.A.M. Rizv (2010), Ontology and Web Usage Mining towards an Intelligent Web Focusing Web Logs. 2010 International Conference on Computational Intelligence and Communication Networks (CICN 2010), Bhopal, India 26-28 November 2010, pp.443-448.
9. Han, Jiawei., et al., (2004), Mining Frequent Patterns without Candidate Generation: A Frequent-Pattern Tree Approach. Data Mining and Knowledge Discovery, 8(1), pp.53-87.