International Journal of Management, IT & Engineering
Vol. 9 Issue 3, March 2019,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Scientometric Data Challenges in the Digital World

M.Sivamani
Vellalar College for Women
Thindal. Erode638012. TN
India

Abstract: In the era of data driven world, all human activities are centred towards data. People depend data and use data for all walks of life. The present study viewed the challenges and issues in scientometric data and its reliability. The work also reviewed the summarizes the significance of scientometric data in the modern data-dependant society.

1. Introduction and Background

Data processing aims at the automatic discovery of underlying non-trivial knowledge from datasets by applying intelligent analysis techniques. The interest in this research area has experienced a considerable growth in the last years due to two key factors: (a) knowledge hidden in organizations' databases can be exploited to improve decision-making; and (b) the large volume of data managed by organizations makes it impossible to carry out a manual analysis.

These issues are also becoming frequent in data and information fusion as a result of the increasing number of sensors used, the paradigm shift from lower-level object recognition to higher-level situation assessment, and the incorporation of heterogeneous sources to the fusion process (including soft information in textual form). Data processing and knowledge discovery methods can be used to extract from datasets elaborated knowledge that can be afterwards fused with varied data and other information. Furthermore, data processing and knowledge discovery methods can be applied on fused data to achieve better inferences towards situation and threat assessment. Consequently, the Information Fusion community can benefit from well-established approaches and new advances in data processing and knowledge discovery –such as machine learning and pattern recognition algorithms, imprecise and uncertain knowledge management formalisms, big data analysis tools, natural language processing techniques, etc.– to develop fusion systems able to exploit more information sources more efficiently.

As the world becomes more digitized and interconnected, the door to emerging threats and leaks has opened wider. Today, there are billions of RFID tags for items including products, passports, buildings and animals. With more than two billion Internet users and cellular phone subscriptions passing the 5 billion mark at the end of 2010, nearly one in three people worldwide surf the Internet.1 More than 50 billion objects are expected to be digitally connected by 2020, including cars, appliances and cameras. Intensifying this complex mix, the amount of digital information created and replicated in the world will grow to an almost inconceivable 35 trillion gigabytes by 2020.

2. Challenges

2.1 Data Growth

The world is experiencing a data revolution, or "data deluge" [1] Whereas in previous generations, a relatively small volume of analog data was produced and made available through a limited number of channels, today a massive amount of data is regularly being generated and flowing from various sources, through different channels, every minute in today's Digital Age.[2] It is the speed and frequency with which data is emitted and transmitted on the one hand, and the rise in the number and variety of sources from which it emanates on the other hand, that jointly constitute the data deluge. The amount of available digital data at the global level grew from 150 exabytes in 2005 to 1200 exabytes in 2010.[3] It is projected to increase by 40% annually in the next few years, [4]which is about 40 times the much-debated growth of the world's population.[5] This rate of growth means that the stock of digital data is expected to increase 44 times between 2007 and 2020, doubling every 20 months.

2.2 Data in the Smart Era

As the virtual and real world increasingly merging together, data produced could be consumed in a 'smart' way. To ensure smartness we stress on the quick consumption of correct data. This ensures smartness; the society that does fail in smart consumption will lag behind.

Better data and the volume and speed with which it is now becoming available, affords new possibilities to understand people and places more deeply to inform design and how it is delivered. By bringing big data together with planning and design we have the power to

International Journal of Management, IT & Engineering
Vol. 9 Issue 3, March 2019,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

transform cities into places that are more responsive to the public's needs and aspirations while also strengthening social capital and engendering digital inclusion.[6]

Data smartness refers to the speed in which the data is consumed. People are smart when they use and adopt technology, design, data and information quickly. We can measure how countries, societies and people are smart with respect to be smart.

2.3Metadata driven content

To achieve ease of information access in the data warehouse, people look at the standardised way of content description. This is the standard way of information access. But to control the activities that are performed in the assessment of data standard and quality, functionalities and responsibilities with continuous quality improvement are considered. To control the quality assessment activities consider two things are considered. 1. Quality planning - Here select the criteria for quality assessment, classify it and assign priorities to the controlling activities. 2. To measure the quality quantitatively. In Meta data quality control system, first the information requirements or demands for quality from the users are collected and then transform these requirements into specifications. The Meta data Quality Control system (MQC) comprises the whole data warehouse architecture and the quality will be measured along the data flow. The following is the architecture of a good metadata based information processing system.[7]

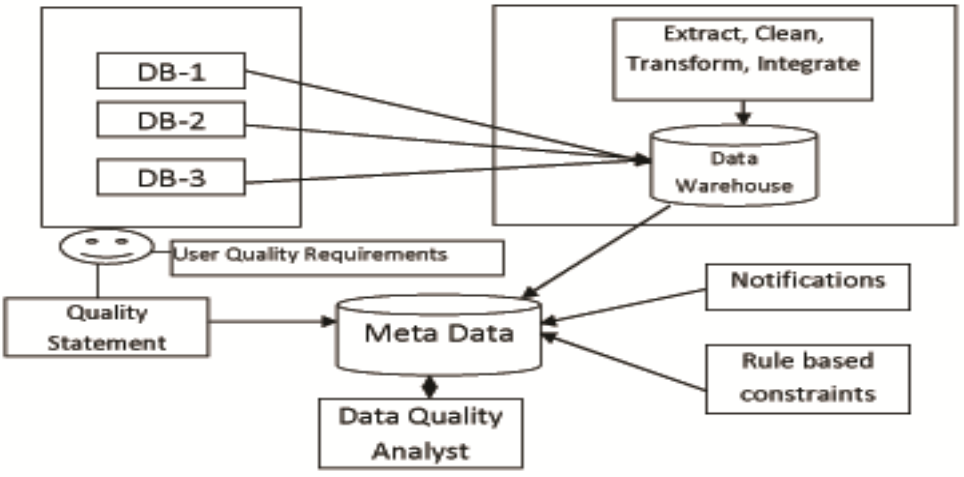

Figure 1. Metadata processing

International Journal of Management, IT & Engineering
Vol. 9 Issue 3, March 2019,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

3.Scientometric Data Challenges

Scientoemetric data can be sourced from two ways. One is the primary data collection where the sources are directly accessed. Second is the use of datasets or databases where secondary data is derived. But secondary data sources offer comprehensive data by providing access to large data sources. The issue in the database dependent data is the overlapping, accuracy and missing data values. When we use the data design science principles, the developed tools offer how the theoretical knowledge from scientometrics and data mining practices can be deployed to solicit reliable data. The application can be used by researchers can lead to gain new insights in many domains of research. The parameters used are applied in practice and beneficial in many respects. . At the same time, the automated data sourcing should only be used in addition to the normal literature search methods. Nevertheless, the developed application can be seen as an enhancement to the traditional methods and hence it offers recent trends and discovers undetected outcomes by the use of not only scientific contributions, but also information from the web=based sources or altmetrics (examples include linked in, library use, Facebook, Twitter, etc).

Some the significant issues are reviewed by the below mentioned studies. Franceschet (2009) presented a correlation analysis to limit quantitative, bibliometric indicators for scientist evaluation. The analysis by them consist as many as more than 10 indices. The volume of papers (for productivity assessment), the number of citations (for impact assessment), the mean citation scores per paper (for relative impact assessment) and the long time data with big volume (for long-term impact assessment) are identified as the most important indicators. Another major analysis is advocated through the trend analysis. Tseng et al. (2009) reviewed data challenges using several trend indices. They documented that the linear regression is useful for timeline analysis, which supports the extensive usage of this method. Guo et al. (2011) in a major study used several data indicators that includes the increase of specific word usage, amount of new authors in research field and amount of interdisciplinary citationsin a complex model.

4.Summary

The efficient data processing for Development can fulfil its immense potential to enhance the greater good, then the answer is clearer. What is needed is both intent and capacity to be sustained and strengthened, on the basis of a full recognition of the opportunities and challenges. Specifically, its success hinges on two main factors. One is the level of institutional and financial support from public sector actors, and the willingness of private corporations and academic teams to collaborate with them, including by sharing data and technology and analytical tools.

References

1."The Data Deluge." The Econ omist. 25 Feb 2010. <http://www.economist.com/node/15579717> and Ammirati, Sean. "Infographic: Data Deluge – 8 Zettabytes of Data by 2015." Read Write Enterprise. <http://www.readwriteweb.com/enterprise/2011/11/infographic-data-deluge---8-ze.php>

2. King, Gary. "Ensuring the Data-Rich Future of Social Science." Science Mag 331 (2011) 719-721. 11 Feb, 2011 Web. http://gking.harvard.edu/sites/scholar.iq.harvard.edu/files/gking/files/datarich_0.pdf

3. Helbing, Dirk , and Stefano Balietti. "From Social Data Mining to Forecasting Socio-Economic Crises." Arxiv (2011) 1-66. 26 Jul 2011 http://arxiv.org/pdf/1012.0178v5.pdf.

4. Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh and Angela H. Byers. "Big data: The next frontier for innovation, competition, and productivity." McKinsey Global Institute (2011): 1-137. May 2011. http://www.mckinsey.com/mgi/publications/big_data/pdfs/MGI_big_data_full_report.pdf

5. " World Population Prospects, the 2010 Revision." United Nations Development Programme. <http://esa.un.org/unpd/wpp/unpp/panel_population.htm>

6. DESIGNING WITH DATA: SHAPING OUR FUTURE CITIES, In: http://www.architecture.com/RIBA/Campaigns%20and%20issues/Designingwithdata/Designing withdata.aspx)

7. Ramesh Babu Palepu, K V Sambasiva Rao. META DATA QUALITY CONTROL ARCHITECTURE IN DATA WAREHOUSING International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.4, August 2012

8. M. Franceschet: A Cluster Analysis of Scholar and Journal Bibliometric Indicators. Journal of

the American Society for Information Science and Technology, 60(10):1950–1964, 2009

9.Y.-H. Tseng, Y.-I. Lin, Y.-Y. Lee, W.-C. Hung and C.-H. Lee: A comparison of methods for detecting hot topics. , 81(1):73–90, 2009.

10. H. Guo, S. Weingart and K. Börner: Mixed-indicators model for identifying emerging research areas. Scientometrics, 89(1):421–435, 2011.