Aids to Analyse the Knowledge Building Portals

Amit Arjun Verma

February 10, 2021

Abstract

With the success of crowdsourced portals like Wikipedia, Stack Overflow, Quora, GitHub, etc., a class of researchers is driven towards understanding the dynamics of knowledge building on these portals. Despite the fact that collaborative knowledge building portals are known to be better than expert-driven knowledge repositories, limited research has been performed to understand the knowledge building dynamics in the former. Moreover, the relatioship between knowledge aquisition on online portals and fundamentals of knowledge building is yet to be understood. This is mainly due to two reasons; first, unavailability of standard data representation format. Second, the lack of proper tools and libraries to analyze knowledge building dynamics.

The aim of this thesis is to create an ecosystem of resources which will help researchers analyze corwdsourced portal's data easily. We describe Knowledge Data Analysis and Processing Platform (KDAP), a programming toolkit which is easy to use and provides high-level operations for analysis of knowledge data. We propose Knowledge Markup Language (Knol-ML), a standard representation format for the data of collaborative knowledge building portals. KDAP can process the massive data of crowdsourced portals like Wikipedia and Stack Overflow efficiently. As a part of this toolkit, a data-dump of various collaborative knowledge building portals is published in Knol-ML format. The combination of Knol-ML and the proposed open-source library will help the knowledge building community to perform benchmark analysis.

1 Introduction

In 1906, Francis Galton surprised the statisticians by accuratly calculating the weight of an Ox using the collective guesses of the crowd [1]. The experiment required each person to guess the weight of the Ox by mearly watching it and by averaging all the guesses, a result closer to the actual weight was observed. With this experiment, Galton showed how crowd intelligence can be used to solve real problems accurately. However, Galton's experiment was not the first time when collective intelligence was used in practice. The first example of collective intelligence comes from Thucydides's description on war [**?**]. He described how statisticians were using the collective knowledge of the crowd to calculate the

approximate height of the castle's outer wall. Using this height, generals used to build ladders to breach the wall and attack the inner area. Over the years, people have leveraged on the collective knowledge of the crowd to solve real world problems or gather knowledge. A few exmaples include using network of people to search the entire United States to find a set of balloons in less than nine hours [2] and using the knowledge of millions of people to develop the most expansive encyclopedia in human history. Collective intelligence in practice have existed for a very long time and researchers from various domains have tried to understand the different forms of collective intelligence. Collective Intelligence is generally defined as shared or group intelligence that imerges from the collaboration. The word "intelligence" appears simple but has a complex history. In medieval times, the intellect was measured with our understanding of the God. The evolution of wisdom with time changed the understanding of intelligence. For example, Ren Descartes described the brain as hydraulics where all the other parts of the body are connected by the fluids. The invention of radio and electricity gave us the definition of brain as "network of wires" working on same wavelength. The present senario provides us a metaphore of processing and algorithic thinking, and the brain as computer.

One of the first historical accounts of collective intelligence is Thucydidess description of how an army went about planning the assault on a besieged town. They first made ladders equal in length to the height of the enemys wall, which they calculated by the help of the layers of bricks on the side facing the town, at a place where the wall had accidentally not been plastered. A great many counted at once, and, although some might makemistakes, the calculation would be more often right than wrong; for they repeated the process again and again, and, the distance not being great, they could see the wall distinctly enough for their purpose. In this manner they ascertained the proper length of the ladders, taking as a measure the thickness of the bricks. [**?**]

Similar examples are avialable throughout the history. Understanding how we work togetherthe collective part of collective intelligencehas been a central concern of social science for several centuries. Some mechanisms allow individual to contribute their choices without requiring any conscious collaboration. The recent experiments with digital collective intelligence like Wikipedia is a modern example of such mechanims. Collective intelligence is as old as civilization. But in modern times, it takes different forms. Some are designed to observe better. Dove satellites, about the size of a shoe box, sit around 250 miles above the surface of the earth. They have shown that in Myanmar, for example, the spread of night-lights suggests slower economic growth than the World Banks estimates. In Kenya, they count up the number of homes with metal roofsone indicator of how fast people are moving out of poverty. Moreover, with the emergence of Web-2.0, collective sharing has become easier and more prominent. In Wikipedia, thousands of people around the world have collectively created a large intellectual product of high quality with almost no centralized control and with mostly volunteer participants.

Our world is filled with the the applications of collective intelligence. However, understanding the dynamics of collective intelligence has always been an

interesting topic of research. Although Collective Intelligence overlaps with various domains, we are interested in the Computer Science aspects that involves intelligent behaviour by groups of people, computers, or both. In this article, we highlight the importance of studying collective intelligence and challenges involved in understanding the collective intelligence in present scenario. We focus on online collaborative portals and how these portals can be used to unravel the dynamics of collective intelligence. To understand the collaborative understanding of the crowd, we need to understand how knowledge is build on these online portals.

Knowledge building is generally viewed as a discursive activity intended to enhance collective understanding [3]. It requires participants to engage in the collaborative solution of the knowledge problems in such a way that responsibility for the success of the effort is shared by them [4]. It has been observed that more collaboration accelerates the knowledge building process. For an instance, in scientific community, those who collaborate more tend to have more publications [5, 6, 7, 8]. In some research domains, collaboration seems to be indispensable as it brings special expertise and knowledge not otherwise available but crucial to research outcomes [9]. Despite these good reasons to expect that scientific collaboration will enhance the knowledge building process, the relationship between these two is not well understood.

The reason behind this could be the presence of vast amount of dataset. For an instance, in order to understand the collaboration dynamics in scientific community, one may require to collect all the collect all the articles which were ever pulished. This is infeasible in the sence that collecting and processing such a huge dataset is non-trivial. A variety of platforms are available on the Internet which fosters computer-supported cooperative work (CSCW). The goal of many such environments, such as Wikipedia[1], is knowledge building.

With progress in computational power, research in various domains is primarily based on the availability of data and appropriate tools for analysis. Open access to libraries and data has enhanced the ease and pace of research [**?**].

References

[1] F. Galton, "Vox populi," 1907.

[2] G. Pickard, W. Pan, I. Rahwan, M. Cebrian, R. Crane, A. Madan, and A. Pentland, "Time-critical social mobilization," *Science*, vol. 334, no. 6055, pp. 509–512, 2011.

[3] C. Bereiter, *Education and mind in the knowledge age.* Routledge, 2005.

[4] M. Scardamalia, "Collective cognitive responsibility for the advancement of knowledge," *Liberal education in a knowledge society*, vol. 97, pp. 67–98, 2002.

[1] http://www.wikipedia.org

[5] A. J. Lotka, "The frequency distribution of scientific productivity," *Journal of the Washington academy of sciences*, vol. 16, no. 12, pp. 317–323, 1926.

[6] D. d. S. Price and D. Beaver, "Collaboration in an invisible college," *American psychologist*, vol. 21, no. 11, pp. 1011–1018, 1966.

[7] H. Zuckerman, "Nobel laureates in science: Patterns of productivity, collaboration, and authorship," *American Sociological Review*, pp. 391–403, 1967.

[8] B. Godin and Y. Gingras, "Impact of collaborative research on academic science," *Science and Public Policy*, vol. 27, no. 1, pp. 65–73, 2000.

[9] S. Lee and B. Bozeman, "The impact of research collaboration on scientific productivity," *Social studies of science*, vol. 35, no. 5, pp. 673–702, 2005.