Predicting the Intelligence of Web 3.0 Search Engines

Josiane Farah

Abstract—As the Web is growing exponentially, it needs to provide a standard mechanism so that individuals can easily obtain data, reports and knowledge about any topic posted on it. Even though Web 3.0 is more theory than reality, it is expected that Web 3.0 will be the future of Web. Implementing Artificial Intelligence in different Web applications specifically search engines should be the next big thing and also one of the features of Web 3.0.

Index Terms—Artificial intelligence, intelligent web agents, search engines, web 3.0.

I. INTRODUCTION

With the big advance in technology and the growth of the amount of content on the Internet, it has become difficult for users to find and utilize information and for content providers to classify and catalog documents [1]. The result of a query from traditional Web search engines is hundreds or thousands of Web pages or links [1], which is time consuming for users to browse. With the growth of online libraries, search engines, press release archives and others it became difficult and costly to categorize every document manually. In order to deal with these problems, researchers look toward automated methods of working with Web documents so that they can be more easily browsed, organized, and cataloged with minimal intervention [1]. Currently, the results provided by Internet search engines make use of sophisticated software to search the internet looking for content to index. This sophisticated software is known as "crawler", "robot" or "spider" [2]. Crawlers are passive and they may sometimes fail to find and index important online contents so they may return results that do not always meet or satisfy the user needs. To solve this problem, Google initiated the idea of Sitemaps which is a supplementary way of listing the content available on a Website in a simple, open and "crawler-friendly" format. This idea is now supported by all the main search engines. Sitemaps do not replace existing crawl-based mechanisms, but they enable crawlers to do a better job of indexing a Website [2]. Because sitemaps exhibit some limits, the need for implementation of Artificial Intelligence in making search engines more efficient is very imperative and is expected to be one of the features of Web 3.0.

This paper answers the question: How intelligent will Web 3.0 search engines be.

II. EVOLUTION OF THE WEB – FROM 1.0 TO 3.0

Like any other technology, the Web is evolving [3] and the main stages of Web evolution are: Web 1.0, Web 2.0 and

Web 3.0.

The first generation of the Internet, Web 1.0, was built around content. Web 1.0 provided information, but users couldn't interact with the Web pages and documents they found online [4]. The lack of active interaction of users with the Web leads to the second generation of the Internet, known as Web 2.0.

The significant characteristic of Web 2.0 (2004 – present) is interactivity. Users interact with each other and create profiles of themselves at social networking Web sites such as MySpace; they express their opinions through blogs; they share videos at YouTube; and they collaborate on creating encyclopedia entries on Wikipedia [4]. However, when it comes to search, Web 2.0 is way to be intelligent since keyword-based search results in an information overload and when you make a search on the Internet, the information that is displayed to you lacks personalization.

It is believed that the next generation of the Web, Web 3.0, will be the future of Web even though its definition varies greatly. According to Nova Spivack, the Chief Executive Officer of Radar Networks, "*Web 3.0 is a set of standards that turns the Web into a big database*" [5]. While Conrad Wolfram has argued that Web 3.0 is where "*the computer is generating new information, rather than humans*" [6], Steve Spadling in [7] defines Web 3.0 as "*Highly specialized information silos, moderated by a cult of personality, validated by the community, and put into context with the inclusion of meta-data through widgets.*" From the latter definition we can understand that Web 3.0 will be about semantic Web (or the meaning of data), personalization (e.g. iGoogle), intelligent search and behavioral advertising among other things.

The Web was always great area of application for Artificial Intelligence. Search engines, Web usage data analysis, finding and delivering information are some examples. But with Web 3.0 Artificial Intelligence is getting new focus and a lot more applications as Jay M. Tenenbaum in [8] is showing.

Many people consider the use of advanced Artificial Intelligence as the next big breakthrough on the Web [9]. In addition, some Internet experts believe that **Web 3.0** will revolutionize the usage of Web by making the searching tasks easier. The search engines will gear towards the user rather than towards the keywords [10]. Keywords will be searched based on the culture, region, and jargon of the user [10]. For example, when going on a vacation instead of conducting separate searches for your airline ticket, hotel reservations, and car rental, you can do one simple search with Web 3.0. The search engine will present the results in a comparative and easily navigated way to the user.

There is a lot of work going into the idea of semantic Web, which is a Web where all information is categorized and stored in such a way that a computer can understand it as a

human. This can be viewed as a combination of Artificial Intelligence and the semantic Web. The semantic Web will teach the computer what the data means and this can evolve into Artificial Intelligence that can utilize that information [11]. Tim Berners-Lee, the inventor of the World Wide Web and director of the World Wide Web Consortium ("W3C"), has described the semantic Web as a component of 'Web 3.0' [12][13]. The internet community as a whole tends to find the two terms "Semantic Web" and "Web 3.0" to be at least synonymous in concept if not completely interchangeable. Before predicting the Intelligence of Web 3.0 in terms of searching, we have to understand what do Artificial Intelligence and intelligent Web agent mean.

III. An Overview of Artificial Intelligence

The term Intelligence is the capacity of learning, reasoning, and understanding, grasping truths, relationships, facts and meanings (Dictionary.com). In particular, we can say that Intelligence is the ability to act as human beings, solve problems and think rationally.

Barr and Feigenbaum in [14] define Artificial Intelligence to be "*the part of computer science concerned with designing intelligent computer systems, that is, systems that exhibit characteristics associated with Intelligence in human behavior – understanding language, learning, reasoning, solving problems, and so on.*"

Dobrev in [15] has defined Artificial Intelligence in one sentence as "*a program which in an arbitrary world will cope no worse than a human.*"

Different people approach Artificial Intelligence with different goals in mind. Some are concerned with thought processes and reasoning while others are concerned about behavior [16]; i.e., should the Artificial intelligent system model human or be ideal?

Unlike humans, computers have difficulty understanding specific situations, and adapting to new ones. Artificial Intelligence aims to improve machine behavior in dealing with such complex tasks. In this article, we'll adopt the view that Intelligence is concerned with rational action. For a system to be rational, it should have the ability to reason, and produce the best outcome or, in case of uncertainty, the best expected outcome.

IV. Intelligent Web Agents

This age of information overload and ever-growing contents on the World Wide Web presents some unique problems such as real-time recommendations, data mining, abstracting useful information, and search optimization based on ones' unique profile [17]. Intelligent agents with their ability to work with enormous amount of data, scalability, robustness, and capability to learn from the environment make them a promising candidate to solve these problems [17].

An agent acts in an environment. It perceives its environment through sensors and acts upon that environment through actuators. A performance measure is used to evaluate the behavior of an agent in an environment.

In the previous section, we defined Intelligence as the ability to act as human beings, solve problems and think rationally. A rational agent can be anything that makes decisions, such a person, firm, machine, or software, acts so as to maximize the expected value of the performance measure, given the percept sequence it has seen so far. This means that a rational agent is an agent which has clear preferences, models uncertainty via expected values, and always chooses to perform the action that results in the optimal outcome from among all feasible actions.

There are many criteria that can affect the action taken by a rational agent. This includes the preferences of the agent, the agent's information of its environment (which may come from past experiences), the actions, duties and obligations available to the agent, the estimated or actual benefits and the chances of success of the actions.

Based on the definitions and knowledge covered in this article so far, we can define the intelligent Web agent as a system that use machine learning methods to retrieve and/or extract textual information from the Web. That agent can accept user preferences in form of instructions and adapt its behavior as it encounters new information [18].

Intelligent Web agents will be programs that crawl through the Web, scanning and interpreting information on Web pages. Intelligent Web agent can be trained by the user to search the Web for specific types of information resources. The agent can be personalized by its owner so that it can build up a picture of individual likes, dislikes and particular information needs [19]. It will optimize the search based on the user's unique profile. Once trained, an agent can then be set free to roam the Web, getting useful information sources, while the user gets on with more urgent tasks, or even goes off line [19]. Each user's software agent will learn more about its user by electronically observing his or her activities.

V. How Intelligent Will Web 3.0 Search Engines Be?

Even though Web 3.0 is more theory than reality, it is expected that Web 3.0 search engines will be more and more user specific and produce precise search results. While Web 2.0 was about data, Web 3.0 will be about knowledge and information [20]. This part of the article predicts the intelligence of Web 3.0 search engines.

According to some Internet experts, Web 3.0 will allow the user to sit back and let the Internet do all the work for them [21]. This can be achieved because Web 3.0 will primarily focus on improving the functionality and usability of search engines [22]. The integration of Artificial Intelligence with Web 3.0 complement accurate search queries by giving the internet users improved relevance of information looked up online [23]. The more frequently the intelligent agents as searching tools are used, the more useful they become. Over a period of time, an agent is capable to build up an accurate picture of a users' information needs. It will learn from past experiences, as the user will have the option of reviewing search results and rejecting any information sources which are not relevant or useful. This information will be stored in a user profile and can be used by the agent when performing a search. So, an agent can also learn from its initial attempts on the Web, and return with a more tightly defined searching agenda if requested.

Why search, when you can find Ontologies? As discussed earlier, intelligent Web agents crawl the Web searching for relevant information. This can be done because Web 3.0 will encompass collection of information called ontologies.

Ontologies are neither knowledge nor information [24]. Ontology is a file that defines relationships among a group of terms. For example, the term "sister" refers to the familial relationship between two persons sharing one parent. For Web 3.0 search engine to be effective and intelligent, ontologies must be thorough and comprehensive. Constructing ontologies is a very hard work and that may be one of the big obstacles that Web 3.0 may face.

There is no doubt that the integration of Artificial Intelligence with Web 3.0 can improve the functionality of search engines; however, some users may not feel at ease to adopt new technology and the reasons may vary…

Can an agent be adequately intelligent when using Social Bookmarking? Social Bookmarking as a search engine can provide more intelligent results compared to traditional search engines such as Google. The reason is that when you use social bookmarking as a search engine, you get Websites that have been voted on by humans, so you have a good chance at hitting a good Website. However, because of the human factor, the results can be manipulated. For the Web 3.0 search engine to be intelligent, the Web agent should be able to separate the good results from the bad results. By bad results we mean the Websites or articles that have been voted in order to make them popular.

What about our privacy on the Web? A Web agent can be trained by its owner so that it can build up a picture of individual likes and dislikes. That seems pretty intelligent but can this affect our privacy on the net? Can other people learn things about us by looking at our results? Can our activities on the Web become public knowledge? The lack of confidence in using Web 3.0 may ban or restrict the user to expose his/her personal information on the Web. This will result in a lack of training for the Web agent. This means that there will be no proper user profile that the agent can use while searching. Therefore, the result of the search query may not be interesting and complete. Addressing privacy issues by introducing new enhancing mechanisms may solve the issue of confronting the use of intelligent agents in the future.

Is the behavior of a rational agent always reasonable to people? Rational agents have the ability to reason and produce the best outcome or, in case of uncertainty, the best expected outcome. However, the behavior of a rational agent may sometimes be counter-intuitive to many people as in the Traveler's dilemma.

VI. CONCLUSION

No one knows how a future technology such as Web 3.0 will turn out. Given that people are willing to put effort to make comprehensive ontologies for their Web sites, search engine agents are intelligent enough not to return manipulated results, users are guaranteed privacy on the Web, and Web agents behave reasonably according to human's perspective, we are pretty sure that the integration of Artificial Intelligence with Web 3.0 will be a big step in enhancing the searching capabilities of the Web. We hope that by the time Web 3.0 gets here, all our fears, worries and doubts about this new technology will be faded.

Based on Internet experts' predictions and our research findings, Web 3.0 is of no doubt going to be the outlook of the Web. This technology, supported by Artificial Intelligence will make searching easier and interesting for Web users. Doing a search query that gives much more appropriate information will help in saving time, giving more knowledge and helping in disseminating that information to the people seeking for it.

Let us all look forward to an intelligent Web.

REFERENCES

[1] A. Schenker, A. Kandel, H. Bunke and M. Last, *Graph-theoretic techniques for web content mining*, 3rd ed. Singapore: World Scientific Publishing, 2005, ch. 1, pp. 1-3.
[2] *Exposing your website to search engines*, Central Office of Information, UK, 2010.
[3] B. Singh, "Difference between web 1.0, web 2.0, and web 3.0 – with examples," Ezine Articles, February 2, 2010.
[4] Web 2.0 arrives and shadows of web 3.0 appear on the horizon*, Trends Magazine*. Issue 49, May 2007.
[5] S. Nakate, "What is web 3.0," Buzzle.com, June 21, 2010.
[6] N. Kobie, "Conrad Wolfram on communicating with apps in Web 3.0," IT PRO, March 2010.
[7] S. Spadling, "How to define web 3.0," howtosplitanatom.com, July 14, 2007.
[8] J. M. Tenebaum, "AI meets web 2.0: building the web of tomorrow today," *AI Magazine*, 27(4), 47-68, 2006.
[9] D. Nations, "What will web 3.0 be like," *About.com*.
[10] Semantic Technology Institute International. [Online]. Available: http://www.sti2.org
[11] *W3C Semantic Web Frequently Asked Questions*. W3C.
[12] H. Daniel, *Web 2.0 evolution into the intelligent Web 3.0,* London, UK: Emereo Pty Ltd., 2008, pp. 48-49.
[13] V. Shannon (August, 2006). A 'more revolutionary' web. *International Herald Tribune* [Online]. Available: http://www.nytimes.com/2006/05/23/technology/
[14] A. Barr and E. Feigenbaum, *The handbook of artificial intelligence*, vol. 1. Los Altos, California: William Kaufmann, Inc., 1981, ch.1.
[15] D. Dobrev. (2005), "Formal definition of artificial intelligence," *International Journal Information Theories and Applications* [Online] vol. 12, no. 3, pp. 277-285. Available: http://www.foibg.com/ijita/vol12/ijita12-3-p12.pdf
[16] S. Russell and P. Norvig*, Artificial intelligence: a modern approach*. Upper Saddle River, New Jersey: Prentice Hall, 2003, ch. 1, pp.1-5.
[17] Z. Usmani, *Web intelligence and intelligent agents,* InTech, 2010, ch.2.
[18] T. Rad, "Building intelligent agents that learn to retrieve and extract information," PhD Dissertation, Dept. Comp. Sc., University of Wisconsin-Madison, USA, 2001.
[19] T. Stanley, "Intelligent searching agents on the web," Search Engines Corner. *Ariadne*, issue 7, January 1997.
[20] M. Watson, *Scripting intelligence: web 3.0 information gathering and processing.* Springer, 2009, ch. 1.
[21] J. Strickland (March, 2008). How web 3.0 will work: web 3.0, how will it work, features. HowStuffWork.com [Online]. Available: http://computer.howstuffworks.com/web-30.htm
[22] T. Rad and J. Shavlik, *Intelligent exploration of the web,* Heidelberg, Germany: Physica-Verlag, 2003, pp. 255-274.
[23] M. Perry, "Artificial Intelligence - AI improvements impact - web 3.0 semantic interpretations," Ezine Articles, December 31, 2008.
[24] M. Fawzi, "Wikipedia 3.0: the end of Google," *Evolving Trends*, June 2006.

Josiane Farah is born in Lebanon April 1977. Josiane got her PHD degree in computer engineering from Atlantic University, Florida, USA in March 2009. She earned her Master of Science degree in computer engineering July 1997 from university of Balamand, El-Koura, Lebanon and her Bachelor of Science degree in computer engineering June 1994 from University of Balamand, El-Koura, Lebanon.

Currently she is the acting dean of the Information Technology faculty at the Royal University for Women, Bahrain. She has more than eight years of teaching experience at international universities in the Gulf region for master and bachelor level. Prior to her academic job, she was working for three years in the industry as a telecom and software engineer in an international company named Integro. During these years, she was supervising big project for Deutch Telecom and France Telecom (currently named Orange).

Dr. Farah is a member of the order of Engineers, Lebanon. She is an active researcher in the field of computer security and Web intelligence. Currently Dr. Farah is working on publishing her first book about Contactless Payment.