WEB EVOLUTION - THE SHIFT FROM INFORMATION PUBLISHING TO REASONING

Abdulelah A. Algosaibi¹, Saleh Albahli², Samer F. Khasawneh³, and Austin Melton⁴

¹Department of Computer Science, King Faisal University, KSA.
²Department of Information Technology, Qassim University, KSA
³Department of Computer Science, Walsh University, USA.
⁴Department of Computer Science, Kent State University, USA.

ABSTRACT

The Web, as communication channel, has had variety of development that allows information to be published and accessed in a scaleable approach. With the revolution of the information, some research studies have conducted to boost the present situation and propose advance version of the Web. Therefore, it is important to look into the new version of the Web in order to improve the way that information is expressed, to make more intelligent choices and to obtain a better meaning of the information over the Web. That is, future web would require specific architecture in order to support the extracting of better meaning or "reasoning". With Web 1.0 and Web 2.0, the current information over the Web is not understandable for the machines. Understanding is big shift for wide open door for innovation and reasoning. In this work, we research the progress of the Web from Web 1.0, Web 2.0, Web 3.0, Web 4.0, to Web 5.0. We are pointing out document types and technologies employed to understand the changes from Web 1.0 to Web 3.0 and to predicate the future of the Web (Web 4.0 and Web 5.0). Also, we present the current status and concerns about the Web as an information source and communication channel.

KEYWORDS

Web Generations; Web 1.0; Web 2.0; Web 3.0; Web 4.0; Web 5.0; World Wide Web; WWW; Semantic Web; WebOS; Intelligent Agent.

1. INTRODUCTION

The trend and behavior of publishing and manipulating information have been evolved as more technology innovations are developed lead by technology company as Google, Youtube, Yahoo, Facebook, Twitter… etc. Accessing information also has changed, and more people are relying on the Web as a primary source of information. This information can be obtained from different places such as websites, blogs, online publications, social networks, databases and much more. However, they are not machine understandable. Indeed, the Web is a massive information exchange platform that was introduced by Tim Burners-Lee [1]. Basically, the idea was to link documents over the internet. These days, it affects social life, economics, news, politics, and many other normal life aspects. Now, with the evolution of the technologies, not only can we connect documents, but we also can understand documents. Documents, in general, come in three categories [2]. The first type is structured documents. Here the formation of the document and the inner data are structured in a way that for each piece of information, it is explicitly known how that piece of information fits with the other data. This leads to the retrieval of more relevant data. Examples of structured documents include databases or spreadsheets. The second type of documents is semi-structured documents. In this type of document, the data are structured, at least in part, based on semantics, but the underlying structure and the semantics are not explicitly
given. This kind of document, if carefully handled, can produce relatively rich semantic information because the semantics is embedded in data and the data’s structure. Examples of semi-structured documents are HTML documents, WordNet [3], and XML documents. The third type of documents is unstructured documents where the knowledge is available only in the data and not in the structure of the documents. Examples are standard text files.

The Web, as an information source, is holding enormous amounts of information in variant document structures. To automate this process, absorbing this diversity is vital to understand the current trend of the improvement. For example, in Big Data aims to effectively assist in efficient data processing that beyond trivial computing power especially with “Variety” and “Variability” characteristics [4][5][6][7]. As defined by [8] it is “the amount of data just beyond technology’s capability to store, manage and process efficiently”. New technologies as Apache Hadoop and MapReduce [9] were proposed to overcome the issues related with the massive prevalence data [8][10][11]. In first generation (Web 1.0) and second generation (Web 2.0), the Web was limited by the available technologies at the time. But that was enough to generate a lot of data in variant structure and format. Peter Norvig, Google’s director of research mentioned that “We don’t have better algorithms. We just have more data.” [12]. This makes the vision of the Semantic Web [13] (Web 3.0) an urgent task. That is Semantic Web or Web 3.0 allows computers to semantically process data. And, it allows for intelligent choices such as semantic selecting aggregation, sharing, searching, organizing or browsing. This idea is the base line for the future generation of the web as will be discussed in details in the following sections.

In this research, the progress and advancement of these generations with respect to documents types and technologies are presented. Also, a discussion of the foundation of the Semantic Web concept and technology are observed. Later, the expectations of the future trends in the next Web (Web 4.0 and Web 5.0) and outline the Web’s current status and some concerns are showed. In section 4, we are addressing some related work. Finally, the conclusion is in section 5.

2.WEB GENERATION

As mentioned earlier, documents come in three categories. Each one of these categories has received attention at certain generations. Table 1 shows when each kind of document received higher attention. These documents are processed by given technologies. The processing of these documents was limited by the availability of Web technologies at particular Web generations. Since progress was made to improve each generation, the related technologies have improved simultaneously within certain periods [14].

<table>
<thead>
<tr>
<th></th>
<th>Unstructured Documents</th>
<th>Semi-Structured Documents</th>
<th>Structured Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web1.0</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Web2.0</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Web3.0</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

In [17] the web versioning was categorized based on stages. For example, Web .5 (The Rise of Tim Berners-Lee’s Vision), Web 1.0 (Growth of the Web: The First Mainstream Websites), Web 1.5 (The Web of Experts), Web 2.0 (The Social Web), Web 2.5 (The Mobile Web), Web 3.0 (Semantic Web) … etc. Taking software engineering as an example, software or system versioning is used to differentiate a stage or status of a software [18][19][20]. The trend of the evolving of the World Wide Web technologies has shown a unique upgrade that differ one generation from the other. Semantic Web vision is an ambitious upgrade for a current problem. It requires a number of sources to share their information in a semantically structured manner so
that it allows machines to process these sources semantically. It is a technology that would have a
great impact over the service provided by Web Wide Web. According to [10], “The Semantic
Web provides a common framework that allows data to be shared and reused across application,
enterprise, and community boundaries.” Table 2 shows a comparison among Web generations
based on the most popular feature of each generation, some of the most technologies advanced
each generation, the scale of the functionality and finally, major differences between each one of
them. With this evolution, the Web infrastructure technologies become more advanced and
complex, yet more promising.

Table 2: Main characteristics of each Web generations

<table>
<thead>
<tr>
<th>Web Theme</th>
<th>Web 1.0</th>
<th>Web 2.0</th>
<th>Web 3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web Theme</td>
<td>Static Web</td>
<td>Dynamic Web</td>
<td>Semantic Web</td>
</tr>
<tr>
<td>Some Major Technologies</td>
<td>HTTP, HTML, XML</td>
<td>JavaScript, PHP, JSP, ASP, JAVA</td>
<td>OWL, RDF, SPARQL</td>
</tr>
<tr>
<td>Level of Functionality</td>
<td>Read Web</td>
<td>Read-Write Web</td>
<td>Read-Write-Execute Web</td>
</tr>
<tr>
<td>Key Difference</td>
<td>Information source but limited features.</td>
<td>Functional but enourmance amount of information.</td>
<td>Semantic linkage of knowledge.</td>
</tr>
</tbody>
</table>

With this evolution, the Web infrastructure technologies become more advanced and complex yet
promising, as Fig 1 [21] shows the partial Web architecture.

Figure 1: Partial Web architecture [21]

2.1. Web 1.0

Tim Berners-Lee views Web 1.0 as what is so called “read-only” Web. Web 1.0 is the beginning
and the first stage of the World Wide Web. It was limited to allow users to read and search for
information. No interaction was happening at all; the architecture of Web 1.0 made it to be static. The user interaction and content distribution was very limited as can be seen in figure 2.

![Publisher Content Client](image)

Figure 2: Web 1.0 is a one-way platform

Web 1.0 was simple, as shown in figure 3, in terms of information and how it was represented. It was invented to overcome the distributed information over computers and provide shared infrastructure that allow information sharing [22]. The Universal Document Identifier (UDI) now Uniform Resource Identifier (URI) [22] allows for identifying resources.

![HTML CSS PNG GIF JPEG](image)

![HTTP URL](image)

Figure 3: snapped from [21]

In fact, it was considered a “static web” on the Web. Web 1.0 was limited to the features it provided and did not exceed the layout representation of a website. In this scope, the medium was documents linked via hyperlinks over the Internet. Most of the information on Web 1.0’s websites was on the webpage itself. It was about a websites that do not interact with visitors to offer implicit functions. A good example of a Web 1.0 website is a professor’s homepage that has only information about courses, publications, or pictures. Web 1.0 was not implemented to offer a service. Webmasters need to manage website in terms of removing, adding, updating the content and hyperlinks [23].

2.1.1. Document Type

Most of the attention was on unstructured documents. As a result, Web technologies were simple in terms of the power of processing these documents (e.g., a web directory for navigating through a list of websites). Semi-structured documents had limited availability and did not exceed Hyper Text Markup Language (HTML) documents or data describing EXtensible Markup Language (XML) documents.

2.1.2. Technologies in Place

Transmission Control Protocol (TCP) and Domain Name System (DNS) were avalible before the Web got invented. Hypertext Transfer Protocol (HTTP) is an application-level protocol. It provides stateless operation, index search, retrieve document globally by name, reduce processing time, and allow for Pipelining [24][25][26]. Also, as mentioned earlier, Uniform Resource Locator (URL) or the web address was invented to g kobally locate resources address [27].
Technologies were focused on how to locate documents "understructure documents". HTML, Cascading Style Sheets (CSS), XML, or Web browser technologies such as Erwise, Viola, Cello and Mosaic [22] were a clear attempt to enhance the limitations of functionality that users experiences. HTML is the publishing language of the World Wide Web [28]. HTML tags on data give presentation capability that is processed by a web browser. HTML has been evolving since the invention of the Web. These tags structure a webpage. This structure is used for representing information in a webpage. Lately, in HTML5, some semantic tags have proposed (e.g., <section>). However, it is still on structural stages. A webpage, in Web 1.0 that was a user-end interface consisted of HTML tags along with text. This version of the Web interfaced with a web directory [29] that created a methodology to browse websites. Later, technologies such as “crawler” and “spider,” which store text within web pages were proposed [30][31]. Thus, processing Web 1.0 was limited by the technologies available.

2.2. Web 2.0

Web 2.0 is an extension of Web 1.0. It is more enhanced in terms of the features, functions, services and usefulness than Web 1.0. Web 2.0 is considered as offering “dynamic web” (e.g., social networking sites, wikis, video sharing sites, online shopping, and web applications). It is a web-as-participation-platform [32]. As defined by [17] Web 2.0 is "Web 2.0 services are user-oriented, content-sharing (upload, edit, and download), social networking (personal data), or static mashup services based on technologies supporting dynamic micropages that harness collective intelligence. They may support an open API with closed data and closed ID in order to use the Web as a distributed file system (user-generated content) or collaboration system (networking effects). Typical examples are YouTube, Flickr, Digg, Del.icio.us, LinkedIn, or MySpace, as well as basic supportive tools, such as Wikis or blogs." In Web 2.0, business models had to be modified in order utilizes the influence of this new trend in technology [33]. Given that, if a desktop-application software is available, then with Web 2.0 it is possible to have another version be a web-based application. Web applications provided with no code on the local machine but on Web server and access the application from Web browser. Web applications are slightly slower than desktop applications. However, they provide wide range of services that are not available on the desktop version. Web 2.0 is bidirectional communication [34]. All of the Web generations with respect to attention on technologies and document types features have a number of back configurations to make them work. Web 2.0 often follows software engineering principles in order to build scalable web applications. Following these principles will help in maintaining and enhancing web applications.

2.2.1. Document Type

With availability of Web Content Management System (WCMS) or Web Content Management (WCM) as in [35][36][37][38][39], unstructured documents have a massive amount of information that increases the number of services done over the Web. Semi-structured documents allow for more precise processing. Also, semi-structured documents get increased attention. The technology can more easily process these kinds of documents to include Natural Language Processing [3] and HTML documents parsing [40].

2.2.2. Technologies in Place

Technologies in this Web generation enhanced the processing functionality of Web documents. The technologies drove Web 1.0 to become more interactive. The technologies allowed the Web to be more dynamic (e.g., Web forums, Social network, online banking). The Web technologies allowed Web applications to be in place (e.g., JavaScript, PHP, Python, JSP, ASP.NET and JAVA). Also, Web 2.0 became more mature to give the user the ability to choose the architecture
of the Web application whether it was client-side such as JavaScript or server-side as JSP. SOAP (Simple Object Access Protocol) and REST (Representational State Transfer) allow for powerful services over the Web as in Yahoo, Google, eBay and Amazon. There are some web editors that help make the creation of Web 2.0 much easier and simpler (e.g., Microsoft Visual Studio that works with ASP.NET and C#). In fact, with the evolution of Web 2.0, allowed for richer Web content. As a result, navigation through the Web, particularly within the website, became harder. Navigation methodologies (e.g., sitemap and mature search engines such as Google) were invented to help speed up the process of finding a desired content. Also, with this evolution at the Web 2.0 generation, security principles including Confidentiality, Integrity, and Availability (CIA triad) became a real concern for individuals, companies, and governments. Because one characteristic for a Web 2.0 application is to encourage greater user involvement, the exposure of the individual user or client to security threats and vulnerabilities increases. Web applications demand a great effort of information security techniques for protecting users privacy[41][42].

With the continuous progress and growth of the Web 2.0, cloud computing became widely accepted as a replacement of many rigid IT infrastructure including hardware, software, man power, energy, offices. Cloud computing as defined by [43] as “Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models.” Moreover, expanding the severvies of a company may bump into the compatiabilty issues that prevents from fully utilization of IT infrastructure. It is allowing delivery of IT infrastructre, platform, and software (applications) which is refered as as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) Cloud computing also allows small and medium size company to overcome the issues with expensive hardware or software installment [44][45][46][47][48]. Big data and cloud computing has sort of relationship as big data utilized cloud serviveses and infastrucutre [49][50]. With its current tendency in the adancement of archticutre, cloud computing has a strong potentional to lead the advancment of Web 2.0.

2.3. Web 3.0

Semantic Web (SW) is another type of Web that builds above the existing version of the Web as shown in Fig 2.5 and is the vision for the coming World Wide Web [13]. It needs structured documents that a machine can process by querying or inferencing to derive more precise information. Tim Berners Lee, the inventor of the Web, coined the term Semantic Web for a Web of meaning that makes it understandable to machines rather than just readable by machines. As such, the Web has been developing toward this vision by embedding huge quantities of machine-processable metadata, structure and different semantic Web technologies into the current Web [51]. Basically, the Semantic Web tries to shift the thinking of published data in the form of Webpages (i.e., HTML documents) to allow machines to understand the contents. The content of Web 1.0 and Web 2.0 suffered from a number of issues, including the amount of information and how to access it and enable delegation [52]. They were provided for humans rather than for comprehension by machines. It therefore was not easy to automate data across the Web. To this end, the key idea behind the Semantic Web is to identify and link the content of the Web in a way that allows machines to understand and derive meaning from the data. Recognizing this vision requires new approaches, languages, technologies and data representation models to be built [53]. For this reason, a variety of semantic languages and standards are maturing, and different applications, tools, and services are developing as shown in Figure 3 and Figure 4. These technologies were organized in strata which are increasingly dependent on their complexity. It is like a layer cake as presented by Tim Berners-Lee where top layers are based on the lower ones.
The base semantic layer shown in Fig 3 is called the Unicode or URI that is inherited from the current web and provides a unique identifier for each semantic object. It also supports international characters using IRI [4]. The XML (eXtensible Markup Language) layer provides a syntactical structure of documents. It defines a set of rules for encoding documents in a format which is both human-readable and machine-readable. It is defined by the W3C's XML 1.0 Specification. It is used for interchanging data over the Internet. It provides flexible text format and used in numerous Software development activities.

On top of this is Resource Description Framework (RDF). RDF is a simple modeling language providing a standard model for data interchange on the Web. RDF generates the linking structure of the Web to use URIs to name the relationship between things as well as the two ends of the link which is usually referred to as a “triple”. It uses URI to identify the Web resources and forms a graph model describing the relation between the resources. Resource Description Framework

![Diagram showing Web fusion and generations](image)

Figure 3: Web generation with respect to attention on technologies and document types [15]

Schema has been added additionally with RDF as a layer in the later versions RDF provides better intelligent solutions than XML. Although RDF is built on top of XML, it is not based on XML, but it may use the syntax of XML [2]. With RDFS, which is short for RDF Schema, hierarchies of classes and properties are provided with domain and range restrictions. RDF schema can be seen as a limited ontology language. In light of this, a robust ontology language (like OWL) provides a much more knowledge and it sometimes requires working with complex web resources. All these data need a query language to retrieve information. A standard semantic language called SPARQL is used as a RDF query language. The rest of the Semantic layers are still unstandardized ones from the W3C. However, they have to be implemented to recognize the
The Rule Interchange Format (RIF) was evolved in parallel with OWL to get the support of rules and relations that cannot be defined utilizing the description logic used in OWL. The cryptography layer is significant to make sure that RDF statements are used under a trusted source. This can be performed by digitally signing RDF statements. The top layers of the stack which are “Logic, Proof and Trust, are currently being researched and simple application demonstrations are being constructed” said (W3C, [6]). The highest layer is the user interface that allows using applications of the Semantic Web. In the case of implementing the Semantic Web for a single website, this implementation does not help much. Rather, the multiple of sources should be semantically structured to continue the chain of multiple websites in order to reach the required information that serves the ultimate goal of the Semantic Web. Data on Web 1.0 and Web 2.0 are about connecting information. However, with Web 3.0, it is about connecting knowledge and semantically structuring documents.

2.3.1. Document Type

The most attention on Web 3.0 was devoted to work on structured documents. Structuring documents includes machine-processable format such Resource Descriptive Language (RDF) or Web Ontology Language (OWL). RDF acts as a data model used to manage, structure and reason about the data found on the Web, and to show how the data relate in reality [54]. In this manner, RDF is a way to represent a small chunks of knowledge and how they are related to each other something lacking from the Web 2.0 and XML particularly. RDF is a graph data model (Fig 5). In RDF, the graph serves to define the massive collection of triples in graph form. It is labeled directed graphs that represent statements as a set of nodes forming a network of information and

![Figure 4: Semantic Web Layer Cake (Ref: W3C [13])]
how they are related to each other. Ontology languages such as RDF Schema (RDFS) and OWL are semantic Web languages that provide semantic meaning for RDF data. From the document type points of view, OWL represents semantics in the same way as RDF (structured way) but has more classes and properties to add semantic richness to RDF.

Therefore, RDFS and OWL are data modeling for displaying RDF data in a structure ontology-based format. Web Ontology Language (OWL) [1] (and its new version OWL 2) is a common semantic language for creating and representing ontologies. OWL had been started in 2002, and the final W3C recommendation was released in 2004. It is utilized as a fast, effective, and flexible data modeling language to add semantic richness to RDF. Moreover, it works efficiently in terms of automated inferencing/reasoning. However, comparing it to RDFS, “in many cases we need to express more advanced, more ‘expressive’ knowledge for example, that every person has exactly one birth date, or that no person can be both male and female at the same time,”[2]. OWL actually adds extensive inferencing capabilities to go beyond the simple semantic of RDFS, with three sublanguages: OWL Lite, OWL DL, and OWL Full. OWL Lite, provides a simple level of OWL dialect with lightweight reasoning capabilities aiming to give classification hierarchies. It also provides easy implementation and fast results. OWL DL (DL stands for Description Logic), supports enough expressivity of logic, but has a decidable inference procedure. Decidability means an approach can return a Boolean value which is different from semidecidability or undecidability that have no guarantee to terminate. OWL Full, is the most expressive OWL dialect, but provides no guarantee that reasoning statements can end up in finite time.

2.3.2. Technologies in Place

![RDF/S graph](image)

The existing Web 2.0 utilizes natural language, multimedia, files, graphics, and much more so that it is easy for people to read and process information; however, with Web 2.0 it is difficult for machines to derive meaning from the information. It is not an easy task for computers to traverse the data meaningfully and even for humans to locate related information. In this context, technologies must be developed to realize the vision of Web 3.0. Some important technologies in the Semantic Web era are SPARQL Protocol and RDF Query Language (SPARQL), LINKED DATA, RDFaCE (RDFa Content Editor), JSON, and DBpedia Spotlight. SPARQL, Along with
previous ontology languages (RDFS and OWL), is one of the three main semantic technologies. It is a query language for semantic datasets that works via pattern matching. Moreover, SPARQL not only provides facilities to query against data, but it also queries against the semantic schema. This query language can work with disconnected datasets (e.g., Linked Data) that are already mapped by RDF in order to retrieve data and obtain results. In this way, Linked Data, although it is still being developed, links different RDF datasets. Furthermore, one of the important Web 3.0 technologies is RDFa (RDF in Attributes). It is a technology for embedding and serializing data between XHTML tags. Therefore, it links the technologies from Web 2.0 with Web 3.0. Recently, big search engines like Google, Yahoo and Bing have been using RDFa to collect and integrate data from web sites. Hence, RDFa helps, for example, Google to provide rich semantics data to express concepts to the user in a few lines under every search result (Google Rich Snippets). Another company using RDFa is Facebook. It uses Open Graph Protocol to display information contained on web sites. Having considered RDFa, it is also reasonable to look at JSON (JavaScript Object Notation). JSON is used to interchange and process data and appears to be an alternative to XML. Consequently, it is another technology that can combine the two eras (Web 2.0 and Web 3.0) by loading data that is represented in Web 2.0 (e.g., Ajax, XML, XHTML) more quickly to avoid delays in site rendering.

Some solutions to manage Web 3.0 technologies have been proposed such as Protégé which is an ontology editor for facilitating intelligent solutions. It analyzes a set of classes and axioms for a given ontology graphically to infer new information from existing information. It also provides an interactive box that allows users to query them using SPARQL. It has a large number of suggested plugins for editing ontologies. Additionally, Protégé can be used as backends storage to Sesame and Jena [3]. Both of Sesame and Jena are frameworks utilize for storing, inferencing and retrieving large number of RDF data. They are quite similar but the inference support, scalability and performance are different. Also, each one of them does complex queries with different query languages. Sesame supports SPARQL and SeRQL, where Jena supports SPARQL and ARQ. However, unlike Protégé framework which works as an editor to RDF triples and ontologies.

Some of the key aspects that make the SW very crucial in the near future mentioned by [86] are the social Web where people share their feelings, ideas, and thoughts using the Web 3.0. it is considered as an well-organized and attractive way of connecting people around the globe. Also, Web 3D that is virtual worlds have been very popular for several communities including education and gaming to improve the experience and activities. Moreover, search engines would be able to take media input elements such as an audio file, and search for objects that are similar to those. Some IT experts anticipate the practical part of Web 3.0. For example, Conrad Wolfram says that Web 3.0 is where computers will be generating new information rather than humans [88]. Google Chairman, Eric Schmidt mentioned Web 3.0 will be “applications which are pieced together – relatively small, the data are in the cloud and it can be run on any device, very fast, very customizable and distributed virtually” [85]. Yahoo founder, Jerry Yang states that Web 3.0 is a collection of tools and techniques for creating programs and online applications, which will make no distinction between different user groups who are mostly nonIT professionals [87].

It is worth noting that as the Web has been progressing towards Web 3.0, security aspects need always to be kept in mind. Various issues related to scalability, security, and performance which were present in earlier Web versions are still valid in Web 3.0, and IT professionals have numerous challenges to ensure the integrity of the data. Due to ubiquitous nature of the data, and the interactivity nature of the Web, access standards have to be in place to validate data access. Furthermore, there is a lack of data standards for having control over metadata and data privacy. RDFS and OWL use URI to represent data which might exist in a database or a data warehouse, without specifying access policy or trust boundaries. This makes Web 3.0 vulnerable and attackers may have unwanted access to sensitive data.
3. **FUTURE TRENDS: CAN THE FUTURE TRENDS BE PREDICATED?**

Ten years from now, “semantic technology could be as ubiquitous as the Web is today, and combined with the capabilities of Web scalability, real-time reasoning, and world-scale knowledge management, there are both exciting new possibilities ahead as well as brave new challenges.” [52].

### 3.1. WEB 4.0

Although as of yet there is no clear technology for Web 4.0, but it is widely believed that Web 4.0 will be a “symbiotic Web” or “WebOS” that is a read-write-execution-concurrency Web [14] [55] that will handle structured documents. In the evolution of wireless technologies, new major shift to connect object anytime and anywhere that support real-time interaction between individuals and physical/virtual worlds [33]. Research in [14] mentions that the Web is moving into artificial intelligence. Given that, personalized agents can work smoothly with users to improve user experiences. These agents will be able to make a decision based on what a user wants, based on current conditions, along with other information to create an ideal response.

While the current attention is on Web 3.0, the development of Web 4.0 is expected to build from 2020 to 2030. With the advancements of artificial intelligence, nanotechnology, telecommunications and controlled interfaces, the machines are aimed to act clever on reading the contents of the web and deciding what content to be executed. Web 4.0 will be termed as “read write concurrency web”. An exciting example can be locating our vehicle key by searching our home at Google. As how the mobile was connected to the internet 24/7, our home will be connected to internet. Chips implanted to restore the vision of blind people or sensors on the brain’s motor cortex for controlling a computer with thoughts (possibly an avatar within a 3D world) can be the future scenarios. WebOS will act as a middleware, which is parallel to the human brain and entail a massive web of highly intelligent interactions [17][84]. This all can be done with the power of data modeling, such as RDF/OWL, that allows reuse, integration, and reasoning. Web 4.0 is expected to handle complex intelligent interactions over the Web. As defined by [23], one of the most critical developments of Web 4.0 will be the migration of online functionality into the physical world. Thus, it would improve the applications of machine-to-machine (M2M) communication and interface.

Daniel Burrus [56] mentioned that Web 4.0 is beginning already; it is about an “ultra-intelligent electronic agent” and personalized intelligent agent anticipated to be on every device. For example, according to Burrus, this personal intelligent agent says: “Good morning. You're flying to Boston today; take a raincoat, it's raining. By the way, that fight you were taking, it's already been canceled. Don't worry about it. There was a mechanical problem. I've already booked you on a new one, I'll tell you about on the way to the airport. But remember you're going to exercise everyday and I'm here to remind you that you're going to exercise.” And you might say, “I don't know if I want to exercise today, and it'll show you a nude profile of yourself. And you'll say, ‘You know what, I think I'm going to exercise today.’” The agent would be able to tell you what you have not asked for but what you should have asked for.

Having considered the rich ontologies and knowledge sharing in Web 4.0, it is also reasonable to look at automated and on-the-fly reasoning. It is expected to be an essential part of the ubiquitous Web to enhance the scalability of reasoning whether using heavyweight or lightweight reasoning. This advance will lead to more flexibility and effectiveness in cloud computing with new operating systems (OS), called WebOS or OS in the cloud [52]. WebOS with the OS and all its functionalities and contents such as data, applications and documents in the cloud, will all be accessible in one place. For example, Google bought Nest Labs, a company for home automation,
and recently introduced the Brillo Operating System that is an extension to the physical world [57]. We are expecting that it would open the prospects for the Web 4.0 era. Therefore, some challenges, such as large-scale Web, will be decreased as a result of applying a heavyweight reasoning with real-time reasoning over the WebOS.

It is important to understand that Web evolution is based on needs. That is, Web 2.0 complemented Web 1.0 by adding more functionality. Web 3.0 solves major issues in Web 2.0. Thus, Web 4.0 should enhance the user experience of Web 3.0. As Web 2.0 enhanced the functionality of Web 1.0, Web 4.0 is expected to enhance the functionality of Web 3.0. The main reason is that Web 4.0 will more affectively process semantic structure. Web 4.0 will merge semantic Web technologies to produce agent-based applications that provide products that work as operating systems to facilitate the interaction between humans and machines. Web 4.0 is expected to introduce intelligent applications based on Semantic Web technologies. We think that Web 3.0 is providing the infrastructure to develop Web 4.0.

3.2. WEB 5.0

Paper [58] thinks of the fifth generation (Web 5.0) as a “Sensory Emotive Web.” Web 5.0 takes into account the feelings of the user. It is guided by technologies that already exist to measure feelings and their effects. As an example, a company called Emotiv Systems [59]2 works in the field of neurotechnology. With headphones, the human brain can communicate with a machine. The machine can read conscious thoughts, emotions, facial expressions and head rotations. Research in [23] describes Web 5.0 as a “Symbionet Web.”. They mention it is “emotionally” neutral and notes that it does not count on a user’s feelings. In this regards, author of [33] mentioned that people “feel” and “think” and they do not necessarily act reasonably. We think it would consider structured data along with structured documents. Also, Web 5.0 may take advantage of data fusion algorithms and applications to merge with pervious Web generations.

The data fusion field has already proven successful in a number of domains, including discovery science and business intelligence, and it may be able to work with multiagent systems [60]. We think that Web 5.0 is about “Fusion Web” where machines and people will process data in forms that they can deal with, interact with, and make decisions with. The Web 5.0 is expected not just about helping humans to be better at the things we can already do; the Web 5.0 is potentially about helping humans to do things which we cannot currently do. An analogy is two hundred years ago, people could move pretty fast, for example, in a train, but we could not fly. Currently, people can think well (as we currently understand thinking), but if the Web 5.0 becomes what it seems it could become, then our thinking may take a jump comparable to going from running to flying. No doubt the Web 5.0 will not become all that we might dream that it could become, just as AI has not become all that people thought it might become. However, if the Web 5.0 becomes just part of what we think it might become, then it might lead to things which we currently have a hard time imagining. We may see Web 5.0 or “Fusion Web” join Web For All [61] to support people with special needs. Web 5.0 could be a read-write-execute-concurrent-fusion Web. Web 5.0 is expected to be built on the power of Semantic Web “Web 3.0” and Symbolic Web “Web 4.0.”.

3.3. Web Current Status and Concerns

At the last W3C 20th anniversary symposium [62], a number of topics were covered, including a long-term view of the World Wide Web and Access for All, and a panel discussion session titled “The Future of the Web and How It Is Run.” On this panel, Web fragmentation was a hot topic. Web fragmentation is a topic that concerns Web leaders, including Sir Tim (Web inventor) and Vint Cerf (Internet co-father) and other field pioneers. Technically, Internet fragmentation is
about having more than one Internet that results in Web fragmentation. In business, The World
Web Consortium (W3C) has considered the Digital Rights Management (DRM) in HTML, which
includes Encrypted Media Extension (EME) [63] specifications to prevent fragmenting the Web
into free-Web and charged-Web. Jeff Jaffe mentioned that “It is W3C’s overwhelming
responsibility to pursue broad interoperability, so that people can share information, whether
content is protected or available at no charge.” Fragmenting the Web also was mentioned at a
government level when German Chancellor Angela Merkel called for the European Union to
create its own regional Internet for security concerns [64]. Eugene Kaspersky, chairman and CEO
of Kaspersky Lab, mentioned that Internet fragmentation means the end of the Web by saying,
“But I fear that we are at a turning point for the internet, and may even be going into reverse. The
utopia of a borderless digital global village may be coming to an end. Fragmentation of the world
wide Web is already taking place—along national borders” [65].

Also, it is worth mentioning that the Semantic Web is growing exponentially, and its technologies
are refining and developing fast from many different sources every day. According to [66], the
Semantic Web technologies introduced as monolithic architecture. That is, different technology
built upon different technology to advance the power of Semantic. Given that, as shown in Figure
6, the author mentioned that the “practical” semantic technologies are already in business.
However, future and fantasy technologies are still in research labs. Some visionaries [52] of the
Semantic Web predict that in the next few years, Semantic technologies will be matured enough
to act as valuable sources of structured documents for different applications and wider integration
of Web content. In addition to do inference on the given knowledge for those applications,
semantic technologies also will have the ability to provide data provenance, which will allow for
tracing and verifying the sources of information. Consequently, we foresee that different types of
application domains will be enhanced by semantic research, including health care, smart life,
mobile technology, energy, and knowledge discovery. In this context, it is a matter of time before
the Web is getting empowered by robust reason-based architectures as in knowledge
representation & reasoning and AI that would have a bright reflect on some new emerging
technologies as Internet of Things (IoT) that is a system to interconnect computers e.g. notebooks,
tables, smartphones, and PDAs to objects e.g. devices or machines over the internet in order to
allow exchanging of the data [67][68]. As a reason for developing the Semantic Web technologies
and for increasing the amount of structured document is, we expect that most Web content will be
semantically marked up so that metadata will become increasingly important to reach the vision
of the Semantic Web. Therefore, applying ontologies over these structured data will be a new
gold to do better reasoning.

Figure 6. A practical view of the semantic stack [66]
Finally, the Semantic Web raises some new issues and challenges in research, such as the availability of content, scalability, multilinguality, visualization, ontology availability, development and evolution, and the stability of Semantic Web languages [69]. Some progress has been made in overcoming them as in ontology learning [70][43][44][19][30], storing RDF graph [54][71][72][73][74][75] and standardizing necessary technology [76]. However, some remaining challenges are still open for research, such as maintaining machine processable data and providing some mechanism that supports engineering tasks for ontologies [69][77][78]. Also, some challenges are mentioned in [79], such as vastness with data redundancy, inconsistency when ontologies from different parties are combined, deceit, and vagueness. Also, some researchers [80][81][82] have showed that some semantic security policies are needed for the evolving Semantic Web. Last, we think that in order to rebound the growth of the Semantic Web, current and new Websites need to seriously consider publishing the semantic part of the Website along with its deployment.

4. RELATED WORK

It is worth mentioning that studying the World Wide Web progress has been researched in some works [5][6][7][8][9][83]. Research [14] has surveyed the Web generations’ background evolution from Web 1.0 to Web 4.0. They studied the characteristics of these generations and provided some comparisons. Research paper [23] has mentioned the fifth generation of the Web (Web 5.0). Web generations also were described by Weber and Rech [17]. They proposed a definition of each generation. Key differences between the first two generations were studied in [34]. Diana, Marta, Carlos and Alberto [58] have discussed the possibilities for the fifth generation. Websites can be part of many Web generations [34]. In this work, we investigate about the development trend of the Web from Web 1.0, Web 2.0, Web 3.0, Web 4.0, to Web 5.0. We identify the attention on document types and technologies employed in order to understand the progressions from Web 1.0 to Web 3.0 and to predicate the future of the Web (Web 4.0 and Web 5.0). Furthermore, we show the present status and concerns about the Web as an information source and communication channel.

5. CONCLUSIONS

In conclusion, conversion from Web 1.0 to Web 2.0 is optional, as it depends on the need to have a dynamic website. Rather, we think conversion from Web 1.0 or Web 2.0 to Web 3.0 is necessary to employ for future development and advancement of the Web. In this regards, some solutions have been proposed to help in reaching Webs (e.g., RDFaCE). In the next Web (Web 4.0) or symbiotic Web, knowledge is expected to be structured well. Therefore, ontologies will be the new gold; namely, ending up with an enriched and efficient conceptualization analysis of a specific area of interest. This will lead to high knowledge sharing over databases and documents on the Web. Therefore, agent-based applications will be able to intelligently reason and perform tasks based on metadata. Moreover, experts in the semantic Web foresee that intelligent software agents will be interacting with humans in symbiosis [52]. Thus, the interplay between Web 3.0 and Web 4.0 will be a key technology for the growth of knowledge and for making the Web more useful for humanity.

REFERENCES


Authors

Abdulaleh A. Algosaibi earned his B.Sc. degree (with Dean-list) in Computer Science from College of Computer Science and Information Technology, King Faisal University, KSA, 2008. He earned his M.S. and Ph.D. degrees in Computer Science from Kent State University, USA, 2011 and 2015, respectively. He is interested in Semantic Web, Ontology Engineering, Natural Language Processing, Semantic Analysis, Artificial Intelligence, and Knowledge Representation and Management. He is currently working as Assistant Professor at the Department of Computer Science, College of Computer Science and Information Technology (CCSIT), King Faisal University, Kingdom of Saudi Arabia.

Saleh Albahli is an assistant professor of Computer Science at Qassim University. He defended his Ph.D in Computer Science from Kent State University in Spring 2016. He earned his Master of Information Technology from the University of Newcastle, Australia. Doing research primarily in data and information management, especially as they relate to Semantic web technologies and the integration of Semantic Web in database systems, graph data management, Big data management.

Samer F. Khasawneh holds a Ph.D. in Computer Science from Kent State University, Kent, OH. He received his master’s degree from University of Akron, Akron, OH and his bachelor’s degree from Al-Yarmouk University in Jordan; both of these degrees are also in Computer Science.

Austin Melton’s research areas include category theory, Galois connections, many-valued mathematics, programming language semantics, security, the Semantic Web, software engineering, and topological systems. Melton enjoys trying to understand and solve problems from a multi-disciplinary perspective. This perspective leads to trying to understand a problem abstractly and then maybe being able to apply the solutions in multiple settings. Melton is currently a full professor of computer science and mathematics at Kent State University. He has held faculty positions at five universities in the States and visiting positions at universities in Denmark, Germany, and South Africa. He has organized international conferences and workshops in category theory, many-valued mathematics, programming language semantics, and software measurement.