Individual Report
Hypertext and Web Technologies for Masters: A Survey of Hypertext on the Web

Chaohai Ding

School of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, UK
cd8e10@ecs.soton.ac.uk

Abstract. Even though the World Wide Web (Web) is the most powerful distributed hypertext system in the world currently, it still exposes some initial weakness in its architecture. This paper mainly presents a brief history of hypertext and its development on the Web. It also proposes the possible future of the Web after evaluating the open hypermedia systems and the Web.

Keywords: Hypertext, Open Hypermedia, Semantic Web

1 Introduction

The World Wide Web (Web) is the most powerful distributed hypertext system in the world currently, but it still exposes some initial weakness in its architecture. Through evaluating the development of hypertext on the Web, there would be an approach to improve basic hypertext functionality of the Web. This paper mainly presents the development of hypertext as well as open hypermedia on the Web. Section one presents a brief introduction about the history of hypertext on the Web, while Section two evaluates and compares the Web with other two hypermedia systems. Section tree analyses the current web technologies which would contribute to the future of open hypermedia on the Web.

2 Historical Overview of Hypertext on the Web

Before the invention of the Web, scholars had proposed the hypertext systems for a long time. Twenty years after the first time conceived by Vannevar Bush in 1945[1], Doug Engelbart implemented oN-Line System (NLS), a report sharing system. Then Ted Nelson coined the word "hypertext" in 1965 and started an ambitious distributed hypertext project, Xandadu, which was a world-wide, no links breaking and versioning hypertext system. In 1967 the first implemented hypertext system, Hypertext Editing System was developed by Andries van Dam in 1968. Moreover, scholars started to develop other complex functionality in hypertext such as the well-documented hypertext in the NoteCards and the

second generation hypertext which was used in Personal Computers like the Guide System described in [2]. In 1987, Halasz proposed the seven issues [3], which provided a standard to be compliant for hypertext systems. From then on, hypertext came into a growth stage.

2.1 The Web and Web2.0

With the development of early hypertext systems and the Internet, in 1989, Tim Berners-Lee proposed a solution for hypertext information losing in[4] and published the global universal hypertext system, the Web. The early Web was a simple approach to hypertext depended on a stateless protocol of HyperText Transfer Protocol (HTTP) and Hypertext Markup Language (HTML) that provided a way to capture the hypertext links and target document. The early Web browser NeXTStep implemented a simple hypertext functionality which provided a function for authoring users own links. This function was not inherited in the later browsers such as Mosaic. Due to the lack of arbitrary link authoring in the browsers, various third-party linking servers were appearing at that time to provide services of separated linking. Extensible Markup Language (XML) and Document Type Definition (DTD) defined a standard data format to exchange various data format between different applications. The early Web just allowed users to access to the textual or image document which was not the multimedia. From 2004 afterwards, the appearance of Web 2.0 allows the collaboration and interaction for multiple users. With comparing the some mainstream websites in Web2.0, some shortcoming of Web2.0 such as supporting content search but not structure search, dynamic content supporting, little version supporting and no supporting for typed n-ray links was presented in[5]. As a platform, it is still a closed hypertext, although Web2.0 has the primary features such as openness, community and interaction. And all these features are applied on the top of the Web, not the fundamental feature of the Web.

2.2 The Semantic Web

In 1999, Tim Berners-Lee[6] proposed the assumption of the next generation Web called the Semantic Web, which means that the data on the Web could be readable and understood by machines. Although the HTML performs effectively in the context search, machines should search the context through the Semantic Web that performs more effective. With the rapid development of semantic technologies like XML, Resource Description Framework (RDF) and Web Ontology Language (OWL), the Semantic Web offers a platform for automatically collecting and extracting the information in the real world by machines[7]. Ontological hypertext described as a fundamental hypertext in the Semantic Web, provides a structure approach to search and navigate the Web. Although, there are also some challenges for the Semantic Web such as the stability of the expression language and information overload[8], the Semantic Web promises a comparable bright future of the Web.

3 Architecture of the Hypermedia Systems on the Web

After 1990, although the Web is experiencing a rapid growth with the development web architecture and browsers, the early Web performs as a closed hypertext system due to its linking and hypertext functionality. According to its simple hypertext model that is not adequate for the development of information, scholars start to search some solutions from early open hypermedia systems such as Hyper-G and Microcosm, which indicated an advanced performance in dynamic linking and structured searching.

3.1 The Web

According to W3C Recommendation in[9], the architecture of the Web is primarily designed in the aspects of identification, interaction, data formats and extensibility, which is the most compiler with Halasz's 7 issues[3]. The URI is a cornerstone of Web architecture which is the identification of the local or remote resource. URIs are analysed by the browser that performs the representation of the resource through the protocols like HTTP, FTP or SMTP. A main factor for the success of the Web is that it allows an embedded reference to other resources via the URIs, which provides a simple way to create and retrieve the hypertext as well as the interaction between agents, but this is also a limitation of the Web. The format data structure such as XML enhances the new applications to use the information of the Web and displays an advanced performance of scalability. The client/server model provides an easy way for multiusers. Although the Web is the most powerful hypertext system, there are still some weaknesses such as the insufficient description of error handling, linking missing and limitation of HTTP, a closed system, a stateless protocol as well as embedded elements and no semantic links in HTML.

3.2 Hyper-G

However, Hyper-G[10] is an advanced multiuser, multi-protocols hypermedia system integrated the functionality on the top of Web, which enhances the speed, flexibility and security. As an extension of the Web, Hyper-G proposes a simple client but a complex server model to handle the external protocols. The local Hyper-G server consisted of a full text server, link server and document server performs as a super server that understands the information stored in other servers[11]. Compared with the unidirectional links in the Web, Hyper-G indicates more flexible in its linking which is bidirectional and could automatically update the missing or broken links. Moreover, the document and database are treated as a collection which provides a build-in function of content search across other servers rather than the dependency of search engines in the Web. Hyper-G is also a distributed system which provides a large scope for information exchanging as well as a flooding algorithm used for servers to achieve the scalability but not scalable than the Web. In the aspect of security, Hyper-G presents a complex authorisation mechanism for users to deal with hypertext.

Hyper-G supports multimedia such as image, video and 3D comparing with the early Web that just supports text and image. Overall, Hyper-G is a global hypermedia system aiming to build a more universal and controlled information platform based on the environment established by the Web.

3.3 Microcosm

With the main problem faced for early hypertext systems, Microcosm is an open hypermedia system dealing with the dynamic linking[12]. Although it is initially developed as a desktop-based system, the architecture performs more suitable for distributed system. Comparing with the early Web, this system corresponds the requirement of the open hypertext system described in [11], which performs a flexibility and openness. According to the architecture presented in [11], Document Control System (DCS) primarily manages the document for the viewer to give the presentation of different resources. The Filter, as the core hypertext functionality in Microcosm, is a message processing model that is available for accessing by other query hypertext systems. The whole system provides a flexible and open interface for other hypertext systems and information integration[13]. In the aspect of data linking, compared with the URL in the Web which is not available (404 Error) if the document reference is removed, but document identifier in Microcosm performs an improvement in maintaining the linking with remote resources. Unlike the Web embeds the data anchor in the texts which need to rewrite documents, Microcosm indicates its position in external linkbases, which means more convenient to add a new linkbase to the text. Comparing with the simple node-links mechanism to deal with the global hypertext in the Web, a generic link mechanism, separated links and nodes, is applied in Microcosm which presents a collection of resources and plays a significant role in increasing the scalability. Document viewer solves the problem of dead end linking in the Web because multiple kinds of documents are presented in the viewer. However, as a later coming system, Microcosm provides the compatibility for supporting some existing systems and protocols. Overall, Microcosm is initially designed to process a large scope hypertext and then becomes an open hypermedia model that supporting other existing closed hypertext systems.

4 The Future of Open Hypermedia on the Web

Open hypermedia as an area of specification has discussed for a long time. The primary responsibility of Open Hypermedia System (OHS) is to manage the hypermedia links which are considered as the first-class citizen and managed and stored in the linkbase[14]. As the weakness of the Web described above, research of this area trends to integrate the OHS with the Web, such as Microcosm and HyperDisco[15]. With the development of web technologies, such as XML, a structured data format provides a low level and standard data format for information exchanging. The existing XML Linking technologies provide an effective way to navigation the target fundamental hypertext. XPath use to find

the existing anchors presented in XML document and XPointer use to navigate arbitrary anchors. Both RDF that provides a standard framework for stating the relationship between different objects and OWL that presents the object on the Web contribute to the representation of real world for the open hypermedia on the Web. These technologies propose a standard way to navigate and search the structured hypertext, such as the Linked Data project[16]. The Semantic Web treats the whole Web as a machine readable database as well as human beings, which collects a huge amount of structured information. Although the Semantic Web is still in progressing, the integration of conceptual open hypermedia with the Semantic Web would propose a bright future of OHS on the Web.

5 Conclusion

After describing the history of hypertext on the Web and evaluating the Web compared with other advanced open hypermedia systems, the trend of the open hypermedia on the Web is described as the conceptual hypertext or the Semantic Web. The Semantic Web proposes a possible future of the Web which connects the data through the whole Web with semantic linking and other hypertext functionality recommended by open hypermedia. This trend would immediately contribute to the usage of effective information extracted by machines.

References

1. V. Bush and A. Think, "The Atlantic Monthly," *As we may think. http://www. theatlantic. com/unbound/flashbks/computer/bushf. htm*, 1945.
2. J. Nielsen, *Multimedia and hypertext: the Internet and beyond*. Morgan Kaufmann, 1995.
3. F. Halasz *et al.*, "Reflections on NoteCards: Seven issues for the next generation of hypermedia systems," *Communications of the ACM*, vol. 31, no. 7, pp. 836–852, 1988.
4. T. Berners-Lee, "Information management: A proposal," 1989.
5. D. Millard and M. Ross, "Web 2.0: hypertext by any other name?" in *Proceedings of the seventeenth conference on Hypertext and hypermedia*. ACM, 2006, pp. 27–30.
6. T. Berners-Lee and M. Fischetti, *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. Harper San Francisco, 1999.
7. T. Berners-Lee, R. Cailliau, J. Groff, and B. Pollermann, "World-wide web: The information universe," *Internet Research*, vol. 20, no. 4, pp. 461–471, 2010.
8. V. Benjamins, J. Contreras, O. Corcho, and A. Gomez-Perez, "Six challenges for the semantic web," *KR2002 (ISOCO White Paper)*, 2002.
9. I. Jacobs and N. Walsh, "Architecture of the World Wide Web, Volume One (W3C Recommendation 15 December 2004)," *World Wide Web Consortium*, 2005.
10. K. Andrews, F. Kappe, and H. Maurer, "The Hyper-G network information system," *Journal of Universal Computer Science*, vol. 1, no. 4, pp. 206–220, 1995.
11. F. Kappe, K. Andrews, J. Faschingbauer, M. Gaisbauer, M. Pichler, and J. Schipflinger, "Hyper-G: A new tool for distributed hypermedia," 1994.

12. A. Fountain, W. Hall, I. Heath, and H. Davis, "MICROCOSM: An open model for hypermedia with dynamic linking," in *Hypertext: Concepts, Systems and Applications. The Proceedings of The European Conference on Hypertext, INRIA, France*, 1990.
13. H. Davis, W. Hall, I. Heath, G. Hill, and R. Wilkins, "Towards an integrated information environment with open hypermedia systems," in *Proceedings of the ACM conference on Hypertext*. ACM, 1993, pp. 181–190.
14. S. El-Beltagy, W. Hall, D. De Roure, and L. Carr, "Linking in context," in *Proceedings of the 12th ACM Conference on Hypertext and Hypermedia*. ACM, 2001, pp. 151–160.
15. K. Anderson, "Integrating open hypermedia systems with the World Wide Web," in *Proceedings of the eighth ACM conference on Hypertext*. ACM, 1997, pp. 157–166.
16. C. Bizer, T. Heath, K. Idehen, and T. Berners-Lee, "Linked data on the web (LDOW2008)," in *Proceeding of the 17th international conference on World Wide Web*. ACM, 2008, pp. 1265–1266.