Postprint of: Olewniczak S., Boiński T., Szymański J., Towards Extending Wikipedia with Bidirectional Links, 31st ACM Conference on Hypertext and Social Media, 13-15.07.2020, Orlando, Stany Zjednoczone, pp. 239-240, DOI: 10.1145/3372923.3404841

Towards Extending Wikipedia with Bidirectional Links

Szymon Olewniczak, Tomasz Boiński, Julian Szymański

Faculty of Electronics, Telecommunications and Informatics
Gdańsk University of Technology, Poland
{szymon.olewniczak, tomasz.boinski, julian.szymanski}@eti.pg.edu.pl

Abstract

In this paper, we present the results of our WikiLinks project which aims at extending current Wikipedia linkage mechanisms. Wikipedia has become recently one of the most important information sources on the Internet, which still is based on relatively simple linkage facilities. A WikiLinks system extends the Wikipedia with bidirectional links between fragments of articles. However, there were several attempts to introduce bidirectional fragment-fragment links to the Web, WikiLinks project is the first attempt to bring the new linkage mechanism directly to Wikipedia.

I. Introduction

The hypertext research has a long tradition. During nearly six decades of its history, it resulted in many interesting concepts and systems. The most popular form of hypertext is called navigational hypertext, which is based on the idea of splitting information into chunks (called nodes), that are arbitrarily connected with links.

Although many hypermedia systems from the pre-Web era implemented advanced linkage mechanisms, the World Wide Web became the lowest common denominator of hypertext with simple embedded unidirectional links and without anchors overlapping [2].

II. Related Work

When it became clear that the Web would dominate the Internet as a global hypermedia system, many researchers have started proposals for extending its limited linkage capabilities by some external solutions. This new wave of hypertext systems was called open hypermedia systems (OHSes) [3].

In the context of the Web, OHSes implemented more advanced linkage mechanisms on top of existing web pages. There are two possible approaches to this integration. First is server-side integration, where the OHS is an extension to the HTTP server, that augments the Web content (eg. Chimera OHS [1]). Second is client-side integration, which relies on browser extensions or scripts that provide new linkage capabilities (eg. Arakne Framework [4]).

Current solutions allow overcoming some limitations of the Web linkage mechanisms but none of them addresses the specific nature of Wikipedia, with its revisions mechanisms and continuous evolution of content.

III. WikiLinks System

WikiLinks may be classified as OHS that integrates with Wikipedia on the server-side level. The system is implemented as a server application that communicates with clients using dedicated REST API (Figure 1).

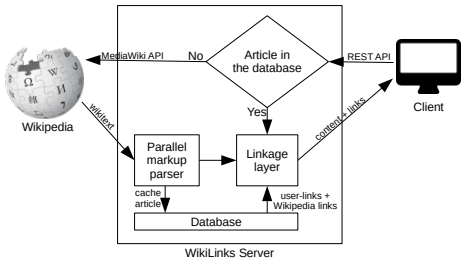

Figure 1: *The architecture of WikiLinks.*

Whenever the user requests a Wikipedia article, WikiLinks tests if it is available in a local cache. If it is not, the article is downloaded from Wikipedia using MediaWiki API and parsed by the parallel

markup parser. Finally, the links are applied and the content with the links is sent to the user.

WikiLinks differs from other Web OHSes in two main aspects. Firstly, it works on the wikitext level, not on the HTML level, which gives it better control over the augmentation process. Secondly, it supports link anchors adjustments on successive articles' revisions, through the link evolution mechanism, which is Wikipedia's specific process.

i. Parallel markup parser

The introduction of an additional linkage layer to the Web is connected with two main challenges. Firstly, the Web anchors are embedded in page content, so adding additional links requires page modification. Secondly, HTML does not allow tags overlapping, nor multiple anchors for the same fragment.

One possible solution to overcome those problems is parallel markup [5]. The main idea of this concept is to keep raw text content and formatting information separately. In the parallel markup approach, tags are not embedded, like in HTML but kept in a separate collection, with information about their length and position in the text. This approach makes applying additional anchors to existing content straightforward.

The parallel markup parser module in WikiLinks takes wikitext as an input and extracts the text content from it with the accompanying collection of formatting tags and anchors. In WikiLinks, formatting tags are treated like links with only one anchor.

The WikiLinks client receives the content with the corresponding links collection, translates it to HTML, and displays it in a browser window. This requires some adjustments, since HTML does not support overlapping tags, but can be achieved by splitting parallel tags into smaller, non-overlapping parts. The WikiLinks server itself is independent of HTML and can also be accessed by clients that are not browser-based.

ii. Links

Links in WikiLinks are first-class objects with a unique identity. The links are immutable objects and cannot be changed after creation. Each link contains an arbitrary number of anchors, called link sides. The links with only one side are called one-sided and are used primarily for content formatting. The links with two or more sides are used to establish connections between articles. By convention, in two-sided links, the first side is called *source* and the second *destination*. Each link can also store an arbitrary number of attributes.

The link anchor can point to the entire article or to the article fragment. The fragment cannot exceed one line of wikitext, which is imposed by the links evolution mechanism. The parallel markup parser parses standard Wikipedia web-links as two-sided links that connect fragment of one article with another article.

We distinguish two sources of links in WikiLinks. The first is called Wikipedia links and contains links created from wikitext by the parallel markup parser. The second is user-links which contains links created by the system users and applied to the Wikipedia content. All links in WikiLinks are bidirectional, which means that they can be accessed from all sides.

In addition to the unique identifier, each link has also a special *lineage* parameter that determines the link family. Using this parameter, we can select links ancestors and successors generated by the link evolution procedure (Figure 2).

Link v.1	
id	f777aa0c-83fd-4091-88fe-6463f2645b42
type	rel
lineage	d6394d9a-b36e-4005-a2f7-c2f120fd6142
author	user
timestamp	2020-06-05 10:45:35.357175
sides	**uuid:** 06b9b429-ad38-4e74-b9c3-a345cdd59f87 **title:** Green **line:** 3 **start:** 1 **length:** 130
	uuid: 2ae110f0-6ab4-4d0f-8c23-9c85e53a1670 **title:** Blue **line:** 1 **start:** 1 **length:** 108

Link v.2	
id	1fd33bb2-ba24-4fa1-b4a4-7d4f85dfad3d
type	rel
lineage	d6394d9a-b36e-4005-a2f7-c2f120fd6142
author	user
timestamp	2020-06-05 10:45:35.362318
sides	**uuid:** 2acae2af-e56a-4d13-9147-2a43ebea31f0 **title:** Green **line:** 2 **start:** 1 **length:** 130
	uuid: 2ae110f0-6ab4-4d0f-8c23-9c85e53a1670 **title:** Blue **line:** 1 **start:** 1 **length:** 108

Figure 2: *Two successive versions of a link after modification of the article: "Green". We can see the same lineage parameter in both links.*

iii. Versioning

One of the significant differences between Wikipedia and standard web pages is explicit versioning support. Wikipedia articles are changed constantly, which raises a need for user-links evolution procedure, which will keep the user-defined links valid through the successive revisions.

In WikiLinks, it is a two-staged process. When a new revision of the article is created, the links evolution procedure matches the modified lines between the previous and the current version, using line diff algorithm. Then for each matched line, it adjusts the anchors accordingly, using charter diff. After adjusting the anchors, a new link is created with new anchors but with the same attributes and lineage as its parent.

The proposed mechanism for two-sided links leads to a situation where we have two links pointing to

the same anchor. When this happens, the system shows only the newest link version.

IV. Future work

There are two main goals for developing a WikiLinks system. Firstly, we want to provide proof-of-concept for a new type of wiki parser that allows extending Wikipedia with new linkage facilities. Secondly, we treat WikiLinks as a research tool for testing new linkage concepts on Wikipedia articles.

We have already conducted a study that tested the usefulness of associative links for Wikipedia which results are very promising. In the future, we are also planning to conduct the experiments for transclusions and narration splits.

The parallel markup parser supports currently only a small subset of MediaWiki syntax. We are planning to extend it in the future, to make it a competitive alternative for the default approach.

Finally, we are also researching for a better evolution mechanism for links, since the current heuristic approach, based on diff algorithm, leads sometimes to errors in proper anchors identification.

References

[1] Kenneth M. Anderson. "Integrating Open Hypermedia Systems with the World Wide Web". In: *Proceedings of the Eighth ACM Conference on Hypertext*. HYPERTEXT '97. Southampton, United Kingdom: ACM, 1997, pp. 157–166. ISBN: 0-89791-866-5. DOI: 10.1145/267437.267454.

[2] Claus Atzenbeck and Mark Bernstein. "Interview with Andy Van Dam". In: *SIGWEB Newsl.* Winter (Mar. 2018), 1:1–1:7. ISSN: 1931-1745. DOI: 10.1145/3183639.3183640. URL: http://doi.acm.org/10.1145/3183639.3183640.

[3] Claus Atzenbeck et al. "Revisiting Hypertext Infrastructure". In: *Proceedings of the 28th ACM Conference on Hypertext and Social Media*. HT '17. Prague, Czech Republic: ACM, 2017, pp. 35–44. ISBN: 978-1-4503-4708-2. DOI: 10.1145/3078714.3078718.

[4] Niels Olof Bouvin. "Augmenting the Web through open hypermedia". In: *New Review of Hypermedia and Multimedia* 8.1 (2002), pp. 3–25. DOI: 10.1080/13614560208914733.

[5] Theodor Holm Nelson. "Embedded Markup Considered Harmful". In: *World Wide Web J.* 2.4 (Nov. 1997), pp. 129–134. ISSN: 1085-2301.