

HTML5

A step towards the Semantic Web?

Dominik Kundel
d.kundel@jacobs-university.de

January 16, 2014

Abstract

Sir Tim Berners-Lee describes in an article, published by Scientific America in 2001, that the current Web will evolve into a Semantic Web, also known as Web 3.0 [BLHL]. That this evolution is desired, is supported by the fact that the World Wide Web Consortium (W3C) founded a special group working on specifications towards a Semantic Web.

This year another big step in the history of the web will be made through the publication of the final version of the HTML5 specification. HTML5 is the first update of the HTML specifications made by the W3C since 1999 [GMEG13] and earned already in 2010 major publicity through the "Thoughts on Flash" by Steve Jobs [Job10]. The specification introduces a variety of different elements to HTML. These changes include the introduction of a variety of new HTML5 specific tags including `<article>`, `<section>`, `<nav>`, `<header>`, `<footer>`, etc.¹. The members of the W3C are additionally working on the introduction of a standard for custom tags, called Web Components [Gla14]. The problem of the current specifications (HTML 4.01) is that they resulted in an extensive use of `<div>` tags, especially for layouting and styling purposes. A perfect example of this is Gmail that uses just in the display of the inbox more than 1000 of these tags. This makes a systematic parsing of websites, for example for screen readers, extremely difficult. Another support that got introduced with HTML5 is the support for drawing SVG graphics using specified SVG tags.

The importance of the new tags, introduced by HTML5, for documents got additionally highlighted when the IDPF announced the support of HTML5 in their EPUB 3.0. The consortium decided to move away from their own defined XML structure for the list of contents to the newly introduced `<nav>` tags [GMEG11].

¹Detailed Overview at MDN: https://developer.mozilla.org/en-US/docs/Web/Guide/HTML/HTML5/HTML5_element_list

In this paper we cover in which way the introduction marks a new step towards a Semantic Web and improves our interaction with digital documents such as eBooks. For this we focus especially on the, still experimental, Web Components and their potential influence. We create documents using different states of the HTML specification (4.01, 5, 5+Web Components) and compare them on a basis of ease of implementation, compatibility and the ease of information extraction.

References

- [Bid13] Eric Bidelman. Custom Elements - defining new elements in HTML. <http://www.html5rocks.com/en/tutorials/webcomponents/customelements/>, 2013. HTML5Rocks, Visited: Jan 16, 2014.
- [BLHL] Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *Scientific America*. <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>, Visited: Jan 16, 2014.
- [Gla14] Dimitri Glazkov. Custom elements. <http://w3c.github.io/webcomponents/spec/custom/>, 2014. W3C Editor's Draft, Visited: Jan 16, 2014.
- [GMEG11] Markus Gylling, William McCoy, Erika J. Etemad, and Matt Garish. Epub content documents 3.0. <http://www.idpf.org/epub/30/spec/epub30-contentdocs.html>, 2011. Visited: Jan 16, 2014.
- [GMEG13] Markus Gylling, William McCoy, Erika J. Etemad, and Matt Garish. Html5 - A vocabulary and associated APIs for HTML and XHTML. <http://www.w3.org/TR/html5/>, 2013. W3C Candidate Recommendation, Visited: Jan 16, 2014.
- [Job10] Steve Jobs. Thoughts on Flash. <http://www.apple.com/hotnews/thoughts-on-flash/>, 2010. Visited: Jan 16, 2014.