Software Languages Practinar Summer Semester 2012

A Taxonomy of HTML5

by
Tobias Keweloh
Jan Rüther

07/22/12

Contents

1 Introduction

1.1 Problem definition

HTML is a living standard used in the web, historical independent from the W3C-specifications. Its a permanent developing language and in most cases a Recommendation by W3C only determines features being used in the web already.

In Addition WHATWG and W3C are both specifying HTML5. Having finished HTML 4.01 W3C officially stopped developing HTML and tried to replace it with XHTML based on XML. According to Kröner (2011) browser producers considered that not to be in actual step with actual practice and founded WHATWG to develop HTML5. By now W3C resumed working on HTML and both organizations are not always following the same directions.

Most publications covering HTML5 are practical, not covering the scientific background of HTML. In particular there is only one Taxonomy by Mavrody (2012) to be found and in our point of view it leaves essential questions unanswered. E.g. without consulting the book, its sometimes not obvious at all what features are addressed.

The frequent changes on HTML5 make it essential to publish a taxonomy on the web and summing up there isn't any by now.

1.2 Objective

As a result of our research we would like to publish an online taxonomy for HTML5, which is easy to maintain and extend. In the same way different views should allow that developers as well as scientists can benefit from the taxonomy. For an intuitional understanding we base our visualization on Spence's (2007) fundamentals representation, presentation and interaction.

1.3 Course of investigation

Starting with the problems arising from Mavrody's taxonomy, we decided to develop different views. We listed the features collected in the W3C and WHATWG specifications and tried to categorize them attribute-based. Following this we derived views by using the most promising attributes.

The kind of visualization was an highly important criteria for publication as well. First conclusion was that HTML5 is the central element for any view. Step two was to focus on a certain attribute by putting them in the second layer and sort the features accordingly into the third layer.

Each time we do the sorting, we have a different kind of tree, always having HTML5 as root and the Categories as first-child-node. Knowing it to be a kind of Tree-Structure we decided to visualize it via Sunburst[1]. This Visualization is based on the Treemap Visualization Technology by Shneiderman/Johnson. According to Stasko et al. (2000) users “preferred the Sunburst tool [to visualize trees], primarily due to its more explicit portrayal of the hierarchical structure”.

Working like this Spence's representation is covered by using a tree-structure and presentation by visualizing it with Sunburst. Leaves us with interaction, a part easily fulfilled in the web. Offering different views, switchable by mouse-click, offers the possibility to choose the specific view, focus features and categories and get supporting information and links as well.

The best way to maintain our data was table based in a conventional word processing program. To

1 we used: http://thejit.org/static/v20/Jit/Examples/Sunburst/example2.html

convert the data to a format we could visualize, we programed a converter transfering it to desired JSON data structures. We built a web page, where the sunburst and additional descriptions are displayed.

2 Visualization of HTML5

The essential result of our work is the visualization of the taxonomy itself. The more important a feature, the longer the related description is. Following we tried to distribute the visualization-size to that effect. In the following we'll describe the motivation for each view and category as done on the particular web pages.

2.1 Fundamental attributes

Because of HTML5 being a living standard its important to ensure that its obvious on which version the taxonomy refers to. Therefore the date of the corresponding W3C and WHATWG specifications is listed. Users can look-up both specifications directly per hyperlink, to allow best effort working.

2.2 Category

With the category attribute the features are sorted according to their main aspect from our point of view. The data-category includes all features regarding the storing, tagging and standardization of data. Features giving an audio-visual output to the user are represented by the visualization-category. On the opposite all features allowing user-input are allocated to the access-category. The communication-category contains all features covering interaction between different programming-languages via web.

2.3 Implementation Status

Implementation Status sorts the features according to how far they are implemented in different browsers as categorized by the WHATWG. We added in the wild and the status for those features not covered by WHATWG. The highest level is 'implemented', followed by 'heavy testing', 'in the wild' and 'not implemented'.

Implemented features have passed all possible test-cases and are running in all browsers. If there are not covered test-cases the feature belongs to the heavy-testing-category. In the wild are those features for which there are implementations in the web but they are not yet in heavy testing mode regarding WHATWG.

2.4 Orientation

HTML is associated with lots of languages. There are languages embedded via XML. Very important for HTML is ECMAScript called JavaScript. Last but not least there are pure HTML tags.

Resulting from our research it emerged to which category a feature fitted best. Therefore a feature belonging e.g. to JavaScript could have important aspects in HTML as well.

2.5 WHATWG vs. W3C

HTML5 is being standardized by W3C and WHATWG. Both of them have specifications for HTML5. This view covers if they are the same or not. The categories are 'yes', 'nearly', 'no' and 'only W3C'.

If a feature is categorized by 'yes' the specifications are exactly the same. In the nearly-category are those where e.g. one hast an extra column, but differences aren't crucial. If they are sorted to 'no' WHATWG specification is only use-oriented and W3C specification covers all aspects. Some features are only covered by W3C specifications and therefore in the only W3C-category.

2.6 W3C-Status

According to the W3C (2012) “the W3C technical report development process is the set of steps and requirements followed by W3C Working Groups to standardize Web technology.”

This view displays the three steps being essential in our point of view. The first step in specification-process is the 'Working Draft'. A Working Group is frequently updating the specification. After the Working Draft the specification becomes 'Candidate Recommendation' Status. In this step public review is expected. 'Recommendation' is the final status. W3C reached consensus that this specification should be implemented.

2.7 Combined View

This View combines W3C-Status in the inner circle and ImplementationStatus (WHATWG) in the outer circle. The idea is to point out where the development is concurrent and where not.

3 Conclusion and future prospects

Our Visualization of the taxonomy should be usable by scientists as well as developers. The different views on HTML5 allow each user fast overview from different angles. Most important for us was to publish it on the web, so it could be kept up to date as easy as possible and links to WHATWG and W3C specifications are included.

There are many features not collected yet and a lot of work is needed to carve them out from WHATWG and W3C specifications. On the contrary its easy to include these features in the visualization, just as it is a matter of seconds to generate other two or three layered views.

4 Bibliography

Kröner, Peter (2011): HTML 5 – Webseiten innovativ und zukunftssicher, Open Source Press, München, 2011.

Mavrody, Sergey (2012): HTML5 & CSS3 – Quick Reference, Belisso Corp., Chicago, 2012.

Spence, Robert (2007): Information Visualization – Design for Interaction, Pearson Educational Limited, Harlow, 2007.

Stasko, John; Catrambone, Richard; Guzdial, Mark; McDonald, Kevin (2000): An evaluation of space-filling information visualizations for depicting hierarchical structures, in: International Journal of Human-Computer Studies 53, S. 663-694, 2000.

W3C (2012): About W3C-Standards, http://www.w3.org/standards/about, retrieved 22.7.12.