How modern Web Browsers work[1]

Prof. Dr. Stefan Zander

07. March 2018

[1] Special Topic for the Module "Entwicklung Webbasierter Anwendungen"

Objectives:

- *Learn about the basic building blocks of modern Web browsers*
- *Get acquainted with the processing internals and the DOM building logic*
- *Understand what happens inside the browser when you type in an URL*

Contents

---

1 Preface

This lecture note is a revised summary of the excellent article "How Browsers Work: Behind the scenes of modern web browsers" published by Tali Garsiel and Paul Irish in 2011. The original article is available at: `https://www.html5rocks.com/en/tutorials/internals/howbrowserswork/`.

There is also a video available at vimeo about Tali's talk: `http://vimeo.com/44182484`.

> *In the years of IE 90% dominance there was nothing much to do but regard the browser as a "black box", but now, with open source browsers having more than half of the usage share, it's a good time to take a peek under the engine's hood and see what's inside a web browser. Well, what's inside are millions of C++ lines...*
>
> —Tali Garsiel

The following information and facts about the internal operation principles of WebKit and Gecko is the result of extensive research done by the Israeli developer **Tali Garsiel**. Over a few years, she reviewed all the published data about browser internals and spent a lot of time reading Web browser source code. Tali published her research on her site[2]. In the following years, her research results have been revised and republished on numerous occasions and provided insights to a larger audience.

[2] See `http://taligarsiel.com/`

Why should you learn about browser internals?

Learning the internals of browser operations helps you make better decisions and know the justifications behind development best practices. It also helps you to identify performance bottlenecks and build lightning fast websites. As we will see, page loading time has an influence on the Google page rank—a page loading time > 2 sec. results in a lower rank in the Google search results and the Google

TODO: Add refs

crawler also crawls such pages less frequently, meaning that search index terms are less frequently updated and the time until new or updated page content will be considered by the Google search engine is extended.

2 *Introduction*

Web browsers are the most widely used software. This lecture explains their fundamental operation principles so that students get an understanding about the things that happen internally when a website is requested, i.e., the time from typing in a website's URL in the browser's address bar until it is rendered by in the browser's viewport.

The complexity of Web browser software has significantly changed over recent years.

TODO: Add image of first NEXT browser; compare it with inspector of Google Chrome

3 *The Browser's Main Functionality*

The main function of a browser is to present the web resource you choose, by requesting it from the server and displaying it in the browser window. The resource is usually an HTML document, but may also be a PDF, image, or some other type of content. The location of the resource is specified by the user using a URI (Uniform Resource Identifier).

The way the browser interprets and displays HTML files is specified in the HTML and CSS specifications. These specifications are maintained by the **World Wide Web Consortium** organization[3], commonly denominated as **W3C**, which is the standards organization for the web. For years browsers conformed to only a part of the specifications and developed their own extensions. That caused serious compatibility issues for web authors. Today most of the browsers more or less conform to the specifications.

[3] https://www.w3.org/

Browser user interfaces have a lot in common with each other. Among the common user interface elements are:

- Address bar for inserting a URI
- Back and forward buttons
- Bookmarking options
- Refresh and stop buttons for refreshing or stopping the loading of current documents
- Home button that takes you to your home page

Strangely enough, the browser's user interface is *not specified* in any formal specification, it just comes from good practices shaped over years of experience and by browsers imitating each other. The HTML5 specification does not define UI elements a browser must have, but lists some common elements. Among those are the address bar, status bar and tool bar. There are, of course, features unique to a specific browser like Firefox's downloads manager.