Spezialthema Internet

Lukas Prokop

June 20, 2009

"To lead the World Wide Web to its full potential by developing protocols and guidelines that ensure long-term growth for the Web" (The W3C-Mission)
"The Internet? We are not interested in that" (Bill Gates 1993)
"I just had to take the hypertext idea and connect it to the TCP and DNS ideas and – ta-da! – the World Wide Web" (Tim Berners-Lee)
"The Web is a tool for communicating" (Tim Berners-Lee)

"Working for Google" http://xkcd.com/192/

Contents

1. ARPANET

After the occurence of modern computers in the 60s, computer scientists tried to connect computers together. This could be used for wide-range-communication (eg. between America and Europe) and military proposes (communication with soldier bases). The research led to the development of several packet-switched networking solutions like the ARPANET ("Advanced Research Projects Agency Network"). Packet-switching is the idea to group data in small packets for transmission. The main advantage: if one of the packets gets lost during the transmission, just a small packet has to be resent. It's the task of the sender's computer to divide those data into packets. With packet-switching it was possible to merge separate networks to a super-network without technical troubles. Other ideas and specifications followed and once there were different huge computer clusters[1] in universities, the Internet was born.

---

[1] mass of connected computers

In the movie "The Net", one of the leading researcher in the ARPANET-project – Robert Taylor (NASA-Engineer at Pentagon) – was asked about the project.

> "The ARPA was founded in 1957, maybe 1959, as a result of Sputnik. The Sputnik occured in october 1957 and it greatly surprised the United States; they had no idea. In 1958 president Eisenhower asked the Department of Defense to set up a special agency called ARPA, the "Advanced Research Projects Agency" to look for research projects that had a longer term expectation. So in the hope, that we will not get surprised again, like the Russians surprised us. The initial ARPA-programs were all space-related, not computer. [...]

> ...the ARPA has developed the ARPANET for a case of nuclear disaster. No! That's not the reason, why we developed the ARPANET. We developed the ARPANET to enable people in different places, who had common interest, to share those interests. The internet is simply an evolution of the ARPANET. Both in philosophy *and* technology. [...] That was a lot of fun!"

X.25 and UUCP were the names of other projects which contributed into the growth of the internet. But the ARPANET became the technical core of what we know as the Internet today.

2. The Internet – Definition and Overview

The Internet is basically a global system of **inter**connected (computer) **net**works and technically it's the implementation and usage of a collection of protocols and specifications. Different ideas from the past came together and a communication system was built.

First of all, the ARPANET was only a theoretical thing, used by universities on big clusters. But the aim of connecting computers was to serve different services. Different protocols enabled that. The first well-known service was the email. SMTP and POP are the responsible protocols for email transmission; for webpages it's the HTTP. But before talking about protocols, let us go into more detail. What is communication? How does it work?

To explain parts of the internet, I will use the terms clients and servers. Clients are people or their computer (like your one at home) and servers provide different services. Some servers will tell you the current datetime and some servers will give tell the ISBN of the book "1984" by George Orwell. There are DNS-, SMTP and POP-servers (and loads of others). But the best known of all servers is the webserver, which offers webpages. Protocols define the language, they use while communicating.

3. DNS – Domain Name Service

People are bad at remembering numbers. But computers are identified by an IP address (specified in the **I**nternet **P**rotocol), which consists of 4 numbers between 0 and 255; separated by dots. For people it's much more easy to remember youtube.com instead of 213.33.99.70. The DNS protocol enables to make an association between domains (youtube.com) and IP addresses (213.33.99.70).

So if you want to connect with youtube.com, you have to ask a DNS server for its IP address.

4. URI – Uniform Resource Identifier

I claim: emailaddresses and webaddresses are the same. You might know emailadresses like admin@example.org and webaddresses like www.youtube.com. Technically, I am right, because those addresses are specified in the URI and both of them use the URL (Uniform Resource Locator), which is a subclass of URI. The URL gives us the possibility to find a resource through an exact specified address. This looks like that (the terms in angled brackets are optional):

$$\text{protocol://[user@][subdomain.][domain].tld/[resource][?variables][\#anchor]}$$

If you use want to use an address youtube.com, you leave out the user, the subdomain, the variables, the anchor and the resource. The protocol is obvious (http) and you want the main resource (homepage of a website). At the same time you can use an emailaddress like admin@example.org. There you leave out the subdomain, the variables, the anchor and the resource. The protocol (smtp) is obvious again.

With the URI/URL specification it's possible to address resources.

5. Email – SMTP and POP

The email was the first very important service of the internet. People liked the way of sending messages transatlantically in a very short time. The email is an available service since the ARPANET. Different RFC[2]-standards define the way of communication between servers. One important adaption was the invention of attachments (files, that are attached to an email).

One email at least contains the address of the author, the destination address and the content. SMTP (Simple Mail Transfer Protocol) is responsible for sending messages to

---

[2]Request for Comments – an organisation collecting memorandums on Internet systems and standards

another mailserver and POP (Post Office Protocol) is responsible for offering the received mails to someone.

6. HTTP – Hypertext Transfer Protocol

Alright. . . Email seems to be very nice, but much more important nowadays is the Web (World Wide Web). The history of the Web is very interesting and can be traced back to CERN (Organisation Européenne pour la Recherche Nucléaire) researchers Tim Berners-Lee and Robert Cailliau. HTTP stands for **H**yper**T**ext **T**ransfer **P**rotocol and Hypertext is simple text with references to other hypertexts. Those references were the most important innovation of the Web and are also called hyperreferences, hyperlinks or simply links. Tim Berners-Lee put all those ideas into practice and developed the first webserver "CERN httpd" on a NeXT Computer (sold by Steve Jobs' company) and the first webbrowser "WorldWideWeb". All the stuff is based on the problem of CERN: the CERN building is separated by the borders of Switzerland and French and CERN was searching for a way to transfer documents between those buildings easily.

> "I just had to take the hypertext idea and connect it
> to the TCP and DNS ideas and – ta-da! – the World Wide Web."
> – Tim Berners Lee

On 6th of August 1991[3] he wrote a short description of the WWW project on the alt.hypertext newsgroup and this date marks the debut of the Web as a publicly available service on the Internet. The Web became the way to share things for free. It was not part of the Web to pay fees for participating (basically pre-projects of WWW were commercial). The idea of freedom is related to the revolutionary time of hippies and OpenSource[4]. After creating the web, Tim Berners-Lee founded the World Wide Web Consortium (W3C) which publishes web standards to "lead the Web to its full potential"[5].

7. HTML – Hypertext Markup Language

In difference to the other sections, the title is not a protocol. We have talked about connecting computers and looked into some protocols. When we talk about HTML, we talk about the (hyper)text served by webservers and rendered by web browsers[6]. If you want to write websites, you have to learn HTML.

---

[3]Remarkable date: on 21th (15 days later) of August 1991 Linus Torvalds announced his operating system Linux.

[4]OpenSource is sharing your programming code for free to other people

[5]"The World W3C develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential" is the slogan of W3C

[6]programs, which transform hypertexts to graphical websites. Eg. Internet Explorer or Mozilla Firefox

HTML consists of so-called tag elements. The current version of HTML is 4.01 and version 5 is highly expected. An example for a tag is <p>. If you surround text with the <p>tag element</p> this text is identified as paragraph (p is an abbreviation of "paragraph"). If you add an a element, the element is identified as a hyperlink. Basically the syntax of HTML looks like this (again, the terms in angled brackets are optional):

<tagname[property="value"][property="value"]>[content]</tagname>

In the example above, *a* is the tagname, *href* is the property and *http://google.com/* is the value. The syntax of HTML is based on SGML, but nowadays it's a mixture of SGML and XML.

The idea of tags and resources is quite interesting. For example Google thought "the more people refer to one website, the more important it seems to be". If everybody publishes a reference to the homepage of the American government, it seems to be more important than a small website with bomb construction manuals. Google made some mathematical research and created a formula[7] which rates each website based on the occurences of hyperlinks to that website; the rating is called PageRank. PageRanks became the core "engine" of Google search engine. If someone is looking for the term "Webdesign" a list of websites is served to him with the most linked one at the top. Nowadays the criteria for rankings have changed (it's more about keywords than links), but the webmaster continue to spread links to their website.

Tim Berners-Lee himself dreams of the "Semantic Web". I mentioned tags. If you write down Text (em stands for emphasize), you add the meta-information "emphasize" to the text "Text". By analysing such meta-information you can create a Web, where everything is connected to other stuff and relations show "more truth" than simple text. Google has some nice features. If you type "define:algorithm" into the search engine, it outputs a definition of "algorithm". This definition (or there are several definition) is the result of analysis of website contents. If the same definition is repeated more often on the Web it seems to be more reliable to Google than the definition of Wikipedia.

People often call the Semantic Web "Web 3.0". Another well-known term is "Web 2.0". Web 2.0 is the collection term for all the social and technical developments during the years of the World Wide Web. People started to present themselves on the internet, share interests and send messages. New forms of websites are boards (topic discussions), Weblogs ((un)regularly entries like in a diary), microblogging services (small entries with less than 140 characters; answering the question "What do you do currently?"), newsgroup (like boards but more technical and older) and wikis (everybody participates in writing one article together). As a result the people perceive the internet as a much more personal network.

---

[7]the so-called "PageRank algorithm"

8. Webbrowsers and Browserwars

I have talked about HTML. Webbrowsers have to render HTML and display websites. The first one (by Tim Berners-Lee) was "WorldWideWeb" and several followed. In 2001 there were 2 important webbrowsers: Netscape and Internet Explorer. When people started to use the Web in the 90s, it was a struggle for power of webbrowser programmers. Each one wanted to offer more opportunities and features to webmasters[8]. It was the target of W3C to define <p> as a paragraph and keep that standard. But the webbrowser programmers were unhappy with the W3C developments. They didn't want to wait so long to become part of the web standard; they simply implemented it in their browser and offered it to the people. W3C had to less power to define rules. Because there were several browsers, (eg.) one said <p> is a paragraph and the other said it's a header.

As a result webmasters didn't know how to write websites. They have to choose between the browsers and optimize website only for one browser. You probably know small notes on websites like "Optimized for Internet Explorer 5". This war between webbrowsers is called "browserwar" today and ended in 2001. Microsoft (with their webbrowser "Internet Explorer") pre-installed the IE at Windows. Because it was installed anyway and you would have to install Netscape manually more and more people started to use Internet Explorer. Netscape fought that pre-installation at court, but has lost the case. Microsoft continued to spread Internet Explorer and Netscape Communications got bankrupt.

After this first browserwar Netscape opensourced their program code and the Mozilla community created a new webbrowser (out of Netscape's code) "Mozilla Firefox". Nowadays Internet Explorer and Mozilla Firefox are the dominant browsers at the browser marktplace. They realized the impossibility to create a good Web for webmasters without webstandards. So they started to listen to W3C again and nowadays W3C is the most important developer of the World Wide Web.

9. The basics of websites

A webpage is a simple text file written in HTML. But HTML only defines the structure of websites (what the title of the page is, what a header is ... and so on). The real design (colors, margins, paddings, alignment, adjustment, decoration, borders, ...) is defined in CSS ("**C**ascading **S**tyle**s**heet"). The decision to separate the content/structure and style of a website was a made after the firstbrowser war and is part of the XML philosophy. It was always important to make the Web accessible for all the people around the world. Some do not have the capability to use keyboards or can't see websites (then websites are read out with "screenreaders" for them) and some might use unusual devices

---

[8]Creators of websites

with unusual displays (a webpage should not have a fixed width). This is the issue of accessibility.

So HTML is for structuring and CSS is for designing a webpage. You know popups and "jumping" elements on websites: Javascript is responsible for that and makes websites more dynamic. In HTML there are tags to include CSS and Javascript files.

Another technology is PHP ("PHP Hypertext Preprocessing"). HTML files are sent out by the webserver at request. But sometimes (it gets more and more popular) it is more comfortable to generate websitesi before sending them out. For example you want to use the same menu several times at different HTML files. In such situation you can write PHP code into HTML files and PHP will include the menu from a global file. The webserver will execute that PHP code before sending out HTML.

PHP and Javascript are programming (scripting) languages. HTML & CSS are markup and stylesheet languages.

So...If you want to create your own website, you need a domain (like google.com, see DNS), you need a webserver (you can rent one googling for the term "hosting provider") and the webpage itself. The webpage consists at least of HTML. If you want to use a more beautiful design, you need CSS. If you want to make your webpage more dynamic, you need Javascript and if you want to generate the content of your webpage on the webserver, you need PHP. You can learn all those languages on the internet in tutorials.

10. Problems and Law

Because it's a way of communication, the governments all over the world insist that the Web is not an unlegislated place. Pornography, copyright violations and censorship are the main topics of discussions. It was the idea of Tim Berners-Lee to make the internet extensible and free, but this does not comply with the legislatives of some countries. The "Great Wall of China" is a huge firewall and censor all websites, which criticizes the Chinese government. A current topic is the German Internet blockade. The German minister of family wants to censor all child pornographic websites. Opponents argue, that this blockade can be used to censor any kind of website (like in China). Several website of opponents of that blockade have already been added to the blacklist. In the opponents' opinion it's more targeted to delete such websites instead of blocking them. Also the copyright issues make troubles on the Web. With the Web it became more and more easy to download texts, images or even movies.

11. This document – Copyleft 2009

I know, that this document is full of (too) short technical descriptions and a lot of important things are not mentioned. But I think, that it's a nice 5-page introduction.

12. Questions

- Discuss the following questions:
    - Can the Internet die?
    - How is responsible for the development of the Web?
    - What roles have Google & Co. to play in future?
- How does communication work? What is required?
- What is the Web related to the Internet?
- The history of the Web
- The history of the Internet
- What is accessibility?
- What is Web 2.0?
- Explain HTML and describe how to build your own website

References

[1] http://en.wikipedia.org/wiki/History_of_the_internet

[2] http://en.wikipedia.org/wiki/Portal:Internet

[3] Movie "The Net" by Lutz Dammbeck