Opdenplatz, Juliette
Haselwanter, Stefan

Web Science

doc.md
rev. 1 / 19.03.2018
Seite 1 von 2

1 Assignment 1

1.1 Evolution of the Web

1.1.1 Web 1.0

The first stage of the Web era *Web 1.0* was mainly made up of web pages which were connected by hyperlinks. It mainly consisted of a set of static websites. Users were limited to consuming the content that was presented. There was little to no interactive component or way to contribute content. This is why experts sometimes refer to *Web 1.0* as "Read-Only" web. The terms "Read-Only" and the corresponding "Read-Write" and "Read-Write-Execute" were by the way coined by *Tim Berners-Lee* also known as the inventor of the world wide web. However even tho most websites were static there already was some kind of interaction on some websites. Those were mostly e-commerce websites whose shopping cart applications are basically falling under the category of *Web 1.0*.

With the development of *Web 1.0* the information age was born. During the *dot-com boom* the amount of websites increased rapidly which lead to a period of excessive speculation and thus to a historic economic bubble.

1.1.2 Web 2.0

The *Web 2.0* era erased the lack of active interaction of common users. When exactly the transition from *Web 1.0* to *Web 2.0* cannot be determined. The change happened more gradually. With the user's interactions and contributions becoming a new component, the web could now be called "Read-Write" web. Due to the now emerging *social-media* it's also known as the *social-web*.

A lot of new concepts like blogs, social-media and media-sharing were introduced and made the once so one-sided relationship of publisher to consumer become more balanced. Even users without any technical background can now contribute.

1.1.3 Web 3.0

Since the web mainly consists of information written in human readable languages, they are most likely not easy to interpret for machines. *Semantic Web* or *Web 3.0* tries to fill this communication gap. It makes data accessible to machines by using *semantic web technologies* such as RDF, OWL, SKOS or SPARQL. Those techniques provide an environment where applications can query some data, draw inferences using vocabularies and so on. This version of the web can be called "Read-Write-Execute" web since applications are now able understand and work with the data.

Semantic markup, which is used to enrich some data with meta data for context, empowers the meaning of web services as we know them. Web services are software systems designed to support computer-to-computer interaction over the internet. However, in the context of *Web 3.0*, they become more important.

Example: Wolfram Alpha is a computational knowledge engine which uses its own internal knowledge base with its own internal semantics and ontology. You can for example enter "3 pretzels vs 1 bread" and you will get comparisons of ingredients and calorie contents.

It does so by taking a large amount of raw information and performing computations using those data. It produces pages of new information that have never existed on the internet. Google however provides you with links to pre-exisiting data. In a nutshell Google provides links to websites which *might* provide answers while Wofram Alpha provides answers. It does so by generating data, for example statistics by processing data provided by different databases which contain structured data. The data has to be structured to enable the knowledge engine to process it.

Opdenplatz, Juliette
Haselwanter, Stefan

Web Science

doc.md
rev. 1 / 19.03.2018
Seite 2 von 2

1.2 Computer Science of the 21st Century

While computer science is all about how computers work, web science is a completely different multi-disciplinary field, in which researchers try to understand what the web is, how it affects us and how we can prepare its future. These are questions which cannot be answered by computer science alone. Web science research will make us design and build more complex human-oriented systems.

Computer science also is *synthetic*. It is made up by humans and all about designing formalisms and algorithms to solve specific problems, by using mathematical models which are themselves made up by humans. To the contrary natural sciences such as chemistry, biology and physics are *analytic* disciplines that aim to find laws that explain observed phenomena. Since Web science is such an interdisciplinary field it follows that there at least have to be some scientific methods that are a combination of both of these paradigms.

1.3 Data Science Web and Social Considerations

Any decision process (human or alhorithm) that decides what to include, exclude, or emphasize in information processing has the potential to include biases. Algorithms can not understand concepts like racism but for example if some measurable criteria they use to exclude some sort of information accidently correlates with race divides, they might appear to have a racial bias.

Example: Google News was one of the first platforms to attempt to aggregate and personalize news with algorithms. Its algorithm ranks and groups news articles by frequency of appearance, source, freshness, location, relevance and diversity. Unless explicitly programmed to do so an algorithm will never favorite one thing over another just because it likes the one thing more than the other. However Google News will show you news according to its criteria. So for example might tell you more about specific political parties if one presidential candidate appears more often in the media (frequence of appearance). So the site might seem biased even though the algorithm isn't, at least not intentionally. Think about the two party system in the USA, there's actually more parties which you hardly ever hear about, unless you're specifically searching for them.

To quote Nick Diakopoulos:

> It can be easy to succumb to the fallacy that, because computer algorithms are systematic, they must somehow be more "objective". But it is in fact such systematic biases that are the most insidious since they often go unnoticed and unquestioned. Even robots have biases.

There are also ways that biases in information that is to be processed, can influence the decision making of algorithms in a way that the algorithm outputs biased information as well. Thus biases in algorithms can occur when an algorithm uses data that is already biased. Biases in news are a good example since the content is created by humans. Those contents are almost always biased. *The Propaganda Model* by Noam Chomsky and Edward S. Herman explains how systematic biases work in mass media. However we're not going to get too political here, but it's definitely worth a read.