☰ Sections

Option C - Web Science (/topic/C-web-science.html)

C.6 [HL] The intelligent web

How the web can be used for gathering collective intelligence, e.g. through the semantic web.

C.6.1 Define the term semantic web

The semantic web, also referred to as the web of data, web 3.0 or linked data web, is an extension of the World Wide Web "in which information is given well-defined meaning, better enabling computers and people to work in cooperation"[1]. It helps computers to read, understand and analyze the information in documents, e.g. through tags and document structure.[2,3]

Summary:

- Extension of WWW
- Information is given meaning
- Meaning of information is readable for computers
- Computers can perform more complex tasks based on meaning of information

C.6.2 Distinguish between the text-web and the multimedia-web

Text web is text-based information, whereas multimedia web consists of the utilization of a combination of different media, mostly motion video, images, sound or text. Examples of tools used for multimedia include "Adobe Photoshop" for content creation, or "Adobe Flash" for multimedia content display.

Summary:

- Multimedia web -> uses mixture of media to provide information, e.g. motion video, sound or text
- Text web consists of plain text information

C.6.3 Describe the aims of the semantic web

The problem with the current web is that information is often difficult to process for machines, because of ambiguity and different data formats. The semantic web aims at expressing information in a way that is unambiguous and easy to process by machines.

This way, data should be able to be linked easily across different websites and enterprises, so that the web can become more collaborative. This way the web could be considered a big database, where autonomous agents have extended possibilities.

To understand this better, it can help to look at how a database usually works. A database stores data, which can be added, modified or deleted by users. However, in contrast to normal documents, data are usually stored in a uniform format. The computer can know this format and can therefore perform complex tasks based on these data as it knows the meaning of the data through the given format. The idea of the semantic web is to create something similar on a larger scale, with data from the internet.

The implementation of such a semantic web requires standards among different websites. This includes vocabulary for giving data meaning, but also data formats. One widely used model for formats, specified by the World Wide Web Consortium (W3C), is RDF, short for Resource Description Framework.

RDF (Resource Description Framework)

RDF (https://www.w3.org/RDF/) is a data model, meaning that it is not a format in itself. It is based on the idea of giving information meaning through "triples". Such triples make a statement about resources, following a structure of *subject-predicate-object* . For example in the statement "Water is a liquid", the subject is 'water', the predicate describes that 'it is something' and the object is 'liquid'. This way, the predicate describes the relationship between the subject and object.

The RDF model applies this idea to web resources, by allowing triple components to be URIs. Wikipedia illustrates this with the example statement "The title of this resource is 'Tony Benn'" about the Wikipedia article for Tony Benn. This can be represented in RDF as a triple:

```
<http://en.wikipedia.org/wiki/Tony_Benn> <http://purl.org/dc/elements/
1.1/title> "Tony Benn" .
```

In this example, each triple component is separated by <> . The first component gives the resource URI (subject); the second component gives an URI to a definition of 'title' (predicate); and the last component states the actual title of the resource (object). Note that RDF is not a format, so that such a triple can be formatted in different ways. [4]

Summary:

- linking and sharing data should be easy
- idea of world wide database
- giving data meaning through uniform formats
- RDF (Resource Descripton Framework) is a data model NOT a format
  - ideas of triples: *subject-predicate-object*
  - data should be linked by associating information with URIs

C.6.4 Distinguish between ontology and folksonomy

Ontology

In the previous section it was mentioned that the semantic web comes with the need for a common vocabulary for giving data its meaning. One reason for this is that different groups of people can perceive the same piece of information in different ways. A computer needs meaning to be explicit, however, in order to be able to work with it. This allows for the following definition of an *ontology*:

> An ontology is an explicit, formal specification of a shared conceptualization. (Thomas R. Gruber, 1993)

Conceptualization here includes the concept of some piece of information itself, but also relations between concepts(e.g hierarchies), attributes and instances. Relationships can describe equivalence or hierarchy.

This said, ontologies are also called "authorative metadata", because they are created centrally by expert knowledge engineers. It creates a formal and structured knowledge organization.

An example use of ontology is DBpedia, which attempts to "extract structured information from Wikipedia" and make this information available for use.

Folksonomy

Folksonomy usually takes the form of tags and is also called "collaborative metadata", because metadata can be added by different users. It provides informal knowledge for classifying and organizing content. Folksonomies help to improve searchability, classification and content visibility and are part of most social networks.

Examples of folksonomy:

- social networks, e.g. Twitter, Instagram, Facebook
- categorization of pictures, e.g. Flickr, Pinterest
- content curation, e.g. ScoopIt, Pocket

Comparison

	Ontology	Folksonomy

Advantages	• structured • high expressive power • explicit meaning • formal specification of knowledge domains • synonyms and other relations can be clearly expressed • created by knowledge engineers	• quick & easy, no expertise needed • collaborative creation • can be used on a large-scale • can include multiple social meanings • importance for the social web
Disadvantages	• laborious • requires expertise • hard to implement on a large-scale • only one perspective on meaning	• no vocabulary control • informal medata • unstructured • low expressive power • ambiguity

Source: Theodosia Togia
(https://www.cl.cam.ac.uk/~aac10/R207/ontology_vs_folksonomy.pdf)

C.6.5 Describe how folksonomies and emergent social structures are changing the web.

- they can improve search results
- they can be used to detect trends, as for example in Twitter
- they can be used to discover new content: similar tags can be identified and be used for content suggestions
- can be used for a more individual experience, as web services learn what tags a user likes
- customized advertising by analyzing user preferences and interests through tags

C.6.6 Explain why there needs to be a balance between expressivity and usability on the semantic web

One of the aims of the semantic web is to give information more expressive power, so that computer systems can understand the meaning of information and its relation to other pieces of information. In theory this should allow for more powerful applications and something similar to a global database. For this reason, expressivity is the guiding factor to how semantic information will be.

However, giving information more expressive power may come at an expense of usability. A language like RDF comes with high expressive power, but requires the user to specify relations to other information and to classify a given piece of information - in this case through triples. This might be feasible and convenient for building shared scientific databases, but for common user interactions, such as in social networks, the process will be counterproductive.

It is therefore essential for the semantic web to be easy to use - ideally the markup being similar to natural language. For non-expert user generated content folksonomies probably offer the best usability as they are as easy as typing in related tags. Folksonomies provide a very low level of expressive power, but they still allow systems to suggest similar content to the user and to identify certain trends. For scientific knowledge databases, more expressive power provides users with better searchability and a potentially improved work flow - usability can usually be sacrificed, because users are experts in their fields anyway.

A common tool for web masters to structure data on their websites is Schema.org (http://schema.org/), which is currently supported by major search engines, such as Google, Bing, Yandex and Yahoo! . This tool provides common vocabulary for on-page markup, which helps search-engines to understand the information on a given page and provide richer search results. For instance it could be used to display contact information of a company website right in the web results or to show thumbnail pictures of an article. As not all content needs to be marked up in Schema.org, this allows web masters to only mark up as much as is usable for them, while providing a high degree of expressive power.

C.6.7 Evaluate methods for searching for information on the web

A search engine in general is defined as an information retrieval system. As such it can be used for navigational search or for research. These differ in that the user usually knows what document he/she is looking for, while research is focused on finding new, previously unknown documents. Apart from this difference, search engines can differ in their scope and target audience. For instance web search engines such as Google, Bing, Yahoo! or DuckDuckGo aim at searching the entire publicly accessible web (World Wide Web). Other search engines may limit their scope. For instance, Google Scholar searches for academic papers only, while Cloud Kite is a search engine for searching cloud documents from Google Drive. The YouTube search bar can be considered a search engine for videos from the same platform only, while the similar concept applies to online shopping websites.

Search engines might use crawlers or if the search is more specific, it might be based on database queries.

C.6.8 Distinguish between ambient intelligence and collective intelligence

Ambient intelligence

Ambient intelligence is closely related to ubiquitous computing (see C.3.) and is based on the idea of a computing being integrated unobtrusively in the environment, providing intelligent services as people need it. Characteristics commonly attributed to ambient intelligence include (Zelkha et al. 1998; Aarts, Harwig & Schuurmans 2001):

- embedded: integration of devices into the environment
- context aware: personal identification and situational needs of the user
- personalized: different responses according to personal interests/preferences
- adaptive: they can change in response to you
- anticipatory: providing information when adequate

Probably the most common application of ambient intelligence would be the concept of a smart home, including automation and smart assistants such as Amazon Echo↗ (https://www.amazon.com/gp/product/B00X4WHP5E/ref=as_li_tl? ie=UTF8&camp=1789&creative=9325&creativeASIN=B00X4WHP5E&linkCode=as2&tag=csibrev-20&linkId=04e76f6da77c317c5a2f265cb5c52773). The Internet of Things provides a potential platform for making ambient intelligence possible.

Collective intelligence

Collective intelligence is the intellectual outcome of a group of people working together, which could range from families to international companies. The internet plays an important role as it can connect people that wouldn't have done so otherwise. The fact that people can share information on their websites, so that others can find it through search engines like Google greatly contributes to collective intelligence. Wikipedia is another example of how the internet can bring people together to create high-quality intellectual content. In other areas of knowledge there exist different tools that improve collaboration in a similar way. More examples include:

- open source software platforms like GitHub (https://github.com/); facilitate the creation of open source software like Linux
- Splice (https://splice.com/), an application that allows musicians to collaborate and share ideas easily
- photo sharing platforms
- cloud application

C.6.9 Discuss how ambient intelligence can be used to support people

Pros

- home care systems for elderly or handicapped people: emergency assistance, autonomy enhancement, comfort
- real-time shopping: improved experience for consumers, who could buy a good when they see it (e.g. some piece of clothes they see someone else wearing)
- personal information provides better means for risk assessment for insurance companies: personalized insurance fees

Cons

- privacy concerns: how many data are collected and who controls them?
- surveillance
- discrimination: is a personalized insurance fair? Particularly concerning genetic conditions for health insurance
- risk of automating too much: if computer systems fail, this may have a very large impact
- reliability, maintainability
- compatibility between systems
- dystopian ideas of *Brave New World* ?

It is also important to consider the devices and technologies that are necessary for ambient intelligence, especially regarding different types of sensors that allow the system to act based on context. This may include cameras for facial recognition, RFID tags and sensors for accounting (e.g. smart fridges), but possibly also nanotechnology or biometrics.

C.6.10 Discuss how collective intelligence can be applied to complex issues

A few examples are given here.

Climate change

Climate CoLab (https://climatecolab.org/) is a project by the MIT Center For Collective Intelligence where people work with each other to "create, analyze, and select detailed proposals for what to do about climate change".

Finance

In one paper (http://www.sciencedirect.com/science/article/pii/S1877042811023895) scientists have analyzed collective trends from Twitter posts in order to try to predict stock market indicators, like Dow Jones, NASDAQ and S&P 500. Emotional outbursts appeared be a good indicator in this early research.

Astronomy

Galaxy Zoo (https://www.galaxyzoo.org/) is a project where people contribute to classify a large number of stars and galaxies.

Reddit Place

Reddits Place (https://www.reddit.com/r/place/) was a project run by Reddit during 3 days, where users of the platform could color a single pixel on a 1,000 by 1,000 pixel canvas every 5 minutes. The result is quite amazing, considering different interests of different users. While this is certainly not a very serious project, nor is it a particularly complex issue, it might be considered a creative case of collective intelligence.

As a more general concept crowdsourcing (https://en.wikipedia.org/wiki/Crowdsourcing) can be considered a model for creating collective intelligence.

Sources:

- [1] W3C Presentation (https://www.w3.org/RDF/Metalog/docs/sw-easy)
- [2] Google Definition of Semantic Web (https://www.google.co.uk/?gws_rd=cr&ei=STaAWMyCJIrT8QWv8pDwDQ#q=semantic+web+definition)
- [3] Cambridge Semantics (http://www.cambridgesemantics.com/semantic-university/introduction-semantic-web)
- [4] Wikipedia RDF (https://en.wikipedia.org/wiki/Resource_Description_Framework)

Home (https://www.cs-ib.net)   About (/about)   Privacy policy (/privacy)