

[Inspiratron.org - Natural language processing, machine learning and cybersecurity](https://inspiratron.org)

Brief introduction to Linked Data

by Nikola Miloševi? - Saturday, March 22, 2014

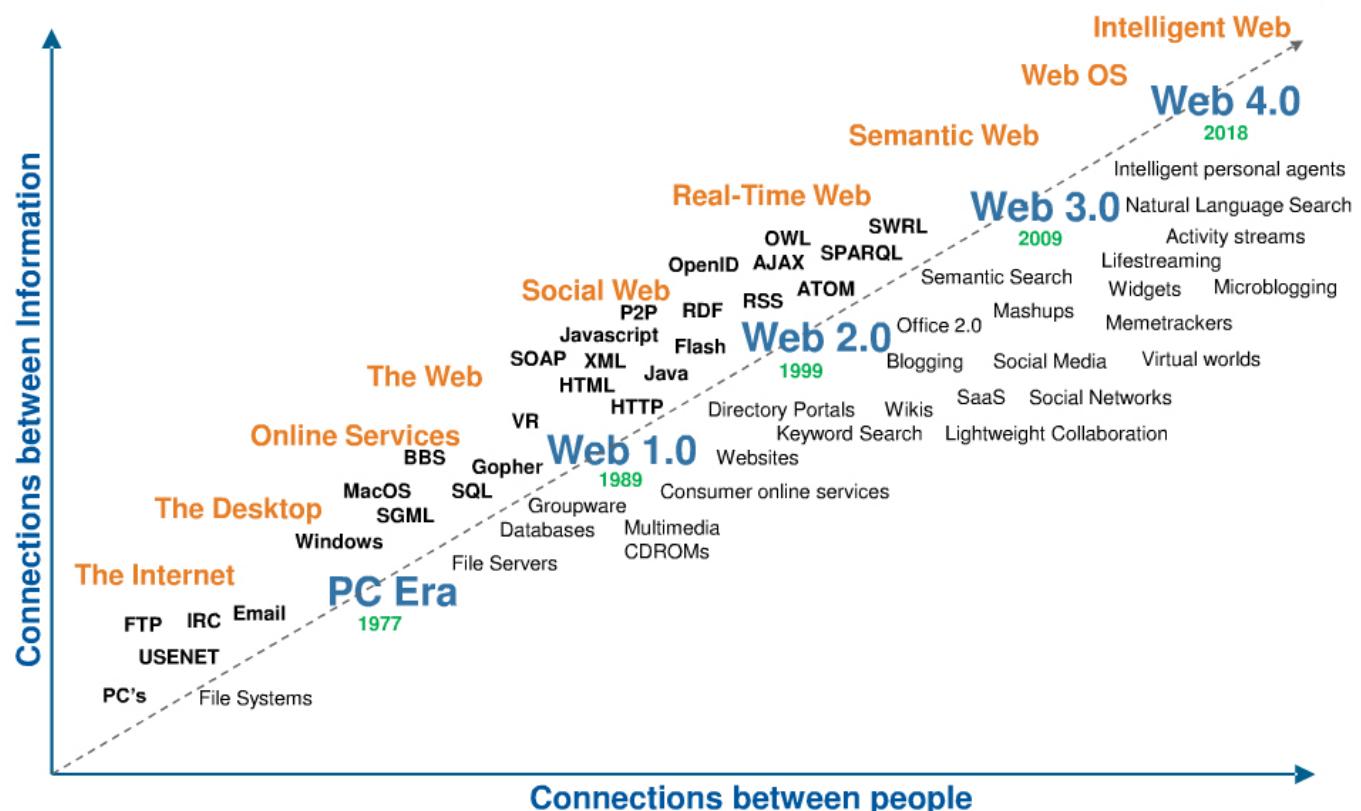
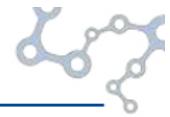
<https://inspiratron.org/blog/2014/03/22/brief-introduction-linked-data/>

My recent research brought me to linked data as quite interesting concept. I will here write some brief introduction and notes on linked data. Probably at some future point of time I will go deeper into standards and usage.

Current state of the Web

Internet have revolutionized the way we communicate and how we access and handle data. It becomes very hard to imagine world without internet, even it exist only 25 years. Many things have changed even on internet over these years. From the first html only pages linked with each other, we moved to Web 2.0, where everyone can contribute to quality and amount of data. However, these concepts are built for humans, which is ok, since humans are primary readers of these data. But there exist emerging need to provide good access to data for the machine processing. What machines need for processing? They need structured data. HTML and current state of the web provides tools for structuring documents, but not data.

The Intelligence is in the Connections



Maybe years are not right, but this is the evolution of the web

In the past several years there is trend in building API based applications. These applications provides structured data, from provider's database. However, even API has its drawbacks. If your application relies on several different data sources, you need to learn how different APIs work. API provide structured data, but they do not provide links to related data. So if we want to browse data, we would need to query API interface again and again. And most of the data information in APIs are bound to some IDs that makes sense only for the particular data provider.

Motivation for linked data

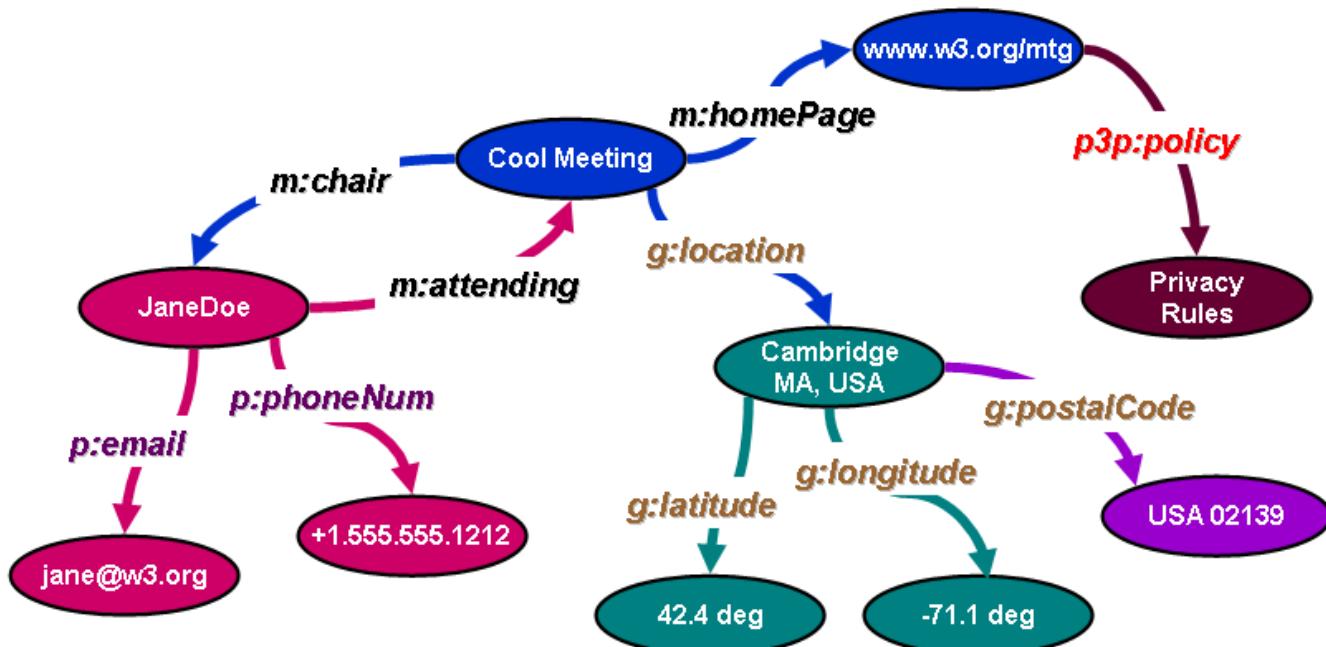
Limitation of current state of the web is basic motivation for the linked data concept on the web. Basically research is done to enable semantic linking of concepts and data on the web. By doing this, data will be structured and semantically linked and meaningful. This data would be easier to process and make quite advanced reasoning. As we said before, by using API based approach, that can return structured data in XML, JSON or some other format, is step towards structured and semantic data. But, APIs can only provide data from one provider and data is not linked with related data, but related data has to be queried again.

Lets imagine utopia or ideal world for data sharing. In ideal world we would be able to access data from as many data providers as we need by using same interface. Data would be structured as possible, and for each piece of data, we can follow links to further explanation, specifications or generalizations. Data would have between semantically defined relationships, which will give all these data meaning for reasoning and processing. So in this world we would know that Michael Jackson is a person, who was singer, and we would be able to follow links to his albums, songs, concert dates or movies and video clips on the web. This is the rationale behind Linked Data.

How Linked Data work?

Main concept behind Linked Data is to represent data as triples of subject, predicate and object. Subject and object can be two entities, while the predicate defines relationship between them. Object of one triple can be subject of the other triple. So this is how web of data is built. This kind of approach will describe the real world and will overcome all the issues current web have with its semantics. However, to do this correctly, there has to be standardized way of representing things and making it available. Resource Description Framework (RDF) is a standardized way to represent this kind of data introduced by W3C in 2004. There are still left a lot of freedom in RDF how to represent data,

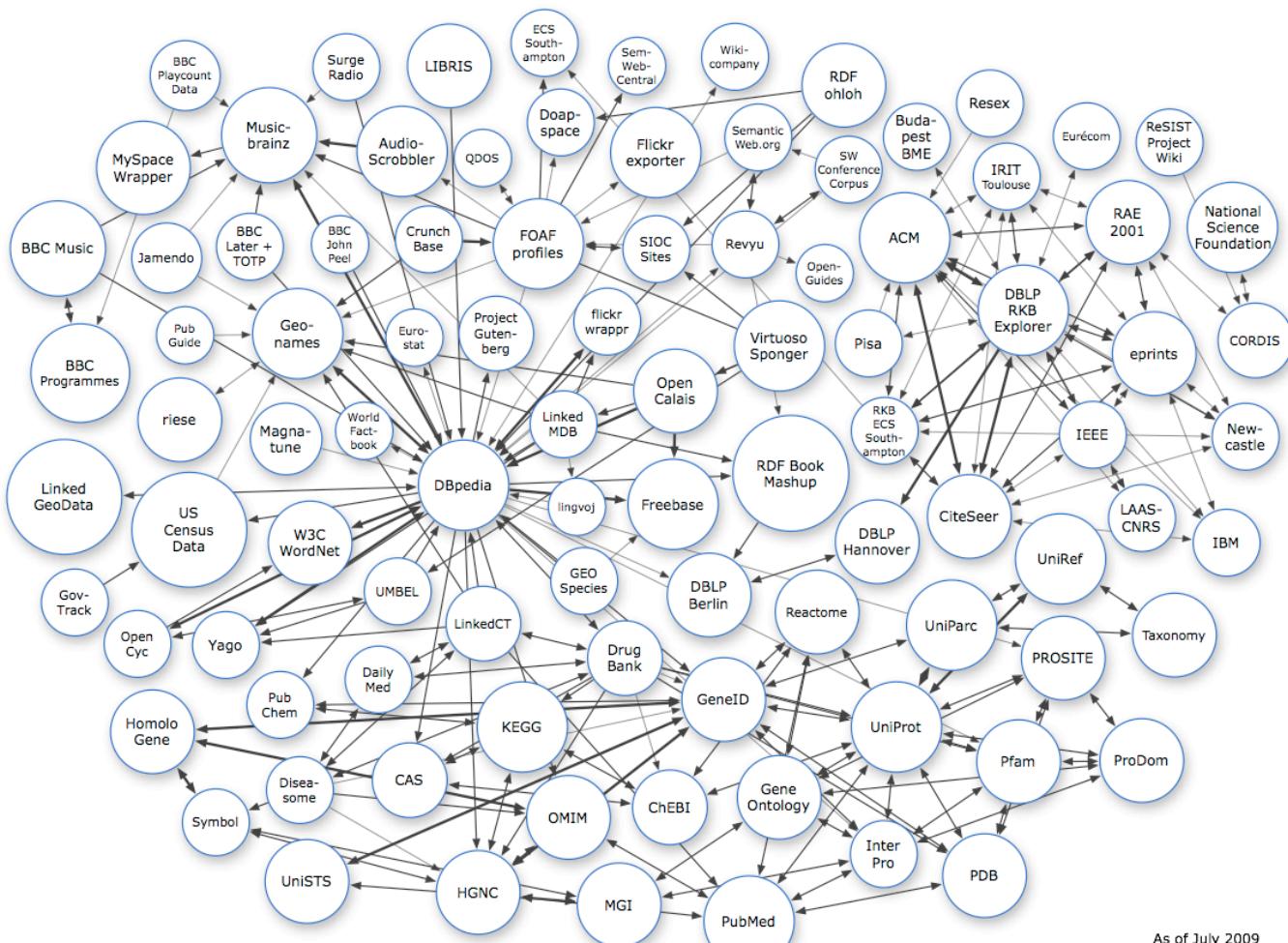
and this is the reason of possible redundancy between different data providers. By standard, entities are represented as Unique Resource Identifiers (URI). One question remains is how to create URI. Each data provider can have its own standard for doing this. Because of this, it becomes sometimes hard to cope with different data sources. We can however define that some entity is same as the other entity in some well known data source. This will need just one more triple of described data. The second problem in Linked Data is how unambiguously describe relationships. Currently, users are free to even make their own ontology and name entity relationship how they prefer. This would not be certainly a good idea, so W3C recommended Web Ontology Language (OWL) as a mean for describing entity relationships. The data described by an ontology in the OWL family is interpreted as a set of "entities" and a set of "property assertions" which relate these entities to each other. An ontology consists of a set of axioms which place constraints on sets of entities (called "classes") and the types of relationships permitted between them. These axioms provide semantics by allowing systems to infer additional information based on the data explicitly provided. Using OWL we may say that Michal Jackson is a person, by using `is_a` property between entity Person and Michael Jackson.



Thing that should be mentioned is that there is a query language called SPARQL. SPARQL is an RDF query language, that is, a query language for data sources, able to retrieve and manipulate data stored in Resource Description Framework format. It can be viewed as SQL for data on the Semantic Web data sources. However, with SPARQL you can query all available RDF data sources, not only your own database.

Challenges in Linked Data

As we mentioned before, there is a challenge on defining URIs, because not all data providers would agree on using the same way of creating URIs. This makes overhead of linking data that is same. Other challenge is the use of semantics. There is in place Ontology, but people would still need to learn how to do it right way. This can be quite challenging, since people are quite comfortable using things they know. This is quite large ontology, which needs quite some attention to be used right. One more issue is privacy and security. Web of data is created for sharing data on mind. However, some data should remain private, secured from the view of unauthorized individuals. This can be done, and some portion of research is done on this field, but it remains one of the challenges.



As of July 2009

Next time...

Next time I decide to write about linked data, I will probably go deeper in explaining RDF format and some parts of OWL ontology language. Also I will probably have a look at some applications of Linked Data. It is interesting to mention that many media houses are using Linked Data, such as BBC and The New York Times. Also, EU government is trying to use this type of data as much as possible. There are also many interesting biomedical applications. Interesting source to mention is DBpedia, which is Linked Data Wikipedia.

All rights reserved and copyrighted by inspiratron.org and Nikola Milosevic