

Engenharia de Dados e Conhecimento

2019/2020

Dados Semânticos:
Fontes e Publicação

Fontes de Dados



- A ontologia FOAF é usada para representar informação acerca de pessoas, como:
 - os seus nomes, datas de nascimento, fotografias, blogs e especialmente outras pessoas suas conhecidas.
- Os grafos FOAF são particularmente bons para representar dados que aparecem nas redes sociais
 - Várias redes sociais permitem o acesso a informação FOAF dos seus utilizadores

FOAF



- Especificação

- FOAF Vocabulary Specification 0.99
- Namespace Document 14 January 2014 - Paddington Edition
- Link: <http://xmlns.com/foaf/spec/>

- Exemplo de dados FOAF

- Link:
<https://github.com/RDFLib/rdflib/blob/master/examples/foaf.rdf>

FOAF



- Importante

- tal como a web tradicional é construída a partir da ligação entre múltiplos documentos
- também a web semântica é construída a partir da ligação entre múltiplos dados
- a diferença principal reside na “entendimento” muito maior das máquinas sobre a informação processada

- Assim

- quando uma máquina analisa um documento contendo dados FOAF, não só consegue informação acerca da pessoa descrita nesse documento, como consegue informação acerca das pessoas ligadas a essa pessoa
- isto é possível porque a máquina pode seguir as URLs (URIs) que identificam as pessoas ligadas

Wikidata



- A Wikidata é um projeto da Fundação Wikimedia, que tem por objetivo reunir dados para suporte da Wikipedia, Wikimedia e outros.
 - Url – <http://www.wikidata.org>
 - Livre – os dados são publicados sob a licença Creative Commons (CCO 1.0) e podem ser usados livremente, mesmo para efeitos comerciais.
 - Colaborativo – os dados são inseridos e mantidos por editores da wikidata.
 - Multilingue – a navegação, edição e consumo dos dados é automaticamente multilingue.
 - Segundada por Bases de Dados – para além dos dados, encontra-se também informação sobre as suas fontes o que aumenta a possibilidade da sua verificação.
 - Dados Estruturados – ao contrário da Wikimedia e da Wikipedia, a Wikidata reúne dados estruturados de uma certa forma.

Wikidata



- Acesso

- para aceder à Wikidata, de forma programática, é possível utilizar vários meios:
- Wikibase API – <https://www.mediawiki.org/wiki/Wikibase/API>
- SPARQL endpoints – <https://query.wikidata.org/>
- Wikidata Bots – <https://www.wikidata.org/wiki/Wikidata:Bots>

Publicação



- Embutindo semântica
 - Uma das maiores críticas feitas à web semântica é a de que os formatos e standards, como o RDF/XML, por exemplo, são demasiado complicados para serem utilizados pelos webmasters
 - Muita informação já existe em HTML e fazer a sua duplicação é um grande investimento
 - Os micro-formatos e o RDFa aparecem para ajudar a resolver este problema

Publicação



- Micro-formatos

- São um modo simples de adicionar dados semânticos às páginas web, utilizando o familiar atributo class do HTML
- Existe uma variedade de microformatos que podem ser consultados em <http://www.microformats.org>
- Exemplos dos mais conhecidos: vCard e hCalendar
- Exemplo de utilização:

```
<div class="vcard">  
  <div class="fn">Toby Segaran</div>  
  <div class="org">The Semantic Programmers</div>  
  <div class="tel">919-555-1234</div>  
  <a class="url"  
    href="http://kiwitobes.com/">http://kiwitobes.com/</a>  
</div>
```


Publicação



- A publicação de dados semânticos, nunca deve ser feita com recurso a grandes documentos
- Esta deve ser feita:
 - Embutindo fragmentos em documentos HTML e/ou XML já existentes
 - Em documentos RDF, relativamente pequenos
 - Ou usar *triple stores*, com boa performance
- Isto porque o consumo de dados semânticos, é feito de forma incremental e por isso são requeridos apenas pequenos conjuntos de dados de cada vez