

Web a sémantický web (co je sémantický web)



ISKM89 Organizace dat - sémantický web | podzim 2023
Zuzana Nevěřilová | Centrum zpracování přirozeného jazyka

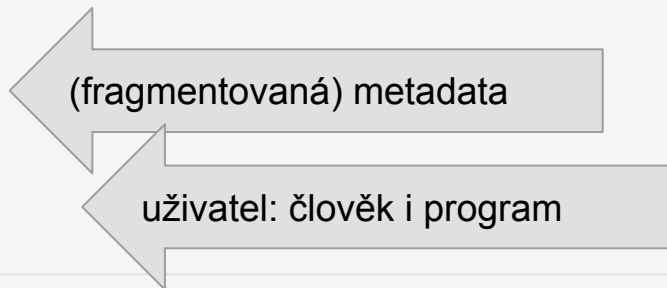
Sémantický web (Semantic Web, Web 3.0)

The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation.

Tim Berners-Lee et al. 2001

https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf

Web 1.0 spojoval **webové stránky**,
Web 2.0 spojoval **aplikace**,
Web 3.0 by měl spojovat **data a znalosti**.



Sémantický web a data

For the semantic web to function, computers must have access to structured collections of information and sets of inference rules that they can use to conduct automated reasoning.

https://www-sop.inria.fr/acacia/cours/essi2006/Scientific%20American_%20Feature%20Article_%20The%20Semantic%20Web_%20May%202001.pdf

Dva hlavní typy dat:

- Sémantická metadata (Semantic Metadata)
- Otevřená provázaná data (Linked Open Data)

Sémantická metadata (Semantic Metadata)

Idea: označkovat koncepty na stávajících webových stránkách

Např. adresa má komponenty: ulice, číslo domu, poštovní kód, jméno města, země atd.

Sémantická metadata umožňují více porozumění při vyhledávání (Paris vs. Paris Hilton).

mikrodata a mikroformáty

Sémantická metadata - příklad mikrodat v HTML

```
<div itemscope itemtype="https://schema.org/SoftwareApplication">
  <span itemprop="name">Angry Birds</span> - REQUIRES ANDROID
  <span itemprop="operatingSystem">ANDROID</span>
  <link itemprop="applicationCategory"
    href="https://schema.org/GameApplication" />
  <div itemprop="aggregateRating" itemscope
    itemtype="https://schema.org/AggregateRating">
    RATING:
    <span itemprop="ratingValue">4.6</span>
    (<span itemprop="ratingCount">8864</span> ratings)
  </div>
  <div itemprop="offers" itemscope itemtype="https://schema.org/Offer">
    Price: $<span itemprop="price">1.00</span>
    <meta itemprop="priceCurrency" content="USD" />
  </div>
</div>
```

Angry Birds - REQUIRES ANDROID
RATING: 4.6 (8864 ratings)
Price: \$1.00

Sémantická metadata (Semantic Metadata)

Používané (a vyhledávané třídy): schema.org (Google+Yahoo+Microsoft+Yandex):

Person, LocalBusiness, GeoCoordinates, Review, Place, Organization, JobPosting, CreativeWork...

... we found structured data within 0.9 billion HTML pages out of the 2.5 billion pages contained in the crawl (37.1%)

<http://webdatacommons.org/structureddata/2018-12/stats/stats.html>

Přehled nejpoužívanějších schemat:

http://webdatacommons.org/structureddata/2018-12/stats/schema_org_subsets.html

Sémantická metadata v praxi - Google Knowledge Graph

The image shows a Google search interface. The search bar contains the text "ready player one year published". Below the search bar are navigation tabs for "Obrázky", "Videa", "Zprávy", "Knihy", "Mapy", "Lety", and "Finan...". The search results show a Knowledge Graph for "Ready Player One". The graph includes the date "August 16, 2011" and the text "Ready Player One". To the right of the graph, there is a detailed description of the movie: "Ready Player One: Hra začíná je americký dobrodružný sci-fi film z roku 2018, který natočil Steven Spielberg dle stejnojmenné knihy z roku 2011 od Ernesta Clina. Ten se také podílel na scénáři k filmu. [Wikipedie](#)". Below the description are several key facts: "Datum uvedení: 29. března 2018 (Česko)", "Režisér: [Steven Spielberg](#)", "Adaptace díla: [Ready Player One: Hra začíná](#)", and "Tržby v kinech: 592,2 milionů USD".

Google

ready player one year published

Obrázky Videa Zprávy Knihy Mapy Lety Finan...

Přibližný počet výsledků: 697 000 000 (0,64 s)

August 16, 2011

Ready Player One

Ready Player One: Hra začíná je americký dobrodružný sci-fi film z roku 2018, který natočil Steven Spielberg dle stejnojmenné knihy z roku 2011 od Ernesta Clina. Ten se také podílel na scénáři k filmu. [Wikipedie](#)

Datum uvedení: 29. března 2018 (Česko)

Režisér: [Steven Spielberg](#)

Adaptace díla: [Ready Player One: Hra začíná](#)

Tržby v kinech: 592,2 milionů USD

Otevřená propojená data (Linked Open Data)

Opět Tim Berners-Lee (2006): Čtyři pravidla pro otevřená propojená data

- použij URI pro jména věcí
- použij URI se schématem HTTP, aby lidé věci našli na Internetu
- když někdo požaduje URI, přidej užitečné informace (s využitím standardů RDF a SPARQL)
- přidej odkazy na další URI, ať mohou ostatní objevovat další věci

<https://www.w3.org/DesignIssues/LinkedData.html>

Otevřená propojená data (Linked Open Data) - co jsou dobrá data

- ★ Dostupná na webu s otevřenou licenci
- ★★ Dostupná jako strojově čitelná, strukturovaná data (např. tabulka v Excelu místo skenu tabulky)
- ★★★ Navíc ještě neproprietární formát (např. CSV místo Excelu)
- ★★★★★ Navíc ještě s využitím otevřených standardů W3C (RDF a SPARQL),
kde jsou identifikované věci (aby mohli uživatelé na věci odkazovat).
- ★★★★★ Navíc ještě jsou data provázaná s jinými daty, takže uživatelé dostanou i kontext.

<https://www.w3.org/DesignIssues/LinkedData.html>

Otevřená propojená data (Linked Open Data) - příklady

- faktografické databáze: např. Warsa (<https://www.ldf.fi/dataset/warsa>)
2. sv. válka ve Finsku
- ontologie - schémata/třídy konceptů (osoba, kniha, adresa), např. schema.org
- někdy ontologie obsahují jak popisy tříd, tak fakta (individua/instance tříd)

Ontologie obsahuje:

- třídu a její popis (osoba)
- atributy třídy (datum narození)
- relace k jiným třídám (potomek jiné osoby)

Otevřená propojená data a znalostní grafy

- Tím, že jsou data z různých zdrojů propojitelná, vzniká (obrovský) graf propojených URI
- Podívejte se na něj <https://lod-cloud.net/>

Sémantický web

Od provázaných dokumentů (webu) k provázaným informacím (strukturovaným datům) zase zpět k dokumentům, které mají autoři motivaci obohatit mikrodaty.

Stejné standardy jako má Web: URI, jmenné prostory

Nové standardy oproti webu (zatím jen zmíněny): RDF a SPARQL

Přestože se z celého webu nestal sémantický web ani po 20 letech, část webu sémantizovaná je a výhody sémantického webu užíváme.