

Standardy sémantického webu

RDF(S)



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

Sémantický web kolem nás

Web obsahuje stránky v HTML

mikroformáty a mikrodata: do stránky v HTML přidáme sémantická metadata

Výhody:

- přesnější odpovědi na vyhledávací dotazy
- odkazy na další, sémanticky propojené objekty
- agregace dat (např. vyhledávač zboží, sumarizované recenze)

HTML rendering:

- prohlížeč dostane instrukce, jak stránku zobrazit pro člověka
- mikrodata se nezobrazují, ale slouží pro komunikaci s jinými počítači

Standardy pro web a sémantický web

URI - světově unikátní identifikátor, obsahuje schéma, může obsahovat jmenný prostor

namespace - jmenný prostor pro uvedení do kontextu

HTML - jazyk pro hypertextové dokumenty

XML - zobecnění HTML, jazyk pro popis hierarchií

RDF - jazyk pro popis objektů a relací mezi nimi

SPARQL - dotazovací jazyk pro data v RDF

Resource Description Framework - RDF

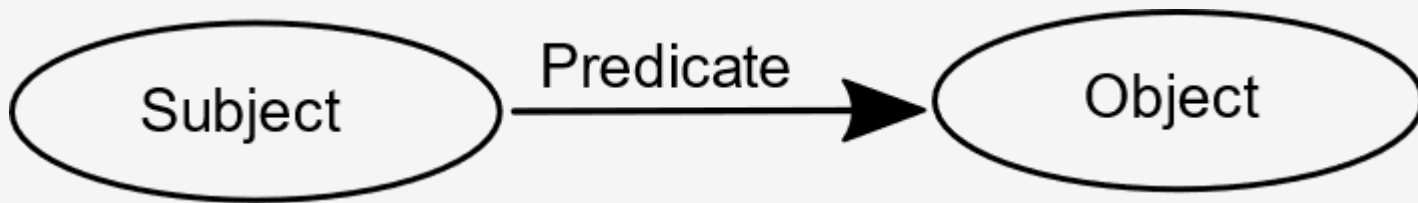
= model pro výměnu dat mezi počítačovými programy

Ideje:

- nevadí, že jiný návrhář použil jiný model pro stejnou situaci
- nevadí, že potřebujeme časem modelovat data podrobněji
- stávající data není třeba kvůli změnám v návrhu měnit
- každý objekt, který modelujeme, má URI
- relace mezi objekty jsou definované objekty (tudíž mají URI)
- ne všechna data musí být strukturovaná

Resource Description Framework - graf

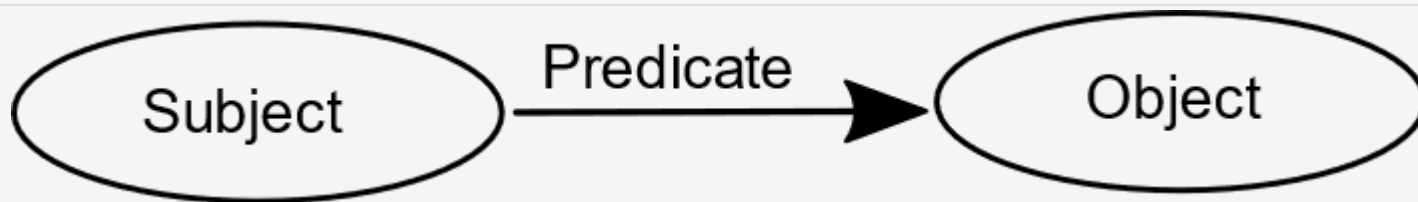
Film je umělecké dílo.



<Movie> <is a> <Creative Work>

<<https://schema.org/Movie>> <<http://www.w3.org/2000/01/rdf-schema#subClassOf>> <<https://schema.org/CreativeWork>>

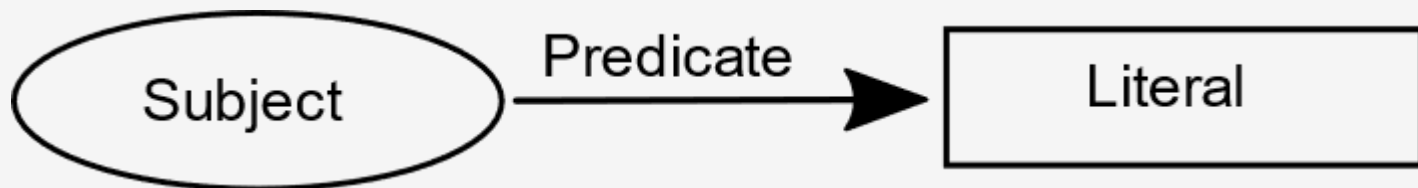
Resource Description Framework - graf



Trojice subjekt-predikát-objekt je tvrzení (statement).

Tvrzení má pravdivostní hodnotu.

Tvrzení je způsob, jak uložit strukturované i nestrukturované informace.



Příklad grafu

Bob is interested in The Mona Lisa .

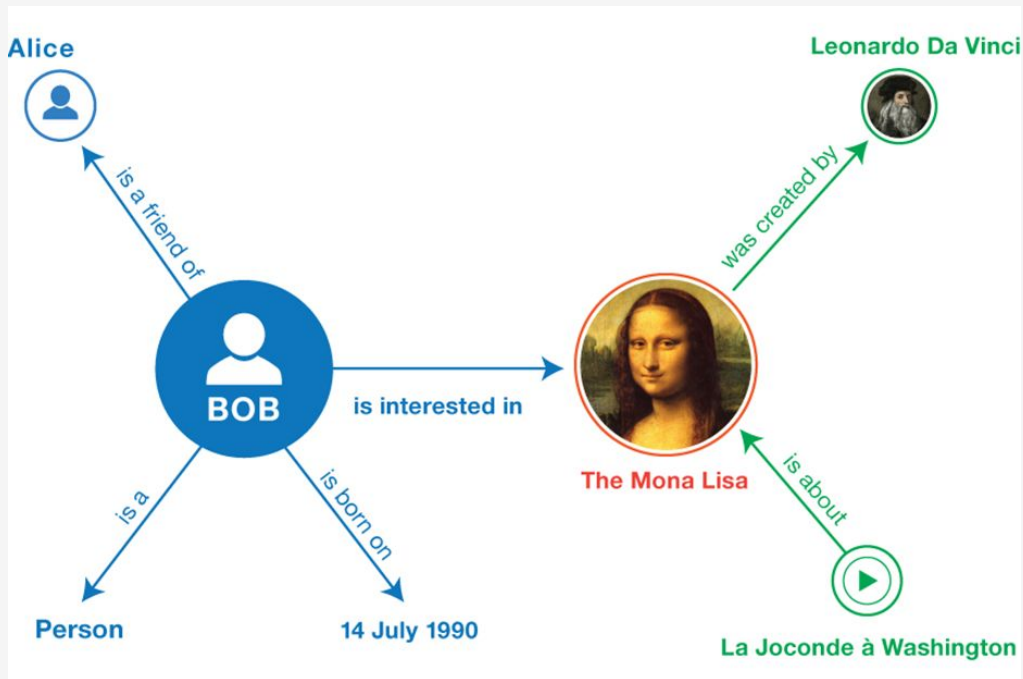
Bob is born on 14 July 1990 .

Bob is a Person .

Bob is a friend of Alice .

La Joconde a Washington is about The Mona Lisa .

The Mona Lisa was created by Leonardo Da Vinci .



Příklad grafu

Bob is interested in The Mona Lisa . `<Bob> <interestedIn> <TheMonaLisa> .`

Bob is born on 14 July 1990 . `<Bob> <DoB> "14 July 1990" .`

Bob is a Person . `<Bob> <isA> <Person> .`

Bob is a friend of Alice . `<Bob> <friendOf> <Alice> .`

La Joconde a Washington is about `<LaJocondeAWashington> <about>`
The Mona Lisa . `<TheMonaLisa> .`

`<TheMonaLisa> <createdBy>`

The Mona Lisa was created by `<LeonardoDaVinci> .`

Resource Description Framework - model a také schéma

RDF je způsob, jak modelovat znalosti (strukturované informace).

RDF nabízí slovník (vocabulary) se základními koncepty.

RDF Schema (RDFS) je rozšířením tohoto slovníku.

Zpravidla se používá jak RDF, tak RDFS (<https://www.w3.org/TR/rdf12-schema/>)

RDF	RDFS
Statement	subClassOf
type	label
List	Literal

RDF(S) - jmenné prostory

Namespace prefix	Namespace IRI	RDF vocabulary
rdf	http://www.w3.org/1999/02/22-rdf-syntax-ns#	The RDF built-in vocabulary [RDF11-SCHEMA]
rdfs	http://www.w3.org/2000/01/rdf-schema#	The RDF Schema vocabulary [RDF11-SCHEMA]
xsd	http://www.w3.org/2001/XMLSchema#	The RDF-compatible XSD types

<https://www.w3.org/TR/rdf11-concepts/#vocabularies>

Modelování znalostí a RDF(S)

<Movie> <is a> <Creative Work>

<<https://schema.org/Movie>> <<http://www.w3.org/2000/01/rdf-schema#subClassOf>> <<https://schema.org/CreativeWork>>

<<https://schema.org/Movie>> <rdfs:subClassOf> <<https://schema.org/CreativeWork>>

rdfs:subClassOf je QName (kvalifikované jméno)

Je zvykem používat URI:

- US-ASCII
- nepoužívat mezery
- pro víceslovné identifikátory používat camelCase
- objekty začínají velkým písmenem, vlastnosti malým písmenem

Další slovník (kvůli literálům)

XSD - XML Schema Definition

(<https://www.w3.org/TR/rdf11-concepts/#dfn-rdf-compatible-xsd-types>)

koncepty datových typů: xsd:string, xsd:integer, xsd:gYear

```
<Bob> <DoB> "14 July 1990" .
```

```
<Bob> <DoB> "14 July 1990"^^xsd:date .
```

Literály jsou jen označení

Označení jsou v přirozeném jazyce.

Jazyky mají standardní zkratky.

```
<TheMonaLisa> <rdfs:label> "The Mona Lisa"@en .
```

```
<TheMonaLisa> <rdfs:label> "La Joconde"@fr .
```

```
<TheMonaLisa> <rdfs:label> "Mona Lisa"@cs .
```

```
<TheMonaLisa> <rdfs:label> "La Gioconda"@it .
```

RDF(S) jsou standardy

<https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/>

O čem jsme ještě nemluvili:

- jaké znaky smí obsahovat literál
- jak porovnávat literály
- jak porovnávat grafy
- **jak data v RDF ukládat**
- **co graf v RDF znamená? jaká je jeho sémantika?**