

Organizační úvod

Poznámka (Organizační úvod)

Nahrávky budou. (Z minulého roku anglicky, z letoška česky.)

1 Úvod

Definice 1.1 (Strojově čitelný soubor)

Strojově čitelný soubor je vlastnost konkrétního souboru, ne formátu (jelikož do formátu můžu nacpat data v jiném formátu).

Strojová čitelnost se špatně definuje.

Definice 1.2 (Binární soubor)

Binární soubor je takový, kde je struktura popsána na úrovni bitů (bit po bitu). Není čitelný textovými editory.

TODO!!!

2 RDF

Definice 2.1 (RDF – resource description framework)

RDF je formát popisu grafu, kde se každé tvrzení (tedy trojice) má tvar „subjekt predikát objekt“, tj. „kdo co s-čím“. Vše se identifikuje pomocí IRI odkazující na definici (nebo v případě některých objektů (často Stringů/čísel/datumů) – literálem).

Poznámka

Uri budeme často zkracovat (takové zkrácení se zapisuje jako např. `@prefix dcterms: https://...`).
Obecné zkratky lze najít na [prefix.cc](https://www.prefix.cc).

Definice 2.2 (Literál)

Literál má dvě části – text odpovídající formátu a uri na ?XML schéma toho typu. Nebo je tvaru `"text"@jazyk`.

Například

Nejčastější predikát je `rdf:type` – „je typu“.

Definice 2.3 (Blank node)

Existují i nepojmenované uzly.

Definice 2.4 (RDF serializace)

(Jak zapsat RDF do textu.)

- RDF 1.1 N-Triples = každá trojice se zapíše jako `<uri> <uri> <uri> . # comment.`
- RDF 1.1 Turtle = použijí se prefixy, středníky na shodný subjekt a čárku na shodný subjekt i predikát + se používají relativní IRI (base se definuje pomocí `@base` IRI, implicitní je URL dokumentu) + multiline stringy a odescapeované znaky + `rdf:type` má zkratku `a` + blank nody se píšou pomocí hranatých závorek + běžné literály nemusí mít typ.
- RDF 1.1 N-Quads = místo trojice se kóduje i pojmenování grafu.
- RDF Trig = Turtle + pojmenované grafy (jsou reprezentovány jako bloky).

Definice 2.5 (Reifikace)

Pokud chci něco říct o naší trojici, můžu to udělat tak, že si definuji (zase pomocí trojic) objekt, který jako subjekt bude mít subjekt, atd. a navíc bude mít doplňující informace. Tato metoda se nazývá reifikace.

Definice 2.6 (Pojmenovaný graf, dataset)

Vztahy lze seskupit do tzv. pojmenovaného grafu.

Pojmenované grafy + defaultní graf se nazývá dataset.

Definice 2.7 (RDFS)

Nadstavba RDF, které umožňuje definovat třídy a dědičnost. `rdfs:Class`, `rdfs:subClassOf`, `rdf:Property`, `rdfs:range`, `rdfs:domain`, `rdfs:subPropertyOf`.

Oproti OOP není třeba definovat třídy, lze definovat property jako takové.

Také umožňuje label, comment, seeAlso: `rdfs:label`, `rdfs:comment`, `rdfs:seeAlso`, `rdfs:isDefinedBy`?

Definice 2.8 (rdf:List a jiné kolekce)

Ve specifikaci RDF je přímo definován spojový seznam (`rdf:List` + anonymní prvky + `rdf:nil`).

`rdf:_i`, kde *i* je libovolné číslo jsou predikáty náležení do kolekce (`rdf:TODO`).

Definice 2.9 (Open World Assumption (OWA))

Tvrzení může být pravdivé, i když to nevíme. (Tj. máme i odpověď nevím.)

TODO!!!