

Úvod do znalostních grafů a Wikidat

Josef Šlerka

Co je znalostní graf?

- Znalostní graf je síť propojených informací.
- Uzly představují entity (osoby, místa, objekty).
- Hrany představují vztahy mezi entitami.

Reprezentace vztahů v SPARQL

- SPARQL je dotazovací jazyk používaný pro práci se znalostními grafy.
- Vztahy mezi entitami se v SPARQL reprezentují pomocí trojic (subjekt, predikát, objekt).
- Příklad: „Jan Novotný je autorem knihy“ se zapíše jako trojice: `?author wdt:P50 ?book .`

Srovnání s Datalogem

- **Datalog** je deklarativní logický jazyk používaný v databázích a znalostních systémech.
- V Datalogu se vztahy také reprezentují pomocí pravidel a faktů.
- Na rozdíl od SPARQL se Datalog zaměřuje spíše na deduktivní dotazování a umožňuje definovat nové vztahy na základě existujících.
- SPARQL se používá především pro dotazování nad existujícími daty, zatímco Datalog klade důraz na inferenci nových informací.

Trojice v SPARQL

- **Subjekt:** Entita, o které se mluví (např. Jan Novotný).
- **Predikát:** Vztah mezi entitami (např. je autorem).
- **Objekt:** Entita, která je s prvním subjektem propojena (např. kniha).
- Každá trojice definuje vztah mezi dvěma entitami v grafu.

Příklad vztahů mezi entitami v SPARQL

```
SELECT ?author ?authorLabel ?book ?bookLabel
WHERE {
  ?book wdt:P50 ?author.  # Kniha je napsána autorem
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }
}
```

- Tento dotaz vrátí seznam autorů a jejich knih, přičemž ukazuje vztahy mezi nimi.

Co jsou entity?

- Entity jsou jednotlivé objekty nebo koncepty, které znalostní graf popisuje.
- Mohou to být osoby, místa, věci nebo abstraktní pojmy.
- Každá entita má jedinečné ID, které umožňuje její identifikaci a propojení s dalšími informacemi.

Co jsou vztahy?

- Vztahy spojují entity v rámci znalostního grafu.
- Každý vztah popisuje, jak jsou dvě entity propojeny (např. „autor knihy“, „město v zemi“).
- Vztahy dávají smysl jednotlivým datovým bodům a umožňují vytvářet sémantické souvislosti mezi různými entitami.

Proč jsou znalostní grafy důležité?

- Organizují a strukturovaně ukládají data.
- Usnadňují vyhledávání a analýzu informací.
- Zjednodušují opakované využívání informací.

Co jsou Wikidata?

- Otevřená, strojově čitelná databáze.
- Obsahuje strukturované informace z různých zdrojů.
- Wikidata jsou jako „chytrá Wikipedie“ pro stroje.
- Každá položka má své unikátní ID pro snadnou manipulaci a propojení.

Výhody Wikidat

- **Otevřenost:** Každý může přispívat a editovat.
- **Propojenost:** Každá položka má jedinečné ID, což umožňuje propojení napříč různými systémy.
- **Aktualizace v reálném čase:** Změna na jednom místě se promítne do všech relevantních oblastí.

Reálné využití Wikidat

- **Vědecký výzkum:** Agregace dat z různých vědeckých článků.
- **Příklad dotazu:** „Které knihy byly napsány v 19. století?“
- **Wikidata Query Service:** Umožňuje pokládat dotazy v jazyce SPARQL.

Wikidata Query Service (WDQS)

- **WDQS**: Nástroj pro interakci s daty ve Wikidatech pomocí SPARQL.
- **SPARQL**: Jazyk pro pokládání otázek ke znalostním grafům.
- Příklad jednoduchého dotazu: Vyhledání knih od konkrétního autora.

Příklad dotazu v SPARQL

```
SELECT ?book ?bookLabel
WHERE {
  ?book wdt:P50 wd:Q12345.  # knihy od konkrétního autora
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }
}
```

- Dotaz vrací seznam knih napsaných konkrétním autorem.
- WDQS umožňuje zobrazit výsledky přímo v prohlížeči.

Shrnutí

- **Znalostní grafy:** Propojené informace, které zjednodušují analýzu.
- **Wikidata:** Otevřená databáze umožňující snadnou práci s daty.
- **WDQS:** Umožňuje pokládat dotazy na data ve Wikidatech.

Doporučené Zdroje

- Interaktivní tutoriál: [Wikimedia Israel WDQS](#)
- Dokumentace k Wikidatům a SPARQL
- Příklady dotazů a inspirace pro výzkum

Přechod k další části

V další části se naučíme, jak psát složitější dotazy v jazyce SPARQL a jak získat užitečné informace z Wikidat.

Děkuji za pozornost 🙏