

Triplifikace dat

Co je triplifikace?

Převod dat do RDF.

Vstup: xml, csv, xls, databáze...

Výstup: RDF soubor, Sparql endpoint,...

Jak triplifikovat data?

1. vstupní soubor -> mapování -> výstup

výhody: jednoduchost

nevýhody: jiné vstupní soubory obsahující stejná data - nový mapovací soubor

2. vstupní soubor -> databáze -> mapování -> výstup

výhody: možnost naplnění databáze

nevýhody: pracnější (ne o moc :))

Postup s databází - <http://bit.ly/soviqP>

1. vybrat soubory pro import
2. vybrat nástroj pro import
 - použijeme Access databázi (odkazy na importovací nástroje na konci prezentace)
3. upravit vstupní soubor
4. nahrát do Accessu
 - přejmenovat tabulky
 - skontrolovat řádky databáze
5. upravit mapovací soubor
6. spustit server
7. vytvořit dump

1. vybrat soubory pro import

Celkový seznam ekologických podnikatelů:

http://cz.ckan.net/dataset/complete_list_of_entrepreneurs_in_the_field_of_organic_farming

Záložka: Výrobci biopotravin

2. vybrat nástroj pro import

Databázy: MySQL, PostgreSQL, Oracle
- je potřeba nástroj pro import

Access: jednoduchý import vstupních dat

3. Upravit vstupní soubor

Tabulku Výrobci biopotravin upravit tak aby bylo možné první řádek použít pro pojmenování sloupců tabulky

4. Nahrát do Accessu

- nahrát soubor
- upravit počet sloupců
- vytvořit ODBC zdroj

5. upravit mapovací soubor

- jazyk D2RQ

<http://www4.wiwiss.fu-berlin.de/bizer/d2rq/spec/#specification>

- existuje

R2RML <http://www.w3.org/2001/sw/rdb2rdf/r2rml/>

- implementace není dostatečně rozšířena

V souboru: <http://bit.ly/soviqP>
zápis mapovacího souboru

6. nainstalovat a spustit server

- Stáhnout server: <http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/>
- Rozbalit
- skopírovat soubory
- spustit server: `d2r-server mapping.n3`
- vytvořit dump: `dump-rdf -m mapping.n3 -b http://localhost:2020/ > vyrobci.rdf`