Framework for Extraction of Wikipedia Articles Content Diplomová práce

Oleksandr Husiev Vedoucí práce: Ing. Milan Dojčinovski, Ph. D.

> Fakulta informačních technologií České vysoké učení technické v Praze

> > 07. 12. 2021



- 1 Thesis goals
- 2 Motivation
- 3 Sekce 1
- 4 Analýza
- 5 Realizace 1
- 6 Design neceho
- 7 Testování

Thesis goals

DBpedia is a crowd-sourced community effort that aims at extraction of information from Wikipedia. Vast amounts of information are not yet extracted from the Wikipedia texts. The main goal of the thesis is to develop a framework for extraction of Wikipedia articles content, structure and annotations which can be further split into those subgoals:

- Accept and process input data in the form of Wikipedia XML dumps
- Extract Wikipedia context, page structure, and links.
- Provide outputs for context, links and page structure in the form of N-Triples.
- Implement language extensibility.
- Provide a user interface to interact with the framework.

Motivation

- Motivace 1
- Motivace 2
- Motivace 3
- Motivace 4
- Motivace 5

Sekce 1

- item 1
- item 2
- item 3
- item 4

Snimek s obrazkem



Analýza

- Analyza 1
- Analyza 2
- Analyza 3
- Analyza 4
- Analyza 5
- Analyza 6

Realizace 1

podtitulek realizace 1

- Dokonceno 1
- Dokonceno 2
- Dokonceno 3
- Dokonceno 4

Realizace 2

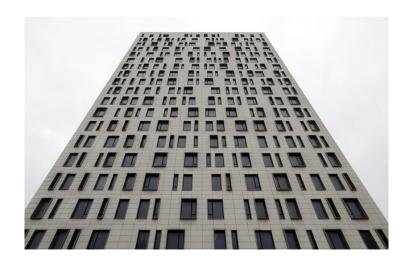
podtitulek realizace 2

- Dokonceno 1
- Dokonceno 2

podtitulek design neceho

- Pozadavek na design 1
- Pozadavek na design 2
- Pozadavek na design 3
- Pozadavek na design 4

design ceho



design ceho



design ceho



Testování

- Testování neceho 1
- Testování neceho 2
- Testování neceho 3
- Testování neceho 4

Shrnutí

- shrnuti 1
- shrnuti 2
- shrnuti 3
- shrnuti 4
- shrnuti 5
- shrnuti 6

Konec

Děkuji za pozornost!

Otázky oponenta

Otázka první: Proč?

Odpověď: Prostě proto.

Otázky oponenta

Otázka druhá: Proč?

Odpověď: Prostě proto.