

# 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

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.

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

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

# Snimek s obrazkem



- 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

# Design neceho

podtitulek design neceho

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

# Design neceho

design ceho



# Design neceho

design ceho





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

- shrnutí 1
- shrnutí 2
- shrnutí 3
- shrnutí 4
- shrnutí 5
- shrnutí 6

Děkuji za pozornost!



Otázka první: Proč?

Odpověď: Prostě proto.

Otázka druhá: Proč?

Odpověď: Prostě proto.