Abstract

Abstract

Contents

Abstract **i**

1 Introduction **1**

2 Analysis **2**
2.1 Problem analysis . . . 2
2.2 Literature analysis . . . 2
2.3 Lithuania's parliament open data analysis . . . 2
2.4 Method analysis . . . 2
2.4.1 MDS . . . 2
2.4.2 Unsupervised learning: *k*-means clusterization . . . 2
2.5 Visualizations . . . 2

3 Software design **3**
3.1 Components of system . . . 3
3.1.1 Data flow diagram . . . 3
3.2 Tools . . . 3
3.2.1 Programming language: *S*cala . . . 3
3.2.2 Database: *M*ySQL . . . 3
3.3 Downloader . . . 3
3.4 Coordinator . . . 3
3.5 API server . . . 3
3.6 Repository . . . 3
3.7 User interface (Frontend) . . . 3

4 Description of experimental research **3**
4.1 Data . . . 3
4.1.1 Data semantics . . . 3
4.1.2 Data statistics . . . 3
4.2 Experiments . . . 3
4.2.1 MDS . . . 3
4.2.2 Unsupervised learning: *k*-means clusterization . . . 3
4.3 Results . . . 3

A Appendix **i**

1 Introduction

Introduction text

2 Analysis

2.1 Problem analysis

2.2 Literature analysis

2.3 Lithuania's parliament open data analysis

2.4 Method analysis

2.4.1 MDS

2.4.2 Unsupervised learning: *k*-means clusterization

Due missing labels of this dataset we are forced to look into unsupervised machine learning methods. *k-means* clusterization [**?**] is one such method and it fits well with our features. This is due TFIDF which transforms text into vectors be used with *k-means*. For distance between comments we can use cosine similarity [**?**] due its nature of performing better when comparing texts. To compare we can try with Euclidian and Manhattan distances too.

2.5 Visualizations

3 Software design

3.1 Components of system

3.1.1 Data flow diagram

Data flow diagram

3.2 Tools

3.2.1 Programming language: *S*cala

3.2.2 Database: *M*ySQL

3.3 Downloader

3.4 Coordinator

3.5 API server

3.6 Repository

3.7 User interface (Frontend)

4 Description of experimental research

4.1 Data

4.1.1 Data semantics

4.1.2 Data statistics

4.2 Experiments

4.2.1 MDS

4.2.2 Unsupervised learning: *k*-means clusterization

4.3 Results

References

[1] J. Ratkiewicz, M. D. Conover, M. Meiss, B. Goncalves, A. Flammini, F. Menczer *Detecting and Tracking Political Abuse in Social Media*. In proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011.

A Appendix