

Strona tytułowa

Title page

Oświadczenie

statement

Dedication

Abstrakt

ABSTRACT

TABLE OF CONTENTS

Abstract	12
List of figures	15
List of tables	17
1. Introduction	19
1.1. Context.....	19
1.2. Motivation.....	19
1.3. Objectives	19
1.3.1. Advisor	19
1.3.2. Parallen environment framework.....	19
1.4. Hypothesis	19
1.5. Outline	19

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This chapter introduces the dissertation titled "*Text documents classification in parallel environment*" by presenting its context, motivation, project's objectives, hypothesis and an outline on sections 1.1, 1.2, 1.3, ??, ??out respectively.

1.1. Context

1.2. Motivation

The motivation for this work is to provide a framework for text document categorization in parallel environment.

1.3. Objectives

In this project a tool for fast and scalable text document categorization will be developed and implemented. It will consist of two parts described shortly below.

1.3.1. Advisor

The first part of work is to prepare implementation of advisor for text document categorization. Advisor will be responsible for generating decision on class of given text document in relevance to known set of documents. It will be based on a tiered model with layers specified as follows:

1. First layer will be responsible for general decision that restricts the set of potential classes. It will be implemented using simple minimal distance kNN classifier.
2. second layer will return more precise decision and will be implemented using more complex Support Vector Machine approach.

1.3.2. Parallen environment framework

The second part of work consist of designing and implementing generic framework, which would allow users to run different advisors in parallel environment, speeding up total time needed for text document categorization. This includes designing of systems

1.4. Hypothesis

1.5. Outline

The rest of the document is structured as follows:

1. **Chapter ??** - Analysis of the state of knowledge in the field of: parsing text documents, counting theirs statistics, finding distance and relevance between them, methods of distributing data across cluster and running scalable tasks on it, comparing and evaluating effectiveness of these methods.
2. **Chapter ??** - Methodology, choice of algorithms for automatic analysis of facial expressions and analysis of the possibilities of their realization in HTML5 WebRTC technology.
3. **Chapter ??** - Design of the system, aim of the project, potential users, system requirements.
4. **Chapter ??** - Description and implementation of the system.

5. **Chapter ??** - Development of the test scenarios, performing the tests and conducting the experiments in different measuring conditions. Description of the achieved results.
6. **Chapter??** - Drawing conclusions from the results of work and possibilities of the future work.

State of the art Methodology Design Implementation Tests Conclusions

Bibliography Appendix