INTENTS

List of Abbreviations
List of Figures
List of Tables
Abstract

CHAPTER       PAGE
1. INTRODUCTION  1
  1.1 INTRODUCTION  1
  1.2 NECESSITY    2
  1.3 OBJECTIVES   3
  1.4 THEME        3
  1.5 ORGANIZATION OF THE THESIS  4
      REFERENCES  1
2. LITERATURE SURVEY  8
  2.1 DATA PREPARATION  9
  2.2 FIELD MATCHING TECHNIQUES  9
  2.3 RECORD MATCHING TECHNIQUES  25
  2.4 DATA MINING WITH ONTOLOGIES  27
      2.4.1 DOMAIN ONTOLOGY  29
      2.4.2 METADATA ONTOLOGIES  30
      2.4.3 ONTOLOGIES FOR DATA MINING PROCESS  30
      2.4.4 THESAURUS AND ONTOLOGIES  31
      2.4.5 ELABORATING ONTOLOGY  32
      2.4.6 FROM THESAURUS TO ONTOLOGY  34
      2.4.7 HUNSPELL FOR THESAURUSES  36
      REFERENCES  37
3. ALGORITHM FOR FINDING DUPLICATES IN DISTRIBUTED  43
      DATABASE USING ONTOLOGY AND SHORTCUTS DICTIONARY
      3.1 HOMOGENEOUS DISTRIBUTED DATABASE SYSTEM  43
      3.2 HETEROGENEOUS DISTRIBUTED DATABASE SYSTEMS  44
      3.3 CLIENT/SERVER DATABASE ARCHITECTURE  44
      3.4 SQMD: SINGLE QUERY MULTIPLE DATABASE ARCHITECTURE  45
      3.5 WEB SERVICE  46
      3.6 SHORTCUT DICTIONARY  48
      3.7 THE ARCHITECTURE THAT IS DEVELOPED  49
      REFERENCES  51
4. ENHANCEMENTS IN THE ALGORITHMS FOR FINDING THE  52
      WEIGHTS IN NUMERICAL DATA
      4.1 NUMBER SEARCH IN DOCUMENTS  53
      4.2 NUMBERS IN DATABASES  54
      4.3 ALGORITHM BASED ON COSINE SIMILARITY  56
      4.4 PROPOSED ALGORITHM FOR PURE NUMERIC FIELDS  57