

**Белорусский государственный университет
Факультет социокультурных коммуникаций**

Кафедра компьютерной лингвистики и лингводидактики

**Реферат дипломной работы
«Алгоритмы контекстно-зависимого аннотирования и кластеризации
документов при информационном поиске»**

**Раздёрина Янина Ивановна,
Руководитель Воронович Валерий Викторович**

2018 год

РЕФЕРАТ

дипломной работы Раздёрной Я.И.

АЛГОРИТМЫ КОНТЕКСТНО-ЗАВИСИМОГО АННОТИРОВАНИЯ И КЛАСТЕРИЗАЦИИ ДОКУМЕНТОВ ПРИ ИНФОРМАЦИОННОМ ПОИСКЕ

Дипломная работа включает 55 страниц, 31 рисунок, 3 приложения и основана на 37 источниках.

Ключевые слова: кластеризация документов, информационный поиск, аннотация, сниппет, контекстно-зависимое аннотирование, перцептрон, интеллектуальный поиск, скрипт, кликабельность, поисковая система.

Цель работы – разработка программы для составления автоматической аннотации при информационном поиске.

Объект исследования – процессы контекстно-зависимого аннотирования и кластеризации текстовых документов.

Предмет исследования – алгоритмы контекстно-зависимого аннотирования и кластеризации текстовых документов при информационном поиске.

Актуальность работы – распределение документов при информационном поиске, их грамотное отображение в поисковой выдаче и возможность найти пользователю необходимую информацию в кратчайшие сроки актуальны как никогда.

Основные методы исследования – аналитический, классификационный, обобщающий, дедуктивный, метод аналогии.

Теоретические новизна работы состоит в создании словарей синонимов и словоформ, которые используются разработанной в процессе исследования программой для предоставления качественных результатов по запросам.

Практическая значимость. На базе файлов синонимов и словоформ была разработана программа на языке программирования BASH, при ее усовершенствовании и добавлении необходимых словарей и корпуса текстов можно усовершенствовать программный код и разработать систему автоматического реферирования текстов.

Материалы дипломной работы были апробированы на 74-ой научной конференции студентов и аспирантов БГУ, Минск, 18.05.2017, 75-ой научной конференция студентов и аспирантов БГУ, Минск, 17.05.2018.

Автор работы подтверждает, что приведенный в работе материал объективно отражает состояние исследуемого процесса, а все заимствования из источников сопровождаются ссылками на их авторов.

подпись студента

РЭФЕРАТ

дыпломнай працы Раздзёрынай Я.І.

АЛГАРЫТМЫ КАНТЭКСТНА-ЗАЛЕЖНАГА АНАТАВАННЯ І КЛАСТЭРЫЗАЦЫІ ДАКУМЕНТАЎ ПРЫ ІНФАРМАЦЫЙНЫМ ПОШУКУ

Дыпломная праца ўключае 55 старонак, 31 малюнак, 3 дадатку і заснаваная на 37 крыніцах.

Ключавыя словы: кластарызацыя дакументаў, інфармацыйны пошук, анатацыя, сніпет, кантэкстна-залежнае анатаванне, персептрон, інтэлектуальны пошук, скрыпт, клікабельнасць, пошукавая сістэма.

Мэта даследавання – распрацоўка праграмы для складання аўтаматычнай анатацыі пры інфармацыйным пошуку.

Аб’ект даследавання – працэсы кантэкстна-залежнага анатавання і кластарызацыі тэкставых дакументаў.

Прадмет даследавання – алгарытмы кантэкстна-залежнага анатавання і кластарызацыі тэкставых дакументаў пры інфармацыйным пошуку.

Актуальнасць працы – размеркаванне дакументаў пры інфармацыйным пошуку, іх пісьменнае адлюстраванне ў пошукавай выдачы і магчымасць знайсці дапамогу ў самых кароткія тэрміны актуальныя як ніколі.

Асноўныя метады даследавання – аналітычны, класіфікацыйны, абагульняючы, дэдуктыўны, метады аналогіі.

Тэарэтычныя навізна працы складаецца ў стварэнні слоўнікаў сінонімаў і словаформаў, якія выкарыстоўваюцца распрацаванай ў працэсе даследавання праграмай для прадастаўлення якасных вынікаў па запытам.

Практычная значнасць. На базе файлаў сінонімаў і словаформаў была распрацавана праграма на мове праграмавання BASH, пры яе ўдасканаленні і даданні неабходных слоўнікаў і корпуса тэкстаў можна удасканаліць праграмны код і распрацаваць сістэму аўтаматычнага рэферыравання тэкстаў.

Матэрыялы дыпломнай працы былі апрабаваны на 74-ай навуковай канферэнцыі студэнтаў і аспірантаў БДУ, Мінск, 18.05.2017, 75-ая навуковай канферэнцыі студэнтаў і аспірантаў БДУ, Мінск, 17.05.2018.

Аўтар працы пацвярджае, што прыведзены ў працы матэрыялы правільна і аб’ектыўна адлюстроўвае стан даследнага працэсу, а ўсе запазычаныя з літаратурных і іншых крыніц суправаджаюцца спасылкамі на іх аўтараў.

подпіс студэнта

SUMMARY

of the diploma paper of Yanina Razdzioryna

ALGORITHMS OF CONTEXT-DEPENDENT ANNOTATION AND CLUSTERING OF DOCUMENTS FOR INFORMATION SEARCH

The diploma paper includes 55 pages, 31 drawings, 3 applications and is based on 37 sources.

Key words: document clustering, information retrieval, abstract, snippet, context-dependent annotation, perceptron, intellectual search, script, clickability, search engine.

The purpose of the work is to develop a program for compiling an automatic annotation for information retrieval.

The object of research is submitted by processes of context-dependent annotation and text documents clustering.

The subject of the study is submitted by algorithms for context-dependent annotation and clustering of text documents for information retrieval.

The relevance of the work shows the distribution of documents in the information search, their competent display in the search result and the ability to find the user the necessary information in the shortest possible time to be as relevant as ever.

Research basic methods – analytical, synthetic, the method of classification, generalizing, deductive, the method of analogy.

The theoretical novelty of the work is the creation of dictionaries of synonyms and word forms, which are used by the program developed in the process of research to provide qualitative results for queries. Theoretical bases of methods of context-dependent annotation and clustering of documents in information search are used.

Practical significance. Based on the file of synonyms and word forms, a program was developed in BASH programming language, with its improvement and addition of dictionaries and the corpus, we can improve the program code and develop a system of automatic text abstracting.

Materials of the diploma paper were approved at the 74th Scientific Conference of Students and Postgraduate of BSU, Minsk, 18.05.2017, 75th Scientific Conference of Students and Postgraduate of BSU, Minsk, 18.05.2018

The author of the paper confirms that the material correctly and objectively reflects the state of the investigated process and all borrowed theoretical, methodological and methodical concepts from literature and other sources are accompanied by links to their authors.

signature of the student