

**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**  
**Факультет радиопизики и компьютерных технологий**  
**Кафедра интеллектуальных систем**

Аннотация к дипломной работе

**«Алгоритм поиска фактов в текстах на естественном языке с использованием GLR-парсера»**

Колодник Илья Анатольевич

Научный руководитель - старший преподаватель, Адуцкевич И. А.

## Реферат

Дипломная работа: 56 страниц, 8 рисунков, 9 таблиц, 10 источников.

ЕСТЕСТВЕННЫЙ ЯЗЫК, МАШИННОЕ ОБУЧЕНИЕ, GLR-АЛГОРИТМ, ИЗВЛЕЧЕНИЕ ФАКТОВ, МОРФОЛОГИЧЕСКИЙ АНАЛИЗ, СИНТАКСИЧЕСКИЙ АНАЛИЗ.

*Цель работы* – анализ существующих систем извлечения фактов из текстов на естественном языке, выявление их недостатков и построение системы по разработанному алгоритму.

В работе проведено исследование общих алгоритмов работы естественно-языковых систем и описаны методы, используемые на каждом из этапов анализа. Выполнено сравнение систем работы с текстами на английском и русском языках, описаны возможности каждой из систем.

Предложены способы устранения имеющихся недостатков, реализована система выделения фактов из текста на естественном языке и разработан web-сайт выступающий в качестве интерфейса для системы.

В заключительной части работы описаны возможные сферы применения систем извлечения фактов из текстов на естественном языке, отмечены возможные направления развития построенной системы.

## Рэферат

Дыпломная праца: 56 старонак, 8 малюнкаў, 9 табліц, 10 крыніц.

НАТУРАЛЬНАЯ МОВА, МАШЫННАЕ НАВУЧАННЕ, GLR-АЛГАРЫТМ, ВЫМАННЕ ФАКТАЎ, МАРФАЛАГІЧНЫ АНАЛІЗ, СІНТАКСІЧНЫ АНАЛІЗ.

*Мэта працы* – аналіз існуючых сістэм здабывання фактаў з тэкстаў на натуральнай мове, выяўленне іх недахопаў і пабудова сістэмы па распрацаваным алгарытме.

У працы праведзена даследаванне агульных алгарытмаў працы натуральна-моўных сістэм і апісаны метады, якія выкарыстоўваюцца на кожным з этапаў аналізу. Выканана параўнанне сістэм для англійскай і рускай мовы, апісаны іх асноўныя магчымасці.

Прапанаваны спосабы ліквідацыі наяўных недахопаў, рэалізавана сістэма выбару фактаў з тэкстаў на рускай мове і распрацаваны web-сайт які з'яўляецца інтэрфейсам да сістэмы.

У заключнай частцы апісаны магчымыя сферы ўжывання сістэм здабывання фактаў з тэкстаў на натуральнай мове, адзначаны напрамкі развіцця пабудаванай сістэмы.

## **Abstract**

Thesis: 56 pages, 8 figures, 9 tables, 10 sources.

NATURAL LANGUAGE PROCESSING, MACHINE LEARNING, GLR-ALGORITHM, FACTS EXTRACTION, MORPHOLOGICAL ANALYSIS, TEXTS PARSING.

*Main purpose of the work* - analysis of the existing fact extraction systems for the texts in natural language, identifying their flaws and building such system according to the developed algorithm.

The papers contain research of general algorithms for the natural language processing systems and description of methods that are used on each step of analysis. The comparison of systems, which work with texts in English and Russian, including features of each system was done.

This document provides ways of resolving existing drawbacks. Facts extraction system with a web-interface was realized as a part of practical work.

Final part of the papers is for the description of the scope of facts extraction systems and possible directions of program extensions.

