

**БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ**

**Факультет прикладной математики и информатики**

**Кафедра компьютерных технологий и систем**

Аннотация к дипломной работе

**ГРУППИРОВКА ТЕЛЕЗРИТЕЛЕЙ ПО БОЛЬШОМУ НАБОРУ  
СОЦИАЛЬНЫХ ПРИЗНАКОВ**

Глушук Филипп Викторович

Научный руководитель - профессор кафедры КТС, кандидат  
физ.-мат. наук Казаченок В.В.

**Минск 2020**

## РЕФЕРАТ

Дипломная работа содержит, 62 страниц, 43 иллюстрации (рисунков), использованных литературных источников 11.

**Ключевые слова:** БОЛЬШИЕ ДАННЫЕ, КЛАСТЕРИЗАЦИЯ, ПОИСК ПОДОБНЫХ, АВТОМАТИЗАЦИЯ, PYTHON, APACHE SPARK.

**Объектом исследования** являются процесс построения системы для поиска групп подобных объектов при использовании больших данных.

**Целью дипломной работы** является исследовать особенности, рассмотреть возможные проблемы и решения, возникающие при построении модели поиска подобных объектов при использовании больших данных.

**В результате исследования** получены следующие результаты:

подготовлен аналитический обзор этапов, выделяемых при построении моделей для поиска групп подобных объектов, рассмотрены наиболее популярные методы в рамках выделенных этапов, выявлены достоинства и недостатки, а также найдены основные ограничения, при которых могут использоваться рассмотренные методы, реализовано приложение и продемонстрирована его работа на наборе реальных данных.

**Методы исследования** – изучение предметной области, обзор существующих подходов, аналитическое сравнение, объектно-ориентированное программирование.

**Областью применения** является поиск подобных, кластеризация данных, большие данные.

Дипломная работа выполнена автором самостоятельно.

## РЭФЕРАТ

Дыпломная праца ўтрымлівае, 62 старонак, 43 ілюстрацыі (малюнкаў), выкарыстаных літаратурных крыніц 11.

**Ключавыя словы:** ВЯЛІКІЯ ДАДЗЕННЫЯ, КЛАСТАРЫЗАЦЫІ, ПОШУК ПАДОБНАГА, АЎТАМАТЫЗАЦЫЯ, PУТНОН, АРАСНЕ SPARK.

**Аб'ектам даследавання з'яўляюцца** працэс пабудовы сістэмы для пошуку груп падобных аб'ектаў пры выкарыстанні вялікіх дадзеных.

**Мэтай дыпломнай працы з'яўляецца** даследаваць асаблівасці, разгледзець магчымыя праблемы і рашэнні, якія ўзнікаюць пры пабудове мадэлі пошуку падобных аб'ектаў пры выкарыстанні вялікіх дадзеных.

**У выніку даследавання** атрыманы наступныя вынікі:

падрыхтаваны аналітычны агляд этапаў, што выдзяляюцца пры пабудове мадэляў для пошуку груп падобных аб'ектаў,

разгледжаны найбольш папулярныя метады ў рамках выдзеленых этапаў, выяўлены вартасці і недахопы, а таксама знойдзеныя асноўныя абмежаванні, пры якіх могуць выкарыстоўвацца разгледжаныя метады,

рэалізавана прыкладанне і прадэманстравана яго праца на наборы рэальных дадзеных.

**Метады даследавання** - вивучэнне прадметнай вобласці, агляд існуючых падыходаў, аналітычнае параўнанне, аб'ектна-арыентаванае праграмаванне.

**Вобласцю ужывання** з'яўляецца пошук падобных, кластарызацыя дадзеных, вялікія дадзеныя.

Дыпломная праца выканана аўтарам самастойна.

## ABSTRACT

Diploma work contains 62 pages, 43 illustrations (figures), 18 references.

**Key words:** BIG DATA, CLUSTERING, SEARCH FOR SIMILAR OBJECTS, AUTOMATIZATION, PYTHON, APACHE SPARK.

**Object of research** is the process of building a system to search for groups of similar objects when using big data.

**The goal of the work** is to investigate features, consider possible problems and solutions that arise when building a system for searching for similar objects when using big data.

**As a result of the work**, the following results were obtained:

an analytical review of the stages allocated during the construction of models for searching for groups of similar objects was prepared,

considered the most popular methods in the framework of the selected stages, advantages and disadvantages are identified, and the main limitations are found under which the considered methods can be used,

implemented the application and demonstrated its work on a set of real data.

**The research methods** - the study of the subject area, a review of existing approaches, analytical comparison, object-oriented programming.

**The scope of application** is search for similar, data clustering, big data.

Diploma work is made by the author himself.