МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені І.І.МЕЧНИКОВА

(повне найменування вищого навчального закладу)

Факультет математики, фізики та інформаційних технологій

(повне найменування інституту, назва факультету (відділення))

Кафедра математичного забезпечення комп'ютерних систем

(повна назва кафедри (предметної, циклової комісії))

Дипломна робота

на здобуття освітньо-кваліфікаційного рівня «бакалавр»

(освітньо-кваліфікаційний рівень)

на тему Система аналізу кластерної структури природних об`єктів

Analysis system of the natural objects nomenclature cluster structure

Виконав: студент денної форми навчання

напряму підготовки 6.050102 – Комп'ютерна інженерія

(шифр і назва напряму підготовки, спеціальності)

Зубрицький Кирило Борисович

(прізвище, ім'я, по-батькові)

Керівник

ст. викл. Трубіна Н.Ф.

(науковий ступінь, вчене звання, прізвище та ініціали, підпис)

Рецензент

ст. викл. Берков Ю.М.

(науковий ступінь, вчене звання, прізвище та ініціали)

Рекомендовано до захисту:

Протокол засідання кафедри

№___ від «___»_____ 2019 р.

Завідувач кафедри

_____ _Є.В. Малахов_

(підпис) (прізвище, ініціали)

Захищено на засіданні ЕК № ___

протокол № __ від «___» ____ 2019 р.

Оцінка _____/_____/_____

(за національною шкалою, шкалою ECTS, бали)

Голова ЕК

_____ _О.О. Арсірій_____

(підпис) (прізвище, ініціали)

Одеса - 2019

АНОТАЦІЯ

У роботі розглядаються методи рішення задач систематизації природничої номенклатури за допомогою кластерізації. Проведено аналіз сучасних методів рішення задач систематизації. Проведено аналіз найбільш ефективних та сучасних методів кластерізації які дозволяють вирішувати поставлену проблему найбільш швидким та ефективним способом. На підставі проведеного аналізу розроблено відповідне програмне забезпечення, яке дозволило реалізувати прототип систематизації природної номенклатури.

ABSTRACT

In the thesis are consideredthe methods of solving problems of systematization of the natural nomenclature with the help of clusterization The analysis of modern methods of solving systematization problems is carried out. The analysis of the most effective and modern methods of clusterization that allows solving the problem in the most rapid and effective way is carried out. Based on the analysis conducted, the software was developed, which allowed to implement the prototype of systematization of the natural nomenclature.

АННОТАЦИЯ

В работе рассматриваются методы решения задач систематизации естественной номенклатуры с помощью кластеризации. Проведен анализ современных методов решения задач классификации. Проведен анализ наиболее эффективных и современных методов кластеризации которые позволяют решать поставленную проблему наиболее быстрым и эффективным способом. На основании проведенного анализа разработано соответствующее программное обеспечение, которое позволило реализовать прототип систематизации природной номенклатуры.

ЗМІСТ

ВСТУП

Практично у кожній галузі людської діяльності використовуються та чи інша систематизація. Під систематизацією тут розуміється опис і розміщення в системі всіх об’єктів деякої предметної області.

Найліпших та найточніших результатів досягли в вивчені природничого комплексу. На сьогодні існують стандарти класифікації. Наприклад, ієрархія таксонів і правила найменування рослин (номенклатура) регулюються обов'язковим для всіх ботаніків Міжнародним кодексом ботанічної номенклатури [1]. Вносити зміни в цей документ мають право тільки міжнародні ботанічні конгреси. Незважаючи на вже усталені стандарти, зміни до систематизації вносяться по сьогодні.

Важливим завданням систематизації є діагностика (визначення, тобто знаходження місця об’єкту в системі). Під діагностикою розуміють передусім складання таблиць для визначення об’єктів, так званих визначників. Розроблюється багато визначників різноманітного направлення: регіонального, міжвидового, для вивчення, польових умов та інш.[2-8].При цьому встає питання формальної перевірки якості побудованих таблиць визначення. Наприклад в визначнику «Плантаріум»[8] за очевидними зовнішніми ознаками невдається визначити таку всім відому рослину як кульбаба. В зв’язку з цим стає необхідність в інструменті, який дозволяв би виконувати формальну перевірку в якості визначника. Можливо припустити, що для добре розробленого якісного визначника отримаємо результат дуже близький до вже існуючі системи класифікації.

Кластеризація − задача розбиття заданої вибірки об'єктів (ситуацій) на підмножини, так, щоб кожен кластер складався з схожих об'єктів, а об'єкти різних кластерів істотно відрізнялися.

Помітимо, майже усі класифікацій природних об’єктів мають ієрархічний характер. Багато спільного є також у методах побудови ключів визначення. Тому головним завданням даної роботи є дослідження методів систематизації

з метою створення такої програми, яка могла би використовуватися для різних природних об’єктів лише за рахунок незначних змін та наповнення бази даних, або незначних змін у самій базі.

Інструмент який дозволяє виявити кластерну структуру номенклатури об`єктів, що міститься у визначнику або довіднику, та порівняти її з існуєчую класифікацію вельми цікава в усіх тих сферах, де знаходиться величезна кількість систематизованої інформації.

Для досягнення поставленої мети необхідно вирішити наступні задачі:

- провести дослідження предметної області;
- розглянути методи класифікації та побудови ключів визначення;
- виконати огляд існуючих електронних аналогів;
- провести проектування функціональної моделі і вибрати апаратну платформу для її реалізації;
- вибрати інструментальне середовище розробки і виконати програмну реалізацію інформаційної системи;
- проаналізувати роботу та знайти похибки у системі (якщо вони є);

ВИСНОВКИ

В ході виконання даної дипломної роботи мною була проаналізованна эфективнітсть кластерізації системи номенклатури природничих об’єктів .

Під час виконання данної роботи була організована, описана і база данних, було описано специфікацію вимог до бази данних. Була розроблена кластеризація системи 2 методами. Були обрані на мою думку найбільш актуальна для даної теми технології, які дозволили досягнути головної мети даного проєкту.

Проєкт можливо було вдосконалити використав єдину модель та підхід створена система таксономічної ідентифікації для предметної області та використавши більш доскональні індефікаційні ключи.

Мета яку мав данний проект була досягнута та реалізована в повному обсязі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Міжнародний кодекс ботаніної номенклатури (Венский кодекс). пер. с англ. – КМК,2009 – 288 с.

2. Глобальний інформаційний фонд з біорізноманіття (GBIF) [Електроний ресурс]. – Режим доступу: htt p://www.gbif.org/

3. ITIS - інтегрована таксономическая інформаційна система [Електроний ресурс]. – Режим доступу: http://www.itis.usda.gov/

4. Міжнародний індекс назв рослин (IPNI) [Електроний ресурс] – Режим доступа: http://www.ipni.org/

5. Определитель Израильских растений "Flora of Israel Online" [Електроний ресурс]. – Режим доступа: http://flora.org.il/en/plants/

6. Определитель 408 видов птиц, обитающих в Великобритании [Электронный ресурс]. – Режим доступа: https://www.rspb.org.uk/birds-and-wildlife/wildlife-guides/identify-a-bird

7. Определитель птиц "Birds in Backyards" [Електроний ресурс]. – Режим доступа: http://www.birdsinbackyards.net/finder

8. Определитель растений "Плантариум" [Електроний ресурс]. – Режим доступа: http://www.plantarium.ru/page/find.html

9. Шаталкин А. И. Таксономия: Основания, принципы и правила / Зоологический музей МГУ. — М.: Товарищество научных изданий КМК,2012. — 600 с.

10. Лобанов А.Л. Логический анализ и классификация существующих форм диагностических ключей // Энтомологическое обозрение. 1972. Т51,вып.3 С. 668–681

11. Цифровой иллюстрированный атлас-определитель растений средней полосы России [Електроний ресурс]. – Режим доступу: http://school-collection.edu.ru/catalog/rubr/23f0d487-f462-4575-887b-1f731c55a849/

12. T.Petrushina, N.Trubina, Quality analysis of the computer identifier based on a unified approach, PROCEEDINGS of the 3d International Conference on

Computer Algebra and Information Technologies, August 20 – 25, 2018 Odessa, Ukraine, p.188-192

13. Энциклопедия комнатных растений[Електроний ресурс]. – Режим доступу: http://rus.gflora.com/

14. ARTHUR, D., VASSILVITSKII, S. k-means++: The advantages of careful seeding // Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms. — Society for Industrial and Applied Mathematics, 2007. — P. 1027–1035.

15. CHARIKAR, M. ET AL. Incremental clustering and dynamic information retrieval // SIAM Journal on Computing. — 2004. — Vol. 33, №. 6. — P. 1417–1440.