Облікова картка ДіР

I. Загальні відомості

Державний обліковий номер: 0220U102891

Державний реєстраційний номер: 0117U002946

Особливі позначки: відкрита

Дата реєстрації: 22-06-2020

II. Етап виконання ДіР

Номер етапу: 1

Назва етапу: Розробка алгоритмів та методів автоматизації процесів консолідації інформації з відкритих джерел

Початок етапу: 05.2017

Закінчення етапу: 04.2020

Вид звітного документа: Остаточний звіт

III. Відомості про виконавця ДіР

Повне найменування юридичної особи (або ПІБ фізичної особи): Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського"

Код за ЄДРПОУ: 02070921

Місцезнаходження: проспект Перемоги, 37, м. Київ, Київська обл., 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Розмір організації:

Телефон: 380442367989, 380442044862

IV. Відомості про співвиконавців ДіР

V. Відомості про замовника ДіР

Повне найменування юридичної особи (або ПІБ фізичної особи): Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського" Інститут енергозбереження та енергоменеджменту

Код за ЄДРПОУ: 247571500

Місцезнаходження: вул. Борщагівська 115, м. Київ, Київська обл., 03056, Україна

Форма власності:

Сфера управління: Міністерство освіти і науки України

Ідентифікатор ROR: Не застосовується

Розмір організації:

Телефон: 0442049690

VI. Джерела, напрями та обсяги фінансування ДіР

Підстава для проведення ДіР: 43 - власна ініціатива (якщо робота виконується з власної ініціативи за кошти виконавця НДР або безкоштовно)

Напрям фінансування: 2.7 - інше

Джерела фінансування

7706 - безплатно (договір про науково-технічне співробітництво, тощо)

VII. Відомості про ДіР

Назва роботи українською:

Розробка алгоритмів та методів автоматизації процесів консолідації інформації з відкритих джерел

Назва роботи англійською:

Development of algorithms and methods for automating the processes of consolidation of information from open sources

Реферат українською:

Звіт містить 73 аркушів машинописного тексту, 22 рисунка 30 бібліографічних джерел. В роботі виконано: 1. Розглянуто задачу консолідації даних з відкритих джерел, з застосуванням спеціальних підходів та інструментів керування даними. Розроблено кількісні показники оцінки інформації у запитах. 2. Розглянуто 5 основних моделей пошуку інформації: булеву модель; модель нечітких множин; векторну модель; латентно-семантичну модель; ймовірнісні моделі, на основі яких створено систему базових елементів консолідації інформації і алгоритм кількісної оцінки інформації. 3. Розроблено програмний алгоритм, представлений у блок-схемі, формування релевантних документів у інформаційних джерелах для тестування алгоритмів консолідації. Наведено приклад згенерованого тестового середовища з 5-и інформаційних джерел. 4. Досліджено ефективність системи консолідації інформації з відкритих джерел на основі стохастичного автомату, що дозволяє констатувати підвищення ефективності в 3–и рази і більше. 5. Досліджено використання засобів парсингу при консолідації інформації з відкритих джерел. Виявлено і обґрунтовано, що найбільш відповідним та ефективним може бути використання засобів парсингу зі зберіганням ієрархії елементів. 6. Розглянуто застосування штучного інтелекту, у консолідації інформації, на основі триграм аналізу текстової інформації. 7. Розроблено систему пошуку патентної інформації. 8. Розроблено систему консолідації інформації при геокодуванні інформаційних запитів. Розроблено схему системи, розглянуто компоненти пошуку інформації, компоненти аналізу. 9. Розглянуто особливості пошуку інформації в INDUSTRY 4.0. Розглянуто метрики Хеммінга, Левенштейна і Дамерау Левенштейна. Запропоновано алгоритм системи консолідації інформації в INDUSTRY 4.0. Ключові слова: КОНСОЛІДАЦІЯ ІНФОРМАЦІЇ З ВІДКРИТИХ ДЖЕРЕЛ; ШТУЧНИЙ ІНТЕЛЕКТ; СТОХАСТИЧНИЙ АВТОМАТ; ПАРСИНГ; ГЕОКОДУВАННЯ; INDUSTRY 4.0.

Реферат англійською:

The report contains 73 sheets of typewritten text, 22 Figure 30 bibliographic sources. The work performed: 1. The problem of consolidation of data from open sources, using special approaches and data management tools is considered. 2. 5 main models of information retrieval are considered: bleft model; fuzzy set model; vector model; latent-semantic model; probabilistic models, on the basis of which a system of basic elements of information consolidation and an algorithm for quantitative evaluation of information are created. 3. The software algorithm presented in the block diagram, formation of relevant documents in information sources for testing of consolidation algorithms is developed. An example of a generated test environment from 5 information sources is given. 4. The efficiency of the system of consolidation of information from open sources on the basis of the stochastic automaton is investigated that allows to state increase of efficiency in 3 times and more. 5. The use of parsing tools in the consolidation of information from open sources is studied. It is revealed and substantiated that the most appropriate and effective may be the use of parsing tools with the preservation of the hierarchy of elements. 6. The application of artificial intelligence in information consolidation on the basis of trigrams of textual information analysis is considered. 7. A patent information retrieval system has been developed. 8. A system of information consolidation in geocoding of information requests has been developed. The scheme of the system is developed, the components of information search, the components of the analysis are considered. 9. Features of information retrieval in INDUSTRY 4.0. Metrics are considered Hemming, Levenstein and Damerau Levenstein.

The algorithm of the information consolidation system inINDUSTRY 4.0. Keywords: CONSOLIDATION OF OPEN SOURCES INFORMATION; ARTIFICIAL INTELLIGENCE; STOCHASTIC AUTOMATIC MACHINE; PARSING; GEOCODING; INDUSTRY 4.0.

Індекс УДК: 001.89:004, 004.622

Коди тематичних рубрик: 50.53

Керівники роботи

Власне Прізвище Ім'я По-батькові: Кузьміних Валерій Олександровис

Науковий ступінь:

Вчене звання:

Ідентифікатор ORCID ID:

Додаткова інформація:

VIII. Наукова (науково-технічна) продукція (НТП)

Назва НТП українською: Програмні засоби аналізу патентної інформації

Назва НТП англійською: Software tools for patent information analysis

НТП, яку передбачалося створити:

Причини, через які НТП не було створено:

Отримані результати: Технології

Галузь застосування: J 63.11

Реєстраційний номер картки технології:

Опис НТП: На даний момент в Україні існує декілька баз патентів, які дозволяють здійснювати пошук патентів. Але в них є значні недоліки, такі як: не зручний інтерфейс, повільний час обробки запиту, обмеженні можливості сортування пошукової видачі. Розроблено систему програмних засобів для пошуку, аналізу та збереження для наступної обробки впорядженої консолідованої інформації у сфері патентної інформації. Така система має дати аналітику чітку інформацію про тематику заявки відповідно до цілей пошуку (пошуку плагіату, пошуку патентів на винаходи в суміжних областях, перевірки на предмет існування подібної заявки, визначення особливостей нового продукту, визначення інших сфер застосування нового продукту, пошуку винахідників або компаній, які отримали патенти на винаходи в тій же галузі, отримання інформації про приватних осіб, які мають патенти на схожі винаходи і т.ін.). Основними перевагами програмної системи засобів збору та обробки патентної інформації є: – можливість сортування пошукової видачі за всіма можливими параметрами; – можливість пошуку за синонімами ключових слів; – можливість повторного пошуку за ключовими словами; – накопичення інформації за результантами пошуку; – впорядження запиту по мірі релевантності за кількісною оцінкою результатів. Задачі, що вирішуються за допомогою даної системи: - аналіз існуючих систем пошуку патентної інформації; - підключення до існуючої патентної бази з метою отримання первинної патентної інформації; - вилучення патентної інформації за запитом; - аналіз вилученої інформації відповідно до запиту; - надання впорядженої інформації по запиту в зручному вигляді з можливістю аналізу конкретного патенту. Описана програмна система відрізняється від своїх аналогів тим, що крім виконання основного пошукового запиту, ведеться також пошук за синонімами ключових слів, вказаних в пошуковому запиті.

Соціально-економічна спрямованість НТП: Економія енергоресурсів, Підвищення продуктивності праці, Підвищення автоматизації виробничих процесів, Забезпечення промисловості чи населення новим видом інформаційно-комунікаційних послуг

Вплив НТП на довкілля:

Впровадження НТП: Впроваджено

Практична реалізація НТП

Початок етапу:

Закінчення етапу:

Споживачі продукції:

Перспективні ринки:

Характер співробітництва з інвестором

Потрібний обсяг інвестицій, тис. грн.:

Права, що надаються інвестору після завершення роботи:Права не надаються

Наявність бізнес-плану: Ні

Техніко-економічне обгрунтування: Ні

Потенціальний обсяг продажу, тис. грн.:

Очікуваний термін окупності (років):

Додаткова інформація:

IX. Бібліографічний опис

1. Черняк Л. Большие Данные — новая теория и практика (рус.) // Открытые системы. СУБД. — М.: Открытые системы, 2011. — № 10. — ISSN 1028-7493

2. Шаховська Н.Б. Методи опрацювання консолідованих даних за допомогою просторів даних / Н.Б. Шаховська // Пробл. програмув. — 2011. — № 4. — С. 72-84.

3. Растригин Л. А. Автоматная теория случайного поиска / Л. А. Растригин, К. К. Рипа. – Рига: Зинатне, 1973. – 344 с. – (ЛатАН).

4. Ухоботов В. И. Избранные главы теории нечетких множеств/ В. И Ухоботов // Учеб. пособие. – Челябинск : Изд-во Челяб. гос. ун-та, 2011. – 245 с.

5. В. О. Кузьміних. Стандартизація навчальних ресурсів на базі об'єктно-орієнтованого підходу / Д. В. Хаустов, В. О. Кузьміних, О. М. Шевченко // V (XXIX) Міжнародна міжвузівська школа семінар «Методи і засоби діагностики в техніці та соціумі (МіЗД ТС-201)» : Збірник матеріалів / Д. В. Хаустов, В. О. Кузьміних, О. М. Шевченко. – Івано-Франківськ: Видавництво «Факел» ІФНТУНГ, 2015. – С. 81–85.

6. Певцов Г.В. Анализ методов консолидации информации и особенностей их применения / Г.В. Певцов, Э.Г. Фастовский, М.А. Олейник // Вісник Національного технічного університету «ХПІ»: зб. наук. пр. Тематичний випуск: Інформатика і моделювання. – Х., 2007. – № 39. – с.145-153 с.

7. Кузьміних В. О. Система автоматичного геокодування інформаційних запитів / В. О. Кузьміних, О. С. Бойченко // V (XXIX) Міжнародна міжвузівська школа семінар «Методи і засоби діагностики в техніці та соціумі (МіЗД ТС-201)» : Збірник матеріалів / В. О. Кузьміних, О. С. Бойченко. – Івано-Франківськ: Видавництво «Факел» ІФНТУНГ, 2015. – С. 12–17.

8. Кузьміних В. О. Система автоматизованного геокодування запитів користувачів / В. О. Кузьміних, О. С. Бойченко // Екологічна безпека територіально-виробничих комплексів: енергетика, екологія, інформаційні технології / В. О. Кузьміних, О. С. Бойченко. – Київ: "МП Леся", 2015. – (НТУУ "КПІ"). – С. 217–222.

9. Koval O., Kuzminykh V., Khaustov D. Using stochastic automation for data consolidation // Research Bulletin of NTUU "KPI".Engineering. – 2017. – №2. – С. 29–36.

10. Schrenk M. Webbots, Spiders, and Screen Scrapers 2nd edition: A Guide to Developing Internet Agents with PHP/CURL. No Starch Press Inc. 2012. P. 362.

11. Бенчмарк HTML парсеров [Електронний ресурс] // Сайт «Хабрахабр» – Разработка. 2012. 26 декабря. URL: https://habrahabr.ru/post/163979/.

12. Морган К. [Morgan C.] XML для PHP-разработчиков: Часть 2. Расширенные методы парсинга XML: пер. с англ. [Електронний ресурс] // Сообщество developerWorks. 2010. 15 апреля. URL: http://www.ibm.com/developerworks/ru/library/x-xmlphp2/index.html.

13. Расти Хэролд Э. [Rusty Harold E.] Синтаксический анализ XML в PHP: пер. с англ. [Електронний ресурс] // Сообщество developerWorks. 2007. 11 октября. URL:

http://www.ibm.com/developerworks/ru/library/xpullparsingphp/index.html.

14. Key Marketing Trends for 2017 [electrinic resourse]. – Access mode: https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN

15. Introduction to Deep Neural Networks (Deep Learning) [electrinic resourse]. – Access mode: https://deeplearning4j.org/neuralnet-overview

16. Learning Deep Structured Semantic Models for Web Search using Clickthrough Data [electrinic resourse]. – Access mode: https://www.microsoft.com/en-us/research/publication/learning-deep-structured-semantic-models-for-web-search-usingclickthrough-data/

17. Люгер Ф. Дж. Искусственный интеллект: стратегии и методы решения сложных проблем. 4-е издание. Пер. с англ. – М.: «Вильямс», 2003. – 864с. ISBN 5-8459-0437-49 (рус.)

18. Слащева А. Патентознавство: навч. посібник для студ. Вищих навч. закл. / Донецький національний ун-т економіки і торгівлі ім. Михайла Туган-Барановського. - Донецьк: ДонНУЕТ, 2008. - 230с.

19. Брижко В. Патентознавство як самостійна наукова дисципліна / Національне агентство з питань інформатизації при Президентові України. — К., 1996. — 184с.

20. Єгоров І.Ю. Наука та інновації в процесах соціально-економічного розвитку: Монографія; Центр дослідні. наук.-техн. потенціалу та історії науки ім. Г.М. Доброва НАН України. - К., 2006. - С. 129-136

21. Кузнєцов Ю.М., Косенюк Г.В., Данильченко М.Г. Ін телек туальна власність. Навчальний посібник / За ред. проф. Кузнєцова Ю.М. — Тернопіль: Економічна думка, 2006. — 419 с., ISBN.

22. Google Developers [Електронний ресурс]. — Режим доступу: https://developers.google.com/maps/documentation/geocoding/?hl=ru

23. Tech Yandex [Електронний ресурс]. — Режим доступа: - https://tech.yandex.ru/maps/doc/jsapi/2.0/ref/reference/geocode-docpage/

24. MSDN Microsoft [Електронний ресурс]. — Режим доступа: https://msdn.microsoft.com/en-us/library/ff701713.aspx

25. Developer Yahoo [Електронний ресурс]. — Режим доступа: https://developer.yahoo.com/boss/geo/

26. Журкин И. Г., Шайтура С. В. Геоинформационные системы. — Москва: Кудиц-пресс, 2009. — 272 с. — ISBN 978-5-91136-065-8

27. Дравица В. Промышленная революция Indusry 4.0 / В. Дравица, А.Курбацкий // Наука и ин- новации. 2016, №3. С. 13–16. Мосалев, П.М. Обзор методов нечеткого поиска текстовой информации / П.М. Мосалев // Вестник Московского государственного университета печати. – 2013. – № 2. – С. 87-91.

28. Зеленков Ю.Г., Сегалович И.В. Сравнительный анализ методов определения нечетких дубликатов для WEB-документов // Труды 9-ой Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» RCDL'2007: Сб. работ участников конкурса, том 1. Переславль-Залесский, Россия: «Университет города Переславля», 2007. С. 166-174

29. Guha R. Semantic search / R. Guha, R. McCool, E. Miller // Proceedings of the 12th international conference on World Wide Web. – N.Y. ACM Press, 2003. – P. 700–709

30. Christopher D.M. Introduction to information retrieval / D.M. Christopher, R. Prabhaka, S. Hinrich. – Cambridge University press, 2008. – 504 p.

X. Заключні відомості

Керівник юридичної особи Письменний Євген Миколайович

Перелік осіб-виконавців

Кузьміних Валерій Олександрович

Тараненко Руслан Анатолійович
(05.13.03)

Відповідальний за підготовку облікових документів Кузьміних Валерій Олександрович

Телефон +38 (095) 012-66-63

Реєстратор

Керівник відділу УкрІНТЕІ, що є відповідальним за реєстрацію наукової діяльності

