

Методи та засоби автоматичної систематизації динамічного науково-освітнього ресурсу глобального електронного інформаційного простору

Методы и средства автоматической систематизации динамического научно-образовательного ресурса глобального электронного информационного пространства

Methods and means for the automatic dynamic scientific and educational resources ordering of the global electronic information environment

- 1. Номер державної реєстрації теми - 0115U000324,**
- 2. Науковий керівник -** д.т.н., проф. Тарасенко В.П., Тарасенко В.П., Tarasenko Volodymyr P.

3. Суть розробки, основні результати. **(укр.)**

В процесі виконання роботи було реалізовано принципово нову архітектуру програмних засобів автоматичної систематизації науково-освітнього сегменту глобального електронного інформаційного простору. Орієнтованість на науково-освітню галузь дозволила використати особливості відповідних даних, у першу чергу той факт, що їх зазвичай можна віднести до класу частково структурованих. Останнє призводить до можливості використання методів структурування, спрямованих на обробку даних цього класу, зокрема методів на базі відбору інформації (Information Extraction).

Також були створені методи та відповідні програмні засоби, які реалізують освітньо-орієнтований метод відбору інформації, що розглядає шукані ресурси з точки зору контекстно-атрибутивної моделі. Подібне подання дозволяє виділити з вхідних даних інформаційні зрізи та виконувати структурування множини документів як для кожного зрізу окремо, так і для їх сукупності з використанням методів Data Mining, Text Mining, Web Mining та Knowledge Discovery in Databases. Однією з важливих переваг такого підходу є можливість обробки гетерогенних даних, зокрема, текстових та мультимедійних (наприклад, відеокурсів, відеозаписів мастер-класів, науково-практичних конференцій тощо). Результатом структурування інформаційних ресурсів є метадані, які описують проаналізовану інформацію, та на основі яких відбувається налаштування пошукових засобів з метою тонкого підстроювання під освітні інформаційні потреби користувача.

Таким чином, завдяки реалізації повнофункціональної системи, яку розроблено в процесі роботи над проектом, та структурно-алгоритмічної організації програмних засобів надається можливість отримання, структурування, аналізу та обробки даних і надання кінцевому користувачеві високоефективних семантико-орієнтованих пошукових механізмів.

(рос.)

В процессе выполнения работы была реализована принципиально новая архитектура программных средств автоматической систематизации научно-образовательного сегмента глобального электронного информационного пространства. Ориентированность на научно-образовательную отрасль позволила использовать особенности соответствующих данных, в первую очередь тот факт, что их обычно можно отнести к классу частично структурированных. Последнее приводит к возможности использования методов структурирования, направленных на обработку данных этого класса, в частности методов на базе отбора информации (Information Extraction).

Также были созданы методы и соответствующие программные средства, которые реализуют образовательно-ориентированный метод отбора информации, рассматривают искомые ресурсы с точки зрения контекстно-атрибутивной модели. Подобное представление позволяет выделить из входных данных информационные срезы и выполнять структурирование множества документов как для каждого среза отдельно, так

и для их совокупности с использованием методов Data Mining, Text Mining, Web Mining и Knowledge Discovery in Databases. Одним из важных преимуществ такого подхода является возможность обработки гетерогенных данных, в частности, текстовмещающих и мультимедийных (например, видеокурсов, видеозаписей мастер-классов, научно-практических конференций и т.п.). Результатом структурирования информационных ресурсов является метаданные, описывающие проанализированную информацию и на основе которых производится настройка поисковых средств с целью тонкой подстройки под образовательные информационные потребности пользователя.

Таким образом, благодаря реализации полнофункциональной системы, разработанной в процессе работы над проектом, и структурно-алгоритмической организации программных средств предоставляется возможность получения, структурирования, анализа и обработки данных и предоставления конечному пользователю высокоэффективных семантико-ориентированных поисковых механизмов.

(англ.)

In the process of work was implemented fundamentally new software architecture and automatic systematization of scientific and educational segment of the global electronic information environment. Scientific and educational orientation allows using the features of the relevant data, especially the fact that they can usually be classified as semi-structured. This leads to the possibility of using structured methods aimed at processing the class, including methods based on Information Extraction.

There were established methods and related software tools that implement education-oriented method of information selection, considering the desired resources in terms of context-attribute model. This presentation reveals to extract the sections of input information and perform a structured set of documents for each slice individually and collectively for their using methods Data Mining, Text Mining, Web Mining and Knowledge Discovery in Databases. One of the important advantages of this approach is the ability to process heterogeneous data, including text and multimedia (video courses, video master classes, scientific conferences, etc.). The result of the information resources structuring is metadata that describes the analyzed information, and on which is tuned search engines for the purpose of fine tuning in education information needs of the user.

Thus, through the implementation of a fully functional system that developed in the project, and software structure and algorithms of the opportunity of obtaining, structuring, and analyzing data and providing highly efficient end-user-oriented semantic search engines.

4. Наявність охоронних документів на об'єкти права інтелектуальної власності.

- Патент на корисну модель № 100006 G06F 7/00. Багаторозрядний суматор зі змінним модулем з груповим переносом / Клятченко Я.М.; Роговенко А.І.; Тарасенко В.П.; Тесленко О.К.; Шепотіннік О.Ю.; НТУУ КПІ; заявл. 21.03.2014; опубл. 10.07.2015, бюл. №13.;
- Патент на винахід №110814. Бортовий спектрометр для виявлення уражених зон рослинності./ Яценко В. О.; Донець В.В.; Лапчук В.П.; Державний реєстр патентів України на винаходи 25.02.2016

5. Порівняння зі світовими аналогами

На фоні кращих світових аналогів програмне забезпечення, яке створено в процесі виконання роботи над проектом, відрізняється підвищеними показниками релевантності пошукових відгуків, високим рівнем оперативності отримання доступу до неструктурованого текстового ресурсу глобального інформаційного простору, а також розширеними можливостями щодо різнобічної аналітичної обробки інформаційних ресурсів, в першу чергу, завдяки систематизації формально та семантично зв'язаних текстових ресурсів науково-освітнього призначення на основі новітніх методів технологій Data Mining та Web Mining, включаючи реалізацію методів класифікації, кластеризації та асоціювання неструктурованої інформації.

6. Економічна привабливість для просування на ринок

Застосування розроблених технологій та обладнання дозволяє значно підвищити якість параметри інформаційно-пошукових систем:

- підвищення на 15 – 20% швидкодії засобів інтелектуального аналізу даних;
- зменшення на 10 – 15 % часу очікування пошукового запиту;
- підвищення релевантності та пертинентності результатів пошукових запитів.

7. Потенційні користувачі (галузі, міністерства, підприємства, організації)

Потенційними користувачами та замовниками розробок є освітні заклади, в першу чергу Міністерства освіти і науки України, навчальні підрозділи інших установ, підприємств та організацій різних форм власності, а також всі науково-дослідні установи НАН України та галузевого підпорядкування.

8. Стан готовності розробки

Розроблено готові для розгортання та експлуатації програмні засоби систематизації та пошуку даних в розподілених гетерогенних інформаційних ресурсах науково-освітнього спрямування.

9. Існуючі результати впровадження

Технологія плазмового різання керованим плазмохімічним струменем впроваджена на Херсонському суднобудівному підприємстві у виробництві заготовок обшивки корпусів і забезпечує високу продуктивність, чистоту різку та хімічний склад поверхні під наступне зварювання. Заплановано сумісне використання технології і пристроїв з ІЕЗ ім. Є.О. Патона НАН України та «Інститутом механізації та електрифікації сільського господарства».

10. Форма участі інвестора (яка краща форма участі в реалізації результатів проекту інвестора: частка в проекті%, частка від прибутку%, інше)

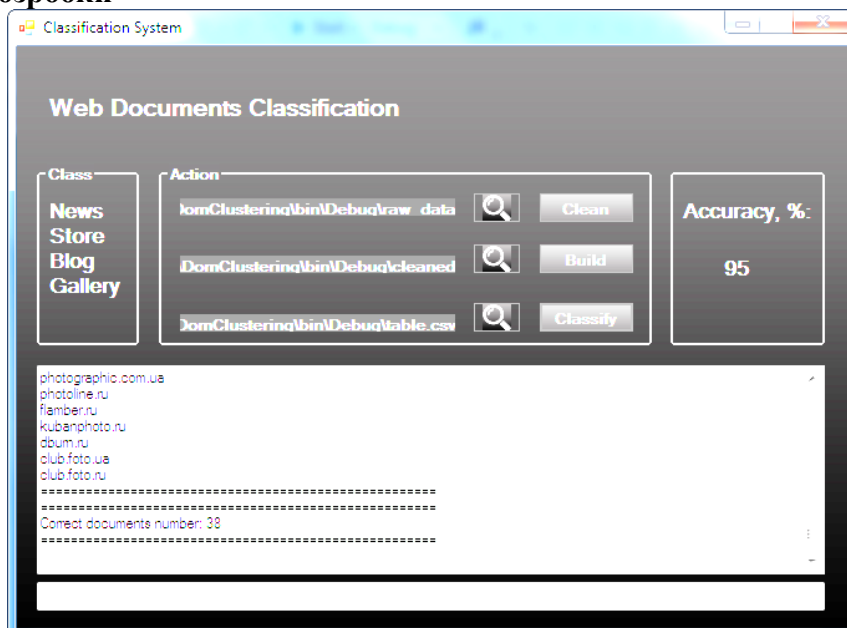
11. Обсяг інвестицій (необхідна для результатів проекту сума інвестицій в доларах США).

12. Мета інвестицій (розширення бізнесу, створення нового підприємства, інше).

13. Назва організації, телефон, E-mail

НТУУ"КПІ", факультет прикладної математики, кафедра системного програмування і спеціалізованих комп'ютерних систем,
(044) 236-32-02, spscs.kpi.ua@gmail.com

14. Фото розробки



15. Перелік публікацій за матеріалами досліджень за період виконання розробки

1. О. Корченко, І. Терейковський, А. Білощицький Методологія розроблення нейромережових засобів інформаційної безпеки Інтернет-орієнтованих інформаційних систем. К. ТОВ «Наш Формат». – 2016, 249 с.
2. Шитий Д.В. Спосіб резервного копіювання гостей операційних систем в Nureg-V / Д.В. Шитий // Комп'ютерно-інтегровані технології: освіта, наука, виробництво. Видавництво Луцького НТУ. – 2015. – № 20. – С. 97 – 100.
3. Михайлюк, А. Ю. Організація системи автоматичного реферування тексту на основі нечіткої ієрархічної нейронної мережі [Текст] / А. Ю. Михайлюк, О. С. Михайлюк, Л. М. Огнівчук, В. П. Тарасенко // Радіоелектронні і комп'ютерні системи. - 2016. - №7.-С. XX 20 стор.
4. Щербина О.А. Experience API – новий стандарт програмних засобів для е-навчання та приклади його практичного використання / О.А. Щербина // Інформаційні технології і засоби навчання, 2016. - Том 53, №3. - С. 150-163.

16. Надати ключові слова до розробки: інформаційний пошук, розподілені гетерогенні ресурси, інтелектуальний аналіз даних