

РОЗРОБКА ІНФОРМАЦІЙНО-ПОШУКОВОЇ ВЕБ-СИСТЕМИ З ВІДКРИТИХ ДАНИХ «ISODATA»

Вінницький національний технічний університет

Анотація

Запропоновано новий підхід до формалізації даних та створення інформаційно-пошукової веб-системи відкритих числових даних з прив'язкою одночасно у просторі, у часі та за ключовими словами на основі онтологічних моделей і семантичних мереж за географічною ознакою та змістовим описом. Продемонстровано працездатність розробленої моделі на прикладі створеної авторами інформаційної веб-системи відкритих даних «ISODATA».

Ключові слова: відкриті дані, онтологічна база даних, семантична мережа, інформаційно-пошукова веб-система, просторово-часові дані.

Abstract

A new approach for data formalization and to development of an information retrieval web-system of open numerical data has been proposed in the paper. This model provides data simultaneously linked in time, in space and on keywords based on ontological models and semantic networks based on geography and descriptions of content. There has been demonstrated the ability to work of the developed model for example the authors created the information web-system of open data «ISODATA».

Keywords: open data, ontological database, semantic network, information retrieval web-system, spatio-temporal data.

Для проведення системного аналізу в будь-якій сфері людської діяльності необхідно, перш за все, зібрати достовірну та актуальну інформацію про об'єкт дослідження, умови його функціонування, вплив на інші об'єкти тощо. В умовах сьогодення така інформація може бути в багатьох установах одночасно, різних веб-порталах та банках даних. Причому, для забезпечення ефективної роботи вона потрібна дослідникам у форматі відкритих даних [1, 2].

Отже, розробка спеціалізованої комплексної моделі відкритих числових даних (ВЧД) інформаційно-пошукової веб-системи, яка буде забезпечувати швидкий комплексний пошук цих даних одночасно за прив'язкою у часі, у просторі та по ключових словах, є актуальною. Метою дослідження є побудова та опис реалізації моделі такої веб-системи.

По-перше, запропоновано, для уникнення проблем з авторськими правами, не копіювати кожний набір ВЧД на окремий веб-ресурс, а для кожного такого набору зберігати тільки гіперпосилання на веб-ресурс, який безпосередньо містить ці ВЧД. По-друге, для забезпечення зручності пошуку, усі ВЧД формалізуються за трьома категоріями: «Час» (час, якого дані стосуються чи час, коли вони були зібрані або завантажені на веб-ресурс); «Простір» (просторові об'єкти, яких ці дані стосуються чи просторова прив'язка установ або організацій, які їх зібрали); «Ключові слова» для опису предметної галузі. По-третє, кожна з категорій формалізується окремо:

- пошук в часі з точністю до днів;
- формалізація просторових об'єктів та ключових слів здійснюється у вигляді окремих семантичних мереж.

Вчетверте, ці семантичні мережі пропонується будувати за принципом mesh-мереж, коли кожен вузол з'єднується з багатьма сусідніми і вже, через них, — з іншими вузлами [3]. Такий варіант топології комунікаційних мереж, як відомо, забезпечує один з найефективніших способів зв'язку [3].

Відповідно до запропонованого підходу авторами створено інформаційну веб-систему відкритих числових даних — інформаційну систему відкритих даних (ISODATA — англійською мовою: Information System for Open DATA) [4].

У цій веб-системі планується розміщувати та здійснювати пошук не лише відкритих числових даних, але у певний момент вона наповнюється тільки числовими даними, оскільки, як показав аналіз, саме вони мають проблеми із систематизацією і забезпеченням зручного і швидкого пошуку. Інтерфейс системи (рис. 1) дозволяє проводити пошук по заданих полях, пошук за обраним форматом, за заданим періодом часу (у днях) та за категорією «Простір» — за заданим регіоном на основі наявних в базі онтологій у вже сформованій семантичній мережі (Європа, Північна Америка, Південна Америка, Антарктида тощо). Також є можливість проводити пошук по ключових словах на основі наявних в базі онтологій у вже сформованих семантичних мережах різного змісту, наприклад: забруднювальні речовини, діоксид сірки, моніторинг довкілля, викиди, викиди забруднювальних речовин тощо [4].

Наприклад: Викиди забруднюючих речовин в атмосферне повітря

Пошук по заданих полях

Пошук по обраному формату

Параметри дати:

Виберіть потрібний регіон
 Європа
 Північна Америка
 Південна Америка
 Антарктида

Виберіть ключові слова
 забруднюючі речовини
 моніторинг довкілля
 викиди
 викиди забруднюючих речовин
 забруднювачі довкілля
 природно-заповідний фонд

Рис. 1. Інтерфейс пошукової системи

Результати пошуку:

Назва: Викиди забруднюючих речовин в атмосферне повітря
Власник: Державна служба статистики України
Дата: 15.06.2016 по 15.06.2017
Географічна приналежність: Європа>Україна
Ключові слова: викиди>забруднюючі речовини>атмосферне повітря
Опис: Дані, зібрані протягом останніх років по всіх регіонах України
Посилання: <http://data.gov.ua/passport/61a562db-30b5-4bf5-988e-037797523fd9> [Не відкривається?](#)

Назва: Викиди забруднюючих речовин та діоксиду вуглецю в атмосферне повітря
Власник: Державна служба статистики України
Дата: 15.06.2016 по 15.06.2017
Географічна приналежність: Європа>Україна
Ключові слова: викиди забруднюючих речовин>стаціонарними джерелами забруднення>пересувними джерелами забруднення>викиди діоксиду
Опис: Дані, зібрані за останні декілька років
Посилання: <http://data.gov.ua/passport/ce595398-2aa0-47da-864c-15746bc40e17> [Не відкривається?](#)

Рис. 2. Приклади результатів пошуку в режимі перегляду

id	Назва	Статус посилання	Дата створення	Редагування
50	Викиди забруднюючих речовин в атмосферне повітря	1	13.12.2016	Edit Delete
51	Перелік територій та об'єктів природно-заповідного фонду загальнодержавного та місцевого значення в розрізі адміністративно-територіальних одиниць	1	13.12.2016	Edit Delete
49	Перелік об'єктів, які є найбільшими забруднювачами довкілля	0	13.12.2016	Edit Delete
48	Викиди забруднюючих речовин та діоксиду вуглецю в атмосферне повітря	1	13.12.2016	Edit Delete
47	Викиди забруднюючих речовин в атмосферне повітря	0	13.12.2016	Edit Delete

Рис. 3. Приклад результатів пошуку в режимі редагування

Висновки

Розглянуто проблему формалізації та ідентифікації інформаційних моделей відкритих числових даних будь-якого призначення для забезпечення їх швидкого комплексного пошуку одночасно за прив'язкою у часі, у просторі та за ключовими словами. Запропоновано новий підхід для такої формалізації, який дозволяє досить швидко проектувати таку систему та забезпечувати за нею ефективний та релевантний пошук відкритих числових даних. Внесено пропозицію будувати семантичні мережі за принципом топології mesh-мереж. Продемонстровано працездатність розробленої моделі та підходів до формалізації відкритих числових даних на прикладі створеної авторами інформаційної веб-системи відкритих даних «ISODATA».

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Олексюк Л. Механізм правового забезпечення електронного доступу до публічної інформації та відкритих даних ; стан і тенденції розвитку / Ліля Олексюк / Національна академія державного управління при Президентіві України // Державне управління та місцеве самоврядування. — 2016. — Вип. 3. — С. 92—99.
2. Юдін О. К. Особливості адміністрування та менеджменту в публічному управлінні державними інформаційними ресурсами / О. К. Юдін, С. С. Бучик, Р. В. Зюбіна, Ю. С. Авраменко // Наукоємні технології. — № 4 (32). — 2016. — С. 395—404
3. Akyildiz F. A survey on wireless mesh networks / F. Akyildiz, Xudong Wang // IEEE Communications Magazine. — Vol. 43. — No. 9. — P. S23—S30. — Sept. 2005. — doi: 10.1109/MCOM.2005.1509968.
4. Мокін В. Б. Розробка комплексної моделі інформаційно-пошукової веб-системи з відкритих числових даних / В. Б. Мокін, С. О. Довгополюк, М. П. Боцула, М. Р. Коханський // Вісник ВПІ. — 2017. — № 1. — С. 62-69.

Мокін Віталій Борисович – доктор технічних наук, професор, завідувач кафедри системного аналізу, комп'ютерного моніторингу та інженерної графіки Вінницького національного технічного університету, Вінниця, e-mail: vbmokin@gmail.com;

Довгополюк Сергій Олександрович – аспірант кафедри системного аналізу, комп'ютерного моніторингу та інженерної графіки факультету комп'ютерних систем і автоматики Вінницького національного технічного університету, Вінниця;

Коханський Максим Русланович — студент факультету комп'ютерних систем і автоматики, Вінницький національний технічний університет, Вінниця;

Боцула Мирослав Павлович — канд. техн. наук, доцент, доцент кафедри системного аналізу, комп'ютерного моніторингу та інженерної графіки Вінницького національного технічного університету, Вінниця.

Vitalii B. Mokin – Prof., Dr Hab. (Eng.), Head of the Department of Systems Analysis, Computer Monitoring and Engineering Graphics, Vinnytsia National Technical University, Vinnytsia, e-mail: vbmokin@gmail.com

Sergiy O. Dovgopoluk – Post-Graduate Student of the Department of Systems Analysis, Computer Monitoring and Engineering Graphics of the Faculty of Computer Systems and Automatics, Vinnytsia National Technical University, Vinnytsia;

Kokhanskyi R. Maksym — Master Student of the Faculty of Computer Systems and Automatics, Vinnytsia National Technical University, Vinnytsia;

Botsula P. Myroslav — Cand. Sc. (Eng.), Assistant Professor, Assistant Professor of the Department of Systems Analysis, Computer Monitoring and Engineering Graphics, Vinnytsia National Technical University, Vinnytsia.