

УДК 004.9

В. Б. Мокін¹
С. О. Довгополюк¹
М. П. Боцула¹
М. В. Коханський¹

РОЗРОБКА КОМПЛЕКСНОЇ МОДЕЛІ ІНФОРМАЦІЙНО-ПОШУКОВОЇ ВЕБ-СИСТЕМИ ВІДКРИТИХ ЧИСЛОВИХ ДАНИХ

¹Вінницький національний технічний університет

Запропоновано нову комплексну модель 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, яка дозволяє досить швидко проектувати таку модель та забезпечувати за нею ефективний та релевантний пошук відкритих числових даних. Побудову семантичних мереж запропоновано здійснювати за принципом топології mesh-мереж, що забезпечує достатньо швидке проектування розгалуженої семантичної мережі. Продемонстровано працездатність розробленої моделі та підходів до формалізації відкритих числових даних 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.

Вступ та вихідні передумови

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].

Згідно з Законом України від 09.04.2015 № 319-VIII «Про внесення змін до деяких законів України щодо доступу до публічної інформації у формі відкритих даних» (ст. 10): «Публічна інформація у формі відкритих даних — це публічна інформація у форматі, що дозволяє її автоматизоване оброблення електронними засобами, вільний та безоплатний доступ до неї, а також її подальше використання. Будь-яка особа може вільно копіювати, публікувати, поширювати, використовувати, у тому числі в комерційних цілях, у поєднанні з іншою інформацією або шляхом включення до складу власного продукту, публічну інформацію у формі відкритих даних з обов'язковим посиланням на джерело отримання такої інформації».

Особливий інтерес дослідників пов'язаний, в першу чергу, з відкритими числовими даними (ВЧД), які найчастіше зберігаються у форматах csv, xls та ін. Іноді вони зберігаються у форматі баз даних mdb, dbase IV, dbf та ін. Також, вони можуть бути в універсальних форматах збереження структурованих даних: xml, JSON та ін.

Існує велика кількість порталів з відкритими даними, у т.ч. числовими [1, 2]. Наприклад, у більшості країн світу є національні портали даних. Наприклад, в Україні таким є Єдиний державний веб-портал відкритих даних: <http://data.gov.ua/>. Однак, хоча за його статистикою станом на 13.02.2017 він містив більше 11 тисяч наборів даних, за твердженням заступника голови Державного агентства з питань електронного урядування України О. Вискуба «не дивлячись на значні позитивні кроки, головними викликами залишаються погана якість даних, закритість багатьох пріоритетних наборів, а також відсутність чіткого розуміння які дані взагалі накопичуються органами влади».

Аналіз показав, що подібні портали мають такі основні недоліки: по-перше, не завжди здійснюється перевірка працездатності і достовірності цих даних; по-друге, переважна кількість цих даних не структурована у форматі числових відкритих даних, наведених вище; по-третє, є складнощі з формалізацією цих відкритих даних — зазвичай вони формалізуються тільки по ключових словах та загальних параметрах типу назви і власника даних. Є ускладненим пошук ВЧД за заданий період часу, для заданих просторових об'єктів по певних ключових словах та у заданому форматі даних.

У роботах [3, 4] запропоновано систематизацію, формалізацію та структурування різних рівнів онтологічних баз даних та семантичних мереж на їх основі для опису просторових об'єктів, але систематизація відкритих даних на основі тільки просторових чи ключових даних є недостатньою.

У світі для створення таких систем використовуються різні технології. Наприклад, в ЄС є спеціальна директива INSPIRE, яка, за даними однойменного порталу <http://inspire.ec.europa.eu/>, спрямована на створення інфраструктури просторових даних країн-членів ЄС для цілей екологічної політики і політики або діяльності ЄС, які можуть вплинути на навколишнє середовище [5—8].

Однак, формалізація відкритих як екологічних, так і інших даних за усіма вимогами INSPIRE є надлишковою. Варто вибрати модель, яка забезпечувала б достатньо широкий пошук даних (і у часі, і у просторі, і за змістом ключових слів), і водночас була б проста у використанні.

Для усунення цих недоліків доцільно створити спеціалізовану комплексну модель відкритих числових даних інформаційно-пошукової веб-системи. Однак, для забезпечення ефективності її функціонування спершу потрібно побудувати її інформаційну модель.

Отже, розробка спеціалізованої комплексної моделі відкритих числових даних 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.

Ідея інформаційної технології побудови системи

Ідея запропонованої технології така.

По-перше для уникнення проблем з авторськими правами, варто 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:

1. «Час» — усі такі дані мають часову прив'язку — чи час, якого вони стосуються (дані моніторингу) чи час, коли вони були зібрані або завантажені на веб-ресурс (різні довідкові і кадастрові дані та ін.);

2. «Простір» — усі дані мають просторову прив'язку — чи просторові об'єкти, яких вони стосуються (дані про просторові об'єкти) чи просторова прив'язка установ або організацій, які їх зібрали (управління, інститути, підприємства тощо);

3. «Ключові слова» для опису предметної галузі — складові довікля, сектори економіки, об'єкти моніторингу, типи задач обробки даних — будь-які слова, за якими користувачі можуть здійснювати пошук даних, але такі, які не характеризують прив'язку набору даних у часі та у просторі.

По-третє, кожна з категорій формалізується окремо. Щоб уніфікувати пошук в часі, пропонується його формалізувати з точністю до днів, тобто 2016 рік описується як «01.01.2016—31.12.2016», січень 2017 — як «01.01.2017—31.01.2017» і т. п. Формалізацію просторових об'єктів та ключових слів пропонується здійснювати у вигляді окремих семантичних мереж. В термінах робіт [3, 4] семантичні мережі для категорій «Простір» та «Ключові слова» утворюють окремі термінополя. Отже, характеристики кожного набору ВЧД будуть формалізуватись як перетин відповідного часового інтервалу та термінополів категорій «Простір» та «Ключові слова».

Вчетверте, семантичні мережі пропонується будувати за принципом mesh-мереж, коли кожен вузол з'єднується з багатьма сусідніми і вже, через них, — з іншими вузлами [8]. Такий варіант топології комунікаційних мереж, як відомо, забезпечує один з найефективніших способів зв'язку [8]. У такий же спосіб пропонується формувати прив'язку кожного ВЧД до семантичних мереж «Простір» та «Ключові слова»:

— за категорією «Простір» — формується набір назв, а потім вказуються просторові об'єкти вищого рівня, які включають повністю чи частково координати цих об'єктів; наприклад, для даних про концентрацію СО в атмосферному повітрі м. Вінниця — це «м. Вінниця» ∈ «Вінницька область»;

— за категорією «Ключові слова» — формується набір ключових слів, а потім вказуються ключові слова вищого рівня, зміст яких включає повністю чи частково зміст цих слів; наприклад, для прикладу вище: «СО» ∈ «парникові гази», «атмосферне повітря» ∈ «довкілля».

Ключові слова чи просторові об'єкти вищого рівня, у свою чергу, будуть мати ще вищий рівень — у такий спосіб з часом автоматично створюються відповідні термінополя по різних предметних галузях.

Вп'яте, пошук організовується одночасно за заданим часовим інтервалом, просторовою прив'язкою та ключовими словами. При цьому, варто зробити можливим задавати повний збіг умов (щоб знайти точно те, що треба) чи частковий збіг (щоб знайти хоча б щось, якщо точний пошук не дав результату), хоча б на один день, хоча з якимись координатами (наприклад, Вінницьку область шукати у басейні р. Південний Буг і навпаки) та хоча б з одним ключовим словом семантичної мережі.

Технологічно час варто задавати з використанням інструментарію для вибору дат типу «Календар», простір — з використанням ГІС-технологій, які дозволяють вибрати заданий шар просторових об'єктів у вибраній точці. А ключові слова — з використанням зручної візуалізації усїєї семантичної мережі або її окремих слів, які можна вибирати зі списків, що випадають, з поступовим їх уточненням, коли перше слово вибирається довільно, а друге та третє — вже тільки з тих слів, які хоча б через якісь слова пов'язані з першим. Хоча варто передбачати і варіант вибору декількох таких перших слів, кожне з яких відноситься до різного термінополя.

Крім вищезазначених категорій, варто кожен ВЧД описувати й іншими характеристиками: назва, власник даних, дата завантаження тощо.

Формалізація моделей різних категорій даних

Побудуємо моделі даних, формалізованих за запропонованою технологією, для деякого i -го набору ВЧД.

Оскільки набір ВЧД може характеризувати декілька часових інтервалів одночасно, наприклад, теплі пори року протягом $w = 20$ років — це буде 20 періодів часу. Отже, пропонуємо формалізувати час як:

$$T_i = [T_{i11}, T_{i12}] \cup [T_{i21}, T_{i22}] \cup [T_{i31}, T_{i32}] \cup \dots \cup [T_{i w1}, T_{i w2}], \quad (1)$$

де T_{ij1} та T_{ij2} — початковий та кінцевий моменти j -го періоду ($j = 1, 2, \dots, w$) i -го набору ВЧД.

Категорії «Простір» та «Ключові слова» пропонуємо формалізувати як семантичні мережі, утворені онтологіями. Як відомо, онтологія O_i описується трійкою [4]:

$$O_i = \langle X, R, F \rangle, \quad (2)$$

де X, R, F — кінцеві множини, відповідно: X — концептів (понять, термінів), які характеризують набір ВЧД за категоріями «Простір» або «Ключові слова», R — відношень між концептами окремої категорії («Простір» або «Ключові слова»), F — функцій інтерпретації X та/або R у певній інформаційно-аналітичній системі.

Особливо варто зупинитись на формалізації найпоширеніших відношень між онтологіями, які можуть бути між характеристиками різних наборів ВЧД та їх прив'язкою у просторі. Відношення $r \in R$ між двома концептами у просторі можуть бути такими:

$$r_1 = \{ \langle I_1 \rangle \cap \langle I_2 \rangle = \emptyset \}; \quad (3)$$

$$r_2 = \{ \langle I_1 \rangle \cap \langle I_2 \rangle \neq \emptyset \}; \quad (4)$$

$$r_3 = \{ \langle I_1 \rangle \cup \langle I_2 \rangle = \langle I_1 \rangle \}, \quad (5)$$

де I_j — множина координат усіх точок, що описує j -й просторовий об'єкт, якого стосується заданий набір ВЧД, r_1 — відношення «немає збігу», r_2 — відношення «є спільні точки», r_3 — відношення «включає» (тобто «один об'єкт розташований на території іншого» чи «один об'єкт включає інший»).

Аналогічні до (3)—(5) відношення можна сформулювати і щодо змісту, який охоплює концепт, не координат у просторі: r_2 — відношення «один об'єкт частково змістовно подібний до іншого» та r_3 — відношення «включає».

Пропонуємо аналогічно (2) позначати онтології прив'язки O_P у просторі:

$$O_P = \langle X_P, R, F_P \rangle \quad (6)$$

та онтології O_K по ключових словах:

$$O_K = \langle X_K, R, F_K \rangle. \quad (7)$$

Отже, побудову онтологічної бази знань для категорій «Простір» або «Ключові слова» пропонується здійснювати за таким алгоритмом для кожного i -го набору ВЧД:

1. Визначається прив'язка в часі за моделлю (1) і зберігається у базі даних, наприклад $T_i = [15.04.2016, 15.10.2016]$.

2. Визначається прив'язка у просторі у вигляді назв просторових об'єктів у кількості N_{Pi} , наприклад ($N_{Pi} = 3$): $x_{Pi1} = \langle \text{«Вінниця»}$, $x_{Pi2} = \langle \text{«Транспортна мережа»}$, $x_{Pi3} = \langle \text{«басейн р. Південний Буг»}$.

3. Для кожного концепту прив'язки у просторі визначаються концепти із загальної множини X_P та визначаються відношення між ними з множини R . Важливо відмітити, що, з урахуванням викладеного у попередньому розділі про те, що пропонується будувати онтологічну базу знань за аналогією з принципом побудови mesh-мереж, відношення слід визначати і формалізувати тільки парні, тобто між кожним x_{Pij} ($j = 1, 2, \dots, N_{Pi}$) та іншим із множини X_P .

4. Задається інтерпретація із множини F_P .

5. Визначаються ключові слова для опису предметної області цього набору ВЧД у кількості N_{Ki} , наприклад ($N_{Ki} = 3$): $x_{Ki1} = \langle \text{«Парникові гази»}$, $x_{Ki2} = \langle \text{«Атмосферне повітря»}$, $x_{Ki3} = \langle \text{«Довкілля»}$.

6. Для кожного концепту прив'язки у просторі визначаються концепти із загальної множини X_K та визначаються відношення між ними із множини R . Важливо зазначити, що з урахуванням пропозиції будувати онтологічну базу знань за аналогією з принципом побудови mesh-мереж, відношення слід визначати і формалізувати тільки парні, тобто між кожним x_{Kij} ($j = 1, 2, \dots, N_{Ki}$) та іншим із множини X_K .

7. Задається інтерпретація із множини F_K .

Отже, вирази (1), (3)—(7) є комплексною моделлю інформаційно-пошукової веб-системи відкритих числових даних. Комплексність зумовлена по-перше, формалізацією і у часі, і у просторі, і за змістом, а по-друге, поєднанням онтологічних моделей для формалізації за категоріями «Простір» та «Ключові слова» з інтервальною моделлю для формалізації в часі.

Визначивши у такий спосіб онтології для наборів ВЧД за категоріями «Простір» та «Ключові слова», з часом вдасться сформувати семантичні мережі по різних пов'язаних між собою онтологіях, наприклад, формалізувати басейни річок різного рівня за басейновим принципом, адміністративно-територіальні одиниці — за адміністративним та ін.

Пошук наборів ВЧД здійснюватиметься за таким алгоритмом:

1. Задається часовий інтервал, просторова прив'язка та ключові слова, які описують набори ВЧД, що необхідно знайти.

2. Здійснюється пошук наборів ВЧД у часі, відповідно до заданого відношення: «збіг хоча б на один день» або «збіг повністю усіх днів».

3. Для знайдених за п. 2 наборів ВЧД здійснюється пошук у просторі відповідно до заданого відношення: «включає» (тобто «один об'єкт розташований на території іншого») або «є спільні точки».

4. Для знайдених за п. 3 наборів ВЧД здійснюється пошук за змістом по ключових словах відповідно до заданого відношення: «включає» або «один об'єкт частково змістовно подібний до іншого».

5. Якщо нічого не знайдено, тоді варто перейти до п. 1 і уточнити запит.

Охарактеризуємо практичну реалізацію запропонованої інформаційної технології.

Приклад реалізації запропонованої інформаційної технології

Відповідно до запропонованої інформаційної моделі та алгоритмів її застосування авторами створено інформаційну веб-систему числових відкритих даних — інформаційну систему відкритих даних (ISODATA — англійською мовою: Information System for Open DATA). У цій веб-системі планується розмішувати та здійснювати пошук не лише числових відкритих 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.

Оскільки система знаходиться на стадії наповнення, то в ній поки передбачена формалізація відношень онтологічних концептів тільки за виразом (4) («є спільні точки» для категорії «Простір» та «один об'єкт частково змістовно подібний до іншого» для категорії «Ключові слова»). Отже,

система побудована на основі моделі (1), (4), (6), (7). Крім того, параметрами наборів ВЧД є не тільки прив'язка у часі, просторі та ключові слова за змістом, а ще вибрано такі: назва даних (за замовчуванням: «Набір даних»); веб-адреса джерела даних; власник даних; період часу, за який ці дані, у форматі з ____ до ____ (тип «Стислий формат дати»); опис (текст до 65535 символів); формат даних (текст з списку, що випадає) та статус посилання (1 — пряме посилання на набір ВЧД або 0 — посилання на сайт, де їх можна скачати у форматі ВЧД, але — вручну, а не програмним шляхом).

Зокрема, 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.

Рис. 1. Інтерфейс пошукової системи

Є два режими функціонування (окрім системного адміністрування): редагування (для зареєстрованих користувачів) (рис. 2) та перегляду (для всіх хто бажає).

викид				
id ▼	Назва	Статус посилання ▼	Дата створення	Редагування
51	Перелік територій та об'єктів природно-заповідного фонду загальнодержавного та місцевого значення в розрізі адміністративно-територіальних одиниць	1	13.12.2016	Edit Delete
50	Викиди забруднюючих речовин в атмосферне повітря	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. 2. Приклад результатів пошуку в режимі редагування

У режимі редагування в результаті пошуку виводиться назва набору ВЧД, його порядковий номер, статус посилання та дата внесення його параметрів у цю систему. При цьому, надається можливість редагування параметрів інформації або видалення цього набору ВЧД, якщо, наприклад, перевірка показала, що гіперпосилання не дійсне. А в режимі перегляду в результаті пошуку виводиться вікно з основними параметрами знайдених наборів ВЧД (рис. 3). Системний адміністратор може керувати списком полів, які виводяться в результаті пошуку.

Подальший розвиток системи планується в напрямку використання ГІС-технологій для формалізації та вибору просторових об'єктів безпосередньо з карти, а не зі списків, що випадають, їх назв, як це показано на рис. 1.

Результати пошуку:

Назва: Викиди забруднюючих речовин в атмосферне повітря

Власник: Міністерство екології та природних ресурсів України

Дата: 01.01.1990 по 31.12.2013

Географічна приналежність: Європа>Україна

Ключові слова: забруднюючі речовини>діоксид сірки; забруднюючі речовини>оксид азоту; забруднюючі речовини>аміак; забруднюючі речовини>тверді частки; забруднюючі речовини>оксид вуглецю; забруднюючі речовини>неметанові леткі органічні сполуки; забруднюючі речовини>стійкі органічні сполуки; забруднюючі речовини>важкі метали

Опис: Даний показник надає уявлення про ступінь наявного та очікуваного антропогенного тиску викидів забруднюючих речовин на навколишнє середовище, дозволяє визначити ступінь досягнення цільових значень. Оцінювання впливу на навколишнє природне середовище окремих секторів економіки, стаціонарних та пересувних джерел викидів. Показник включає дані щодо викидів окремих забруднюючих речовин: діоксиду сірки (SO₂), оксидів азоту (NO_x), аміаку (NH₃), твердих часток (ТЧ10, ТЧ2,5 таабо загального вмісту зважених часток (ЗЗЧ)), оксиду вуглецю (СО), неметанових летких органічних сполук (НМЛОС), стійких органічних сполук (СОЗ, у тому числі поліхлорбіфеніли (ПХБ), діоксини/фурани та поліциклічні ароматичні вуглеводні (ПАВ)), важких металів (кадмію, свинцю і ртуті). В структурі РС-Т-С-В-Р показник «Викиди забруднюючих речовин у атмосферне повітря» (Т) має безпосередній зв'язок з показником «Якість атмосферного повітря в міських населених пунктах» (С). Вимірюється в тис. т, т, або кг відповідної забруднюючої речовини за рік. Для міждержавних зіставлень показник виражається величиною викидів на км² території країни, на одну особу населення або на одиницю валового внутрішнього продукту (ВВП).

Посилання: <http://www.menr.gov.ua/ekologichni-pokaznyky-monitorynhu/3829-pokaznyk-a-1> Не відкривається?

Рис. 3. Приклади результатів пошуку в режимі перегляду

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 нову комплексну модель інформаційно-пошукової веб-системи відкритих числових даних з прив'язкою одночасно у часі, у просторі та за ключовими словами на основі онтологічних моделей і семантичних мереж за географічною ознакою та змістовим описом, яка дозволяє досить швидко проектувати таку модель та забезпечувати за нею ефективний та релевантний пошук відкритих числових даних. Зазначено, що достатньо швидке проектування розгалуженої семантичної мережі можна здійснювати за принципом топології mesh-мереж. Розроблено алгоритми ідентифікації запропонованих інформаційних моделей на практиці та їх використання для пошуку релевантних числових відкритих даних.

Продемонстровано працездатність розробленої моделі та підходів до формалізації відкритих числових даних на прикладі створеної авторами інформаційної веб-системи відкритих даних «ISODATA». Охарактеризовано її інтерфейс та особливості функціонування.

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 Президентові 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. Стрижак О. Є. Формування таксономій шарів карт в ГІС-середовищах на основі онтологій натуральних систем / О. Є. Стрижак, М. А. Попова // Радіоелектронні і комп'ютерні системи. — 2013. — № 4 (63). — С. 46—54.

4. Стрижак О. Є. Онтологічні інформаційно-аналітичні системи / Олександр Євгенійович Стрижак // Радіоелектронні і комп'ютерні системи. — 2014. — № 3. — С. 71—76.
5. The Importance of Open Data and Software for Large Scale Hydrological Modelling [Electronic resource] / [L. Strömbäck, B. Arheimer, G. Lindström, C. Donnelly and others] // Open Water. — Vol. 2. — Iss. 1. — Article 32. — Access mode: <http://scholarsarchive.byu.edu/openwater/vol2/iss1/32>.
6. Frederika W. D. How to assess the success of the open data ecosystem / Frederika Welle Donker, Bastiaan van Loenen // International Journal of Digital Earth. — Vol. 10. — Iss. 3. — 2017. — P. 284—306. — doi: 10.1080/17538947.2016.1224938.
7. van Loenen B. INSPIRE Empowers Re-Use of Public Sector Information / B. van Loenen, M. Grothe // International Journal of Spatial Data Infrastructures Research. — 2014. — No. 9. — P. 86—106. — doi:10.2902/1725-0463.2014.09.art4.
8. 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.

Рекомендована кафедрою системного аналізу, комп'ютерного моніторингу та інженерної графіки ВНТУ

Стаття надійшла до редакції 15.02.2017

Мокін Віталій Борисович — д-р техн. наук, професор, завідувач кафедри системного аналізу, комп'ютерного моніторингу та інженерної графіки;

Довгополюк Сергій Олександрович — аспірант кафедри системного аналізу, комп'ютерного моніторингу та інженерної графіки, e-mail: isergeyq@gmail.com ;

Боцула Мирослав Павлович — канд. техн. наук, доцент, доцент кафедри системного аналізу, комп'ютерного моніторингу та інженерної графіки;

Коханський Максим Русланович — студент Інституту екологічної безпеки та моніторингу довкілля.
Вінницький національний технічний університет, Вінниця

V. B. Mokin¹
S. O. Dovhopoliuk¹
M. P. Botsula¹
M. R. Kokhanskyi¹

Development of the Integrated Model of Information Retrieval Web-System of Open Numerical Data

¹Vinnitsia National Technical University

A new integrated model of the 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 that allows quickly design such a model and provide its effective and relevant search for open numeric data. Building a semantic network has been proposed to implement by the principle of mesh-network topology that provides plenty of quick designing an extensive semantic network. There has been demonstrated the ability to work of the developed model and approaches to formalize open numerical data on the created by the authors information retrieval web-system of open data «ISODATA».

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

Mokin Vitalii B. — Dr. Sc. (Eng.), Professor, Head of the Chair of Systems Analysis, Computer Monitoring and Engineering Graphics;

Dovhopoliuk Serhii O. — Post-Graduate Student of the Chair of Systems Analysis, Computer Monitoring and Engineering Graphics, e-mail: isergeyq@gmail.com ;

Botsula Myroslav P. — Cand. Sc. (Eng.), Assistant Professor, Assistant Professor of the Chair of Systems Analysis, Computer Monitoring and Engineering Graphics;

Kokhanskyi Maksym R. — Master Student of the Chair of Systems Analysis, Computer Monitoring and Engineering Graphics

В. Б. Мокин¹
С. А. Довгополук¹
М. П. Боцула¹
М. В. Коханский¹

Разработка комплексной модели информационно-поисковой веб-системы открытых числовых данных

¹Вінницький національний технічний університет

Предложена новая комплексная модель информационно-поисковой веб-системы открытых числовых данных с привязкой одновременно во времени, в пространстве и по ключевым словам на основе онтологических моделей и семантических сетей по географическому признаку и содержательному описанию, позволяющая достаточно быстро проектировать такую модель и обеспечивать по ней эффективный и релевантный поиск открытых числовых данных. Построение семантических сетей предложено осуществлять по принципу топологии mesh-сетей, обеспечивающего достаточно быстрое проектирование разветвленной семантической сети. Продемонстрирована работоспособность разработанной модели и подходов к формализации открытых числовых данных на примере созданной авторами информационной веб-системы открытых данных «ISODATA».

Ключевые слова: открытые данные, онтологическая база данных, семантическая сеть, информационно-поисковая веб-система, пространственно-временные данные.

Мокин Виталий Борисович — д-р техн. наук, профессор, заведующий кафедрой системного анализа, компьютерного мониторинга и инженерной графики;

Довгополук Сергей Александрович — аспирант кафедры системного анализа, компьютерного мониторинга и инженерной графики, e-mail: isergeyq@gmail.com;

Боцула Мирослав Павлович — канд. техн. наук, доцент, доцент кафедры системного анализа, компьютерного мониторинга и инженерной графики;

Коханский Максим Русланович — студент Института экологической безопасности и мониторинга окружающей среды