

РОЗРОБКА МОДЕЛІ СТРУКТУРНО-ЛОГІЧНОГО ПРЕДСТАВЛЕННЯ ДАНИХ САЙТУ ВИЩОГО НАВЧАЛЬНОГО ЗАКЛАДУ НА ОСНОВІ ІЄРАРХІЧНОГО КЛАСИФІКАТОРА

У роботі на підставі проведеного аналізу визначено та обґрунтовано модель структурно-логічного представлення даних сайту на основі ієрархічного класифікатора. Показано яким чином запропонована модель дозволяє автоматизувати такі функції супроводження сайту, як: виявлення і усунення суперечностей в сторінках сайту; виявлення і видалення застарілої інформації на сторінках сайту з урахуванням специфіки інформації, що міститься на сторінках сайту, адже з одного боку, до змісту інформації, що міститься на сайті висуваються вимоги достовірності, актуальності, несуперечливості, цілісності, а з іншого боку, з моменту створення сайту на ньому збирається великий обсяг інформації, який важко контролювати та не допускати ситуацій з дублюванням даних.

Ключові слова: модель структурно-логічного представлення даних, ієрархічний класифікатор, супроводження сайту.

Вступ та постановка проблеми. Останнім часом в мережі Інтернет сайти різних організацій і фізичних осіб набули такого ж розповсюдження, що і мобільний зв'язок. Разом із стрімким зростанням кількості сайтів збільшується й обсяг накопичуваної інформації на сайті, що призводить до проблем ефективного функціонування сайту.

З одного боку, до змісту інформації, що міститься на сайті висуваються вимоги достовірності, актуальності, несуперечливості, цілісності. З іншого боку, з моменту створення сайту на ньому збирається великий обсяг інформації: додаються нові сторінки, які інколи дублюють за змістом вже існуючі сторінки; вже створені сторінки підлягають оновленню із заміною застарілих даних новою інформацією.

Власне за якість інформації відповідає системний адміністратор сайту, в обов'язки якого входить перевірка інформації (даних) на коректність своєчасного оновлення інформації, видалення (або архівація) застарілої інформації. В той же час, із-за великої кількості сторінок (обсяг обліковується кількома сотнями сторінок), часто стає неможливим прослідкувати, які сторінки оновлювалися, а які – ні. Причина цього криється за рахунок додавання блоків нової інформації або ж корегуванні помилок на сторінці тим самим формально унеможливаючи визначення застарілої інформації з використанням полів «дата останнього редагування». На подібних сторінках може бути повністю відсутні також часові маркери «дати», «числа», «року» тощо.

Таким чином, ефективне функціонування сайту безпосередньо залежить від стану інформаційно-аналітичного забезпечення системи підтримки та ведення сайту, який, в свою чергу, залежить від логічної організації даних сайту. Отже, актуальною є задача побудови моделі структурно-логічної схеми представлення даних сайту, яка б задовольняла вимогам інформаційно-аналітичного опрацювання з боку системного адміністратора сайту.

Метою роботи є обґрунтування й розробка структурно-логічної моделі представлення даних сайту, яка б задовольняла вимогам їх подальшого автоматизованого (автоматичного) інформаційно-аналітичного опрацювання системним адміністратором сайту.

Аналіз досліджень та публікацій. Особливістю представлення інформації на сайтах є те, що дані у переважній більшості представлені у вигляді слабо структурованої інформації. Слабка структурованість проявляється в тому, що інформації притаманні ті ж самі властивості, що й природній мові, а саме: надлишковість (коли за формою подання різні вирази мають однаковий сенс), суперечливість і недостатність (які у формальному плані проявляються

неоднозначно). Отже, розглянемо логіко-лінгвістичні моделі представлення знань про предметну область (в нашому випадку зміст сайту).

До семантичного подання знань про предметну область (ПрО) відносять: семантичні мережі та їх різновиди, фрейми, модель універсального семантичного коду (УСК). До логічного подання відносять продукційні та предикатні моделі.

Семантичні мережі та фрейми складають клас реляційних моделей подання знань [1]. В основі реляційних моделей виступають бінарні відношення подання знань. Мова формалізованого подання знань має загальні закономірності для всього класу реляційних моделей. Для моделювання знань про навколишній світ (ПрО) виділяють декілька класів елементів, які подаються через лінгвістичні (слова та словосполучення) одиниці, які виступають як базові поняття і відношення опису певної ПрО. Так, виділяють імена, відношення, поняття. Поняття поділяються на поняття-класи, поняття-процеси і поняття-стани. Поняття-клас - це сукупність конкретних об'єктів та предметів, які мають визначені властивості (*генератор шуму, стіл тощо*). Поняття-процеси описують групу однорідних процесів (*навантаження, ушкодження тощо*). Поняття-стан визначає стан об'єкта (*нормальний режим, лінія ввімкнена тощо*). Імена слугують для ідентифікації понять (літак *АН-64А*, фірма *"Макдоннел гелекоптер"* тощо). Відношення слугують для встановлення зв'язків на множині понять. В деяких джерелах [1] виділяється до 200 простих (базових) відношень. Найбільш часто вживані з цих відношень наведені в таблиці 1.

Ця таблиця розглядає не мовні відношення, а відношення фізичного світу. Те чи інше відношення в мові може мати різні описи. Так, наприклад, в наступних різних фразах: *"Деталь переміщується по конвеєру до кінцевого бункеру"*, *"Робот йде до складу №4"*, *"Автомобіль наближається до перехрестя"* - реалізується одне відношення R_{31} з табл. 1. Для виводу правильно побудованих формул (ППФ) вводиться сукупність синтаксичних правил.

Таблиця 1

Фрагмент класифікації відношень в предметній області

Тип відношення	Найменування відношення	Позначення відношення
Часові	Бути одночасно	R_{11}
	Бути раніше	R_{12}
Просторові	Знаходитись в <input checked="" type="checkbox"/> -оточенні	R_{21}
	Знаходитись	R_{22}
	Знаходитись позаду	R_{23}
	...	R_{2r}
Динамічні	Рухатися до	R_{31}
	...	R_{3n}
Класифікаційні	Належати до класу	R_{41}
	Мати (властивості)	R_{42}
	...	R_{4k}
Ідентифікуючі	Мати ім'я	\square
Прагматичні	Слугувати для	R_{51}
	Мати стан	R_{52}
	Бути перепоною	R_{53}
	...	

Клас реляційних моделей має дуже широкий спектр їх різновидів.

Семантичні мережі набули активного розповсюдження в системах штучного інтелекту для автоматизації задач управління. Семантична мережа в загальному випадку представляється як: $C = \langle x_1, x_2, \dots, x_s, r_1, r_2, \dots, r_q \rangle$, де x_1, x_2, \dots, x_s – фіксовані множини; r_1, r_2, \dots, r_q – система відношень, визначена на елементах цих множин. Існує доволі багато різних концепцій побудови семантичних мереж, автори яких намагаються як можна повніше

передати формалізоване подання знань про ПрО. Прикладами широко поширених семантичних мереж є мова концептуальних залежностей.

Синтагматичні ланцюги і RX -коди є окремими випадками семантичних мереж. Трійку виду $(X r Y)$, де X і Y – коди понять, r – код відношення, прийнято називати елементарною синтагмою. Формули, які складаються з елементарних синтагм та зв'язків між ними (логічні операції), називають синтагматичними ланцюгами. Над такими ланцюгами вводять алгебру, яка здійснює певні перетворення над ланцюгами. RX -коди фіксують бінарні відношення R між X_i і Y_j . За допомогою RX -кодів будується статистичне подання властивостей об'єктів в ПрО.

Окремий клас реляційних моделей складає фреймове подання [2]. Фрейми бувають структурні та ролеві. Між ролевими фреймами і RX -кодами можна провести аналогію. Так, будь-який RX -код можна розглядати як ролевий фрейм. Вони мають розгалужену ієрархічну структуру, в якій вузли вищого шару відповідають більш загальним поняттям. Поняття в кожному вузлі визначається набором атрибутів, які називаються слотами. Слоти можуть бути атрибутивні і процедурні. Кожний слот пов'язаний з відповідною процедурою обробки, коли міняється інформація в слоті. З кожним слотом можна пов'язати певну кількість процедур. Ці процедури можуть слідувати за приписуванням інформації відповідному вузлу та контролювати, які дії мають бути виконаними при зміні цієї інформації. Ці можливості фреймів корисні в тих предметних областях, де форма подання та зміст даних відіграють важливу роль. Фреймова процедура аналізу використовується в ряді систем, в тому числі і для аналізу природно-мовного тексту. Але фрейм є занадто великою одиницею змісту, що не дозволяє відбити всі особливості ПМТ.

Універсальний семантичний код розроблений Мартиновим [3]. Елементарна структура УСК представляється трійкою: $(S A O)$, де S – підмет в реченні, A – присудок, O – пряме доповнення. Правила перетворення природно-мовного речення до УСК полягають у виділенні елементарних конструкцій у реченні. Над елементарними структурами вводиться алгебра (складання синтагм, інтегрування, розкладання).

Особливе місце в семантичному поданні знань представляють ієрархічні моделі. Структура понять в ПрО переважно ієрархічна за своєю природою. Це зумовило розвиток ієрархічних моделей подання знань. Ієрархічні моделі представляють дерево підпорядкування. На самому верхньому рівні є лише один вузол, який називається корінь. Кожний вузол, крім коріння, пов'язаний з одним вузлом на вищому рівні, який називається вхідним для даного вузла. Жоден елемент не має більш одного вхідного вузла, кожний елемент може бути пов'язаним з одним або кількома елементами на нижньому рівні. Принциповим для ієрархічного подання є те, що кожний запис має своє значення лише в контексті дерева. Підпорядкований елемент не може існувати без свого попередника в ієрархії. Ієрархічні моделі поширені при складанні тезаурусів з ПрО.

Більшість предикатних мов базується на численні предикатів першого порядку [4]. Використовуються багато місцеві та одно місцеві предикати, логічні зв'язки і квантори. Знання подаються у вигляді формул логіки предикатів. Приєднання формул до отриманих раніше дозволяє отримати нові твердження про об'єкти за допомогою правил виводу вірогідних формул в численні предикатів першого порядку. Ця процедура інтерпретується як логічний вивід.

Серед недоліків треба зазначити, що числення предикатів не має достатнього набору засобів для формалізованого подання знань, які містяться в природно-мовних текстах.

Основна частина. Аналіз логіко-лінгвістичних моделей показав, що для організації даних сайту, зокрема, сайту вищого навчального закладу найбільше підходить ієрархічна модель [5], оскільки такі моделі дозволяють автоматично виявляти парадигматичні відношення на сайті, що дуже важливо для заявлених задач з інформаційно-аналітичного опрацювання даних сайту. Графічна інтерпретація запропонованого класифікатору представлена на рис. 1.

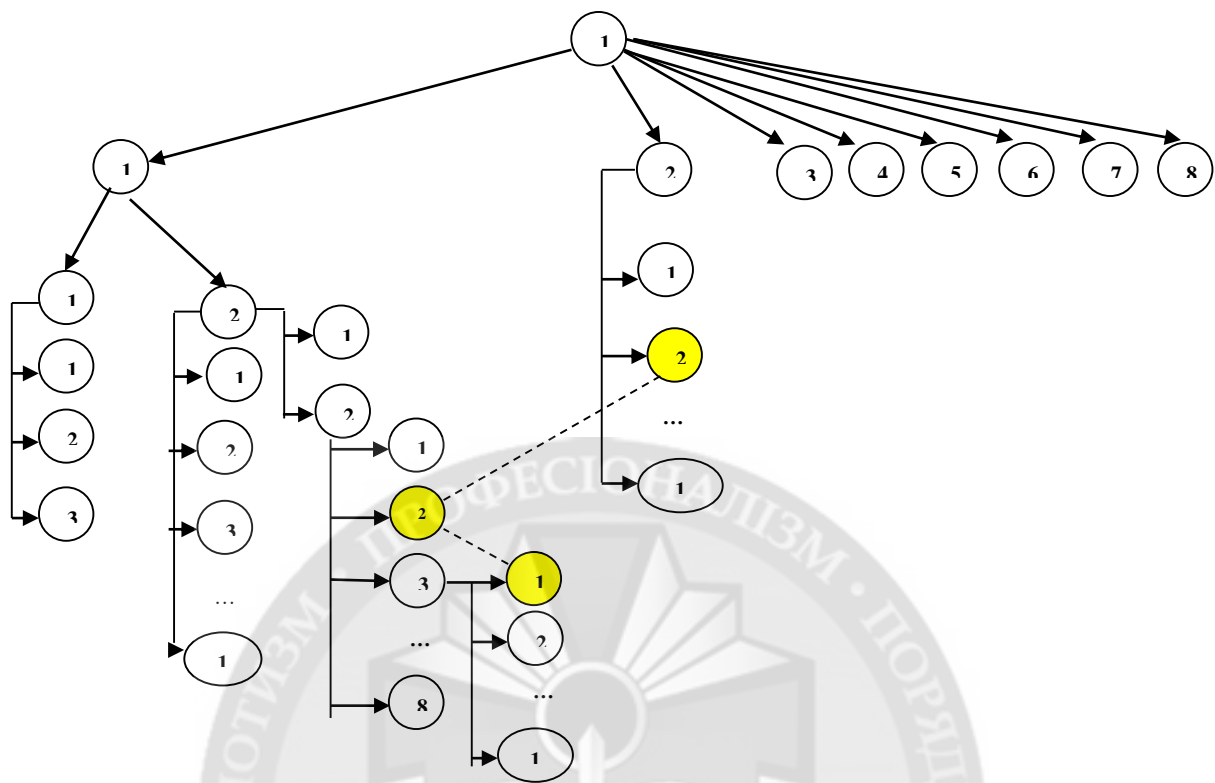


Рис. 1. Графічна інтерпретація ієрархічного класифікатору сайту

Кожен вузол відповідає певній сторінці сайту та має унікальний ключ, який вказує на місце сторінки в ієрархічній системі [6-9]. Так самий верхній вузол має значення: 1. Головна сторінка, на другому рівні ієрархії нумерація до порядкового номера вузла обов'язково додається номер вузла, що знаходиться вище. Так, наприклад, на рис. 1 це сторінки, які визначені в розділах головної сторінки сайту:

- 1.1. Про університет
- 1.2. Вступнику
- 1.3. Навчання
- 1.4. Наука
- 1.5. Студентське життя
- 1.6. Сертифіковані курси
- 1.7. Новини
- 1.8. Фотогалерея

Крапкою виокремлюється рівень ієрархії. В межах підпорядкування певному вузлу всі підпорядковані вузли мають наскрізну нумерацію. Таким чином термінальні вузли у своєму ключі включають всі номери вузлів, які виступають у ролі власника для відповідного термінального вузла.

Така систематизація сторінок надає можливість автоматизувати завдання з інформаційно-аналітичного опрацювання змісту інформації сайту. Так перевірка інформації на суперечливість здійснюється як по вертикалі, так і по горизонталі. По вертикалі перевірка здійснюється зверху донизу. Тобто інформація, яка знаходиться в вузлах, які розташовані нижче має відповідати інформації в вузлах, які розташовані вище. Слід зазначити, що системний

адміністратор не відповідає за достовірність інформації на сторінці сайту, але він може перевірити на суперечливість інформацію, що знаходиться на різних сторінках. При цьому перевіряється лише та інформація, яка є спільною для двох сторінок. Так, наприклад, вузол 1.1.2.2.2. на рис.1 відповідає сторінці «спеціальності факультету», а вузол 1.1.2.2.3.1. – сторінці «загальна інформація кафедри факультету». На сторінці кафедри може повторюватися інформація про спеціальності, за якими іде підготовка на кафедрі та факультеті. При цьому, якщо назви спеціальностей не співпадають, то достовірною вважається та, що знаходиться на вищому рівні. В свою чергу, інформація також має перевірятися з основною інформацією. В даному випадку формально це перевірка по горизонталі. Так вузол 1.1.2.2.2. перевіряється з вузлом 1.2.2. з рис. 1, який відповідає сторінці: «Перелік спеціальностей до вступу» у розділі «Приймальна комісія».

Крім того, така класифікація дозволяє в автоматичному режимі виставляти так звані часові маркери на актуальність інформації в залежності від специфіки сторінки, на якій вона представлена. Так, інформація відділу аспірантури про конференції, семінари, публікації тощо порівнюється з поточною датою, і застаріла інформація автоматично видаляється, інформація про вартість навчання визначається маркерами роботи приймальної комісії, а новини кафедри, факультету, ВНЗ визначаються по ключових словах відповідних сторінок за подію, яка відбулася. Крім того, такі ж маркери можна виставити і для фотогалереї, що дозволить архівувати дані часовими показниками, тим самим не перенавантажувати сайт.

Висновки. Таким чином, запропонована модель структурно-логічної організації даних сайту ВНЗ на основі ієрархічного класифікатора дозволяє автоматизувати такі функції супроводження сайту, як: виявлення і усунення суперечностей в сторінках сайту; виявлення і видалення застарілої інформації на сторінках сайту з урахуванням специфіки інформації, що міститься на сторінках сайту.

ЛІТЕРАТУРА:

1. Поспелов Д.А. Логико-лингвистические модели в системах управления: монографія / Д.А. Поспелов. – М., 1981. - 232 с.
2. Шемакин Ю.И. Тезаурус в автоматизированных системах управления и обработки информации: монографія / Ю.И. Шемакин. – М., 1974. – 192 с.
3. Мартинов В.В. Универсальный семантический код: монографія / В.В. Мартинов. – К., 1977. - 190 с.
4. Мински М. Фреймы для представления знаний: монографія / М. Мински. – М., 1979. - 324с.
5. Онищенко В.В. Алгоритмы і структури даних: Навчальний посібник / В.В. Онищенко, Р.С. Коник. – К., 2017. – 48 с.
6. Замаруєва І.В. Автоматизація аналізу змісту природно-мовних текстів як шлях забезпечення безпеки прийняття управлінських рішень / І.В. Замаруєва, О.В. Барабаш, І.В. Пампуха // Наукові записки Українського науково-дослідного інституту зв'язку: науково-виробничий збірник. – К.: УНДІЗ, 2017. – № 3(47). – 33 – 41 с.
7. Соколов В. Ю. Інформаційні системи і технології: Навчальний посібник / В. Ю. Соколов. – К., 2010. – 20 с.
8. Секаева Ж.А. Тези доповідей дев'ятої міжнародної науково-технічної конференції «Проблеми інформатизації». – К., 2017. – 9 с.
9. Гайна Г.А. Основи проектування баз даних: Навчальний посібник / Г.А. Гайна. – К., 2005. – 19 с.
10. Буйницька О. П. Інформаційні технології та технічні засоби навчання: Навчальний посібник / О. П. Буйницька. – К., 2012. – 13 с.

REFERENCES:

1. Pospelov D.A. (1981). Logiko-lingvisticheskie modeli v sistemah upravlenija [Logico-linguistic models in control systems], Moscow, 232 p.
2. Shemakin Ju.I. (1974). Tezaurus v avtomatizirovannyh sistemah upravlenija i obrabotki informacii [Thesaurus in automated control and information processing systems], USSR, 192 p.
3. Martynov V.V. (1977). Universal'nyj semanticheskij kod [Universal Semantic Code], Minsk, 190 p.

4. Minski M. (1979). Frejmy dlja predstavlenija znaniј [Frames for representating knowledge], Moscow, 324p.
5. Onishhenko V.V., Konik R.S. (2017). «Algoritmi i strukturi danih» [Algorithms and data structures], Kyiv, 48 p.
6. Zamarueva I.V., Barabash O.V., Pampuha I.V. Avtomatizacija analizu zmistu prirodno-movnih tekstiv jak shljah zabezpechennja bezpeki prijnattja upravlins'kih rishen'[Automation of the analysis of the content of natural-language texts as a way to ensure the security of the adoption of managerial decisions] Naukovi zapiski Ukraїns'kogo naukovo-doslidnogo institutu zv'jazku: naukovo-virobnichij zbirnik: UNDIZ[Scientific notes of the Ukrainian Research Institute of Communication: scientific and production collection], 2017, no. 3(47), pp. 33–41.
7. Sokolov V.Ju.. (2010). Informacijni sistemi i tehnologii: Navchal'nij posibnik [Information Systems and Technologies: Tutorial], Kyiv, 20 p.
8. Zh.A. Sekaeva. (2017). Problemi informatizacii tezi dopovidej dev'jatoї mizhnarodnoї naukovo-tehnichnoї konferencii [Problems of informatization of abstracts of reports of the ninth international scientific and technical conference], Kyiv, 9 p.
9. Gajna G.A.. (2005). Osnovi proektuvannja baz danih [Basics of database design], Kyiv, 19 p.
10. Bujnic'ka O.P.. (2012). Informacijni tehnologii ta tehnicni zasobi navchannja [Information technology and technical means of training], Kyiv, 13 p.

Без рецензії.

Aronov A.O.

DEVELOPMENT OF MODEL OF STRUCTURAL-LOGICAL REPRESENTATION OF SITE DATA OF THE HIGHER EDUCATIONAL SITE ON THE BASIS OF THE HIERARCHICAL CLASSIFICATION

In the work on the basis of the analysis, the model of structural and logical representation of site data based on the hierarchical classifier is determined and substantiated. It is shown how the offered model allows to automate such functions of maintenance of a site, as: detection and elimination of contradictions in pages of a site; detection and removal of outdated information on the pages of the site, taking into account the specifics of the information contained on the pages of the site. Because on the one hand, to the contents of the information contained on the site put forward the requirements of reliability, relevance, consistency, integrity, and on the other hand, since the creation of the site, the large amount of information is fill and it is difficult to control and avoid situations with duplication of data.

Keywords: model of structural-logical representation of data, hierarchical classifier, maintenance of a site.

Аронов А.А.

РАЗРАБОТКА МОДЕЛИ СТРУКТУРНО-ЛОГИЧЕСКОГО ПРЕДСТАВЛЕНИЯ ДАННЫХ САЙТА ВЫСШЕГО УЧЕБНОГО ЗАВЕДЕНИЯ НА ОСНОВЕ ИЕРАРХИЧЕСКОГО КЛАССИФИКАТОРА

В работе на основании проведенного анализа определена и обоснована модель структурно-логического представления данных сайта на основе иерархического классификатора. Показано каким образом предложенная модель позволяет автоматизировать такие функции сопровождения сайта, как: выявление и устранение противоречий в страницах сайта; обнаружения и удаления устаревшей информации на страницах сайта с учетом специфики информации, содержащейся на страницах сайта, ведь с одной стороны, к содержанию информации, содержащейся на сайте выдвигаются требования достоверности, актуальности, непротиворечивости, целостности, а с другой стороны, с момента создания сайта на нем собирается большой объем информации, который трудно контролировать и не допускать ситуаций с дублированием данных.

Ключевые слова: модель структурно-логического представления данных, иерархический классификатор, сопровождение сайта.