

САМООРГАНІЗАЦІЯ ЗНАНЬ В КВАЗІІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ НАВЧАННЯ ПІД ВПЛИВОМ ЗОВНІШНІХ ЧИННИКІВ

Данчук В.Д., доктор фізико-математичних наук

Лемешко Ю.С.

Лемешко Т.А.

Постановка проблеми. В сучасному суспільстві вже звичним явищем стала постійна зміна вимог ринку праці (РП) до знань фахівців. Як один з ефективних методів підтримки належного рівня професійних знань широко використовуються електронні системи навчання (СН). Їх мета полягає в постачанні користувачу необхідного контенту на будь-який термінал, що має доступ до серверу СН, незалежно від місцезнаходження в зручній формі і в потрібний час. Проте, в базах знань таких систем може накопичуватись навчальний матеріал (НМ) схожого змісту, але різної актуальності або, навіть, цілком застарілий. Користувач, що вибирає відповідний НМ, або викладач, що керує навчанням в СН, не завжди можуть визначити, який з варіантів подання НМ однакового призначення найбільше відповідає вимогам кон'юнктури РП для даної спеціальності або повинні витратити на це певний час. Така ситуація може створювати невідповідність між знаннями, які отримав користувач в процесі навчання, і знаннями, які повинен мати фахівець, щоб відповідати вимогам РП. Як наслідок, ефективність навчання буде незадовільною, а результати не відповідатимуть поставленим цілям.

Аналіз публікацій та постановка задачі. Більшість інтелектуальних СН (ІСН), що пропонують адаптивні навчальні курси (АНК), різним чином пристосовують процес навчання під потреби користувача у відповідності до параметрів його портрету. Серед технологій адаптивного навчання в ІСН широко застосовуються наступні [1]:

– адаптація контенту: додаткові, попередні і порівняльні пояснення; варіанти пояснень; сортування контенту;

– адаптивна навігація: пряме керівництво; адаптивне сортування посилань; адаптивне приховування посилань; адаптивна нотація посилань; адаптація карт;

– адаптивна фільтрація інформації: фільтрація на основі вмісту; колективна фільтрація.

Проте, як правило, ІСН, що враховують досвід навчання і параметри портрету користувача, не враховують зміну вимог РП, попит на специфічні знання зокрема і самі професії вцілому [2].

Нова парадигма безперервної освіти вимагає всебічної підтримки навчального процесу – інтелектуального аналізу не лише накопичених даних, що забезпечує адаптивність навчання, а й застосування технологій інженерії знань для формування найбільш актуального НМ [3].

Мета роботи. Для актуалізації НМ в ІСН необхідно розробити концепцію, методику і принципи самоорганізації системи знань відповідно до вимог зовнішніх чинників впливу.

Основна частина. В Національному транспортному університеті протягом п'яти років розробляються наукові основи та принципи функціонування квазіінтелектуальної СН (КСН) [4], побудованої на принципах самоорганізації навчального контенту та адаптивності процесу навчання.

В даній роботі розглянуто параметри, зміна яких забезпечує самоорганізацію системи знань в КСН. Параметри зовнішнього впливу РП безпосередньо впливають на структуру КСН і НМ. Вони частково відображають відповідність поточній кон'юктурі РП, а також певну прогнозованість потреб РП на найближче майбутнє. До них можна віднести:

а) статистичні. Визначаються з офіційних і неофіційних джерел статистики. До них, зокрема, належать дані про безробіття та зайнятість населення, структуровані у відповідності до національного класифікатору професій Держспоживстандарту України ДК 003:2005 [5-6]. Вказану інформацію можна отримати з інформаційних представництв компетентних державних органів: Міністерства праці та соціальної політики України (www.mlsp.gov.ua), Державного комітету статистики України (www.ukrstat.gov.ua) та Державної служби зайнятості України (www.dcz.gov.ua), яка представлена у всіх регіонах України і надає послуги з пошуку роботи і працівників (trud.gov.ua). Обсяги щорічного державного замовлення у розрізі регіонів, ВНЗ та спеціальностей доступні на інтернет-сторінці інформаційно-пошукової системи "Конкурс" (www.vstup.info) та на сайті Міністерства освіти і науки України (www.mon.gov.ua/education/higher/dz_specialist/, www.mon.gov.ua/education/higher/dz_magistr/);

б) експертні. Визначаються на основі прямих експертних опитувань, анкетувань і голосувань користувачів КСН. Об'єктивність отриманих результатів безпосередньо залежить від методу експертної оцінки, що використовується, кількості експертів і їх рівня компетенції;

в) рейтингові. Визначаються на основі методик ранжування НМ, які враховують його актуальність і інтенсивність використання;

г) пошукові. Визначаються шляхом автоматизованого збору інформації роботами-аналізаторами зі спеціалізованих інформаційних баз даних, наприклад, сайтів пошуку роботи. Обмін інформацією може здійснюватись, як і у випадку збору статистичних даних, через експортні файли.

Розглянемо детальніше процедуру обміну статистичними параметрами зовнішнього впливу.

На даному етапі збір статистичної інформації і періодичне внесення її до КСН здійснюється в ручному режимі. В перспективі функції збору статистичних даних можуть бути перекладені на окремий портал-каталогізатор – хронологічний накопичувач специфічної інформації.

Реальним прикладом такої реалізації з кінця квітня 2009 року може бути один з групи урядових порталів США, а саме Data.gov, який розроблено в рамках концепції електронного уряду Gov 2.0. Він надає доступ до інформаційних баз, що містять: федеральні дані в машиночитаемих форматах, їх метаописання для здійснення ефективного пошуку, інформацію про спосіб доступу і інструментарій для доступу до інформаційних баз [7].

До задач “сховища”, яке може використовуватись різними КСН, має входити акумулювання експертних висновків з певних предметних областей, пошук і накопичення певних даних роботами-аналізаторами. Даний портал, назвемо його системою статистичних даних (ССД), повинен мати можливість обміну цими даними універсальним способом. В сучасних каталогах це реалізується за допомогою обміну експортними файлами наперед визначеної структури у різних форматах.

Схема взаємодії ССД і КСН через обмін експортними файлами представлена на рис. 1.

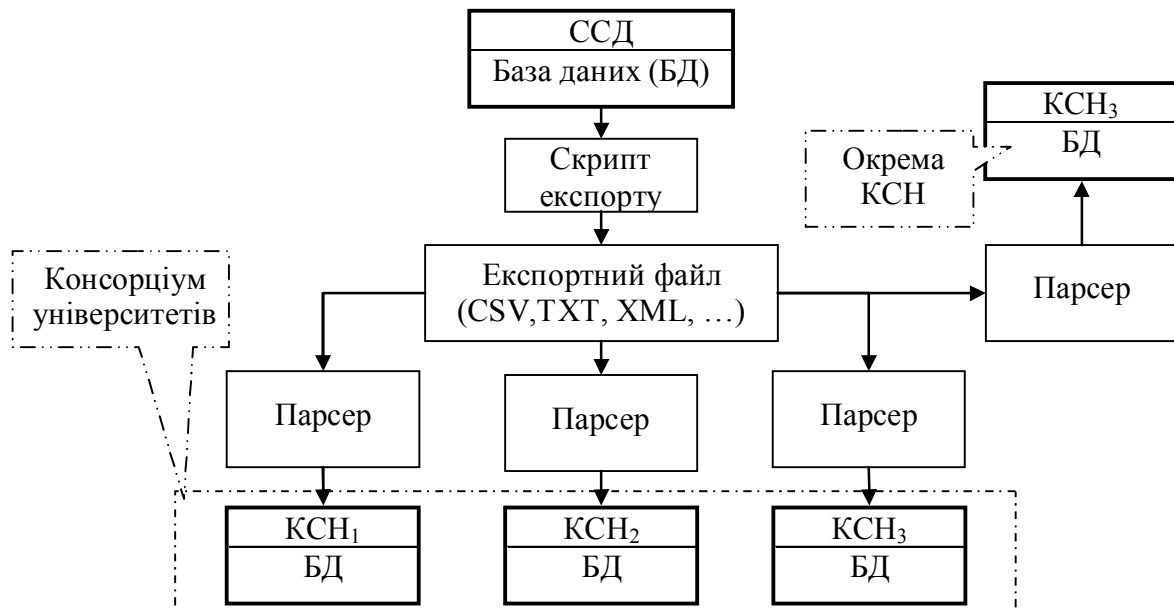


Рис. 1. Схема взаємодії ССД і КСН через обмін експортними файлами

Розглянемо детальніше методику і особливості розрахунку рейтингових параметрів.

В розробленій КСН, серед інших, використовується принцип ранжування НМ за його рейтингом, для чого введено параметр рейтингу НМ кластеру clusterrank (CR). Його мета полягає у представленні найбільш актуального НМ, подібно тому, як це роблять сучасні пошукові системи. Наприклад, пошукова система Google використовує незалежний від параметрів запиту алгоритм розрахунку авторитету (ваги) сторінки pagerank (PR) як один із чинників для визначення рейтингу пошукового матеріалу [8]. За клієнтським запитом видається та частина інформації, яка відповідає параметрам запиту в порядку зменшення PR. Формула визначення ваги PR сторінки A, що використовувалась на початковому етапі розвитку Google, як дослідницького проекту [9], має вигляд:

$$PR(A) = (1-d) + d(PR(T_1)/C(T_1) + \dots + PR(T_n)/C(T_n)), \quad (1)$$

де $d=0,85$ – коефіцієнт затухання, ймовірність того, що користувач слідуватиме за посиланням (хоча б одним зі всієї сукупності посилань) на поточній сторінці A;

$(1-d)$ – ймовірність так званого “випадкового стрибку” – користувач виходить з поточної

сторінки А не за посиланням на цій сторінці, а будь-яким іншим чином;

$T_1 \dots T_n$ – сторінки, що посилаються на сторінку А;

$PR(T_1) \dots PR(T_n)$ – вага PR відповідних сторінок $T_1 \dots T_n$, що посилається на сторінку А;

$C(T_1) \dots C(T_n)$ – кількість вихідних посилань на відповідних сторінках $T_1 \dots T_n$;

$PR(T_n)/C(T_n)$ – вираз показує повтор операції для кожної сторінки, яка вказує на сторінку А.

По аналогії з пошуковою системою, в КСН весь НМ у вигляді навчальних субкластерів (НСК) ранжується за визначеними критеріями і отримує певний рейтинг. Кластери НМ, які складаються з групи НСК і мають свій рейтинг, при формуванні навчальних планів (НП) підготовки фахівців кожного напрямку, що (НП) відповідають наперед визначеним критеріям (вимогам РП), потрапляють до тих чи інших навчальних курсів (НК). Для КСН у відповідності з (1) і [10] визначимо рейтинг $CR(X)$ певного кластеру X :

$$CR(X) = \frac{\varepsilon}{n} + (1 - \varepsilon) + \sum_{(X, Y_h) \in C} \frac{CR(Y_h)}{N(Y_h)}, \quad (2)$$

де $\varepsilon = 0,85$ – коефіцієнт затухання (константа);

n – кількість кластерів в масиві;

C – масив, що включає кластери X і Y_h ;

$CR(Y_h)$ – рейтинг кластеру Y_h ;

$N(Y_h)$ – кількість НСК в кластері Y_h , які використовуються (дублюються) в інших кластерах (в теорії графів, півстепінь виходу Y_h).

Варто зазначити, що PR є параметром, який відноситься до кожної окремої сторінки, а не до всього сайту цілому, тому на одному сайті можуть знаходитися сторінки з різним значенням PR. Так само, в КСН визначається рейтинг НМ кластерів різного рівня, а не лише НСК чи кластерів НК.

З (1) видно, що рейтинг сторінки сайту тим вищий, чим більша кількість і вищий рейтинг (краща якість) сторінок, які посилаються на дану сторінку. Аналогічно, з (2) маємо, що в КСН рейтинг НМ кластеру тим вищий, чим більша кількість і вищий рейтинг НМ інших кластерів, які містять в якості посилань його НСК (рис. 2). Крім того, частина рейтингу CR кластеру вищого порядку передається його субкластеру (вважається, що зв'язок між ними відбувається у відповідності до ієрархічної структури через вихідне посилання від кластеру до субкластеру).

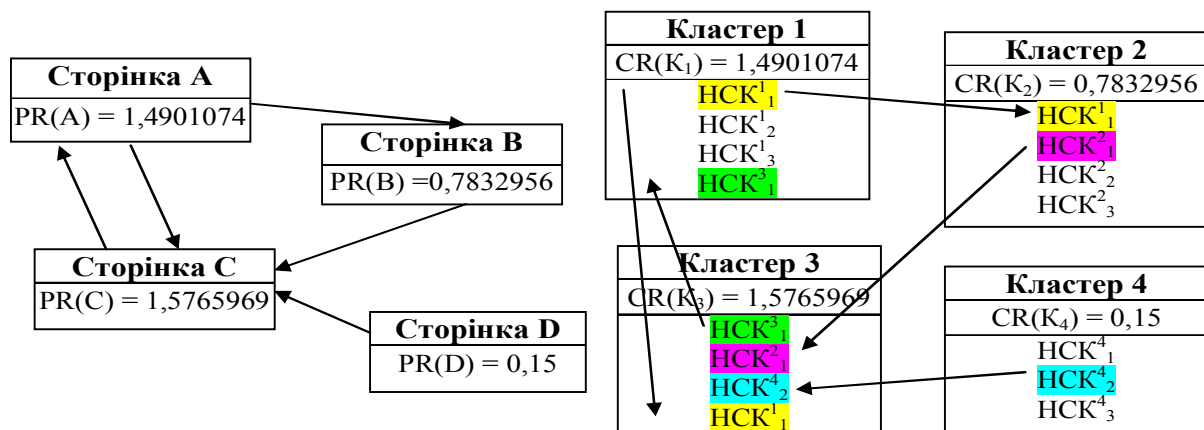


Рис. 2. Приклади визначення PR і CR: а) PR сторінки сайту визначається за вагою і кількістю посилань з інших сторінок, б) CR кластеру в КСН визначається за вагою і кількістю включень його НСК в інших кластерах

На сайті визначити належність посилання певній сторінці (яка сторінка посилається на іншу) просто, бо тег `<a>` кожного посилання має параметр `href = "URL_ресурсу"`, який визначає універсальний локатор ресурсу (адресу сторінки), на який посилається дана. В КСН в якості посилань виступають НСК, які використовуються (дублюються) в різних кластерах і не мають параметра, який би визначав їх базову належність. Для вирішення даної проблеми, одним з варіантів може бути зчитування в масив інформації про належність даного НСК різним кластерам даного порядку і за датою створення (зрозуміло, що дублюючі НСК матимуть пізнішу дату) визначити базовий.

Використовуючи (1-2) і приклади моделей сайту і масиву кластерів (рис. 2) розрахуємо рейтинги PR відповідних сторінок і кластерів. Виділимо сторінки/кластери, з яких посилаємося, в лівій колонці і позначимо курсивом сторінки/кластери, на які посилаємося, у верхньому рядку (табл. 1). Самі посилання позначимо на перетині сторінок через "X". Початковий PR сторінки (CR кластеру) визначимо рівним одиниці і виконаємо 50 ітерацій обчислень згідно з рекомендаціями [9, 11]. Послідовність ітерацій розрахунку буде такою:

- | | |
|--|--|
| 1 A:1.00000, B:1.00000, C:1.00000, D:1.00000; | 2 A:1.00000, B:0.57500, C:1.91375, D:0.15000; |
| 3 A:1.77669, B:0.90509, C:1.80192, D:0.15000; | 4 A:1.68163, B:0.86469, C:1.72718, D:0.15000; |
| 5 A:1.61811, B:0.83770, C:1.67724, D:0.15000; | 6 A:1.57565, B:0.81965, C:1.64386, D:0.15000; |
| 7 A:1.54728, B:0.80759, C:1.62155, D:0.15000; | 8 A:1.52831, B:0.79953, C:1.60664, D:0.15000; |
| 9 A:1.51564, B:0.79415, C:1.59667, D:0.15000; | 10 A:1.50717, B:0.79055, C:1.59001, D:0.15000; |
| 11 A:1.50151, B:0.78814, C:1.58556, D:0.15000; | 12 A:1.49773, B:0.78653, C:1.58259, D:0.15000; |
| 13 A:1.49520, B:0.78546, C:1.58060, D:0.15000; | 14 A:1.49351, B:0.78474, C:1.57927, D:0.15000; |
| 15 A:1.49238, B:0.78426, C:1.57839, D:0.15000; |; |
|; |; |
| 49 A:1.49011, B:0.78330, C:1.57660, D:0.15000; | 50 A:1.49011, B:0.78330, C:1.57660, D:0.15000. |

В результаті отримаємо показові значення PR і CR, що відображають взаємний вплив масиву сторінок (кластерів) один на одного (табл. 1).

Таблиця 1.

Визначення PR сторінок на сайті і CR кластерів в КСН

	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	PR (CR)
	<i>A</i>	X	X		1,490107 4
	<i>B</i>		X		0,783295 6
	<i>C</i>	X			1,576596 9
	<i>D</i>			X	0,15
Сумарний PR (CR): 4,0					

Для деяких сторінок пошукові системи збільшують або зменшують рейтинг в ручному режимі. Так окремим сторінкам наперед визначається високий рейтинг, а деякі навпаки заносяться до списку спамерів з нульовим або негативним рейтингом. Аналогічно, в КСН при створенні певних базових кластерів їм встановлюється завищений рейтинг, що забезпечує активний статус у нових НП, утворених в процесі самоорганізації системи. З іншої сторони, деякі кластери з самого початку можуть визначатись як тестові з негативним рейтингом і не потрапляти в НП.

Пошукові системи використовують багато параметрів для визначення рейтингу сторінок і PR є не єдиним, а його значимість менша, ніж у багатьох інших, таких як тег <title>, метатеги "description", "keywords" або тексту посилань. В КСН так само, для визначення рейтингу навчального кластеру застосовується не лише рейтинг CR, а й інші параметри, наприклад, експертні оцінки користувачів. Проте, очевидно, що для КСН немає сенсу визначати всі параметри рейтингів для кластерів кожного рівня. Наприклад, немає потреби визначати для Кластеру ЗМ рейтинг, що відображає співвідношення кількості вакансій до кількості резюме фахівців певного напрямку на сайтах роботи, проте для Кластеру НС значення даного параметру дуже важливе. Отже, визначимо, які параметри є сенс розраховувати для різних кластерів:

- 1) НСК – експертна рейтингова оцінка НСК в НК; експертна оцінка на відповідність даного НСК певним параметрам (якість, важливість, зрозумілість НМ та ін.);
- 2) Кластер ЗМ – експертна рейтингова оцінка, рейтинг CR;
- 3) Кластер М – експертна рейтингова оцінка, рейтинг CR;
- 4) Кластер НК – експертна рейтингова оцінка, рейтинг CR;
- 5) Кластер НС – експертна рейтингова оцінка, стат. дані Міністерства праці та соціальної

політики України, Державного комітету статистики України, держзамовлення Міністерства освіти і науки України, статистика сайтів пошуку роботи та ін., що відображають потребу РП у певних фахівцях;

б) Кластер СН – експертна рейтингова оцінка.

Інтегральний показник рейтингу кожного кластеру в рамках наближення про незалежність параметрів, що відображають їх рейтинг, визначимо як лінійну комбінацію цих показників:

$$R = \sum_{i=1}^n k_i \cdot R_i, \quad (3)$$

де R_i – i -ий параметр рейтингу;

k_i – i -ий коефіцієнт ваги відповідного R_i -го параметру рейтингу;

Значення коефіцієнтів ваги і перелік (масив) параметрів, що впливають на інтегральне значення рейтингу (3), визначаються в інтерактивному режимі користувачами (експертами) за допомогою методів експертного оцінювання. Крім того, зручним може виявитися використання багатокритерійного методу прийняття рішень, наприклад, MAI (методу аналізу ієрархій). В системі Moodle для цього можна використати модуль анкетного опитування Questionnaire [12].

Висновки. В роботі представлено концепцію, методику і принципи самоорганізації системи знань відповідно до вимог зовнішніх чинників впливу. Особливу увагу приділено розрахунку параметрів рейтингу НМ. Зміна структури НМ відповідно до цих параметрів забезпечує самоорганізацію системи знань в КСН з метою її актуалізації. Використання запропонованого підходу дозволяє КСН ефективно взаємодіяти із зовнішніми СН в рамках консорціуму університетів.

Подальший розвиток. Результати роботи відкривають перспективи використання запропонованої методики і її використання в розробці ІСН нового покоління на базі існуючих систем управління навчанням, таких як Прометей, Atutor та ін..

Подальший розвиток, розробка і взаємна інтеграція таких КСН в рамках консорціуму університетів, а також споріднених порталів їх підтримки, поява систем статистичних даних (по аналогії до порталу інформаційних баз даних (Data.gov) та ін.), глобальних сховищ навчальних об'єктів і НК (по аналогії до сховищ даних (Wikipedia), сховищ файлів (Rapidshare), каталогів інтернет-ресурсів (Catalog.i.ua), електронних бібліотек (Nbu.gov.ua) та ін.), бірж електронного навчання (по аналогії до фінансових (Forex), товарних (London Commodity Exchange), транспортних (Benelog), бірж інтернет-реклами (Google DoubleClick Ad Exchange), аукціонів (eBay) та ін.) призведуть до переходу віртуальної освіти на якісно новий рівень.

Література

1. Гагарін О.О., Тименко С.В. Дослідження і аналіз методів та моделей інтелектуальних систем безперервного навчання // Наукові вісті НТУУ "КПІ". – 2007. – № 6(56). – С. 37-48.
2. Brusilovsky P, Peylo C. Adaptive and Intelligent Web-based Educational Systems // International Journal of Artificial Intelligence in Education 13 (2003) 156–169.
3. Soller, A., & Lesgold, A. (2003). A computational approach to analyzing online knowledge sharing interaction. In U. Hoppe, F. Verdejo, & J. Kay (Eds.), *AI-ED'2003* (pp. 253-260). Amsterdam: IOS Press.
4. Данчук В.Д., Лемешко Ю.С. Синергетична модель квазіінтелектуальних освітніх інформаційних систем нового покоління // Вісник НТУ. – 2007. №15. – С.382-389.
5. Національний класифікатор України. Класифікатор професій ДК 003:2005 зі змінами і доповненнями по 04.08.08р N 270. search.ligazakon.ua/1_doc2.nsf/link1/ed_2008_08_04/FIN19198.html.
6. Електронний класифікатор професій. <http://www.jobs.ua/classifier/>.
7. Data.gov. <http://www.data.gov/about>.
8. Рейтинг сторінки. <http://en.wikipedia.org/wiki/PageRank>.
9. Sergey Brin, Lawrence Page. The anatomy of a large-scale hypertextual Web search engine, Computer Networks and ISDN Systems, 33: 107-17, 1998. <http://infolab.stanford.edu/pub/papers/google.pdf>.
10. Fong A.C.M., Hui S.C., Hong G., Fong B., Multimedia Engineering: A Practical Guide for Internet Implementation. Wiley, 2006.
11. David Austin. Grand Valley State University. How Google Finds Your Needle in the Web's Haystack <http://www.ams.org/featurecolumn/archive/pagerank.html>.
12. Moodle mod Questionnaire. <http://moodle.org/mod/data/view.php?d=13&rid=84>.