

- науки. Вип. 147 / Черкас. нац. ун-т ім. Богдана Хмельницького. – Черкаси : 2009. – С. 146 – 150.
2. Свиридов А.П., Шалобина И.А. Сетевые модели динамики знаний / под ред. Ю.Н. Мельникова. – М.: Изд-во МЭИ, 1992. – 88 с.
  3. Гниденко Б. Д., Коваленко И. Н. Введение в теорию массового обслуживания. – М.: Наука, 1987. – 336 с.
  4. Доррер Г.А. Моделирование процесса интерактивного обучения на базе формализмов раскрашенных сетей Петри / ГА . Доррер , Г.М. Рудакова // Вестн. Краснояр. гос. ун-та. Сер. физ.-матем. науки. — Красноярск: 2004. — № 3 Физико-математические науки. — С. 29-35.

УДК 004.052.42

Руслан Ярмолюк

Хмельницький національний університет

## ПІДХОДИ ДО РОЗРАХУНКУ ІНТЕГРАЛЬНОЇ ОЦІНКИ ЯКІСНОГО НАПОВНЕННЯ ЕЛЕКТРОННОГО КАТАЛОГУ БІБЛІОТЕКИ

© Руслан Ярмолюк, 2011

*В роботі наведено підходи до розрахунку оцінки якісного наповнення електронного каталогу бібліотеки. За допомогою теорії множин та реляційної алгебри представлено схеми розрахунку інтегральної оцінки. Вказано шляхи застосування інтегральної оцінки наповненості при розрахунку актуальної ціни електронного каталогу.*

*Ключові слова: електронний каталог, теорія множин, інтегральна оцінка, ієрархія помилок.*

*The present work demonstrates approaches to calculating the assessment of quality of filling an electronic catalog of the library. With the help of set theory and relational algebra presented integral evaluation scheme calculation. Specified ways to apply an integrated assessment of fullness in the calculation of the actual cost of electronic catalog.*

*Keywords: electronic catalog, set theory, integral evaluation, the hierarchy of errors.*

**Вступ.** Електронний каталог являє собою складну інформаційну систему у структурі бібліотеки. Від якості даних напряму залежить якість інформаційно-пошукових та інтеграційних послуг бібліотеки. На даний час не існує єдиного

підходу, щодо оцінювання наповненості електронного каталогу бібліотеки. Єдиним критерієм наповненості, що широко застосовується є потужність електронного каталогу (відображення бібліотечних фондів), тобто кількість записів у базі даних електронного каталогу.

У свою чергу, при створенні корпоративних електронних бібліотечних каталогів постає проблема оцінки «внеску» окремих бібліотек у загальний фонд бібліографічних записів корпоративного електронного каталогу. Очевидно, що критерій потужності електронного каталогу бібліотеки є малоінформативним і не відображає інформацію про якісний рівень записів. Отже, розробка методів та алгоритмів для розрахунку оцінок якісного наповнення електронного каталогу бібліотеки є досить актуальною задачею.

**Постановка проблеми.** За час свого життєвого циклу у електронному каталозі відбуваються певні інформаційно-технологічні процеси, що на пряму впливають на виникнення помилок в записах бази даних. Метод оцінювання полягає в урахуванні всіх типів помилок (тип помилки, можливість автоматичного виправлення, можливість виправлення без наявності оригінала джерела-інформації, тощо) в значеннях атрибутів, так і ролі даного атрибуту для забезпечення основних функцій (вагові коефіцієнти кожного атрибуту бібліографічного запису).

Сучасні електронні каталоги представляють собою множину бібліографічних записів організованих у схеми реляційних баз даних [1] з відповідними СКБД. Далі за допомогою теорії множин та реляційної алгебри опишемо схеми знаходження оцінок якісного наповнення електронного каталогу бібліотеки.

**Оцінка якісного наповнення за типами помилок.** Отже під оцінкою якісного наповнення електронного каталогу будемо розуміти відношення кількості записів (далі кортежів), що відповідають вимогам (не містять помилок певного виду) до загальної кількості кортежів (потужності):

$$\Omega^r = \frac{|A^r|}{|A|}$$

де  $\Omega^r$  - оцінка якісного наповнення за  $r$ -м типом помилки (основні типи та джерела помилок у записах електронного каталогу представлені у роботі [2],  $r \in \mathbb{N}$ );  $|A^r|$  - кількість кортежів, що не містять помилок  $r$ -го типу;  $|A|$  - загальна кількість кортежів. Відповідно:  $A^r$  - множина усіх кортежів, що не містять помилки  $r$ -го типу, а  $A$  - загальна множина усіх кортежів.

Вимоги приналежності кортежу до множини  $A^r$  за  $r$ -м типом помилок задаються відповідним граничним значенням сумарної оцінки пошуку помилки у кожному окремому значенні атрибуту кортежу:

$$A^r = \left\{ A_i \in A \mid \sum_j f^r(a_{ij}) \geq g^r \right\}, \quad i = \overline{1, \dots, |A|} \quad (1)$$

де  $A_i$  -  $i$ -й кортеж;  $a_{ij}$  - значення  $j$ -го атрибуту  $i$ -го кортежу;  $f^r(x)$  - функція перевірки значення  $x$  наявності помилки  $r$ -го типу;  $g^r$  - граничне значення сумарної оцінки для помилки типу  $r$ .

Функція перевірки  $f^r(x) \in \mathbb{N}$  і приймає значення на проміжку  $[0; 1]$ :

$$f^r(x) = \begin{cases} 0, & \text{якщо значення } x \text{ містить помилку типу } r; \\ 1, & \text{якщо значення } x \text{ не містить помилку типу } r. \end{cases}$$

Граничне значення сумарної оцінки  $g^r$  задає кількість гранично допустимих помилок типу  $r$  у значеннях атрибутів кортежу. Очевидно,  $g^r \in \mathbb{N}$ ,  $0 \leq g^r \leq n$ , де  $n$  - кількість атрибутів у кортежі. При  $g^r = 0 \Rightarrow A^r \equiv A$ , якщо  $g^r = n$ , то  $A^r \equiv A$ , лише у випадку відсутності типу помилок  $r$  у загальній множині записів електронного каталогу.

**Інтегральна оцінка якісного наповнення.** Різні типи помилок у записах по різному впливають на можливість виконання електронним каталогом своїх функцій. Відсутність певних значень атрибутів записів, синтаксичні та стилістичні помилки, наприклад, не впливають на інформаційно-пошукові можливості електронного каталогу.

Для розрахунку сумарної оцінки якісного наповнення потрібно враховувати значення впливу певного виду помилок на можливість виконання основних функцій електронним каталогом, а також можливість автоматичного виправлення певного виду помилок без людського втручання. Тому, для кожного типу помилок задається відповідний коефіцієнт впливу  $c^r \in \mathbb{R}$ ,  $0 \leq c^r \leq 1$   $r = \overline{1, \dots, m}$ . Відсортовані за коефіцієнтом впливу типи помилок задають схему ієрархії помилок.

Якщо задати схему ієрархії помилок, тобто  $A \equiv A^0 \supseteq A^1 \supseteq \dots \supseteq A^r \supseteq \dots \supseteq A^m$ , де  $m$  - кількість типів помилок, що мають вплив на функціональність електронного каталогу

$$A^{r+1} = \left\{ A_i \in A^r \mid \sum_j f^{r+1}(a_{ij}) \geq g^{r+1} \right\}, \quad i = \overline{1, \dots, |A^r|}; \quad r = \overline{1, \dots, m},$$

тоді інтегральна оцінка має вигляд:

$$\Omega_+^{all} = \frac{|A^m|}{|A|}.$$

У випадку відсутності ієрархії за типом помилок:

$$\Omega_-^{all} = \frac{\left| \bigcap_r A^r \right|}{|A|},$$

де  $A^r$  визначається за формулою (1).

Запропоновані оцінки якісного наповнення не враховують вплив (наявність або відсутність помилок) окремо кожного атрибута на функціональність електронного каталогу бібліотеки. Для врахування даного фактору, а також ймовірнісної складової функції потрібно застосовувати теорію нечітких множин [3,4].

### **Застосування оцінки якісного наповнення при економічних розрахунках.**

Бібліотекам потрібні дані про затрати на створення ними окремих продуктів і послуг. З іншого боку для розробки кошторису створення інтегрованих корпоративних каталогів та при прийнятті рішення про роботу зі сторонніми організаціями (аутсорсінг) необхідно вказувати обґрунтовану цінність бібліографічних даних, що пропонуються бібліотекою у вигляді електронного каталогу. Витрати на створення одного запису електронного каталогу (на обробку одного носія інформації) включають в себе затрати на комплектування та каталогізацію носія інформації.

Актуальну ціну електронного каталогу  $S_{current}$ , як упорядкованої бази даних бібліографічної інформації пропонуємо обраховувати за формулою:

$$S_{current} = \delta \cdot |A| - \sum_r \alpha_r (|A| - |A^r|),$$

де  $\delta$  - усереднений показник витрат на створення одного бібліографічного запису;  $\alpha_r$  - усереднений показник витрат на усунення помилки типу  $r$  з бібліографічного запису.

У випадку відсутності інформації про усереднені показники, або при заявленій зі сторони бібліотеки(продавця послуг) ціни  $S_{declared}$  наповненості каталогу покупець може коригувати її на відповідну інтегральну оцінку якісного наповнення, тобто:

$$S_{current} = S_{declared} \cdot \Omega_{+(-)}^{all}.$$

Запропонуємо методику обрахунку усередненого показника витрат на створення одного бібліографічного запису на основі рекомендацій IFLA (<http://www.ifla.org>).

Бібліотека вибирає період з нормальним робочим навантаженням для проведення вибірки. Щоб отримати кількість годин, затрачених на обробку носіїв інформації, необхідний аналіз часових витрат. Оскільки персонал часто виконує різні задачі, час, витрачений ними на обробку, повинен бути

врахований на протязі періоду вибірки. Якщо облік часу неможливий, то потрібно оцінити витрачений час на обробку у процентному співвідношенні[5].

Час, витрачений на обробку усім персоналом на протязі періоду вибірки, додається в годинах. Потім, ця кількість годин множиться на вартість однієї години роботи персоналу (сумарна заробітна плата ділиться на нормовану кількість робочих годин персоналу, що задіяні у роботі) і отримане значення, відповідає витратам на персонал, який виконує обробку носія інформації.

Кількість носіїв інформації, оброблених за період вибірки, також фіксується.

Витрати на обробку одного носія інформації обраховуються шляхом ділення витрат персоналу (який виконує обробку) на кількість оброблених носіїв інформації за період вибірки [5].

Звичайно, на усереднений показник будуть впливати тип обробки, мова оброблених носіїв інформації, відсутність даних для запозичення при каталогізації, рівень бібліографічного опису, ступінь автоматизації і кваліфікація та досвід працівників, тощо.

Очевидно, що даний підхід можливо також застосовувати для розрахунку усередненого показника витрат на усунення певного виду помилок. Однак, у методиці потрібно враховувати також можливість автоматичної корекції помилок, без участі персоналу.

**Висновки.** Запропонований підхід для розрахунку, як узагальненої (інтегральної) оцінки всього каталогу, так і оцінки якості окремих записів (кортежів) електронного каталогу за кожним типом помилок. Дана оцінка може бути застосовна для об'єктивного оцінювання стану електронного каталогу при фінансових операціях та у процесах створення корпоративних електронних каталогів міста, району, області, тощо.

### **Література**

16. Мейер Д. *Теория реляционных баз данных* / Д. Мейер. – М. : Мир, 1987. – 608 с.
17. Ярмолюк Р.С. *Основні типи та джерела помилок у записах електронного каталогу* / Р.С. Ярмолюк // *Вісник Національного Університету «Львівська політехніка» Інформаційні системи та мережі*, - 2010. - № 689. – С. 348-357.
18. Кофман А. *Введение в теорию нечетких множеств* / А. Кофман - М.: Радио и связь, 1982. - 432 с.
19. Заде Л. *Понятие лингвистической переменной и его применение к принятию приближенных решений* / Л. Заде. - М.: Мир, 1976. - 166с.
20. Зауэр К.Д. *Измерение качества в библиотеках* / К.Д. Зауэр. – Мюнхен: Изд. IFLA, 2007. – 215с.