

7. Ефимова М.Р. Рябцев В.М. Общая теория статистики. – Москва: Финансы и статистика, 1991 – 304с.

8. Боровиков, В. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов / В. Боровиков. – СПб.: Питер, 2003.– 688 с.

9. UNECE [Электронный ресурс] <http://w3.unece.org/pxweb/>. (дата обращения: 05.10.2013)

#### Аннотации:

Статья посвящена изучению перспектив дальнейшего развития технологий широкополосного доступа и анализу взаимосвязей с другими показателями, такими как: количество и плотность населения.

**Ключевые слова:** широкополосный доступ, меры рассеяния, корреляционная зависимость, критерии Колмогорова-Смирнова и Лиллиефорса.

Статья посвящена вивченню перспектив подальшого розвитку технологій ширококуткового доступу та аналізу взаємозв'язків з іншими показниками, такими як: кількість і щільність населення.

**Ключові слова** ширококутговий доступ, міри розсіювання, кореляційна залежність, критерії Колмогорова-Смирнова і Лілієфорса.

Article is devoted to study the perspectives of broadband access development and analyze the relationships with the other indexes such as number and density of population are described in this article.

**Keywords:** Broadband access, scattering measures, correlation dependence, Kolmogorov-Smirnov and Lilliefors criteria.

УДК 6081.518:004.451.6

ТКАЧЕНКО Н.О., аспірант (Донецький національний технічний університет)  
ВОРОПАЄВА В.Я., к.т.н., доцент (Донецький національний технічний університет)  
ДУЧЕНКО М.М., к.е.н., доцент (Національний технічний університет України «Київський політехнічний інститут»)

## Принципи функціонування адаптивної інформаційно-пошукової системи зі зворотним зв'язком

---

### Вступ

---

Останнім часом мережа Інтернет та електронні бібліотеки відкривають великі перспективи перед окремими авторами та організаціями. Всесвітня мережа Інтернет дозволила користувачам отримувати різноманітну інформацію у будь-який час, але це призвело до виникнення нової проблеми, а саме отримання релевантної інформації, що відповідає потребам користувача. Відомо, що неймовірно важко знайти інформацію в Інтернет, що задовольняє індивідуальні потреби окремої людини, особливо в ресурсах WWW, кількість яких продовжує зростати. Для вирішення цього завдання

відомі всесвітні організації створюють спеціальні пошукові системи (Google, Yandex), що використовують персональні алгоритми для пошуку релевантної інформації [1]. Але незважаючи на усі переваги таких алгоритмів, всі ці системи мають один великий недолік — вони не враховують індивідуальність конкретного користувача. Тобто, різні люди отримують однакові результати при вводі однакового запиту, незважаючи на їх особисті потреби та інтереси. Однак, в реальному світі окремий користувач має власні інформаційні інтереси щодо свого запиту. Тому, актуальним залишається завдання адаптації пошукових систем до

потреб користувачів з різними інформаційними потребами.

що не дозволяє їх використання у великому пошуковому просторі.

**Постановка проблеми**

Персональний Web-пошук виступає актуальною темою дослідження останніх кілька років. Ключовим моментом персонального пошуку є збір інформації щодо потреб окремого користувача. На сьогоднішній день існує декілька методів [2-4] інтелектуального аналізу даних, що використовують Web log користувачів. На жаль, ці методи не є досконалими при Web персоналізації. Персональний PageRank та HITS алгоритм [5, 6] частково підвищують релевантність пошуку, але зі зростанням кількості ресурсів підвищується складність їх обчислення,

**Вирішення задачі**

**Класифікація інформаційно-пошукових систем.**

Для усунення недоліків існуючих методів було розроблено декілька нових систем, що дозволяють поліпшити результати пошуку інформації згідно потреб користувача. Пропонується здійснювати класифікацію інформаційно-пошукових систем наступним чином: системи зі зворотнім зв'язком, рекомендуючі системи та персоналізовані системи (рисунок 1).

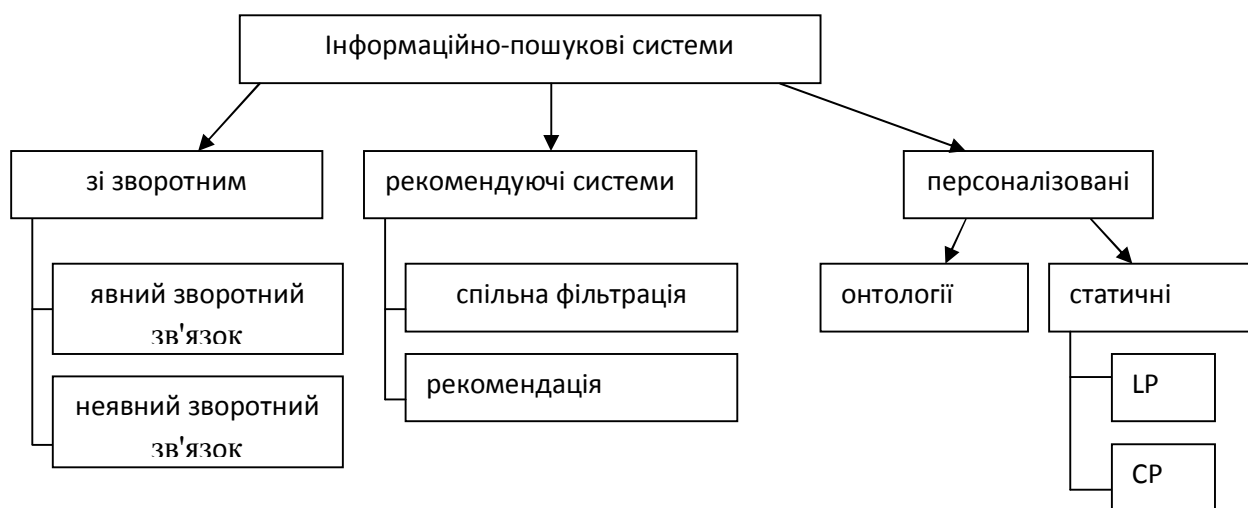


Рис. 1. Класифікація інформаційно-пошукових систем

Системи, що використовують зворотний зв'язок, засновані на залученні користувачів до відбору релевантної інформації. Тобто, для отримання необхідних ресурсів система виконує декілька ітерацій: користувач робить свій запит, система повертає список знайдених документів. Після цього користувач уточнює свій запит, відзначаючи деякі документи як релевантні або

нерелевантні. Використовуючи цей зворотний зв'язок з користувачем, система уточнює набір релевантних ресурсів та повертає їх користувачеві. Використання алгоритму Роккіо є класичним для систем зі зворотнім зв'язком. Для модифікації початкового запиту використовується наступна формула:

$$\bar{q}_m = \alpha \bar{q}_0 + \beta \frac{1}{|D_r|} \sum_{d_j \in D_r} \bar{d}_j - \gamma \frac{1}{|D_{nr}|} \sum_{d_j \in D_{nr}} \bar{d}_j \quad (1)$$

де  $\vec{q}_m$  - модифікований вектор,  $\vec{q}_0$  - початковий вектор запиту,  $D_r$  - релевантні документи,  $D_{nr}$  - не релевантні документи та  $\alpha, \beta, \gamma$  - вагові множники.

У роботах Яковлевої Ю.В. [7] пропонується метод модифікації запиту, що базується на організації зворотного зв'язку та використанні статистичних даних для оцінки релевантності

$$C_i^{\text{int}} = \beta_1 C_{i_1}^0 + \beta_2 C_{i_2}^0 + \beta_3 (\beta_s C_{i_s}^0 + \beta_p C_{i_p}^0) \quad (2)$$

Але запропоновані критерії не в повній мірі відповідають потребам користувача при пошуку у всесвітньому павутинні, оскільки орієнтовані передусім на друкований фонд конкретної бібліотеки.

**Інтегральний критерій оцінки електронних ресурсів.**

Поступовий розвиток електронних публікацій, як окремого наукового ресурсу, та розміщення їх в мережі Інтернет дозволяє науковим організаціям, зокрема університетам вийти на новий рівень співпраці [13]. Але поява таких ресурсів призвела до виникнення нової проблеми – коректної оцінки цих ресурсів. Існуючі критерії оцінки електронних матеріалів мають декілька недоліків, тому авторами [1] у попередніх дослідженнях був запропонований інтегральний критерій оцінки електронних наукових інформаційних ресурсів в мережі Інтернет, що дозволяє користувачеві самому керувати процесом

$$A = \sum \alpha_i a_i = \alpha_1 a_1 + \alpha_2 a_2 + \alpha_3 a_3 + \alpha_4 a_4 \quad (3),$$

де  $\alpha_i$  - ваговий множник,  $a_1$  - h-індекс автора в GoogleScholar,  $a_2$  - i10-індекс автора в GoogleScholar,  $a_3$  - h-індекс

результатів пошуку. В роботі зворотний зв'язок запропоновано організувати двома способами: 1) статистичний аналіз запитів користувачів та 2) багатокритеріальна оцінка інформативності видань, що базується на трьох критеріях (старіння літератури, значущість автора та попит джерела у бібліотеці):

призначення вагових коефіцієнтів та вибору часткових критеріїв. Загальний критерій об'єднує у своєму складі декілька критеріїв (оцінка автора, оцінка організації та оцінка журналу), що мають власні вагові множники.

Ресурси відкритого доступу, що розміщуються у відповідних журналах або інституційних репозитаріях стали потужним конкурентом передплаченим ресурсам. У [14] зроблено висновок щодо безпосереднього впливу таких ресурсів на наукометричні показники авторів чи організацій. Тому при розробці інтегрального критерію об'єднувалися декілька існуючих критеріїв оцінки одночасно.

При оцінці автора враховувалися показники як передплачених наукометричних баз (Scopus), так і баз у вільному доступі (GoogleScholar), що дозволить повноцінно оцінити науковий вклад автора:

автора в Scopus,  $a_4$  - кількість робіт автора в GoogleScholar.

Оцінка сайту університету складається з наступних критеріїв:

$$U = \sum \beta_i u_i = \beta_1 u_1 + \beta_2 u_2 + \beta_3 u_3 \quad (4),$$

де  $\beta_i$  - ваговий множник,  $u_1$  - місце університету у рейтингу Webometrics,  $u_2$  -

оцінка сайту системою Google,  $u_3$  - оцінка сайту системою Yandex.

Загальна оцінка журналу враховуватиме наступні чотири критерії:

$$G = \sum \gamma_i g_i = \gamma_1 g_1 + \gamma_2 g_2 + \gamma_3 g_3 + \gamma_4 g_4 \quad (5),$$

де  $\gamma_i$  - ваговий множник,  $g_1$  - присутність журналу у директорії відкритого доступу DOAJ,  $g_2$  - присутність журналу у директорії відкритого доступу DOAJ,  $g_3$  - імпаکت-фактор журналу,  $g_4$  - місце у національній наукометричній базі даних.

Обчислення загального критерію здійснюється шляхом поєднання

$$\max \varphi(X) = \max \sum \lambda_i \varphi_i = \lambda_1 \varphi_1(X) + \lambda_2 \varphi_2(X) + \lambda_3 \varphi_3(X) \quad (6),$$

де  $\lambda$  - ваговий множник,  $\varphi_1(X)$  - загальна оцінка автора,  $\varphi_2(X)$  - загальна оцінка організації,  $\varphi_3(X)$  - загальна оцінка журналу.

#### Принципи роботи систем з неявним зворотним зв'язком.

Останнім часом все більшу популярність завойовують модифіковані системи зі зворотним зв'язком, а саме системи з неявним зворотним зв'язком (systems with implicit relevance feedback) [8]. Такі системи базуються на аналізі поведінки користувача у всесвітній мережі (кліків, часу, витраченого на сторінці, відстеження очей та інше). Впровадження таких систем частково підвищує загальну релевантність отриманих результатів, але не вирішує проблему їх адаптації під інтереси конкретного користувача.

Рекомендуючі системи. Ці системи забезпечують підвищення релевантності результатів завдяки перевагам користувача. Рекомендуючі системи підтримують зворотний зв'язок з користувачами за допомогою оцінок ресурсів. Існує два поширених підходи до побудови цих систем — система на основі спільної фільтрації та системи рекомендації контенту:

- система на основі спільної фільтрації призначена для передбачення релевантності окремого ресурсу для конкретного користувача на основі

загальних оцінок перерахованих вище критеріїв з врахуванням внеску кожного з них. Для цього використовується один з методів багатокритеріальної оптимізації, а саме метод зважених сум. Таким чином, остаточною формулою матиме вид:

попередніх симпатій та думок інших користувачів.

- система рекомендації контенту висуває рекомендації шляхом порівняння змісту, що представлений в ресурсі з контентом, який цікавить конкретного користувача. Описані системи дають позитивні результати лише у випадку добровільного оцінювання користувачами запропонованих ресурсів. Проте насправді більшість користувачів не бажають оцінювати ресурси, тому точність рекомендацій може бути поганою, бо саме такі рекомендації виступають ключовим елементом формування рейтингу ресурсів.

Персональні веб-сайти. Ці системи умовно поділяють на "Link Personalization" та "Content Personalization".

- Link Personalization базується на виборі посилань, що є найбільш релевантними до потреб конкретного користувача та зміні початкового навігаційного простору шляхом зниження або підвищення відносин між веб-сторінками. Відповідність документу до інтересів користувача обчислюється за допомогою метрики подібності (cosine similarity metric):

$$\cos(i) = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} \quad (7),$$

де  $\vec{a}$  - вектор інтересів користувача,  $\vec{b}$  - вектор знайденого документу, а  $\| \cdot \|$  означає Евклідову норму.

- Content Personalization. Даний метод використовують, коли різні сторінки надають різноманітну інформацію для різних користувачів. Цей метод дозволяє фільтрувати інформацію: користувач вказує свої інтереси на персональному сайті, що дозволяє системі надавати лише релевантну інформацію. Використання таких систем потребує від користувачів їх прямої участі у пошуку інформації (зазначення персональних переваг), що призводить одразу до кількох проблем: 1) навантаження на користувачів збільшується, оскільки такі системи покладаються на дані користувацького профілю та 2) системи не можуть автоматично пристосуватися до змін користувацьких інтересів, якщо вони особисто не вказані користувачем.

#### Адаптивні інформаційно-пошукові системи.

Можливість змінюватися в залежності від окремої ситуації або інтересів конкретної людини дозволили адаптивним системам стати гідним аналогом традиційних систем. Ці системи

$$P = \alpha P^{per} + \beta P^{today} = \alpha P^{per} + \beta x P^{br} + \beta y P^{cur} \quad (8),$$

де  $P^{per}$  - історія відвіданих сторінок за декілька днів,  $P^{br}$  - історія відвіданих сторінок за поточний день та  $P^{cur}$  - поточна сесія пошуку інформації.

В останні роки широку популярність почали завойовувати системи, що використовують онтології для пошуку релевантної інформації. Онтології виступають ефективним засобом моделювання контенту користувача, бо вони представляють опис окремої області знань та можуть бути використані для перегляду та уточнення запиту. У загальному вигляді онтологія являє собою наступну трійку  $O = \langle K, L, F \rangle$ , де  $K$  — множина понять предметної області в онтології,  $L$  — множина зв'язків між цими

поступово впроваджуються в різні напрямки інформаційної діяльності. Наприклад, авторами в роботі [9] розроблено інтелектуальну систему дистанційного навчання, що дозволяє підвищити рівень навчання за допомогою адаптації під конкретного учня. Використання адаптивних систем характерно й для Web-пошуку, Kazunari Sugiyama та ін. [10] був запропонований новий метод отримання релевантної інформації в веб, що враховує інтереси користувачів. Цей метод заснований на адаптації результатів пошуку до потреб користувачів: система адаптується до змін в перевагах користувача, що дозволяє виконати більш точний пошук. Запропонована система відстежує історію відвіданих сторінок користувачів та оновлює його профіль, коли веб-сторінки змінюються. При побудові профілю користувача враховуються не тільки довгострокова історія веб-браузера (тобто історія за декілька днів), але й історія відвіданих сторінок за поточний день. Таким чином, враховуючи усі аспекти, остаточна формула побудови профілю користувача матиме наступний вигляд:

поняттями,  $F$  — множина функцій інтерпретації понять та відношень. Таким чином, будь-яка онтологія може виступати у вигляді графу з наявністю визначеної кількості вершин та дуг.

Такі онтології активно застосовуються в різноманітних інформаційних областях таких, як web-пошук та інформаційний менеджмент. Для пошуку релевантної інформації в Web Gauch та співавтори [11] запропонували систему, що автоматично створює профіль конкретного користувача, враховуючи його інтереси, завдяки оцінці конкретних понять в документі. Такий профіль являє собою онтологію, у якій кожне поняття має свою персональну вагу, що призначається відповідно до інтересів

користувача до цього поняття. Він створюється на основі аналізу пошукового характеру користувача, а саме врахуванням контенту, довжини та часу, витраченого на конкретну відвідувану сторінку.

Оцінкою онтологій при інформаційному пошуку займаються не лише іноземні автори, вітчизняними авторами також розроблено декілька

методів оцінки онтологій для підвищення релевантності пошуку інформації. Наприклад, в роботі Р. Даревича [12] розроблено метод адаптації онтологій при побудові метапошукових систем. Онтологія має вигляд орієнтованого зваженого мультиграфу, що складається з вершин та ребер, що мають свої конкретні ваги (рис.2).

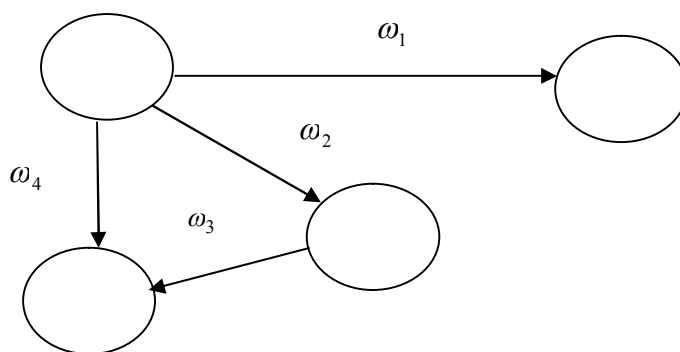


Рис. 2. Мультиграф онтології

**Принципи роботи адаптивних інформаційно-пошукових систем зі зворотним зв'язком.**

Але відсутність зворотного зв'язку з користувачем призводить до проблеми точності оцінки знайденого документу. Тому в даній роботі пропонуються принципи функціонування інтелектуальної адаптивної системи пошуку релевантної інформації в Інтернет на основі онтологій зі зворотним зв'язком. Використання онтологій дозволить враховувати інтереси користувача (за допомогою створення його онтологічного профілю), а доповнення системи адаптивним інтерфейсом (наявність зворотнього зв'язку), що дозволить користувачеві приймати участь у відборі релевантної інформації, підвищить якість інформаційного пошуку. Створення онтологічного профілю користувача відбуватиметься за допомогою аналізу історії відвіданих сторінок веб-браузера та позначенням цим користувачем своїх інтересів. В зазначеному профілі кожна онтологія матиме ієрархічну структуру, в

якій кожному поняттю присвоюється індивідуальна вага. Загальна вага окремої онтології повинна враховувати ваги усіх понять, що входять до її складу та важливість зв'язків між цими поняттями. Таким чином, вага окремої онтології розраховується за формулою:

$$W = \sum \omega * \varphi \quad (9),$$

де  $\omega$  - вага окремого поняття онтології,  $\varphi$  - вага зв'язків між цими поняттями.

Для відбору релевантних документів зі списку знайдених кожен такий документ представляється в якості вектору  $(\omega_1, \omega_2, \dots, \omega_n)$ , де  $\omega$  — вага окремого поняття у цьому документі. Вага кожного поняття в документі розраховується з використанням відомої  $tf * idf$  схеми:

$$\omega_{ij} = tf_{ij} * idf_i \quad (10),$$

де  $tf = \frac{m_i}{\sum m_p}$  - частота слова,  $idf = \log \frac{S}{P}$  - зворотна частота документу.

Для порівняння текстових документів між собою на практиці використовується декілька методів: Евклідова відстань, косинус міра, манхетенська відстань, відстань Чебишева та ін. Відносно високу точність показав саме метод косинус міри, тому доцільне його використання:

$$\cos(\theta) = \frac{\sum D * R}{\sqrt{\sum (D)^2} * \sqrt{\sum (R)^2}} \quad (11),$$

де D – вектор понять в онтологічному профілі користувача, R – вектор понять у знайденому документі.

Таким чином, запропонована система виконуватиме наступні два етапи: створюватиме онтологічний профіль окремого користувача з врахуванням його інформаційних інтересів та надаватиме релевантні документи згідно цих інтересів.

## Висновки

У статті запропонований підхід до класифікації інформаційно-пошукових систем, на базі якого виділяються системи зі зворотнім зв'язком, рекомендуючі системи та персоналізовані системи. Сформульовано інтегральний критерій оцінки електронних наукових інформаційних ресурсів в мережі Інтернет, що дозволяє користувачеві самому керувати процесом. Авторами розроблено принципи функціонування адаптивних інформаційно-пошукових систем зі зворотним зв'язком, що враховують інтереси користувача при пошуку інформації в Інтернет.

## Література

1. Ткаченко Н.О. Розробка критеріїв для оцінки інформаційних наукових ресурсів в Інтернет / Н.О. Ткаченко // Наукові праці Донецького національного технічного університету. Серія: Обчислювальна техніка та автоматизація. - 2013. - Вип. 2 (25). - С.

136-143

2. Cooley R. Data Preparation for Mining World Wide Web Browsing Patterns / R. Cooley, B. Mobasher, and J. Srivastava // Knowledge and Information Systems. – 1999. - 1(1):5–32. – С. 237-244

3. Spiliopoulou M. WUM–A Tool for WWW Utilization Analysis. / M. Spiliopoulou, L. Faulstich // Proceedings of the International Workshop on the World Wide Web and Databases. – 1998. С. 184–203.

4. Wang J. Ranking User's Relevance to a Topic through Link Analysis on Web Logs. / J. Wang, Z. Chen, L. Tao, W.-Y. Ma, L. Wenyin // Proceedings of the 4th ACM CIKM International Workshop on Web Information and Data Management. – 2002. – С. 49–54

5. Haveliwala T. H. Topic-Sensitive PageRank / T. H. Haveliwala // Proceedings of the 11<sup>th</sup> International World Wide Web Conference. – 2002. – С. 517–526

6. Kleinberg. J. M. Authoritative Sources in a Hyperlinked Environment. / J. M. Kleinberg // Journal of the ACM. – 1999. - 46(5). – С. 604–632

7. Хаджинов В.В. Адаптивний пошук як напрям розвитку інформаційно-пошукових систем наукових бібліотек / В.В. Хаджинов, Ю.В. Яковлева // Реєстрація, зберігання і обробка даних. – 2006. - Т. 8, № 2. – С. 53-60

8. Eugene Agichtein Improving Web Search Ranking by Incorporating User Behavior Information / Eugene Agichtein, Eric Brill, Susan Dumais // SIGIR '06 Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. — 2006. - С. 19-26

9. Воропаєва В.Я., Криворучко Д.В. Математичне моделювання процесів дистанційного навчання // Научно-технический журнал «Автоматика. Автоматизация. Электротехнические комплексы и системы» № 2(14) 2004, стр. 11-15

10. Kazunari Sugiyama Adaptive Web Search Based on User Profile

Constructed without Any Effort from Users / Kazunari Sugiyama, Kenji Hatano, Masatoshi Yoshikawa // Proceedings of the 13th international conference on World Wide Web. - 2004. - С. 675 - 684

11. Susan Gauch Ontology-Based User Profiles for Search and Browsing / Susan Gauch, Jason Chaffee, Alexander Pretschner // Journal Web Intelligence and Agent Systems . – 2003. – С. 219-234

12. Р. Даревич Підвищення ефективності інтелектуального аналізу тексту шляхом зважування понять у моделі онтології / Р. Даревич // Штучний інтелект. – 2005. - № 3. – С. 571-577

13. Аноприенко А.Я. Новое познание в новом мире: научные исследования и высшее образование в условиях информационной супермагистрали // TEMPUS/TACIS Conference „Computer Networks in Higher Education“, 26-28 May 1997, National Technical University of Ukraine, Kyiv.

14. Воропаева В. Я., Ткаченко Н.О. Влияние электронного архива ВНЗ на наукометрические показатели // Библиотеки та інформаційні ресурси у сучасному світі науки, освіти та культури : Матеріали наук. - практ. конф., м. Севастополь, 7-10 жовтня 2013 р. Севастополь: Купол , 2013

. – 192 с, С. 13-15

#### Анотації:

У роботі здійснено аналіз існуючих систем пошуку інформації в Інтернет. Запропоновано класифікацію інформаційно-пошукових систем. Розроблено принципи функціонування адаптивної інформаційно-пошукової системи зі зворотним зв'язком.

**Ключові слова:** інформаційно-пошукова система, зворотний зв'язок, адаптивність, релевантність, інформаційні потреби користувача

В работе осуществлен анализ существующих систем поиска информации в Интернет. Предложена классификация информационно-поисковых систем. Разработаны принципы функционирования адаптивной информационно-поисковой системы с обратной связью.

**Ключевые слова:** информационно-поисковая система, обратная связь, адаптивность, релевантность, информационные потребности пользователя

Analysis of existing information retrieval systems on the Internet was given in the article. Classification of information retrieval systems was offered. Principles of adaptive information retrieval system with feedback working was developed.

**Keywords:** information retrieval system, feedback, adaptability, relevance, information needs of the user

УДК 681.391

МОЛОКОВСКИЙ И.А., аспирант (ДонНТУ)

### Исследование особенностей проектирования систем связи с использованием излучающего кабеля

---

#### Актуальность проблемы

---

В подземной части угледобывающего предприятия, как и на многих других промышленных предприятиях, используются различные технологические системы связи. Соответствующие сис-

темы в пределах подземных выработок, на поверхности угледобывающего предприятия, а также между подземными рабочими станциями являются жизненно важной частью правильного функционирования любого подземного средства связи. Отлаженная работа технологических сис-





