

РОЗШИРЕННЯ ОСВІТНЬОГО ІНФОРМАЦІЙНОГО ПРОСТОРУ ЗА РАХУНОК ТЕХНОЛОГІЙ ПОШУКУ

Т. А. Григорова

Кременчуцький національний університет імені Михайла Остроградського
вул. Першотравнева, 20, м. Кременчук, 39600, Україна. E-mail: grital0403@gmail.com

Досліджено складові частини освітнього інформаційного простору і можливість використання пошуку для отримання інформації з усіх наявних ресурсів, щоб організувати наукову, освітню та творчу діяльність. В ході дослідження було виявлено, що системи електронного навчання є основною складовою частиною освітнього простору. Запропоновано розширити ці системи пошуковими модулями або алгоритмами пошуку, при цьому змінивши структуру даних. Додати додатковий модуль автоматизованого перекладу рядка запиту, щоб мати можливість використовувати інформацію, представлену в мережі на різних мовах. Організувати можливість пошуку інформації безпосередньо з системи в мережі інтернет. Додати модуль, що дозволяє звужувати пошук, який ведеться за межами системи. Розширити базу даних системи додатковою інформацією щодо результату пошуку, щоб уникнути повторного пошуку в інтернеті.

Ключові слова: освітній простір, пошук інформації, структура даних.

РАСШИРЕНИЕ ОБРАЗОВАТЕЛЬНОГО ИНФОРМАЦИОННОГО ПРОСТРАНСТВА ЗА СЧЕТ ТЕХНОЛОГИЙ ПОИСКА

Т. А. Григорова

Кременчугский национальный университет имени Михаила Остроградского
ул. Первомайская, 20, г. Кременчуг, 39600, Украина. E-mail: grital0403@gmail.com

Исследованы составные части образовательного информационного пространства и возможность использования поиска для получения информации со всех имеющихся ресурсов, чтобы организовать научную, образовательную и творческую деятельность. В ходе исследования было выявлено, что системы электронного обучения являются основной составляющей частью образовательного пространства. Предложено расширить эти системы поисковыми модулями или алгоритмами поиска, при этом изменив структуру данных. Добавить дополнительный модуль автоматизированного перевода строки запроса, чтобы иметь возможность использовать информацию, представленную в сети на разных языках. Организовать возможность поиска информации непосредственно из системы в сети интернет. Добавить модуль позволяющий сужать поиск, который ведется за пределами системы. Расширить базу данных системы дополнительной информацией о результате поиска, чтобы избежать повторного поиска в интернете.

Ключевые слова: образовательное пространство, поиск информации, структура данных.

АКТУАЛЬНІСТЬ РОБОТИ. Невід'ємною частиною сучасного життя є інформаційний простір. Він складається з сукупності банків та баз даних, технологій їх супроводу та використання, інформаційних телекомунікаційних систем, які функціонують на основі загальних принципів і забезпечують інформаційну взаємодію користувачів та задовольняють їх інформаційні потреби [1]. Основними компонентами інформаційного простору є: інформаційні ресурси, засоби інформаційної взаємодії, інформаційна інфраструктура. Частиною загального інформаційного простору є освітній інформаційний простір. Це відкритий простір з інформацією про наукові дослідження, наукові данні з інструментами для підтримки публікацій, програмним забезпеченням, освітніми ресурсами та інфраструктурою, що забезпечує доступ всім інституціям національних співтовариств. Це багатокомпонентне середовище, що включає навчально-методичні та наукові матеріали, наукоємне програмне забезпечення, тренінгові системи, системи контролю знань, технічні засоби, бази даних і інформаційно-довідкові системи, сховища інформації будь якого вигляду, включаючи графіку, відео й ін., що взаємопов'язані між собою. Основною ціллю освітнього інформаційного простору є надання принципово нових можливостей для пізнавальної творчої діяльності людини. Задачі освітнього інформаційного простору включають:

- підвищення ефективності і якості процесу навчання;
- інтенсифікація процесу наукових досліджень в освітніх установах;
- скорочення часу і поліпшення умов доступу до додаткових знань;
- підвищення ефективності системи освіти в цілому;
- інтеграція національних інформаційних освітніх систем у світову мережу за рахунок доступу до міжнародних інформаційних ресурсів у галузі освіти, науки, культури і взаємодії [1].

Метою роботи було дослідження можливості розширення освітнього інформаційного простору завдяки використанню можливостей сучасних інформаційних технологій.

МАТЕРІАЛ І РЕЗУЛЬТАТИ ДОСЛІДЖЕНЬ. Передумовою до появи освітнього інформаційного простору було створення нових підходів і форм освіти до яких відносяться: дистанційне навчання, змішана форма освіти, STEM – освіта, навчання реєр-2-реєр. Кожна з цих форм освіти з'явилась завдяки появі інформаційного простору і її основною складовою є інформаційні технології. Розглянемо всі перераховані форми освіти з точки зору їх основного призначення і інтеграції в інформаційний простір.

Дистанційне навчання це форма навчання з використанням комп'ютерних і телекомунікаційних технологій, які забезпечують інтерактивну взаємодію викладачів та студентів на різних етапах навчання і самостійну роботу з матеріалами інформаційної мережі. Інтеграція в інформаційний простір відбувається за рахунок систем електронного навчання яке забезпечує:

- надання учбового матеріалу студенту;
- контроль успішності студента;
- інтерактивну співпрацю викладача і студента;
- інструментарій формування навчально-методичного матеріалу.

Змішана форма освіти - поєднує в собі використання традиційних методів освіти та інформаційних технологій. Інтеграція в інформаційний простір відбувається за рахунок систем електронного і мобільного навчання. Останні представляють собою відкриті освітні портали, які призначені для створення і підтримки відкритих навчальних курсів [2].

STEM - освіта поєднує науку, математику, технології і інженерію. Вона спрямована на розвиток креативного мислення і спроможність вирішувати проблеми і розв'язувати задачі. Інтеграція в інформаційний простір відбувається за рахунок створення в Інтернет мережі спільнот, коаліцій, проєктів, віртуальних лабораторій [3].

Навчання реєр-2-реєр це освіта в основі якої лежить принцип колаборації – студенти навчаються самі і допомагають один одному. Інтеграція в інформаційний простір відбувається за рахунок використання репозиторіїв, тематичних чатів і блогів, а також віртуальних освітніх співтовариств, що надають майданчики для співпраці (наприклад Google Educator Group).

Розглянемо основні питання і шляхи розв'язку.

Якщо розглянути всі елементи завдяки яким відбувається інтеграція новітніх форм освіти у інформаційний простір і додати до них платформи, які формують наукометричні бази даних такі як Scopus, Web of Science та інші можна сказати що всі вони є замкнуті на себе за деякими винятками до яких відносяться тематичні чати і блоги. Для того щоб отримати необхідну інформацію з цих порталів потрібно безпосередньо отримати до них доступ через мережу інтернет і тільки після цього з'являється можливість скористатися необхідною інформацією. Таким чином процес пошуку наукової або освітньої інформації займає багато часу і не завжди є ефективним на відміну від пошуку комерційної інформації, яка є відкритою для пошукових машин і де використовуються технології семантичного вебу, штучного інтелекту та інші, щоб постійно підтримувати зацікавленість користувача. Тобто однією із задач розширення освітнього інформаційного простору є задача організації швидкого і зручного пошуку інформації наукового і освітнього призначення, яка зберігається у різних системах і сховищах даних і є багатомовною і неоднорідною за типом даних.

Розглянемо шляхи вирішення цієї проблеми. Якщо дослідити всі нові форми освіти вони використовують ті, чи інші системи, які допомагають орга-

нізувати навчальний процес за рахунок організації взаємодії між його учасниками, які поділяються на тих хто створює навчальні курси і тих хто отримує доступ до інформації, що їх наповнює. Але будь який курс має обмежену кількість інформації і дуже часто не відповідає на всі питання, що виникають у учня, особливо це пов'язано з науковою або дослідницькою діяльністю. Крім того деякі курси вузької направленості потребують додаткових знань з інших предметів. Розробники курсів можуть доповнювати свій матеріал посиланнями на інші ресурси або учні можуть самостійно займатися пошуком в мережі. Але це займає багато часу і у перших і других користувачів ресурсу і не завжди дає необхідний результат.

Розглянемо системи електронного навчання або системи відкритих освітніх порталів. Вони постійно розширюються і наповнюються знаннями, але не одна з цих систем не має внутрішньої пошукової системи, яка б давала можливість шукати додаткову інформацію в межах системи, а також при необхідності розширювала цей пошук за межі системи використовуючи інші портали включаючи наукометричні бази даних. Забезпечення пошуку в межах однієї системи електронного навчання можна виконати за рахунок підключення додаткового модулю, що дасть можливість організувати повнотекстовий пошук інформації у базі даних системи.

Більшість готових рішень повнотекстового пошуку працюють за принципом індексного пошуку. На основі текстових даних будується індекс, який здатний дуже швидко шукати відповідності ключового слова.

Сервіс пошуку складається з двох компонент: пошуковик і індексатор. Індексатор отримує текст на вхід, робить обробку тексту – вирізання закінчень, незначущих слів і т.п. і зберігає все в індексі. Структура індексу дозволяє проводити по ньому дуже швидкий пошук. Пошуковик – інтерфейс пошуку за індексом – приймає від клієнта запит, обробляє фразу і шукає її в індексі. Розглянемо найбільш поширені технології.

Sphinx (англ. SQL Phrase Index) – система повнотекстового пошуку. За замовчуванням підтримує англійську і російську мову. Має інтерфейс для індексування таблиць MySQL. Інтегрується з існуючими СУБД (MySQL, PostgreSQL) і API для поширених мов веб-програмування. Конфігурація ділиться на source і index для визначення джерела даних і параметрів індексу, для пошуку підтримує MySQL протокол.

Конфігурація ділиться на 3 частини: джерела даних, параметри індексації та параметри самого сервера. Індекс включає в себе один або кілька джерел даних і шлях до місця, де цей індекс зберігається. Джерела повинні бути однорідні (однакова кількість стовпців). Після обробки пошукового запиту, Sphinx повертає індекс id знайдених записів. Відповідно в джерелі потрібно вказати, яке поле або поля будуть визначені як індекс id. При великих обсягах даних існує можливість використовувати схему main + Delta індексів для прискорення індексації. Замість одного індексу використовується два: основний, що

індексує всі дані та додатковий (або дельта), що індексує тільки нові дані, що з'явилися з часу останнього ре індексування. Пошукова система Sphinx підтримує Real Time індекси. Вони працюють тільки з SphinxQL, всі операції заповнення, видалення і зміна даних в індекс, відбуваються за SQL запитами. Особливість цього типу індексу полягає в тому, що дані потрапляють одночасно в базу даних і в індекс. Система дає можливість виконувати фільтрацію, сортування результатів пошуку, використовувати додаткові атрибути – групи, тимчасові мітки і т. п., використовувати стоп-слова для скорочення об'єму індексу, виконувати морфологічний пошук для різних мов, а також пошук для кожного документа за умовами wildcard, що передбачає пошук елементів за шаблонами, які замінюють пропущені букви в словах.

Apache Solr це платформа повнотекстового пошуку з відкритим вихідним кодом. Її основні можливості: повнотекстовий пошук; підсвічування результатів пошуку; фасетний пошук, що надає можливість сортування за декількома різнорідними категоріями одночасно, додаючи до пошуку нові параметри; динамічна кластеризація з підтримкою масштабування в кластер, дозволяє працювати з дуже великими обсягами даних і навантажень; інтеграція з базами даних; обробка документів зі складним форматом, наприклад, Word, PDF. Наявність розподіленого пошуку та реплікації надає можливість успішного масштабування системи. Крім звичайного текстового пошуку цей пошуковик може знаходити неточні відповідності, наприклад, при пошуку слів з помилками. Наявність HTTP / XML та JSON API надає можливість використовувати Solr з усіма популярними мовами програмування та гнучко налаштовувати і підключати до нього зовнішні модулі. Solr поєднує пошуковий індекс і сховище документів, що на відміну від Sphinx'a дає можливість зберігати документи цілком, тому їх не потрібно дублювати в базу даних.

Elasticsearch – технологія пошукового серверу. Являє собою повнотекстовий пошуковий движок з HTTP веб-інтерфейсом і підтримкою без схемних JSON документів. Поєднує пошук і сховище даних, має такі інструменти, як візуалізація, система шифрування, збирач логів для обліку змін в системі і т.п., масштабується і дозволяє виконувати вибірки дуже складної форми, що дає змогу використовувати цю технологію для аналітичної платформи. Може використовуватись для індексування та пошуку будь-яких типів документів за рахунок масштабованого пошуку, пошуку близького до реального часу і підтримки мультиарендності – моделі роботи, яка включає частину аспектів архітектури, моделі розгортання, і організацію обслуговування, тобто, безліч баз даних користувача може працювати при використанні однієї копії програмного забезпечення і однієї інсталяції – так званого "контейнера". Технологія Elastic має просунуту систему зберігання даних і протоколів запитів, тому в багатьох випадках її застосовують, як движок для Ad-hoc запитів – вузькоспеціалізованих запитів, що, створені для вирішення одноразової проблеми.

Elasticsearch має можливість розподілення, індекси можуть бути розділені по шардах горизонтальних і вертикальних, при чому кожен шард може мати нуль чи більше реплік. Кожен вузол містить один чи більше шардів і діє як координатор делегування операцій на потрібний шард. Балансування та маршрутизація виконується автоматично.

Всі розглянуті технології використовуються в тому випадку коли потрібно працювати з великою кількістю даних. Якщо кількість даних не є критичною тоді задачу пошуку в рамках системи можна вирішити завдяки більш простим алгоритмам, наприклад алгоритмам прямого пошуку. У них є, наприклад, необмежені можливості по наближеному і нечіткому пошуку. Адже будь-яке індексування завжди пов'язане зі спрощенням і нормалізацією термінів, а отже, з втратою інформації. Прямий пошук працює безпосередньо з оригінальним документом без жодних спотворень.

Крім того, є маса програм, що комбінує індексний пошук для знаходження блоку тексту з подальшим прямим пошуком всередині блоку, наприклад, Glimpse. Але в цьому випадку залишається питання організації гнучкої розширювальної структури даних, тому що будь які системи електронного навчання постійно поповнюються новою інформацією. Розглянемо це питання.

Одним з рішень може бути масштабування бази даних за рахунок створення шардів. Вертикальний шардінг надає можливість створити з великої таблиці даних багато маленьких за будь-яким принципом. Горизонтальний шардінг працює з таблицями, що лежать в різних базах на інших інстансах.

Але враховуючи, що бази даних систем електронного навчання можуть зберігати неоднорідну інформацію, тобто не тільки HTML сторінки, а також файли документів, відео та презентації краще використовувати іншу структуру даних. Наприклад, базу даних, що побудована за допомогою кластерів. Кластер в таких базах даних використовується для фізично спільного зберігання однієї або декількох таблиць, які часто з'єднуються разом в SQL-запитах і створюються за допомогою кластерних індексів [4]. В динаміці тобто в часі кластерні індекси в ході додавання даних розростаються і таблиця починає гілкуватися. Сторінка з даними розділяється на дві, причому в тому вузлі (на тій сторінці), де раніше були дані, тепер розташовується індекс, що охоплює обидві нові сторінки. Конкретний вузол такого дерева зобов'язаний включати в себе індекси всіх дочірніх вузлів або кінцеві дані, якщо вузол останній. Вузли можуть посилатися один на одного тільки в одному напрямку від батька до нащадка рис 1.

Якщо кількість даних буде збільшуватися, дерево буде ускладнюватися і поглиблюватися. І чим більше воно буде, тим більший виграш дасть така схема зберігання даних рис. 2.

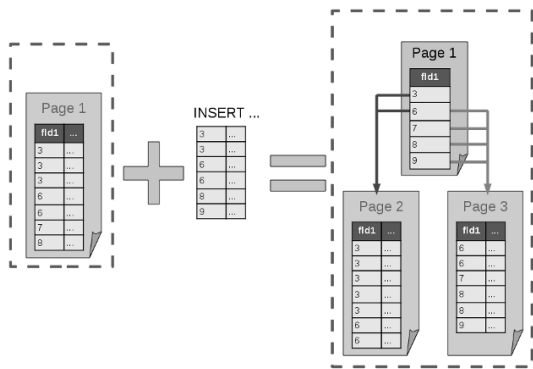


Рисунок 1 – Таблиця 1

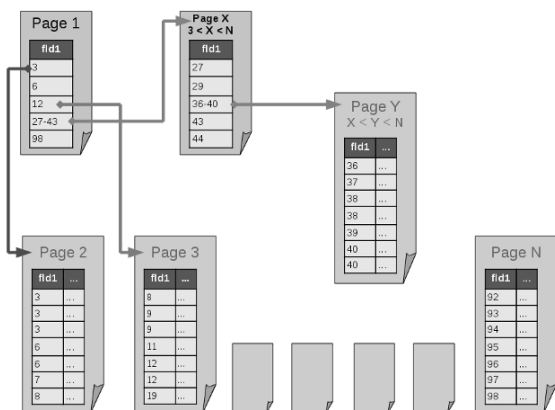


Рисунок 2 – Кластерне представлення таблиці 1

Кластери зберігають взаємопов'язані рядки різних таблиць разом в одних і тих же блоках даних, що дозволяє скоротити кількість операцій дискового введення-виведення і поліпшити час доступу для з'єднань таблиць, що входять в кластер.

Кластери не впливають на вже спроектовану модель даних, їх існування прозоро для користувачів і додатків. Дані, що зберігаються в кластері, обробляються за допомогою тих же інструкцій SQL, що і дані, що зберігаються в некластеризованих таблицях.

Кластери доцільно використовувати для зберігання однієї або декількох таблиць, для яких запити виконують з'єднання даних з декількох пов'язаних таблиць, або витягають пов'язані дані з однієї таблиці.

Кластери можна створювати на основі тематичної подібності [5]. Для цього під час створення кластерів проводиться лінгвістичний аналіз текстової інформації. Його основним завданням є структуризація і формалізація змісту текстів, виявлення понятійного складу предметної області, встановлення парадигматичних, синтагматичних і асоціативних зв'язків між найменуваннями понять і встановлення їх контекстного оточення. Для цього використовуються морфологічні, синтаксичні і семантичні словники, але методи, завдяки яким їх створює людина можна вважати лінгвістичними умовно. Автомати-

зоване складання словників найменувань понять можна виконати за такою схемою:

1. Формально-логічний контроль вихідних текстів з метою виявлення та виправлення орфографічних і синтаксичних помилок у вихідних текстах;
2. Членування вихідного тексту на окремі слова (за пропусками і роздільними знаками між ними);
3. Морфологічний аналіз слів корпусу текстів;
4. Членування корпусу текстів на речення;
5. Семантико-синтаксичний аналіз текстів;
6. Наближений концептуальний аналіз текстів;
7. Виділення найменувань понять;
8. Автоматичне приведення найменувань понять до їх канонічної форми;
9. Формування частотного словника найменувань понять;
10. Лінгвістичний аналіз частотного словника найменувань понять (виняток помилковою малоінформативною лексики);
11. Формування машинного уявлення концептуального словника.

Лінгвістичними або навколо лінгвістичними прийомами можна вирішити наступні задачі:

- автоматичне визначення мови документа;
- токенізація (графематичний аналіз): виділення слів, границі речення;
- виключення неінформативних слів (стоп-слів);
- лематизації (нормалізація, стемінг): приведення словозмінної форм до «словникової» (в тому числі і для слів, що не входять в словник системи);
- поділ складних слів (компаундів) для деяких мов (наприклад, німецької);
- дізамбігуація: повне або часткове зняття омонімії;
- виділення іменних груп.

У використанні кластерів є і свої недоліки. Воно уповільнює операції додавання, оновлення, видалення рядків таблиці в порівнянні зі зберіганням таблиці поза кластером зі своїм власним індексом. Крім того кластери використовують додатковий обсяг дискової пам'яті, оскільки кожна окрема таблиця в кластері займає більше блоків, ніж коли вона зберігається поза кластером. Тому перед створенням кластеру необхідно проводити аналіз з точки зору планованого виграшу в продуктивності запитів, який повинен перевищувати додаткову витрату ресурсів на супровід кластера.

Для організації пошуку додаткової інформації в мережі інтернет за межами системи електронного навчання доведеться використовувати пошукові машини тому що наведені технології працюють з базами даних одного додатку. Для того щоб зробити пошук цілеспрямованим і шукати інформацію не тільки серед інтернет ресурсів представлених веб-документами потрібно провести дослідження інтернет середовища з приводу систематизації інформаційних освітніх і наукових порталів. Наукова і освітня інформація в мережі представлена різними мо-

вами, тому необхідно використовувати механізми швидкого автоматизованого перекладу запитів. Для вирішення цієї задачі розглянемо можливості пошуку у вебі.

Основні алгоритми за якими працюють пошукові машини це алгоритми ранжирування і індексування веб-сторінок [6].

Перший являє собою ітеративний алгоритм у поєднанні з пошуком по лексичній посилань, що вказують на сторінку. Розрахунок статичної популярності включає визначення порядку обходу документів, ранжирування пошуку по тексту посилань і т.д. Формули розрахунку популярності постійно покращуються за рахунок внесення обліку додаткових чинників - тематичної близькості документів, наприклад, популярна пошукова система www.teoma.com, її структури і т.п., що дозволяє знизити вплив непотизму.

Другий використовується для відсікання схожих документів. Він заснований на використанні шинглів - контрольних сум для кожного десятислів'я тексту, які йдуть в нахльст, з перекриттям, для того щоб всі були враховані. Таким чином можна визначити відсоток перекриття текстів, виявляти всі його джерела і т.п.

Пошукові системи для інтернету можуть прискорювати свою роботу за допомогою технік ешелонування і прюнінга [7].

Перша техніка полягає в поділі індексу на свідомо більш релевантну і менш релевантну частини. Пошук спочатку виконується в першій частині, а потім, якщо нічого не знайдено або знайдено мало, пошукова система звертається до другої частини індексу. Прюнінг полягає в тому, щоб динамічно припинити обробку запиту після накопичення достатньої кількості релевантної інформації. У статичному прюнінгу на підставі деяких припущень індекс скорочується за рахунок таких документів, які свідомо ніколи не будуть знайдені.

Для спілкування між пошуковими серверами і серверами, які збирають відгуки і формують сторінку видачі, розробляються спеціальні протоколи щоб забезпечити безперебійну роботу багатокомп'ютерних комплексів, безшовне оновлення індексу, стійкість до збоїв і затримок з відповідями окремих компонентів.

Виконання пошуку інформації можна представити у трьох процесах: сканування, індексування та позиціонування.

Сканування може бути описано, як автоматизований процес систематичного вивчення загальнодоступних сторінок в Інтернеті, виявлення нових або оновлених сторінок і додавання їх в свою базу. Для оптимізації цього процесу використовуються спеціальні програми, які носять назву «боти» або «роботи». Вони відвідують список URL-адрес, отриманих в процесі минулого сканування і доповнених даними карти сайту, яку надають веб-майстри і аналізують їх зміст. При виявленні посилань на інші сторінки під час відвідування сайту, боти також додають їх до свого списку і встановлюють систематичні зв'язки.

Індексація – процес збереження отриманої інформації в базі даних відповідно до різних факторів

для подальшого вилучення інформації. Індексація використовує ключові слова на сторінці, їх розташування, мета-теги і посилання. Основна мета процесу індексації: швидко реагувати на пошуковий запит користувача.

Під час обробки згідно запиту користувача в базі даних виконується пошук, що відповідає умовам і алгоритмічно визначає актуальність змісту. Це призводить до певного рейтингу серед знайдених сайтів. Результати, які вважаються більш доречними для користувача пошукової системи, навмисно отримують більш високий ранг, ніж результати, які мають менше шансів забезпечити адекватну відповідь.

Архітектурно сучасні пошукові системи являють собою складні багатокомп'ютерні комплекси. В міру зростання системи основне навантаження лягає на пошук. Протягом секунди приходять десятки і сотні запитів для обробки яких індекс розбивають на частини і розкладають по десяткам, сотням, тисячам комп'ютерів.

Всі вищеописані процеси сканування, індексування та позиціонування можна представити за допомогою схеми (рис. 3).

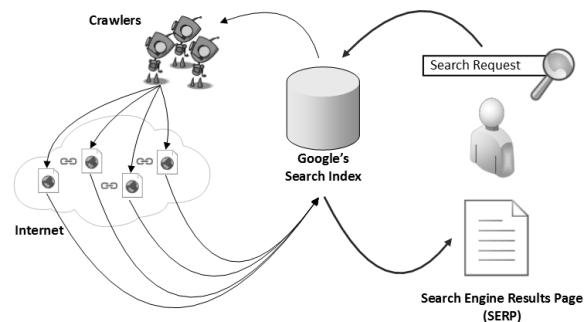


Рисунок 3 – Схема процесів пошуку

Але всі ці алгоритми дають розширений пошук, для того щоб уникнути зайвої інформації потрібно використання алгоритмів звуження пошуку для результатів пошуку пошукових машин. З цією задачею може впоратися штучний інтелект. Він застосовується для самонавчання алгоритмів з метою поліпшення якості пошуку, наприклад Яндекс для оптимізації процесу пошуку, враховуючи оцінки асесорів, використовує Матрикснет. Який будує формули ранжирування з безлічі факторів на основі інформації про сайти і поведінку людей. Фактично, формулу ранжирування пише машина. Ще один алгоритм штучного інтелекту, який використовує Яндекс носить назву «Палех» [8]. Він працює з запитом користувача і заголовком сторінки, порівнює їх на відповідність один одному за змістом. Для цього текст запиту і текст заголовка представляються у вигляді векторів, скалярне множення яких повинно бути більше, ніж релевантні запити на документ з цим заголовком. Нейронна мережа повинна бути навчена таким чином, щоб для близьких за змістом текстів вона генерувала схожі вектори, а для семантично незв'язаних запитів і заголовків вектори повинні відрізнитися.

Google теж застосовує штучний інтелект в пошуку – система RankBrain. Вона обробляє результати пошуку і використовує принципи машинного навчання, що дозволяє системі вдосконалюватися без участі людини або без чіткого слідування заданим алгоритмам. RankBrain являє собою один з факторів, що враховуються в ранжируванні. Першим за важливістю чинником ранжирування в пошуку Google залишаються посилання, а другим – слова, вірніше входження ключів.

Штучний інтелект RankBrain покликаний визначати найбільш релевантні з точки зору запитів користувача, сайти. Він використовується для інтерпретації користувальницьких запитів і встановлення смислових зв'язків між двома абсолютно різними пошуковими запитами, наприклад: взуття для бігу, спортивне взуття, кросівки. RankBrain дозволяє показати користувачеві релевантні сторінки, які не містять прямих входжень ключових запитів.

Штучний інтелект RankBrain отримує в оффлайн режимі групи історичних запитів і вибудовує прогнозовані видачі на основі аналізу отриманих запитів. Прогнози перевіряються на відповідність стандартам якості пошуку, і якщо прогнозована видача релевантна і якісна, запускається нова версія RankBrain. Штучний інтелект вибудовує асоціації між словами запиту, щоб краще зрозуміти «intent» – намір користувача.

Зробивши аналіз процесу пошуку в мережі можна зазначити, що він більш складний ніж пошук у базі даних системи тому пропонується розв'язати ще одну задачу – забезпечення доступу до наукової і методичної інформації без виконання повторного пошуку. Тобто зберігати посилання на ресурси з їх описанням у базі даних ресурсу і під час повторного пошуку спочатку перевіряти чи виконувався цей пошук раніше.

ВИСНОВКИ. В ході роботи було проаналізовано нові форми освіти що з'явилися завдяки використанню інформаційних технологій, які створюють інформаційний освітній простір разом з наукометричними базами даних. Запропоновано шляхи розвитку освітнього інформаційного простору за раху-

нок використання різноманітних технологій пошуку інформації, організації структур даних і використання механізму швидкого автоматизованого перекладу запитів для пошуку необхідної інформації в багатомовному контенті.

ЛІТЕРАТУРА

1. Шахіна І. Ю. Визначення і напрями створення інформаційного освітнього середовища // Проблеми та перспективи формування національної гуманітарно-технічної еліти : зб. наук. пр. – Харків: НТУ "ХПІ". – 2013. – Вип. 36-37 (40-41). – С. 245–255.
2. Heather Staker, Michael B. Horn. Classifying K-12 Blended learning. // Innosight Institute. – 2012. Available: <https://www.christenseninstitute.org/wp-content/uploads/2013/04/Classifying-K-12-blended-learning.pdf>
3. Dennis Vilorio. STEM 101: Intri to tomorrow's jobs. – 2014. Available: <http://www.stem-edcoalition.org/wpcontent/uploads/2010/05/BLS-STEM-Jobs-report-spring-2014.pdf>.
4. Павлов Д. Кластерные СУБД. // Jet Info №12. – 2009. – Режим доступу до ресурсу: <http://www.jetinfo.ru/stati/klasternye>
5. Захаров В., Хорошилов А., Хорошилов А. Опыт создания кластеров документов на основе метода определения их тематического подобия. – 2012. – Режим доступу до ресурсу: http://rcdl.ru/doc/2014/paper/RCDL2014_322-328.pdf
6. Cormen Thomas H., Leiserson Charles E., Rivest Ronald L., Stein Clifford. Introduction to Algorithms. 3rd. – MIT Press, 2009, 1292 p.
7. Raymie Stata, Krishna Bharat, Farzin Maghoul. The Term Vector Database: fast access to indexing terms for Web pages // In Proceedings of the 9th World Wide Web Conference, Amsterdam, Nezerlands, May 2000.
8. Сафронов А. Искусственный интеллект в поиске. Как Яндекс научился применять нейронные сети, чтобы искать по смыслу, а не по словам. – 2016. – Режим доступу до ресурсу: <https://habr.ru/company/yandex/blog/314222/>.

EXPANDING THE EDUCATIONAL INFORMATION SPACE WITH THE SEARCH TECHNOLOGIES

T. Hryhorova

Kremenchug Mykhailo Ostrohradskyi National University

vul. Pershotravneva, 20, Kremenchuk, 39600, Ukraine E-mail: grital0403@gmail.com

Purpose. Carrying out of researches for studying of a way of educational information space expansion using possibilities of modern information technologies. **Methodology.** The research of the components of the educational information space has been conducted and the possibility of using the search for obtaining information from all available resources has been studied in order to organize scientific, educational and creative activities. **Results.** It has been revealed that e-learning systems are the main component of the educational space. It has been proposed by the search modules to expand these systems or search algorithms by changing the data structure in this case. The system adds an additional module to translate the query string to be able to use the information presented on the network in different languages, organize the ability to search for information directly from the system on the Internet, add a module allowing you to narrow the search, which is conducted outside the system. Database system with additional information about the search results to expand, to avoid a second search on the Internet. **Originality.** For the first time, ways of expanding the educational information space have been proposed, through the organization of a special search in the information bases of educational electronic systems and scientometric databases. **Practical value.** The implementation of this proposal will combine educational and scientific knowledge, which are located in the databases of the Internet and due to this, to reduce the search for the necessary educational and scientific information.

Key words: educational space, information search, data structure.

REFERENCES

1. Shahina, I.U., (2013), "Definition and directions of creating an informational educational environment", *The problems and prospects of training national humanitarian-technical elite*, Kharkov: NTU "KhPI", V. 36-37 (40-41), pp. 245-255.
2. Heather, S., Horn, B.M. (2012), "Classifying K-12 Blended learning", *Innosight Institute*, available: <https://www.christenseninstitute.org/wp-content/uploads/2013/04/Classifying-K-12-blended-learning.pdf>
3. Dennis, V. (2014), "STEM 101: Intri to tomorrow's jobs", available: <http://www.stem-edcoalition.org/wp-content/uploads/2010/05/BLS-STEM-Jobs-report-spring-2014.pdf>.
4. Pavlov, D. (2009), "Clustered DBMS", *Jet Info*, №12, available: <http://www.jetinfo.ru/stati/klasternye>
5. Zaharov, V., Horoshilov, A.A., Horoshilov, A.A., (2012), "Experience in creating clusters of documents on the basis of the method of determining their thematic similarity" [Опыт созданиya klasterov dokumentov na osnove metoda opredeleniya ikh tematicheskogo podobiya], available: http://rcdl.ru/doc/2014/paper/RCDL2014_322-328.pdf
6. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C. (2009), "Introduction to Algorithms", 3rd., *MIT Press*.
7. Stata, R., Bharat, K., Maghoul, F. (2000), "The Term Vector Database: fast access to indexing terms for Web pages", *In Proceedings of the 9th World Wide Web Conference, Amsterdam, Nezerlends*.
8. Safronov, A., (2016), "Artificial intelligence in the search. How Yandex learned to apply neural networks to search by sense, and not by words" [Iskusstvennyy intellekt v poiske. Kak Yandeks nauchilsya primenyat' neyronnyye seti, chtoby iskat' po smyslu, a ne po slova], available: <https://habrahabr.ru/company/yandex/blog/314222/>.

Стаття надійшла 06.06.2017.