

## РОЗРОБКА СЕМАНТИЧНОГО ЯДРА ПРОФОРІЄНТАЦІЙНОГО САЙТУ НА ОСНОВІ SOM

К.т.н. О.О. Арсирій, Ю.М. Ларченко

Одеський національний політехнічний університет

Україна, м. Одеса

larec.ostrov@mail.ru

Специфіка об'єкта управління, слабкий розвиток інформаційних систем та загострення проблеми профорієнтації в освітньої галузі роблять актуальним завдання розробки відвідуваних інтернет-ресурсів для випускаючих кафедр вищих навчальних закладів [1], які мають забезпечити доступ до довідкової та оперативної інформації абітурієнтів та зацікавлених осіб. Більшість потенційних абітурієнтів потрапляють на сайти кафедр вищих навчальних закладів при переході за посиланнями у результатах пошукових систем (ПС), робота яких заснована на визначенні «ключових слів» веб-ресурсу [2]. З переліку таких слів складається семантичне ядро сайту (СЯС). СЯС являє собою список ключових слів і їх комбінацій, записаних в метатегах keywords і розподілених в контенті сайту, а саме, у тезі title, в alt-атрибутах, в тексті внутрішніх і зовнішніх посилань, у виділеннях жирним і похилим шрифтом, в назві файлів, в URL та ін. Від повноти і точності розробки СЯС залежить положення сайту в списку видач ПС.

Розробка СЯС складається з ряду інтелектуальних, важкоформалізуємих етапів і процедур, для реалізації яких необхідні великі часові і людські ресурси. На першому етапі для профорієнтаційного сайту необхідно оцінити структуру і цільову аудиторію. Наступним етапом буде створення первинного списку запитів. Для цього можна використовувати різні варіанти назв навчального закладу чи навчального підрозділу, спеціальностей, назв дисциплін що викладаються, назви наукових праць, різні дії, надані користувачам. Потім складається список асоційованих запитів за допомогою засобів статистики пошукових систем (wordstat.yandex, google/insights/search/, та ін) і підраховується частота ключових слів. Ключові слова з найбільшою частотою поміщають в метатеги keywords, з меншою -

розподіляють по контенту сайту. Однак, для сайтів з динамічним контентом, таких як сайт кафедри, де ведеться новинний блог, де змінюються назви й перелік спеціальностей, з'являються нові наукові праці й видання, перераховані етапи розробки СЯС необхідно повторювати досить часто. Тривалість кожного етапу може значно затримувати необхідну періодичність повторення, що призводить до зниження повноти і точності СЯС, і сайт втрачає свої позиції в SERP (Search Engine Result Page) – сторінка видачі пошукових результатів. Для скорочення часу розробки СЯС з динамічним контентом без втрати повноти і точності в даному дослідженні пропонується використовувати самоорганізовані карти Кохонена (Self-organizing map - SOM) [3-4], що дозволяють виконати завдання класифікації текстових корпусів і навчаються без вчителя. SOM - метод, розроблений для відображення багатовимірних даних на двовимірну площину - ще один спосіб візуалізації даних. Параметри класів формуються в результаті спеціального алгоритму навчання мережі, застосовуваного до безлічі документів і спеціальним чином організованих навчальних наборів. Можна сказати, що алгоритм здійснює кластеризацію документів.

В процесі дослідження розроблено відвідуваний інтернет-ресурс для випускаючих кафедр вищих навчальних закладів.

Впровадження автоматизації розробки семантичного ядра сайту дозволить актуалізувати ключові слова. Це зробить СЯС більш точним і повним, що підвищить положення сайту у SERP – більше потенційних абітурієнтів та зацікавлених осіб зможуть потрапити на сайт і отримати довідкову та оперативну інформацію.

#### ВИКОРИСТАНІ ДЖЕРЕЛА

1. Солдатов А. В. Информационная система как основа эффективного управления вузом // Университетское управление: практика и анализ. - 2004. - N 2(31). - С. 116-119.
2. Как работают поисковые системы — сниппет, алгоритм обратных индексов, индексация страниц, особенности работы поисковиков. [Электронный ресурс] / Режим доступа: <http://ktonanovenkogo.ru/seo/search/kak-rabotayut-poiskovye-sistemy-snip-pet-index.html> - 2012 г;
3. Кохонен Т. Адаптивные и интеллектуальные системы // Самоорганизующиеся карты // БИНОМ. Лаборатория знаний.: Москва, 2008. – 665 с.
4. Кириченко К.М, Герасимов М.Б. Обзор методов кластеризации текстовой информации // STAR SPb (Russia) / Режим доступа: [http://www.dialog-21.ru/Archive/2001/volume2/2\\_26.htm](http://www.dialog-21.ru/Archive/2001/volume2/2_26.htm)