УДК 004.94
Козуб В. – ст. гр. СНм-51
Тернопільський національний технічний університет імені Івана Пулюя

ОСОБЛИВОСТІ РЕЙТИНГІВ ВЕБ-САЙТІВ

Науковий керівник: асистент Маєвський О. В.

Для ранжування сторінок *Google* використовує алгоритм PageRank — метод вимірювання «важливості» сторінки. Коли всі інші фактори, такі як тег Title і ключові слова враховані, Google використовує PageRank, щоб відкоригувати результати так, що більш «важливі» сайти піднялися вгору на сторінці результатів пошуку користувача. Теорія Google говорить, що, якщо Сторінка А посилається на сторінку В, то Сторінка А вважає, що Сторінка В - важлива сторінка. Текст посилання не використовується в PageRank. PageRank також впливає на важливість посилань на сторінку. Якщо на сторінку вказують багато важливих посилань, то її посилання на інші сторінки також стають більш важливими.

В свою чергу, *Yandex* розробив для ранжування веб-сайтів власний алгоритм «Снежинск».

Навчання функції ранжування в алгоритмі відбувається за допомогою «жадібних» функцій і полягає у виборі з множини чинників тих, які більшою мірою впливають на релевантність, з огляду на ваги їх класів. На першому кроці за допомогою асесорів (людей, що надають незалежну оцінку релевантності документа) будується еталонна модель, в якій кожному документу з множини документів D = (D1, D2, ...) відповідно до запиту з множини запитів Q = (Q1, Q2, ..., Q3) виставляється релевантність.

Потім за допомогою генетичного алгоритму будується функція релевантності для даного набору, що дає максимально близький до еталонної моделі результат. Отримана функція за визначенням не може дати оптимальний варіант, тому навчання відбувається ітераційно (у кілька проходів). На кожній ітерації вирішується своя локальна оптимізаційна задача, яка наближає результат, отриманий в ході навчання, до еталону шляхом пошуку нових функцій. Напрямок і довжина нової функції визначається градієнтною апроксимацією. В якості аргументів для цих функцій виступають і самі чинники, і функції від цих чинників. Процес навчання безкінечний і при додаванні нової функції загальна формула релевантності буде зазнавати зміни, і загальний внесок кожного фактора в значення функції релевантності буде змінюватися.

Таким чином, функція релевантності є лінійною комбінацією знайдених функцій і виглядає як поліном від N змінних: де N сягає кількох тисяч. На початку навчання алгоритм ранжування видає дещо грубі результати, але в процесі навчання він вдосконалює свої знання, прагнучи глянути на документи очима користувача, імітацією якого виступають асесори.

Найвідомішим рейтингом веб-сайтів наукових установ є *Webometrics.*

Згідно з методикою, розробленою Webometrics, рейтинг наукових установ формується на основі 4 показників — розміру домену університету (S), «видимості» домену (V), об’єму файлів, розміщених в домені (R) та кількості робіт, що містять посилання на домен і проіндексовані системою Google Scholar (Sc).