

Класифікація ключових слів семантичного ядра веб-спільноти

Катерина Слобода

Кафедра соціальних комунікацій та інформаційної діяльності, Національний університет “Львівська політехніка”, УКРАЇНА, м.Львів, вул.С.Бандери, 12,
E-mail: sloboda@misto.ridne.net

Keywords that are part of the same semantic class can replace each other in the process search. That is, if the user requests contain one of the keywords, the search will be issuing all posts that contain the keywords of this class. This approach allows increasing the percentage of completeness and accuracy of search.

Ключові слова – класифікація ключових слів, класифікація тем підфорумів, семантичне ядро, повнота пошукової видачі, точність пошукової видачі, класи семантичного ядра.

I. Вступ

З точки зору інформаційного пошуку, доцільно розрізняти два типи класифікацій інформаційного наповнення сайту веб-спільноти:

- Класифікація ключових слів
- Класифікація тем підфорумів та форумів

Класифікація ключових слів відбувається в межах семантичного ядра і передбачає їх групування на основі семантичних ознак.

Класифікація тем підфорумів відбувається в межах форуму. Ключові слова, які присвоюються дописам в межах дискусій є основою для побудови класів тем підфорумів.

При класифікації ключових слів, в один клас попадають різні за частотністю ключові слова.

Ключові слова, які входять до одного класу семантичного ядра можуть замінити один одного в процесі пошуку. Тобто, при попаданні в запит користувачів одного з ключових слів, в пошуковій видачі будуть всі дописи, які містять ключові слова з цього класу. Такий підхід дозволяє збільшити показник повноти та точності пошукової видачі.

II. Основна частина

Основною проблемою класифікації ключових слів є та, що в більшості випадків вони задані некоректно, оскільки більшість інформаційного наповнення веб-спільноти генерується користувачами цієї спільноти, і для опису оптимальних моделей класифікації неможливо підібрати відповідних критеріїв. Проте, оскільки основною ціллю інформаційного пошуку є видача релевантних даних у відповідь на запити користувачів, тому при розробці семантичного ядра, доцільно зосередити увагу на створенні емпіричних класифікацій ключових слів.

Системи класифікації мають різні формальні властивості, зокрема, класифікація може бути монотетична і політетична [2].

Монотетична класифікація передбачає, що всі ключові слова, що належать до одного класу мають певну спільну ознаку.

Політетична класифікація цього не передбачає. Для семантичного ядра сайту веб-спільноти використання політетичної класифікації є більш доцільним.

Класи ключових слів можуть бути такими, що пересікаються, тобто один клас містить ключові слова, які належать до кількох класів і такими, що не пересікаються. Для семантичного ядра форуму доцільно використовувати обидва типи. В класах, що пересікаються є ключові слова, які вживається в дописах підфорумів різної тематики. В класах, що не пересікаються є ключові слова, які вживаються лише в межах одного підфоруму.

Основними характеристиками класів семантичного ядра: кількість класів в межах ядра, розмір цих класів та їх перетин. Ці характеристики оприділяють ефективність процесу класифікації семантичного ядра з точки зору інформаційного пошуку.

Формування класів семантичного ядра передбачає початковий процес, при якому формується кожен окремих клас семантичного ядра, процес розподілення ключових слів за класами та кінцевий процес, при якому виконуються умови, за яких кожен окремих клас може вважатися повним.

Висновок

Для ефективного позиціонування сайту веб-спільноти, увагу слід зосередити на якісному формуванні семантичного ядра. Класифікація ключових слів в межах семантичного ядра дозволяє суттєво покращити показники релевантності, повноти та точності пошуку інформаційного наповнення веб-спільноти. Класифікація ключових слів є надзвичайно важливим завданням, яке постає перед власниками веб-спільноти в процесі її позиціонування.

Література

1. Пелешин А.М. Модифікація контенту для ефективного позиціонування форуму в середовищі World Wide Web / А.М. Пелешин, К.О. Слобода // Інформаційні системи та мережі: Вісник Національного університету «Львівська політехніка». – 2010. – №637. – С.340–348.
2. Martin T.H., Proceedings of the AFIPS Workshop on the User Interface for Interactive search of Bibliographic Data Bases, Stanford University, Palo Alto, Calif., January 1971
3. Solton D., Dynamic Information and Library Processing, Prentice-Hall, Inc., Englewood, New Jersey, 1975, P. 337-343