

Система класифікації текстів інтелектуального порталу знань

Юр'єв П.М., *аспірант*; Білоцерківець А.М., *студент*
Сумський державний університет, м. Суми

Сучасні портали знань використовують багаторівневий комплекс алгоритмів глибинного аналізу текстової інформації, поданої природними мовами, та інформаційного синтезу класифікаторів або рубрикаторів з застосуванням ефективних інтелектуальних технологій. При цьому інтелектуальна складова може додаватися до практично кожного з існуючих алгоритмів автоматичного оброблення текстів, включаючи семантичний, лінгвістичний, лексикографічний, морфологічний, латентно-семантичний аналіз.

В роботі пропонується використання інформаційно-екстремальної інтелектуальної технології для розроблення математичних моделей та алгоритмів оптимізації функціональних параметрів порталу, який навчається розпізнаванню тематичного розділу знань текстів, що аналізуються. При формуванні вхідного математичного опису портал знань оперує алфавітом класів розпізнавання $\{X_m^o \mid m = \overline{1, M}\}$, який характеризує M тематичних розділів знань, та навчальною матрицею типу «об'єкт-властивість» $\|y_{m,i}^{(j)}\|$, $i = \overline{1, N}$, $j = \overline{1, n}$, де N , n – кількість ознак розпізнавання і реалізації текстів певного тематичного розділу відповідно. При цьому словник ознак розпізнавання для формується шляхом об'єднання множин ключових слів реалізацій текстів. На етапі навчання для апріорно класифікованого алфавіту класів $\{X_m^o\}$, що утворює нечітке розбиття $\tilde{\mathfrak{R}}^{|M|}$, було побудовано шляхом допустимих перетворень в бінарному субпарацептуальному дискретному просторі ознак оптимальне в інформаційному розумінні чітке розбиття $\mathfrak{R}^{|M|}$ еквівалентності класів, яке визначає безпомилкові за $\|y_{m,i}^{(j)}\|$ вирішальні правила для визначення тематичного розділу знань поточного тексту. При цьому як критерій оптимальності параметрів порталу знань використовувалися модифікації ентропійного критерію Шеннона та інформаційної міри Кульбака.

Керівник: Шелехов І.В., *доцент*