УДК 681.518

Д.Ю. Катюха

*Національний авіаційний університет*

ОБ’ЄКТНИЙ ПІДХІД ДО КОМП’ЮТЕРНОГО АНАЛІЗУ
ТЕКСТІВ ЗА ЗМІСТОМ

Порівняльний аналіз текстів машиною (комп’ютером) в наш час є досить складним завданням, адже будь-яка природна мова є надзвичайно складною і містить безліч тонких, не завжди помітних зв’язків, що надає складності її аналізу для машини. Також одним з факторів, що ускладнює розвиток комп'ютерного аналізу текстів, є можливість одного і того ж слова/фрази приймати різні (а іноді і зовсім протилежні) значення; можливість не однозначно трактувати те чи інше твердження; відсутність чіткої структури речення (особливо в розмовній мові). Ще одним важливим фактором, що підвищує складність аналізу є зміна властивостей об’єктів та середовища в часі.

Ідея даного методу полягає в об’єктному підході до аналізу тексту комп’ютером. Кожне речення описує якийсь об’єкт (дію, що теж є об’єктом) і набір властивостей даного об’єкта (і/або дію, що виконується об’єктом, або над об’єктом) та/або зв’язки даного об’єкта з іншими (наприклад, порівняння: “об'єкт1 світліший за об'єкт2”). Таким чином утвориться певна структура об’єктів, пов’язаних між собою, які матимуть набір властивостей, що описує реальний об’єкт. Порівняння і визначення змісту здійснюватиметься за базою синонімів, де слова мають бути представлені у вигляді синонімічних рядів. Причому різні синонімічні ряди можуть бути пов’язані між собою ступенем близькості значення.

Серед недоліків об’єктного підходу до комп’ютерного аналізу тексту за змістом є перш за все складність його реалізації, оскільки доведеться опрацьовувати великі масиви даних, що пов'язані між собою.