

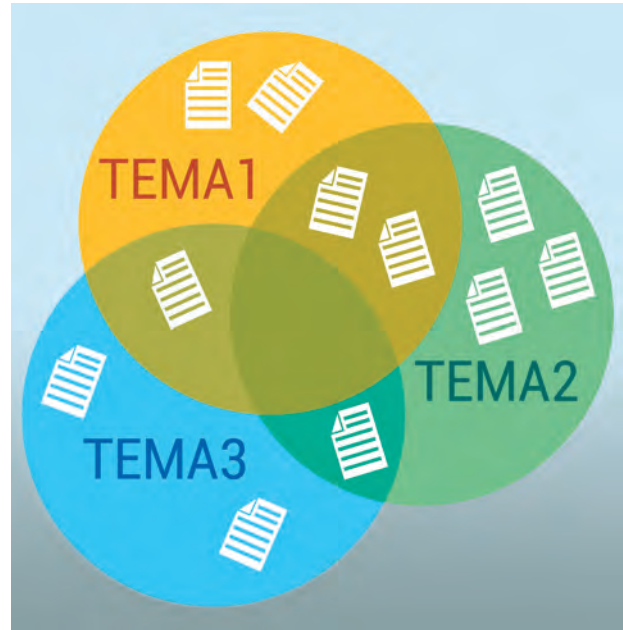
СИСТЕМА АВТОМАТИЧНОЇ КЛАСИФІКАЦІ ТЕКСТІВ DocSorter

Призначення

Система призначена для віднаходження ознак тематичного чи емоційного класу тексту. Основний ринок споживачів — новинні агенції та аналітичні центри, які здійснюють аналіз відкритих джерел

Характеристики

Система визначає елементи тексту, що відповідають певним поняттям. Це дає змогу визначити тематичні елементи, наявні в тексті, а за їх допомогою — теми текстів. Система навчається на прикладах текстів, зібраних у тематичних каталогах, тому може оцінювати різні теми, такі як фізика, математика, політика, спорт тощо. За умови використання наборів текстів, що відповідають емоційно забарвленим обставинам, наприклад відгуки на фільми, алгоритм може бути використаний для визначення емоційного забарвлення тексту. Технічні вимоги: операційна система Windows XP або новіша



Рівень готовності розробки. Пропозиції до комерціалізації

IRL3, TRL5

На замовлення можлива оптимізація системи відповідно до потреб споживача: як до типів текстів, так і до потреб кінцевих користувачів

Переваги

Порівняно з іншими системами DocSorter орієнтований на роботу з англійською, українською та російською мовами. Завдяки використанню семантичного аналізу та, зокрема, лексико-семантичної бази UkrWordNet система орієнтується на змістовну близькість між елементами тексту, а не на безпосередній словник тексту

Охорона інтелектуальної власності

IPR1, IPR3

Контактна інформація

Тарануха Володимир Юрійович, Міжнародний науково-навчальний центр інформаційних технологій та систем НАН України та МОН України, +38 044 502 63 19, e-mail: dep165@irtc.org.ua