Б.Г. Масловський, к.т.н.

*Національний авіаційний університет*

ТЕХНОЛОГІЇ ІНФОРМАЦІЙНОГО ПОШУКУ В НАВЧАЛЬНОМУ ПРОЦЕСІ

Основну частину знань студенти одержують в результаті порівняння, аналізу та синтезу інформації з розрізнених фактів, розміщених у текстах. При роботі з великими потоками документів процес автоматичного структурування текстової інформації заміняє експертний процес виділення фактів і об’єктів, виконуваний вручну. У доповіді розглядаються приклади використання нових технологій витягу знань із текстів українською та російською мовами.

До 85% нових знань студенти одержують, вивчаючи тексти. У найближчому майбутньому найбільш затребуваними стануть системи з максимально автоматизованими процесами структурування контента (extract, transfer, load - «витяг, перетворення, завантаження»). Важливою рисою таких систем буде функція оперативного аналізу інформації, отриманої за запитом для вибору подальшого напрямку дослідження документів.

До найактуальніших засобів інтелектуального аналізу текстів відносяться технології виділення фактографічної інформації про об’єкти з урахуванням анафоричних посилань на них, нечіткий пошук, тематичне і тональне рубриковання, кластерний аналіз сховищ і підбір документів, виділення ключових тем, побудова анотацій, побудова багатовимірних частотних розподілів документів і їх дослідження за допомогою OLAP-технологій.

У сучасних системах використовується двофазна технологія аналітичної обробки. У першій фазі проводиться автоматизований аналіз окремих документів, структуризація їх контента і формування сховищ початкової і аналітичної інформації. У другій фазі – витягання в оперативному режимі знань із сховища або з одержаної за запитом підборки документів.