

ПОИСК И АНАЛИТИЧЕСКАЯ ОБРАБОТКА ТЕКСТОВОЙ ИНФОРМАЦИИ ПРИ АВТОМАТИЗАЦИИ ПОДГОТОВКИ КАДРОВ

С.Ф. Липницкий, Л.В. Степура

Одним из приоритетных направлений совершенствования системы подготовки кадров в различных предметных областях является использование современных технологий поиска в Интернете необходимой научно-технической информации и ее аналитическая обработка. В докладе предложен подход к моделированию

информационного поиска на основе тематических корпусов текстов, т.е. совокупностей текстов по конкретной тематике. Множество всех тематических корпусов образует полный корпус текстов. Предложенный подход к поиску отличается универсальностью, т. е. независимостью от предметной области. Настройка системы на конкретную предметную область сводится к созданию соответствующего тематического корпуса текстов и актуализации словарей базы знаний. Функциональными компонентами проектируемой системы являются три подсистемы: автоматизированное рабочее место (АРМ) эксперта-лингвиста; подсистема поиска текстовых документов в Интернете; подсистема реферирования найденных документов. АРМ эксперта-лингвиста — это программно-информационный инструментарий, предназначенный для визуализации процедур создания и ведения баз данных и знаний, а также управления этими процедурами. В базе данных эксперт-лингвист накапливает и классифицирует по тематическим разделам предметной области электронные варианты текстовых документов для последующего создания на их основе тематических корпусов текстов. На основе информации из базы данных в системе формируются лингвистические словари базы знаний. Основными функциями подсистемы поиска являются индексирование веб-страниц на основе использования словарей базы знаний и поиск текстовых документов по запросам пользователей. Основной функцией подсистемы реферирования является выявление в реферируемом тексте кортежа информативных предложений и синтез реферата.