

0.1. Еримбетова А.С., Ефимова Л.В. Анализ текстов на естественном языке с помощью синтаксического анализатора Link Grammar Parser и семантической компоненты системы Dialing

Работа направлена на исследование базовых понятий, относящихся к формальным онтологиям и тезаурусам и на разработку практических методов по проектированию, созданию и применению онтологий в интеллектуальном программном обеспечении. Центральное место отводится тематике информационного поиска и анализа текстовой информации. Рассматриваются тексты на русском, английском и ряде тюркских языков таких, как: казахский, узбекский (варианты на кириллице и латинице) и турецкий В качестве программного инструментария для создания онтологий используются: синтаксический анализатор Link Grammar Parser [1] и семантическая компонента системы Dialing [2]. В дальнейшем также предполагается использовать редактор Protege [3]. В части, посвященной проектированию и созданию реальных онтологий рассматриваются онтологии для описания социально-экономических процессов. В частности, процессов происходящих в «быстрорастущих» экономиках: Китай, Корея, Тайвань, Индонезия, Япония и др. Предполагается приложить определенные усилия для создания онтологий, которые могут помочь при анализе процессов, происходящих в Центральной Азии. Основой для исследований служит статья [4], в которой авторы рассматривают процесс автоматического резюмирования. При этом из текста могут выделяться фрагменты, соответствующие заданной теме. Фрагменты не обязательно следуют друг за другом. В тексте между ними могут присутствовать вставки на другие темы. Далее выделенные фрагменты могут быть объединены в резюме по данной теме. В общем случае из текста могут быть выделены несколько таких резюме, соответствующих различным темам. Одна из возникающих проблем состоит в том, что перестановка слов в предложении может существенно менять его смысл, что приводит к некорректной работе алгоритмов, оперирующих отдельными ключевыми словами, их частотами и т. д. В упомянутой выше работе предложен метод, позволяющий учесть порядок слов и показавший свою эффективность. Авторы данной статьи предполагают использовать некоторое обобщение данного метода, которое позволяет учесть дополнительно также синтаксические отношения, которые могут быть получены на выходе системы Link Grammar Parser или семантические отношения, выдаваемые семантической компонентой системы Dialing.

Список литературы

- [1] TEMPERLEY D., SLEATOR D., LAFFERTY J. Link Grammar Documentation. Carnegie-Mellon Univ., Pittsburgh, PA, USA. [Electronic resource]. — 1998. Адрес доступа: <http://www.link.cs.cmu.edu/link/dict/index.html> (дата обращения 15.11.2012)
- [2] СОКИРКО А. В. Семантические словари в автоматической обработке текста. Дис... канд. МГПИИЯ, — 2000. — 108 с.
- [3] New Protege Short Course // Stanford Center for Biomedical Informatics Research. Stanford Univ., California, USA. [Electronic resource]. — 2015. Адрес доступа: <http://protege.stanford.edu/> (дата обращения 15.09.2015)
- [4] KUMAR N., SRINATHAN K., VARMA V. Using graph based mapping of co-occurring words and closeness centrality score for summarization evaluation // CICLing Proceedings of the 13th International Conference on Computational Linguistics and Intelligent Text Processing. — 2012. — Vol. 2, P. 353–365.