

## АВТОМАТИЧЕСКОЕ ВЫДЕЛЕНИЕ ЛЕКСИЧЕСКИХ ЯДЕР СЕКЦИЙ КОНФЕРЕНЦИИ

*Сабурова Мария Ивановна*

*Аспирант*

*Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия*

*E-mail: masha-saburova@yandex.ru*

Исследуется задача автоматического выделения лексических ядер секций конференции. Данные конференции содержат названия секций и тезисы статей. Каждый доклад конференции эксперты приписывают ровно к одной секции. В тексте доклада уже выделены термины. Требуется так распределить термины по секциям, что каждый термин либо приписывается ровно к одной секции, либо не распределяется ни в одну из них. Можно сказать, что термины, приписанные некоторой секции, составляют ее лексическое ядро.

Если употребление термина однозначно связывается читателем с определенной секцией, то такой термин принято называть ключевым словом. Большинство исследований ограничиваются тем, что для корпуса текстов выделяют единственное множество значимых слов. Изучаемая в данной работе задача возникает на следующем этапе, когда требуется проанализировать отдельные рубрики в корпусе документов.

Рассматриваемая задача имеет две особенности. Во-первых, предлагается искать ключевые слова для целых групп документов (секций), а не для отдельных текстов. Во-вторых, исходные данные содержат заранее заданный набор рубрик и распределение документов по этим рубрикам. Именно на основе данной информации о распределении текстов по рубрикам и предлагается искать ключевые слова.

Формально задача описывается трехдольной полужесткой моделью. Исходные данные представляют собой трехдольный граф, содержащий три типа вершин: документы, термины, рубрики, – между которыми задано два отношения: между документами и терминами, и между документами и рубриками. Требуется построить функциональную связь от терминов к рубрикам.

Предлагаются и исследуются два метода выделения лексических ядер, основанных на многоклассовом SVM и SVM по схеме “один-против-одного” с линейным ядром соответственно. Результаты проиллюстрированы на данных конференции EURO - 2013.

**Литература**

1. Цыганова С. В., Стрижов В. В. Построение иерархических тематических моделей коллекции документов // Прикладная информатика. 2013. № 1. С. 109–115.
2. Страница конференции «European Conference on Operational Research, EURO–2013»:  
<http://euro2013.org/>