

Тематический поиск схожих дел в коллекции актов арбитражных судов*

Герасименко Н. А.¹, Артёмова Е. Л.², Воронцов К. В.³

nikgerasimenko@gmail.com

¹МАИ; ²НИУ ВШЭ; ³МФТИ

В работе рассматривается задача информационного поиска по коллекции актов арбитражных судов. В качестве запроса поисковой системе может выступать любой документ коллекции. В ответ на поисковый запрос генерируется список документов коллекции, ранжированный по убыванию релевантности. Для решения поставленной задачи построена тематическая модель коллекции актов арбитражных судов с помощью открытой библиотеки BigARTM. При построении модели учтена специфика предметной области: добавлены модальность ссылок на нормативно-правовые акты, а также модальность юридических терминов, выделенных полуавтоматически, с использованием алгоритма TopMine.

Ключевые слова: *LegalTech, BigARTM, тематическое моделирование, аддитивная регуляризация мультимодальных иерархических тематических моделей.*

1 Введение

*Задачу поставил: Воронцов К. В. Консультант: Артёмова Е. Л.