

DOI: 10.24411/1993-8314-2019-10042

М. И. Дли, докт. техн. наук, зам. директора филиала Национального исследовательского университета «МЭИ» в г. Смоленске, midli@mail.ru

О. В. Булыгина, канд. экон. наук, доцент филиала Национального исследовательского университета «МЭИ» в г. Смоленске, Baguzova_ov@mail.ru

П. Ю. Козлов, канд. техн. наук, ассистент филиала Национального исследовательского университета «МЭИ» в г. Смоленске, originaldod@gmail.com

Применение нечетких деревьев решений для рубрицирования неструктурированных текстовых документов небольшого размера¹

Ежедневно на Интернет-порталы органов государственной власти поступает большое количество электронных обращений (заявлений, предложений или жалоб), представленных в неструктурированном текстовом виде. Качество и скорость автоматической обработки указанных сообщений напрямую зависит от правильности их классификации (отнесения к конкретной предметной области), в основе которой лежат их отличительные особенности. Однако специфика таких сообщений (небольшой размер, наличие ошибок, отсутствие четкой структуры и т.д.) не позволяет использовать известные методы рубрицирования текстовых документов. Авторами разработана модель рубрицирования электронных неструктурированных текстовых документов с учетом синтаксических связей и ролей слов в предложениях на основе нечеткого дерева решений. Построение дерева решений основано на анализе степени пересечений словарей рубрик, а также расстояний между рубриками в n -мерном пространстве признаков. Данная модель позволяет более точно рубрицировать электронные неструктурированные текстовые документы в условиях взаимосвязанных рубрик, а также повысить оперативность обработки поступивших документов.

Ключевые слова: электронный неструктурированный текстовый документ, рубрицирование, нечеткие деревья решений

Введение

В настоящее время органы государственной власти активно развивают информационно-коммуникационные технологии взаимодействия с гражданами и орга-

низациями для оперативного обмена электронными сообщениями.

Основными показателями, используемыми для оценки эффективности такого информационного взаимодействия, являются оперативность реакции органов государственной власти на поступающие обращения и удовлетворенность граждан и организаций результатами оказанных государственных ус-

¹ Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта №18-01-00558.