

ход по агрегации схожих вершин в древовидной структуре, представляющей объектную модель документа. Сходными считаются вершины, имеющие одинаковые имена соответствующих HTML-тэгов и имеющие одинаковый родительский узел. На конечных шагах алгоритма использованы простые эвристики, использующие наблюдения о характере текстов, содержащих псевдонимы пользователей и их тексты. При тестировании построенного алгоритма достигнута точность 80 %. Практическая ценность разработанного алгоритма заключается в расширении множества текстовых ресурсов, используемых в качестве источников естественных текстов в задаче построения очень больших корпусов.

AN APPROACH OF AUTOMATIC EXTRACTION OF INFORMATION ABOUT THE AUTHORS AND THEIR TEXTS FROM WEB-FORUMS

Pronin A.K., Kopylov N.Y.

LLC «Abi InfoPoisk», Moscow, Russia (127273, 2B.6, Otradnaya, Moscow), e-mail apronin@abbyy.com

This article describes the approach of automatic extraction of information about the author and his/her texts from web forums. For building the algorithm the concept of style trees was used – approach of aggregating similar nodes in a tree representing Document Object Model. Nodes are similar if they all have the same name of the corresponding HTML-tags and have the same parent node. At final steps, simple heuristics were applied, employing observations about characteristics of texts containing users' pseudonyms and their messages. When testing the developed algorithm 80 % accuracy was reached. Practical value of the developed algorithm resides in expansion of text resources, used as sources for natural discourse, especially when faced with a problem of building very large text corpus.

ОСОБЕННОСТИ СОЗДАНИЯ БАЗ ДАННЫХ НА ОСНОВЕ ВЕБ-ТЕХНОЛОГИЙ ДЛЯ МОДЕЛИРОВАНИЯ ЛЕСНЫХ ПОЖАРОВ (ВЫСОКОБОРСКОЕ ЛЕСНИЧЕСТВО БОРСКОГО РАЙОНА НИЖЕГОРОДСКОЙ ОБЛАСТИ)

**Рылков И.В.¹, Катаева Л.Ю.^{1,2}, Масленников Д.М.¹, Романова Н.А.¹,
Рылкова О.И.¹, Тарнаева С.А.¹, Лошилов А.А.¹**

1 ГОУ ВПО «Нижегородский Государственный Технический Университет им. Р.Е. Алексеева»
(603950, ГСП-41, Нижний Новгород, ул. Минина, д. 24), email: kataeval2010@mail.ru

2 ГОУ ВПО «Московский государственный университет путей сообщения» (МИИТ),
Нижний Новгород, Россия (603011, г. Нижний Новгород, Комсомольская пл., д. 3), e-mail: tu@miit.ru

В статье рассматриваются вопросы создания геоинформационной системы, хранящей данные для расчета распространения лесного пожара на реальной местности. Для моделирования данного процесса, необходимо обладать информацией о карте высот местности и растительности, находящейся на интересующем участке а также её теплофизических и термокинетических характеристик. В данной работе описан программный комплекс, включающий в себя не только лесотаксационное описание и имеющиеся картографические данные, но и типы растительности с их характеристиками. Разработан алгоритм создания базы данных растительности и характеристик, посредством наложения графических изображений, необходимых для моделирования ландшафтных лесных пожаров, что придает системе практическую направленность. При создании геоинформационной системы растительности Высокоборского лесничества Борского района Нижегородской области активно используются современные WEB технологии и библиотеки, такие как HTML5, CSS3, JavaScript, GoogleMapsAPIv3, GoogleElevationServiceAPI, ASP.NETMVC4, KnockoutJS, HighchartsJS, JSON, GeoXML3. Написанный программный продукт реализует современный высокоэффективный MVVM паттерн и имеет клиент-серверную архитектуру.

FEATURES OF CREATING DATABASES BASED ON WEB TECHNOLOGIES TO MODELING OF FOREST FIRES (VYSOKOBORSKY FORESTRY BORSKIY DISTRICT OF NIZHNY NOVGOROD REGION)

**Rylkov I.V.¹, Kataeva L.Y.^{1,2}, Maslennikov D.M.¹, Romanova N.A.¹,
Rylkova O.I.¹, Tarnaeva S.A.¹, Loschilov A.A.¹**

1 Nizhny Novgorod State Technical University n.a. R.E. Alekseev, Nizhny Novgorod,
Russia(603950, GSP-41, Nizhny Novgorod, Minina St., 24), e-mail: kataeval2010@mail.ru

2 Moscow State University of Railway Transport, Nizhny Novgorod, Russia (603011, Nizhny Novgorod,
Komsomolskaya Square, 3), e-mail: tu@miit.ru

The paper considers the creation of a geographic information system, storing the data for calculating the propagation of forest fire on the real terrain. For the simulation of this process, it is necessary to have information on the height map of the terrain and vegetation, located on the area as well as its thermal and thermo-kinetic characteristics. This paper describes a software system that includes not only forest mensuration description and mapping data available, but also vegetation types with their characteristics. An algorithm for filling of database of vegetation and their characteristics needed to model landscape forest fire that gives system a practical focus. When creating a geographic information system of vegetation Vysokoborsky forestry Borskiy district of Nizhny Novgorod region used modern WEB technologies and libraries, such as HTML5, CSS3, JavaScript, Google Maps API v3, Google ElevationService API, ASP.NET MVC4, KnockoutJS, HighchartsJS, JSON, GeoXML3. Developed software implements modern and effective MVVM pattern and has a client-server architecture