

2. Официальный сайт Microsoft Visual Studio 2010 [Электронный ресурс]. – Режим доступа: <http://www.microsoft.com/visualstudio/ru-ru>, свободный.

3. Сайт международной статистики «Rich Internet Application Statistics» [Электронный ресурс]. – Режим доступа: <http://riastats.com/#>, свободный.

4. Официальный сайт Microsoft Silverlight [Электронный ресурс]. – Режим доступа: <http://www.microsoft.com/rus/silverlight/resources/faq/default.aspx>, свободный.

## **Интеллектуальные технологии в моделировании имиджа региона**

***А.А. Пунина**  
АлтГУ, г. Барнаул*

В докладе описывается один из способов исследования и измерения имиджа региона, формируемого средствами массовой информации, с использованием методов и технологий Data Mining.

Объективный имиджевый профиль Алтайского края можно получить, сравнив его с профилями других регионов. Регионами «контрольной группы» стали 11 регионов Сибирского Федерального округа. Данные собирались из трех федеральных сетевых изданий: Лента.ру, РиаНовости, РосБизнесКонсалтинг за 2011 г. (Алтайский край представлен записями последних четырех лет).

Для сбора и последующего преобразования данных были разработаны программы в среде Delphi 7, процедуры которых собирали данные из интернет-СМИ за счет поиска по html-тегам, записывали их в файл программы Excel, и переводили их в форму, удобную для построения классификационных моделей.

Оценка тематики и тональности сообщений СМИ производилась с помощью методов машинного обучения. Для таких методов необходимо обучающее множество. На первом этапе было собрано около девяти тысяч статей, из них релевантными оказались три с половиной тысячи. Половина релевантных статей (порядка тысячи семисот) составили сумму обучающего, тестового и проверочного множеств и были классифицированы экспертом. Вторая половина статей классифицирована специально обученной моделью. Тематика статьи определялась с помощью такой технологии Data Mining, как «деревья решений» (точность модели на проверочном множестве составила 82,6%). Для определения тональности использовалась нейросеть (точность

модели – 83,5%). Рассмотрим подробнее модель определения тональности статьи.

Тональность текста – позитивное или негативное отношение автора сообщения к заданному объекту, фигурирующему в тексте. В нашем исследовании объектом тональности будет тот или иной регион. Сложность задачи классификации текстов можно проиллюстрировать следующим примером: «В Кемеровской области мужчина задержан за попытку изнасиловать девочку». С одной стороны, мужчина задержан – это хорошо, с другой стороны, он пытался изнасиловать девочку – это плохо. Вот и два компонента оценки тональности текстов дают диаметрально противоположные результаты: первый компонент компании Ай-Теко (<http://x-file.su/tm/Default.aspx>) относит этот заголовок к негативным, а компонент Computana.Сентимент–аналитик (<http://computana.ru>) относит этот же заголовок к позитивным. Стоит отметить, что модель, построенная на основе нейронной сети, дает такой же результат как компонент компании «Ай-Теко».

Нейронная сеть имеет следующую структуру: входной слой содержит в себе данные о наличии/отсутствии определенных слов в заголовке и состоит из более чем тысячи нейронов. Выходной слой состоит из одного нейрона – тональности, классифицированной экспертом. Скрытый слой – один, активационная функция – сигмоида, алгоритм – Resilient Propagation (RPROP).

Анализ данных является завершающим этапом исследования имиджа. Так как все данные принадлежат номинальной шкале, традиционной статистикой они обрабатываются достаточно слабо. В достаточной большей части исследований связанных с имиджем подсчет частот и построение графиков является первым и последним этапом анализа данных. В нашем исследовании для анализа данных были использованы сети и карты Кохонена – один из видов нейросетей, использующих обучение без учителя.

Таким образом, используя методы и технологии Data Mining (деревья решений, нейронные сети, карты Кохонена) и современные языки программирования можно провести оценку даже такой качественной и субъективной категории, как имидж.