

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ПРОГРАММНОГО КОМПЛЕКСА МОНИТОРИНГА СРЕДСТВ МАССОВОЙ ИНФОРМАЦИИ В СЕТИ ИНТЕРНЕТ

Губарев И.Д., Болбин С. Н., Митягин С. А.

СПб ГУП «Санкт-Петербургский Информационно-Аналитический Центр»

Аннотация

Данная статья посвящена проблеме сбора и обработки информации, предоставляемой современными электронными средствами массовой информации (далее СМИ) и социальными сетями в Интернет. Статья содержит описание концептуальной модели программного комплекса, основной задачей которого является осуществление анализа подобных данных, их сбор и структуризация для последующей обработки системами поддержки принятия управленческих решений.

Ключевые слова

Мониторинг Интернет СМИ, системы поддержки принятия решений, средства информационной интеграции, облако тегов.

В последние несколько лет в России наблюдается все большее развитие СМИ в сфере Интернет, ввиду широкой доступности и динамичности обновления данных. На данный момент в России зарегистрировано более 6435 СМИ, существенное количество которых обладает собственными официальными сетевыми ресурсами [1]. Стоит отметить, что данный процесс развивается на фоне общего роста популярности Интернет в России. В частности, по данным компании Яндекс в Санкт-Петербурге за последний год количество Интернет пользователей увеличилось на 6% и составило более 71% жителей города, а в Москве – 68% с годовым приростом в 4% [2]. Рост популярности не в последнюю очередь связан с бурным развитием социальных сетей [3, 4, 5]. В социальных сетях практически с самого момента их появления намечена тенденция объединения пользователей в виртуальные группы по различным критериям и интересам.

В связи с этим, информация, получаемая из современных Интернет источников, предоставляет большой интерес для обработки и анализа с целью выявления проблемных сфер в развитии и состоянии общества для последующей поддержки принятия управленческих решений и их исправления. На данный момент большинство систем подобного типа построено на обработке статической информации, формируется на основе трех основных источников:

1. Федеральная Служба Государственной Статистики [6];
2. ведомственная статистика Исполнительных Органов Государственной Власти (далее ИОГВ);
3. социологические опросы населения.

Очевидно, что Интернет представляет собой альтернативный источник данных гораздо более динамичный, нежели перечисленные выше. Стоит отметить, что существующее множество решений для анализа данных социальных сетей и Интернет СМИ предназначено для проведения маркетинговых исследований и выявления целевых аудиторий, например:

1. «Яндекс.Директ» [7];
2. «Google AdSense» [8];
3. «Bing Ads» [9].

В свою очередь, предварительно проведенный анализ показывает, что предоставляемые подобными системами средства программной интеграции (далее СПИ) могут быть использованы для построения программного комплекса мониторинга СМИ в сети Интернет для поддержки принятия управленческих решений.

Основной целью разработки данного комплекса является перспектива получения результатов анализа Интернет СМИ и социальных сетей в качестве дополнительного источника информации для систем поддержки принятия управленческих решений руководством регионов России.

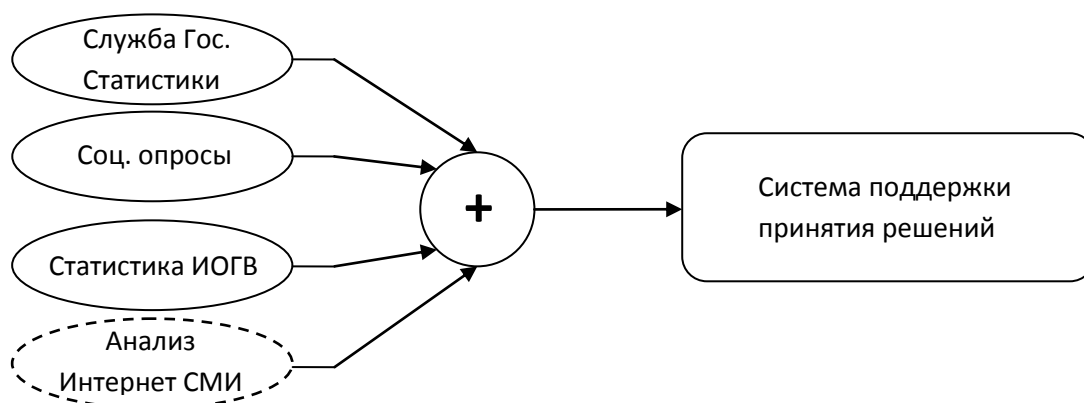


Рисунок 1. Совместное использование различных источников данных в системах поддержки принятия управленческих решений

Для выполнения поставленной задачи необходимо провести следующие шаги:

1. подготовить обзорный перечень источников данных, включая поисковые Интернет системы и социальные сети, на основе СПИ которых разрабатываемый программный комплекс сможет находить интересующую информацию и получать к ней доступ;
2. разработать логическую и функциональную структуру модуля сбора данных посредством использования СПИ выбранных источников;
3. разработать логическую и функциональную структуру модуля обработки полученных данных;
4. разработать и подготовить набор алгоритмов, необходимых для обработки полученных данных;
5. разработать логическую и функциональную структуру модуля структурирования и сохранения данных для их последующего использования.

На рисунке 2 приведена структурная схема разрабатываемого программного комплекса:

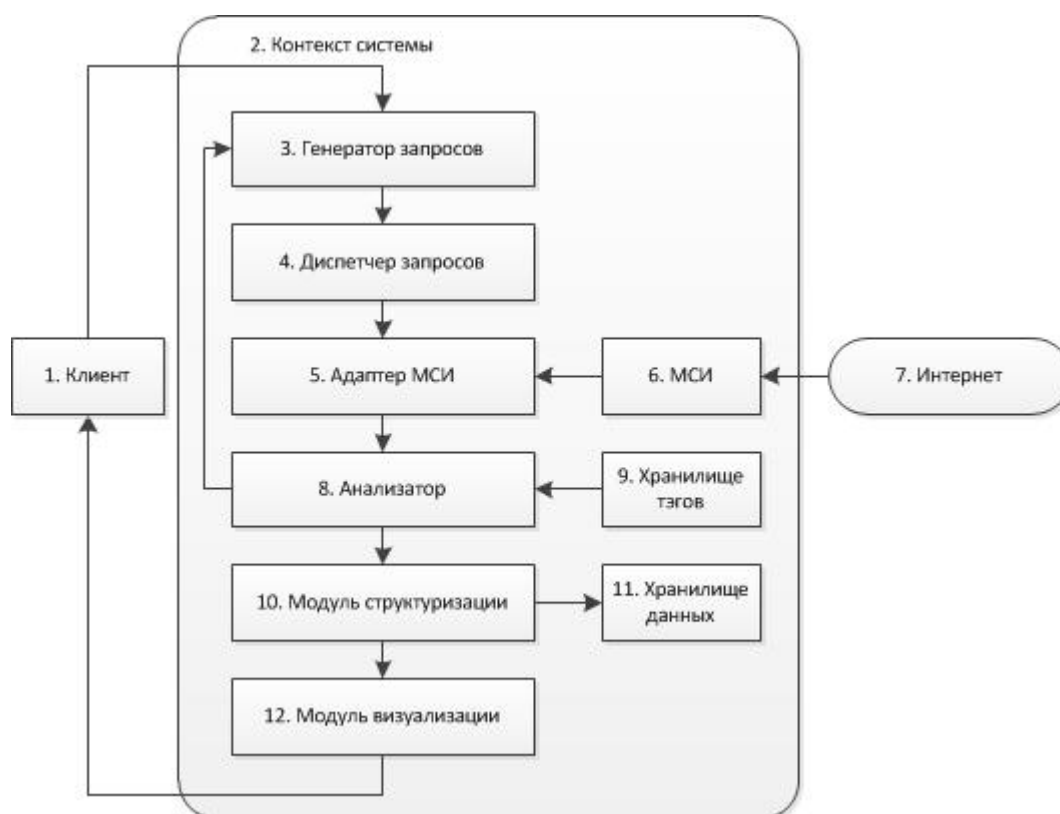


Рисунок 2. Структурная схема программного комплекса мониторинга СМИ в сети Интернет

В качестве клиента (рисунок 2, блок 1) может выступать как конкретный пользователь, так и другая информационная система. Генератор запросов (рисунок 2, блок 3) вырабатывает строго-типизированные обращения на основе сообщений, переданных клиентом в систему, и посредством диспетчера запросов (рисунок 2, блок 4) направляет их

адаптер МСИ (рисунок 2, блок 5). Адаптер МСИ дополняет обращения информацией, полученной из модулей сетевой интеграции (рисунок 2, блок 6) и отправляет их в блок анализа данных (рисунок 2, блок 8). Анализатор, используя хранилище тэгов (рисунок 2, блок 9) вырабатывает утверждения на основе поступающих на вход обращений и передает их в модуль структуризации (рисунок 2, блок 10) для предварительной подготовки обработанных данных перед их сохранением в хранилище данных (рисунок 2, блок 11), включая их классификацию, ранжирование, вычисление агрегированных и базовой статистики. Модуль визуализации (рисунок 2, блок 12) отправляет клиенту отчет, содержащий в себе общее представление упоминаний в Интернет СМИ тегов заданной тематики. Стоит обратить внимание на то, что модуль анализа связан с генератором запросов, что позволяет ему самостоятельно создавать системные обращения для проведения более детального, вложенного анализа.

Основными сложностями на начальном этапе работы являются:

1. создание устойчивых модулей сетевой интеграции (далее МСИ) для сбора информации из Интернет СМИ и социальных сетей, способных функционировать с различными средствами программной интеграции;
2. разработка методик анализа полученной информации и интеграции с показателями социально-экономического развития общества в соответствии с Указом Президента РФ от 21.09.2012 N 1199 [10];
3. реализация интегрированных программных компонент, предоставляющих возможность обработки больших объемов гетерогенных данных.

В связи с увеличивающимся значением сети Интернет в жизни общества, а также со становлением ее в качестве одного из наиболее доступных и предпочитаемых гражданами средств коммуникации, становится очевидно, что для повышения качества принимаемых управленческих решений ИОГВ необходимо учитывать данные из Интернет источников. Разрабатываемый программный продукт является важной частью в системе поддержки государственных решений, так как позволяет получить актуальный информационный срез на основе проанализированных им Интернет СМИ и социальных сетей. В результате его внедрения руководство регионов с активной интернет аудиторией получит динамичный источник данных, отражающей информационную составляющую настроений населения.

Список литературы:

1. Количество печатных и Интернет средств массовой информации в России, <http://www.redom.ru/news/183460>.
2. Развитие интернета в регионах России по данным компании «Яндекс», http://company.yandex.ru/researches/reports/internet_regions_2012.xml.
3. «Одноклассники.ру» - социальная сеть для поиска одноклассников, коллег и знакомых в регионах России, СНГ и за рубежом, <http://www.odnoklassniki.ru>.
4. «ВКонтакте» - поиск людей по их увлечениям, месту учебы и работы, персональным данным, <http://www.vk.com>.
5. «Живой Журнал» - блог-платформа для ведения онлайн-дневников, <http://www.livejournal.com>.
6. Федеральная Служба Государственной Статистики, официальный Интернет ресурс: www.gks.ru.
7. «Яндекс.Директ» - Размещение контекстной рекламы с разными видами привязки к поисковой выдаче Яндекса и на сайтах Рекламной сети, <http://www.direct.yandex.ru>.
8. «Google AdSense» - сервис контекстной рекламы от компании Google, <http://google.com/adsense>.
9. «Bing Ads» - сервис контекстной рекламы от компании Microsoft, <https://secure.bingads.microsoft.com>.
10. Указ Президента РФ от 21.09.2012 N 1199 «Об оценке эффективности деятельности органов исполнительной власти субъектов Российской Федерации», <http://graph.document.kremlin.ru/page.aspx?1623676>.