

# Интеллектуальные технологии анализа общественно-политической обстановки в ситуационном центре Челябинской области

А.С. Козлов  
Министерство информационных технологий и связи Челябинской области  
Челябинск, Россия  
e-mail: info@mininform74.ru

А.В. Мельников  
Институт информационных технологий  
Челябинский государственный университет  
Челябинск, Россия  
e-mail: [mav@csu.ru](mailto:mav@csu.ru)

А.В. Рубцов  
Управление информатизации образования  
Челябинский государственный университет  
Челябинск, Россия  
e-mail: rav@csu.ru

## Аннотация<sup>1</sup>

В данной работе описывается процесс разработки и внедрения в промышленную эксплуатацию информационно-аналитической системы для обработки информационного фонда ситуационного центра губернатора Челябинской области.

## 1. Основные задачи информационного фонда ситуационного центра (СЦ).

«Методические рекомендации по созданию и вводу в эксплуатацию ситуационных центров» определяют следующие основные задачи информационного фонда:

- сбор и загрузка данных из внешних информационных систем (ИС), включая загрузку из государственных автоматизированных систем;
- хранение массива данных, сформированных в процессе работы СЦ, в соответствии с требованиями быстрого доступа к оперативной и ретроспективной информации структурированного и неструктурированного вида;
- поиск и представление информации, включая работу с геоинформационным инструментарием;

- очистка, проверка качества и верификация данных, включая выявление содержательных искажений;
- ведение комплекса показателей, включая показатели общественно-политической ситуации, социально-экономического развития и комплексной безопасности с возможностями агрегации их значений;
- взаимодействие с системой электронного документооборота.

Анализ основных информационных источников показывает, что только в Челябинской области на текущий момент зарегистрировано 828 СМИ, которые ежедневно производят не менее 12 000 публикаций. Один аналитик способен обработать не более 350 публикаций за день, а для обработки всего массива суточной информации необходимо не менее 35 аналитиков. Таким образом, напрашивается решение: автоматизировать обработку информации.

## 2. Выбор информационно-аналитической системы (ИАС).

Был проведен анализ и сравнение следующих ИАС:

- I2 Analyst's Notebook [2]. Система разрабатывалась в Великобритании. Позволяет загружать данные из внешних, структурированных источников и представлять их в виде графа. В графе можно производить поиск, фильтрацию данных; находить кратчайшие маршруты от одного объекта к другому, искать общих соседей. Система не предназначена для анализа текстов на естественном языке.;

---

Труды третьей международной конференции "Интеллектуальные технологии обработки информации и управления", 10 - 12 ноября, Уфа, Россия, 2015

Наименование ИАС обработки неструктури- рованной информации	Страна	Анализ текстов			Сбор информации из сети Интернет				Хранение текстов	Визуализа- ция
		Семантичес- кие	Факто- графический	Эмоцио- нальный	Поисковые системы	Социальные сети	Комментарии	СМИ		
I2 Analyst's Notebook	Велико- британия	X	X	X	X	X	X	X	X	Графовая структура, таблицы
«ДИАрама»	Россия	X	X	X	X	X	X	X	✓	Таблицы
«Семантический архив»	Россия	✓	X	X	X	X	X	✓	✓	Графовая структура, Таблицы
«Кронос»	Россия	X	X	X	X	X	X	X	X	Таблицы
«Арион»	Россия	✓	X	X	X	X	X	✓	✓	Графовая структура, Таблицы
ИСА РАН	Россия	✓	✓	✓	X	X	X	X	✓	Графовая структура, Таблицы

Рис. 1. Сравнение информационно-аналитических систем

- «Семантический архив» [3]. Страна разработки – Россия. Система предназначена для семантического анализа текстов, сбора информации из интернет СМИ. Обеспечивает поиск и представление результатов как в виде графа, так и в табличном виде;
- «CronosPRO» [4]. Страна разработки – Россия. Система предназначена для импорта и хранения структурированных данных. Позволяет осуществлять по ним сложный поиск. Результаты поиска выводятся в виде графов и таблиц;
- «Арион» [5]. Страна разработки – Россия. Функциональные возможности аналогичны системе «Семантический архив»;
- Система, разрабатываемая в ИСА РАН, в лаборатории под руководством Осипова Г.С. [6]. На рынке не представлена, функциональные возможности можно оценить только по демонстрационным страницам в интернет. Основное направление анализа – научные тексты из электронных библиотек. Дополнительно к семантическому анализу позволяет находить похожие тексты, устанавливая авторство с определенной степенью вероятности. Разработаны собственные механизмы эмоционального анализа;
- База знаний органов государственной власти Ярославской области «ДИАрама». Компания-разработчик – АЛАН-ИТ. Система построена на базе Share Point портала,

использует стандартный механизм поиска Fast Search Server 2010 for SharePoint. Интегрирована с системой «Directum». Не может получать информацию из интернет источников: поисковых систем, блогов, СМИ.

Данные системы не удовлетворяют всем поставленным условиям, а именно:

- Семантический, фактографический и эмоциональный анализ русскоязычных текстов;
- Сбор информации не только из локальных источников, но и из интернет источников: социальные сети, блоги, СМИ, поисковые системы.

По результатам анализа было принято решение разрабатывать собственную систему.

### 3. Архитектура и функциональные возможности ИАС «Архивариус».

Система состоит из следующих блоков: crawler, модуль анализа, хранилище данных, модуль поиска и фильтрации, модуль визуализации результатов поиска. Далее будут подробно расписаны функциональные возможности каждого блока.

Crawler предназначен для сбора информации из следующих интернет источников: новостные сайты, RSS, поисковые системы (Google, Yandex), блоги и социальные сети. Из блогов, новостных сайтов, СМИ извлекаются как публикации, так и комментарии к ним. Поисковые системы используются для поиска и предварительной фильтрации ссылок, далее скачиваются публикации по ссылкам.

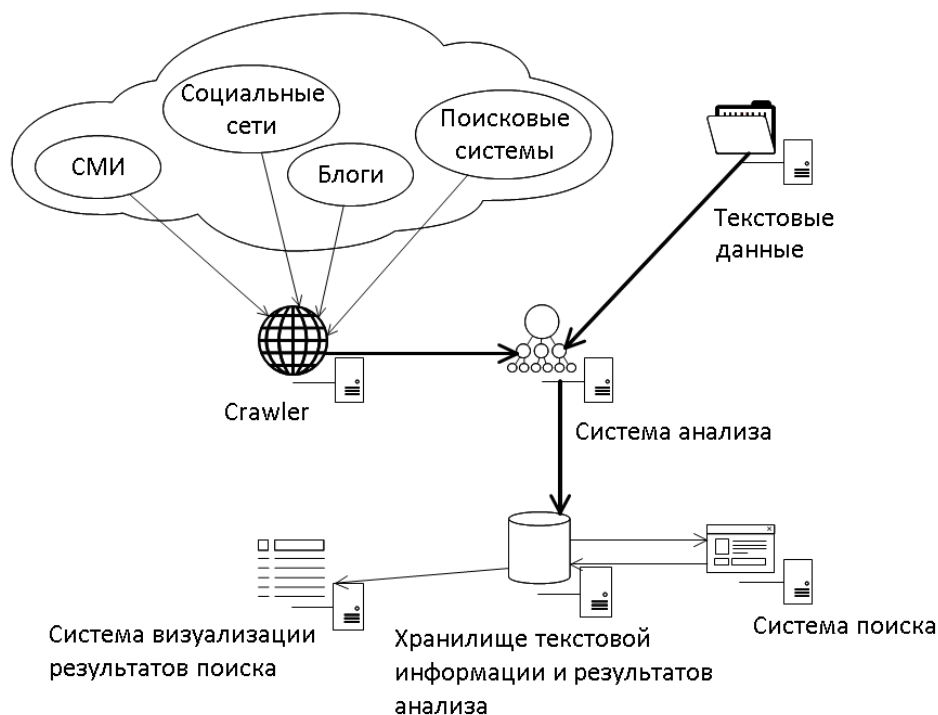


Рис. 2. Архитектура ИАС «Архивариус»

Из социальных сетей собираются данные по персонам, их высказывания и публикации. Crawler осуществляет отбор необходимой информации на основе разработанных для него XML шаблонов. Таким образом для добавления нового типа источника информации требуется разработать соответствующий шаблон.

Модуль анализа осуществляет следующие виды анализа: семантический, фактографический и эмоциональный. Семантический анализ подразумевает выделение различных классов сущностей, упомянутых в тексте: персоны,

организации, данные ГИС, предметы, атрибуты, действия и др. Фактографический анализ осуществляет поиск описаний ситуаций, удовлетворяющих заданным семантическим шаблонам: «покупка акций», «проведение митингов», «назначение на должность» и др. Эмоциональный анализ выделяет как тональность текста целиком, так и фактов, встречающихся в этом тексте. В фактах производится анализ тональности высказываний по отношению к субъектам факта. Данный модуль разработан на базе библиотеки «RCO Fact Extractor SDK».

## Архивариус

Рис. 3. Интерфейс ИАС «Архивариус»

Хранилище данных обеспечивает хранение, резервирование и доступ к данным. Хранилище построено на базе RDBMS Oracle SE.

Модуль поиска и фильтрации предназначен для отбора требуемых документов из хранилища. Интерфейс модуля разработан с использованием Oracle Apex.

Модуль визуализации результатов поиска представляет в виде графа, таблиц, списков и графиков результаты поиска. Модуль состоит из тонкого клиента на базе Apex и из рабочего места аналитика на базе I2 Analyst's Notebook.

#### 4. Результаты внедрения.

В результате внедрения была получена системы, которая обрабатывает ежедневно:

- не менее 440 публикаций от телерадиокомпаний;
- не менее 610 публикаций от печатных СМИ;
- не менее 220 публикаций с новостных сайтов;
- не менее 30 публикаций в блогах.

Осуществляется контроль материалов, освещающих деятельность правительства и губернатора Челябинской области. Выявляются позитивные и негативные публикации. В данный момент в хранилище находится около 1,4 млн. публикаций, общим объемом около 350 Гб.

#### 5. Перспективы.

Модернизация и дальнейшая разработка ИАС «Архивариус» включена в областную программу развития на 2016-2018 гг.

Основными задачами на данный период будут:

- подключение новых источников информации: периодические печатные издания СМИ, информационные выпуски телерадиокомпаний, новостные сайты, блоги и форумы, социальные сети;
- разработка новых модулей анализа: выявление социально-значимых процессов, протекающих в Челябинской области; мониторинг направленности внимания населения Челябинской области; расчет рейтинга социальной стабильности Челябинской области.

Также планируется расширение области применения системы:

- Сбор информации и оценка интенсивности и эмоциональной окраски реакции социума по определенной теме или конкретному событию;

- Обеспечение привязки интегральных оценок к геоинформационной основе;
- Формирование паттернов для выявления первичного источника информации в случае возникновения информационной атаки;
- Автоматизированный сбор информации для портфолио участников праймериз, выборов с получением интегральной оценки по различным источникам.

И расширение возможностей системы:

- Формирование шаблонов для различных вариантов поиска по фактам, объектам, связям и поиска с учетом тональности;
- Формирование шаблонов различных статистических отчетов и диаграмм по сформированной выборке;
- Запуск модуля уведомления о поступлении информации об определенных фактах (комбинаций фактов) в хранилище;

Также планируется перевод системы на Open Source платформу.

#### 6. Заключение

В данной статье представлено краткое описание проекта, который продолжается уже более года. Исходя из анализа достоинств и недостатков как рассмотренных в третьем пункте систем, так и разработанной нами системы, в заключение можно сделать следующие выводы:

- универсальной системы не существует, можно только разработать систему для конкретных конечных пользователей, решающих частные задачи в тех или иных предметных областях;
- незрелая технология анализа текстов вынуждает использовать в проекте разнородное программное обеспечение и ориентироваться на открытую архитектуру;
- основной упор при разработке системы нужно делать на интеграцию с разнородными источниками информации и на возможность использования разных систем и алгоритмов анализа;
- опыт эксплуатации системы показывает эффективность использования интеллектуальных технологий для обеспечения работы аналитиков.

#### Список используемых источников

1. Базенков Н.И., Губанов Д.А. «Обзор информационных систем анализа социальных сетей». «Управление большими системами: сборник трудов» № 41/2013. ББК: 22.18; УДК: 519.87 + 004.94 + 004.89.

2. I2 Analyst's Notebook: [Электронный ресурс]: IBM – Анализ данных: URL: <http://www-03.ibm.com/software/products/ru/analysts-notebook>
3. «Семантический архив»: [Электронный ресурс]: Аналитические бизнес решения: URL: <http://www.anbr.ru>
4. «CronosPRO»: [Электронный ресурс]: КРОНОС ИНФОРМ группа компаний: URL: <http://www.cronos.ru/cronospro.html>
5. «Арион»: [Электронный ресурс]: SyTech: Разработка и внедрение информационно-аналитических систем: URL: <http://sytech.ru/about.php?id=5>
6. Сайт ИСА РАН: [Электронный ресурс] : Лаб. 0-6 «Компьютерная лингвистика и интеллектуальный анализ информации»: URL: [http://www.isa.ru/index.php?option=com\\_content&id=914%3A-0-0-2-&lang=ru](http://www.isa.ru/index.php?option=com_content&id=914%3A-0-0-2-&lang=ru)