

## Содержание:

image not found or type unknown



## Кто такой аналитик данных

Аналитик данных (или дата-аналитик) — это специалист, который собирает, обрабатывает, изучает и интерпретирует данные. Его работа помогает принимать решения в бизнесе, управлении и науке. Обычно такие специалисты работают в компаниях, которые практикуют data-driven подход — ориентируются на данные и их анализ при принятии решений.

Аналитик данных — важный участник бизнеса, потому что обеспечивает уверенность в принятии решений. Создавать новый продукт очень дорого, а ошибка при внедрении новой функции может стоить компании репутации и прибыли. Дата-аналитики проводят A/B-тесты и строят модели, чтобы проверить, как пользователи или клиенты реагируют на нововведения, и оценить перспективы того или иного проекта. Это дешевле и снижает риски бизнеса. Чтобы делать свою работу хорошо, аналитик должен видеть бизнес-процессы. Поэтому важно, чтобы он мог влиять на процесс принятия решения, основываясь на результатах своих исследований. Иначе работа такого специалиста теряет ценность.

Основные этапы деятельности аналитика:

1. Поиск, мониторинг информации – этап на котором аналитику необходимо осуществить поиск требуемой информации в гигантских информационных массивах интернета, печатных СМИ, корпоративных БД. Суть мониторинга заключается в систематическом отслеживании ключевых информационных проявлениях в указанных источниках информации.
2. Структурирование и хранение. Под этим этапом понимается придание каждому информационному сообщению своего признака, атрибута, то есть: источника информации, даты публикации, название объекта. И далее хранение - аккумуляирование собранной информации на предыдущем этапе в одном информационном массиве.
3. Анализ собранной информации и синтез. Как правило, перед специалистом на данном этапе стоит задача установления причинно- следственной связи,

влияние одних фактов на другие, а также прогнозирование развития ситуации.

Практически все современные системы применяющиеся в деловой разведки – универсальные, то есть одна система может обеспечить весь комплекс работ от мониторинга до анализа. Исключение составляют лишь программные продукты попадающие в класс «Shareware» – то есть условно бесплатные. Как правило, подобные программы не предназначены для серьезной работы аналитика, но в отличие от своих собратьев могут применяться в малом бизнесе, которому не под силу содержать отдел деловой разведки. В таком случае программы подобного класса могут использоваться непосредственно руководителями предприятий, дабы оперативно отслеживать все новости. Более подробно о этих программах можно узнать на страницах интернет-проектов «Технологии разведки для бизнеса» -

## **Анализ программных комплексов**

Рассмотрим наиболее популярные программные комплексы, которые может использовать аналитик:

Web-Observer (компания Finport Technologies)

Система на базе Web-Observer представляет собой автоматизированную систему сбора и структуризации информации из открытых интернет-источников и корпоративных хранилищ данных.

Источниками различной политической, общественной, финансово-экономической, нормативной и другой информации служат электронные ресурсы СМИ, информационных агентств, банков, государственных органов, специализированных учреждений, интернет-издания, электронные рассылки и многие другие. Существует возможность добавления практически неограниченного количества информационных источников, которые требуются заказчику для работы.

Система обнаруживает и извлекает появившуюся информацию, разбивает ее на необходимые для конкретного потребителя рубрики (подрубрики) и сохраняет в базе приведенные к единой структуре материалы, формируя единый каталог. Пользователь может отредактировать подачу информации и создать персональные подборки, отвечающие его запросам.

Как только в систему попадут публикации, совпадающие с настройками персональных (клиентских) подборок, механизм уведомлений сразу же может отправить пользователю сигнальное письмо на e-mail или SMS, сообщая о новых поступлениях. Информация поступит в необходимом формате. Свежие материалы пользователь может просмотреть в защищенной паролем части системы или в своем e-mail ящике.

Полученная таким образом информация также будет отправлена в последствии в архив. В результате системой накапливается огромная информационная база, позволяющая эффективно распоряжаться информацией в удобном пользователю виде.

## **Основные функции системы**

- оперативный сбор информации с подключённых источников;
- её последующее разбиение по рубрикам;
- уведомление о нахождении интересующей Вас информации (по e-mail или SMS);
- экспорт выбранных публикаций в Microsoft Word для предоставления отчётности;
- ведение информационного архива;
- предоставление возможности эффективного поиска и обработки информации в накопленном архиве публикаций;
- работа с единым информационным пространством всеми сотрудниками вне зависимости от территориального расположения;
- автоматическое создание лент профильных новостей на веб-сайте.

## **Области применения системы**

- осуществление эффективного мониторинга информации по интересующим темам и объектам (например, название компании, названия торговых марок, персоны);
- возможность сегодня быть в курсе событий, информация о которых выйдет в печатной прессе лишь завтра. Это происходит за счёт того, что интернет версии некоторых изданий появляются на день раньше печатных;
- мониторинг Вашей репутации;

- мониторинг информации связанной с профильным рынком;
- мониторинг конкурентов;
- мониторинг проблемных областей;
- контроль утечки конфиденциальных данных;
- отслеживание первоисточников агрессивной, ложной или некорректной информации.

## **ИАС Астарта (Cognitive Technologies)**

Программный комплекс ИАС Астарта успешно решает следующие задачи:

- Непрерывный информационный мониторинг внешней среды предприятия. Анализируется, систематизируется и аккумулируется текущая и архивная информация по актуальным для компании темам (например рынки сырья и комплектующих, развитие технологий и продуктов в отрасли, информация о конкурентах, информация о потребителях, отраслевая законодательная информация) тематика и состав рубрик формируется пользователем;
- Тематическая фильтрация информации. Такие задачи возникают в случае необходимости построения подборок на заданную тему.

Конфигурация имеет следующие функции:

- Получение информации из большого числа разнородных источников (бумажные периодические издания и е-газеты, Интернет ресурсы, IP каналы, ftp каналы, электронная почта);
- Автоматическая систематизация данных при помощи рубрикатора, обученного экспертом;
- Тематическая фильтрация потока текстовых сообщений;
- Полнотекстовая индексация рубрицированных материалов, размещение информации в базе данных, обеспечивающей быстрый и удобный поиск;
- Автоматическое составление дайджестов;
- Статистический анализ по времени и по тематике информации, находящейся в базе данных.

Эффективность внедрения системы обеспечивается следующими факторами:

- Кардинальные изменения качества обработки информации. Имеется возможность охватить все доступные источники информации и вести

обработку в круглосуточном режиме. За счет этого снижается до минимума вероятность пропуска важной информации. Использование системы опровергает расхожее мнение «за всем не уследишь». Автоматизированный рубрикатор, настраиваемый и обучаемый экспертом, легко адаптирует Астарту к решению задач в любой организации. Мощная система поиска исключает потери времени на изучение архивов. Статистическая обработка данных и модуль генерирования отчетов позволяют решать аналитические задачи любого уровня сложности.

- Снижение расходов на персонал. Производительность системы такова, что позволяет небольшому отделу с персоналом 2-3 человека выполнять работу аналитического подразделения численностью 15-20 человек, работающих в три смены.

## **Краткое описание работы ИАС Астарта**

В зависимости от задач пользователя, система имеет один или несколько каналов получения информации разного типа. Если предполагается работа с печатными периодическими изданиями, то производится сканирование получаемых газет или журналов. Изображения страниц поступают на вход модуля распознавания, который преобразует графические образы в текст, который при помощи специалиста разбивается на отдельные статьи.

Модуль Web-мониторинга позволяет обходить указанные пользователем сайты и производить загрузку обновлений web-страниц.

После получения и предварительной обработки все материалы обрабатываются модулем рубрикации.

На основании экспертных оценок модуль рубрикации производит морфологический и семантический анализ текстов, выделяя основные тематические понятия и анализируя структуру их размещения в тексте. После обучения каждой рубрики на 100-200 сообщениях, модуль формирует достаточный понятийный и семантический ряды, достаточные для работы в автоматическом режиме.

Систематизация данных в автоматическом режиме ведется на основании результатов обучения. Поступающий текст относится к одной или нескольким рубрикам с постановкой степени отношения.

После распределения по рубрикам материалы проходят индексацию по всем словам своего содержания и требуемым реквизитам. Эта процедура обеспечивает гибкие возможности поиска, как по признакам материалов, так и по их содержанию.

Результаты работы системы могут быть представлены в виде тематических дайджестов, которые создаются автоматически в формате Microsoft Word. Модуль статистики позволяет анализировать сообщения по рубрикам и по датам. На основании анализа выявляются наиболее насыщенные информацией темы, что говорит о значимых изменениях в той или иной сфере, которые могут повлиять на деятельность компании.

RCO Fact Extractor (компания Гарант-Парк-Интернет)

RCO Fact Extractor – это интеллектуальная программа для компьютерного анализа текста на русском языке, которая находит в нем описания фактов нужного типа, например, “встречи”, “договоренности”, “приобретение собственности”, классифицирует и упорядочивает их. Основная сфера приложения программы - аналитические задачи из области компьютерной разведки, требующие высокоточного поиска информации, например, автоматический подбор материала к досье на целевой объект или же мониторинг определенных сторон его активности, освещаемых в СМИ. Fact Extractor воплотил в себе наиболее продвинутое технологии искусственного интеллекта. Помимо собственно программы с графическим интерфейсом для Windows, выпускается динамическая библиотека для разработчиков (SDK), на базе которой построен Fact Extractor и которая позволяет включать возможности анализа текста в собственные приложения.

Программа позволяет обрабатывать документы в популярных текстовых форматах из различных источников - файловой системы, базы данных, заданных web-сайтов.

Результат работы программы – таблица, которая содержит информацию о найденных фактах, связанных с объектами мониторинга, и может экспортироваться в html-формат для формирования отчета или для загрузки в стороннее приложение, работающее с уже структурированными данными.

## **Галактика ZOOM**

Система Галактика ZOOM – инструмент для создания хранилища текстовой информации, который обладает уникальными возможностями для проведения эффективного поиска и аналитических исследований.

Система используется в целях повышения эффективности работы специалистов, занимающихся информационными процессами (сбор, хранение, поиск, анализ информации). Применение системы позволяет значительно снизить временные затраты и повысить качество работы сотрудников.

## **Основные возможности**

Руководителям: повысить эффективность работы подразделений и предприятия в целом.

Аналитикам: оперативно составлять отчеты (справки), описывающие ситуацию и тенденции ее развития в политической, экономической, социальной и других сферах.

Маркетологам: исследовать рынок, выявлять источники потенциальных возможностей совершенствования маркетинговой стратегии (в разрезе данных по потенциальным клиентам, партнерам, тенденциям рынка). Контролировать деятельность компаний-конкурентов и оперативно реагировать на их действия.

Специалистам по PR: проводить мониторинг СМИ, отслеживать PR-акции конкурентов, формировать дайджест.

Сотрудникам службы безопасности: выявлять источники угроз, как со стороны внешнего окружения, так и внутри компании. Формировать досье на определенные физические и юридические лица

Галактика ZOOM выполняет следующие задачи – автоматическая загрузка различной информации в систему, хранение архива документов, проведение поиска и аналитических исследований.

В рамках задачи загрузки информации система Галактика ZOOM осуществляет автоматический сбор информации с таких ресурсов как локальная сеть (файловая система, почтовые сервера, системы документооборота, итранет-порталы) и сеть Интернет. Помимо этого существует возможность автоматизированного перевода в электронный вид и дальнейшей загрузки в систему документов с бумажных

носителей.

В рамках задачи хранения информации система Галактика ZOOM позволяет хранить массивы объемом до 10 Тб (10<sup>12</sup> байт). Документы хранятся в виде индексов, что обеспечивает сохранность информации и предупреждает несанкционированный доступ.

В рамках задачи проведения поиска система Галактика ZOOM позволяет проводить поиск по следующим критериям:

- полнотекстовый поиск
- поиск в рамках указанного диапазона текста, в том числе предложения или абзаца
- нечеткий поиск
- поиск с учетом морфологии
- поиск по дате или промежутку времени
- поиск по источнику документа
- поиск по номеру документа
- поиск по автору документа
- поиск по названию документа
- комбинации вышеперечисленных вариантов
- дополнительные критерии поиска, которые заранее определены

После проведения поиска и получения результатов пользователь имеет возможность увидеть авторефераты и главные темы (суть текста) найденных документов (наиболее значимые фрагменты). Для удобства работы с найденными документами, пользователь имеет возможность отсортировать их:

- по дате
- по значимости
- по близости к интересующей теме

Отличительная особенность системы Галактика ZOOM является формирование информационного портрета найденных документов. Информационный портрет – представляет собой главные темы всех найденных по запросу документов, которые расположены таким образом, что можно увидеть связи объекта. Информационный портрет позволяет ориентироваться в найденных по запросу документах, при этом не читая каждый из них.

Информационно-программный комплекс «Тренд»



ИПК "Тренд" позволяет: производить исследования интересующих объектов и процессов, прогнозировать изменения в их состоянии на период до двух месяцев с вероятностью более 63% на основе сравнения информационных образов объектов и процессов, получаемых из потоков любых текстовых сообщений (в том числе СМИ), с эталонными информационными образами (фильтрами).

## **Логика функционирования ИПК "Тренд"**

- 1. Из базы данных формируемой из потока текстов выбирается необходимая информация, относящаяся к исследуемой ситуации и формируется ее "образ";
- 2. Создается набор аналогов ситуаций, позволяющий пользователю-аналитику определить понятное ему тематико-смысловое поле, необходимое для формирования конечного заключения и его интерпретации (например: ситуация развивается в сторону кризиса или в сторону успеха);
- 3. Осуществляется процедура сравнения "образа" ситуации вокруг объекта исследования с требуемым аналогом;
- 4. С использованием располагаемых информационно-аналитической системой наборами правил и критериев формируется, в высокой степени, однозначно интерпретируемый результат.
- 5. Формируется отчет, содержащий следующие таблицы и выводы:
  - о критичности ситуации вокруг объекта
  - распределение публикаций по источникам и временным интервалам
  - аннотированные сообщения, формирующие признаки кризиса
  - аннотируемые сообщения всех публикаций в последнем временном сечении интервала исследования.

ИПК "Тренд" позволяет исследовать и получать данные о внутренних сторонах деятельности объектов или развития процессов, которые явно не отражены в сообщениях СМИ или специальных текстовых данных. Важной особенностью работы ИПК "Тренд" является способность отсекавать заказные информационные выбросы и шумы.

«Тренд» представляет возможность поддерживать как внешние, так и внутренние (формируемые информационными системами пользователя) потоки актуальной текстовой информации.

Структура и набор функций комплекса «Тренд» позволяет осуществлять практически полный перечень работ, выполняемый аналитиком, начиная с ввода потоков информации и развитого поиска по ним, и кончая получением оценок тенденций развития процессов и ранжирования вероятности реализации возможных сценариев.

## **Список источников**

**<https://it.wikireading.ru/13359>**

**<https://it.wikireading.ru/13357>**

**<https://wm->**

**[help.net/lib/b/book/3994518110/96](https://wm-help.net/lib/b/book/3994518110/96)**

**<https://analytics.infozone.pro/chtotakoe-nauka-o-dannyh-analitik-dannyh-data-scientists/>**