4. *Ibragimov N. H.* Azbuka gruppovogo analiza. M.: Znanie. Ser. «Matematika i kibernetika». 1989. № 8. 44 s.
5. *Bluman G. W., Anco S. C.* Symmetry and integration methods for differential equations (AMS 154). Springer, 2002. 418 p.
6. *Polyanin A. D., Zaitsev V. F.* Handbook of exact solutions for ordinary differential equations. 2nd ed. Chapman & Hall / CRC, 2003. 814 p.

А. В. Флегонтов, В. В. Фомин

СИСТЕМА ИНТЕЛЛЕКТУАЛЬНОЙ ОБРАБОТКИ ДАННЫХ

Предлагаются технические решения по структуре, функциям и организации инструментария интеллектуального анализа данных, основанного на технологии WEB систем и облачных ресурсов. Формулируются стратегические направления разработки программной системы для повышения эффективности обработки больших массивов электронных данных для задач интеллектуального анализа: классификации, обучения, прогнозирования.

Ключевые слова: искусственный интеллект, программные системы, интернет, алгоритмы, базы данных и знаний.

A. Flegontov, V. Fomin

SOFTWARE SYSTEM FOR DATA PROCESSING

Technical solutions are suggested for structure, functions and organization of data processing tools based on WEB technology systems and cloud resources. We formulate the strategic direction of the development of a software system for higher-performance processing of large volumes of electronic data for the problems of intellectual structural analysis: classification, machine learning and prognostication.

Keywords: artificial intellect, software systems, internet, algorithms, data and knowledge bases.

Сегодня день можно констатировать, что за последние десятилетия произошло интенсивное развитие автоматизированных информационных систем (АИС), в том числе в таких областях, как сетевые технологии Internet, способы хранения и представления знаний, языки и инструментарий программирования, методы искусственного интеллекта, алгоритмы распределенных и облачных вычислений и т. д.

Научно-технические достижения в области искусственного интеллекта повлияли на формирование новых и трансформацию старых классов информационных систем — интеллектуальные информационные системы, системы интеллектуального анализа данных, экспертные системы, системы поддержки принятия решений и пр. К сожалению, все современные инструментарии в большинстве случаев разрозненны, решают частнопредметные задачи или заточены по одному классу систем. Вместе с тем уровень их автоматизации позволяет сделать вывод о возможности разработки «над-системы», интегрирующей в своем составе все наиболее развитые инструментарии (подходы, методы, моде-

ли, алгоритмы, технологии) в виде интеллектуального хранилища знаний и автоматизации процесса информационной поддержки принятия управленческих решений.

Стремление к повышению эффективности использования накопленной в электронном виде информации (банков данных, репозиториев, баз знаний и др.) путем их интеграции и унификации форматов хранений и процедур обработки привело к развитию Internet-технологий [3], основанных на методах интеллектуального анализа данных (data mining). Становится актуальной идея создания интегрированной Internet-системы интеллектуального хранилища знаний и автоматизации процесса информационной поддержки принятия управленческих решений. Определим основные концепции, рекомендации и технологии, планируемые использовать при создании такой системы, акцентировав свое внимание на программно-аппаратной составляющей — системе интеллектуальной обработки данных (СИОД).

Специфика СИОД накладывает значительный отпечаток на методологию и технологию ее разработки. Технология создания СИОД имеет свои существенные особенности и отличается от процесса проектирования и разработки других информационных и программных систем. Эти отличия в немалой степени определяются тем, что СИОД представляют собой интеллектуальные информационные системы (ИИС), базирующиеся на идеях, принципах и методах искусственного интеллекта (ИИ). Подобные системы требуют для своей реализации специализированные подходы, методы и технологии, в значительной мере отличающиеся от классических способов разработки программных систем. Вместе с тем постоянно возрастающий интерес к искусственному интеллекту способствует все более широкому внедрению и применению интеллектуальных систем в различных прикладных областях. Разработка систем искусственного интеллекта (экспертно-ориентированных систем, систем data mining, перевода, машинного зрения и т. д.) постепенно выходит на промышленный уровень, приобретает черты индустрии. Это приводит к тому, что к подобным системам начинают предъявляться те же требования, что и к традиционным программным продуктам.

Специфика СИОД заключается в том, что разработка и реализация интеллектуальных информационных систем достаточно длительный и трудоемкий процесс [1], который еще полностью не отработан и нередко требует новых идей и решений, влечет использование неординарных подходов, методов, технологий и средств. В результате эффективность и качество получаемой системы в значительной мере определяется талантом и опытом ее разработчиков. Кроме того, проблема реализации связана еще и с выбором подходящих средств разработки, представляющим собой отдельную сложную задачу. Современные аналитические информационные системы приобретают характерные черты и особенности программных систем [2, 4], основанных на алгоритмах интеллектуального анализа данных, экспертных систем, систем искусственного интеллекта и машинного обучения. Выделим следующие фундаментальные признаки СИОД, определяющие ее как инструмент экспертной поддержки принятия решений:

– Способность к решению широкого круга задач в некоторой неформализованной проблемной области.

– Способность извлекать знания из данных и представлять их в виде формализованных моделей знаний.

– Моделирование механизмов интеллектуальной деятельности человека.

– Использование знаний о предметной области.

– Применение эвристических методов решения задач.

– Способность к объяснению полученного решения.
– Высокая производительность. И т. д.

Технология облачных вычислений. Облаком называется вычислительный комплекс на базе виртуальных машин, способный с минимальной задержкой, по требованию, увеличивать или уменьшать свои характеристики.

Сервис удаленных вычислений можно назвать облачным, если он удовлетворяет правилам, сформированным в американском NITS (National Institute of Standards and Technology) (Standards Acceleration to Jumpstart Adoption of Cloud Computing (SAJACC) [http://www.nist.gov]. Все предоставляемые услуги в облаке разделим на три базовые категории:

– *Infrastructure as a servic*e — инфраструктура как услуга. Пользователю предоставляется неограниченный набор виртуальных машин, конфигурацию которых настраивает сам пользователь, также пользователь способен связывать их в сеть и таким образом строить свой виртуальный кластер той конфигурации, которая ему нужна. На данном уровне пользователь не имеет контроля только над железом, все остальное ему подвластно.

– *Platform as a service* — платформа как услуга. Уровень, где пользователь абстрагирован от уровня виртуальных компьютеров, сетей, кластеров. На данном уровне пользователь как бы имеет в своем распоряжении диск бесконечного размера, компьютер с бесконечной памятью и сверхбыстрый многоядерный процессор. На самом же деле платформа строится так, что организует кластер, который внешне выглядит как единая машина.

– *Software as a servic*e — программное обеспечение как услуга. Пользователь удаленно использует какую-либо программу. Существует облачный вариант 1С, Photoshop, оффисных программ. Пользователю доступна только конфигурация его учетной записи в прорамме и конфигурация самой программы в своей сессии.

Платформа интеллектуального анализа данных. Разрабатываемая платформа может относиться к разряду *SaaS Business Intelegence* (BI) или к *AaaS — Analysys as a Service*.

Система также подразумевает клиентское расширение функциональности, и предполагается, что система будет обладать частью функциональности PaaS.

Разрабатываемая платформа обладает следующими возможностями (SaaS BI):

– Как сервис предоставляется рабочее пространство (аутентификация и авторизация, средства загрузки и редактирования файлов).

– Библиотека готовых подсистем анализа и алгоритмов, в том числе: классификация и кластеризация, построение правил и деревьев решений, нейронные сети, генетические алгоритмы, статистические алгоритмы и т. д.

– Средства построения автоматического коллективного решения на основании алгоритмов.

– Средства анализа эффективности обучения на основе данных.

– Импорт/экспорт исходных данных и результатов обучения.

– Визуализация результатов анализа.

К стратегическим планам относится возможность расширения функций пользователя по созданию и корректировке индивидуальных разделов библиотеки алгоритмов (PaaS), в том числе:

– Редактор для составления новых алгоритмов на основе метаязыка.

– Возможность выкладывать алгоритмы в публичный доступ.

– Возможность строить сценарии с использованием алгоритмов анализа.

– Интеграция с бизнес-процессами пользователей/заказчиков (сервер автоматически по завершении анализа или обучения будет отправлять специально сконфигурированные данные на сервер заказчика, тем самым будет организован автоматический вывод данных в систему заказчика).

– Работа с сервером через API (автоматический ввод данных в СИОД).

Стратегический план разработки. Система интеллектуальной обработки данных (СИОД) представляет собой программно-аппаратный комплекс, настраиваемый на использование различных классов распознавания, языков и форматов представления данных и знаний, аппарата синтеза и анализа модельных представлений, а также ориентированный на предоставление услуг в сфере решения научных и прикладных задач обработки разнообразных данных.

Комплекс интегрирует в своем составе высокопроизводительное аппаратное обеспечение (в том числе мощные вычислительные серверы и обширные дисковые хранилища) и эффективное программное обеспечение (ПО), предназначенное для решения широкого спектра задач интеллектуального анализа данных и допускающее настройку и расширение поддерживаемых функций.

Система поддерживает и реализует концепции «публичных облачных вычислений» (*public cloud computing*) и обеспечивает предоставление следующих категорий «облачных» услуг:

– ПО как сервис (*Software as a Service, SaaS*) — пользователи системы посредством сети Интернет получают доступ к специализированному ПО и с помощью веб-интерфейса, функционирующего в любом современном браузере, имеют возможность использовать его для решения своих задач, в том числе ресурсоемких;

– инфраструктура как сервис (*Infrastructure as a Service, IaaS*) — пользователи системы при необходимости могут арендовать на длительный срок ресурсы компьютерной инфраструктуры, например вычислительных серверов или систем хранения данных, а также привлечь для решения своих задач интеллектуальные ресурсы, предоставляемые коллективом, обеспечивающим поддержку функционирования и развитие системы;

– платформа как сервис (*Platform as a Service, PaaS*) — пользователи на базе предоставляемого системой ПО и инфраструктуры имеют возможность дополнять функциональность существующего ПО, а также разрабатывать, тестировать и развертывать собственные веб-приложения, к которым другие пользователи могут получить доступ в рамках схемы SaaS.

Система интеллектуальной обработки данных характеризуется следующими особенностями:

– Доступ к системе обеспечивается посредством сети Интернет. Большинство функций системы может использоваться через веб-интерфейс, работающий в браузере. При необходимости работы с конфиденциальной информацией в системе обеспечивается шифрование передаваемых и хранимых данных (в частности, для доступа может быть использован протокол HTTPS).

– Основная часть услуг и ресурсов системы предоставляется только для зарегистрированных пользователей. Некоторые из ресурсов системы, например общеизвестные ин-

формационно-справочные материалы, открыты для широкого доступа и не требуют регистрации для использования.

– Зарегистрированные пользователи получают возможность доступа ко всему спектру услуг и ресурсов системы. Набор предоставляемых услуг и ресурсов, а также их качественные и количественные характеристики и ограничения определяются в соответствии со статусом пользователя в системе и могут со временем меняться в зависимости от его требований и задач.

– Работа в рамках системы осуществляется в соответствии с концепцией «проектов». Для решения своих задач пользователь с помощью ПО системы организует (создает) один или несколько проектов. Для каждого проекта пользователь может арендовать и использовать (в рамках существующих ограничений) услуги и ресурсы, необходимые ему для решения соответствующих задач. Автор (владелец) проекта может предоставить к нему доступ другим пользователям или их группам и при необходимости задать для каждого пользователя или группы набор прав, определяющих возможности и допустимые операции при работе над этим проектом. Действия всех пользователей, работающих в рамках проекта, протоколируются в системе и могут быть в дальнейшем проанализированы и в ряде случаев отменены.

– СИОД предоставляет программный интерфейс (Web API), позволяющий сторонним приложениям и веб-сайтам взаимодействовать с системой по протоколу HTTP и использовать ее ресурсы по определенным процедурам.

– Пользователи, которым требуется повышенный уровень безопасности и/или которые по каким-либо причинам не хотят или не могут передавать и хранить в системе свои данные, имеют возможность получить и установить на своих компьютерах специальное клиентское ПО, позволяющее в автономном режиме (без подключения к сети Интернет) использовать основные функции СИОД и выполнять обработку и анализ данных локально. При наличии соединения с Интернет клиентское приложение по требованию и под управлением пользователя может осуществлять полную или частичную синхронизацию его проектов с системой.

– Система является расширяемой в аспекте поддерживаемых функций и, соответственно, адаптируемой под различные классы решаемых задач. В состав системы входят инструменты и программные интерфейсы (API), позволяющие ее пользователям добавлять новые функции и модули и делать их доступными для других пользователей.

Основные функции СИОД:

– Система предоставляет средства для решения задач из следующих областей: статистическая обработка данных (корреляционный и регрессионный анализ, дисперсионный и ковариационный анализ и т. п.), распознавание образов (классификация с обучением), кластеризация (классификация без обучения), идентификация (выявление различительных признаков исследуемых объектов), прогнозирование (определение тенденций развития процессов), извлечение знаний из данных (*data mining*) и текстов (*text mining*). Благодаря открытой архитектуре и API системы набор поддерживаемых классов задач может быть достаточно просто дополнен, в том числе за счет использования решающих модулей сторонних разработчиков.

– Система поддерживает импорт данных из разных источников, в том числе из текстовых файлов в формате CSV, XML и HTML, из электронных таблиц в формате Excel и OpenDocument, из реляционных баз данных, а также из веб-приложений и служб, например, таких как таблицы Google, базы данных Zoho Creator или службы Яндекса. Кроме то-

го, данные могут вводиться, изменяться и редактироваться непосредственно через интерфейс системы. Экспорт данных и результатов может быть выполнен в файлы разных форматов, в частности CSV, XML, HTML, PDF, RTF, Excel, JPEG, PNG. Помимо указанных, в систему могут быть интегрированы другие конвертеры импорта/экспорта, обеспечивающие работу со специфическими форматами и источниками данных.

– СИОД содержит инструменты, обеспечивающие возможности для визуализации и графического представления исходных данных и результатов их обработки в различных формах, в том числе в виде всевозможных графиков и диаграмм, а также для формирования разнообразных отчетов, которые могут быть опубликованы в рамках системы или экспортированы для последующего использования вне СИОД.

– Система включает интерактивные учебники, справочники и модули тестирования по тематике интеллектуальной обработки данных, призванные обучить пользователей эффективному решению соответствующих задач с помощью системы, а также повысить их уровень знаний относительно моделей и методов обработки данных. Кроме того, СИОД предоставляет возможности для организации на ее базе электронных обучающих курсов и проведения автоматического тестирования знаний для разных предметных областей.

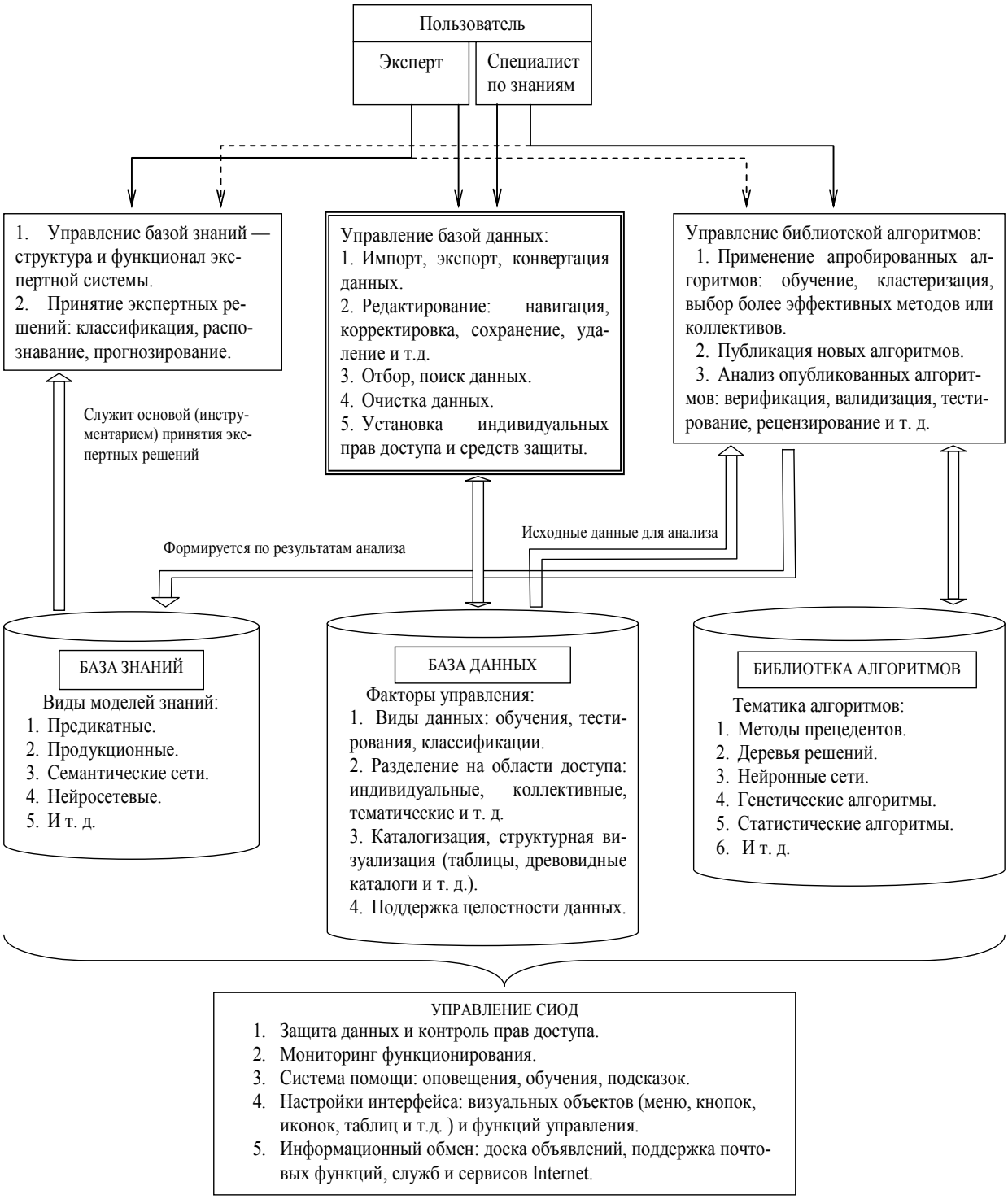
– Для каждого пользователя СИОД автоматически создает веб-сайт, функционирующий в рамках системы и доступный другим пользователям системы. На своем сайте пользователь может агрегировать и публиковать любые материалы, не противоречащие принятым в системе правилам, и при необходимости ограничивать к ним доступ. Помимо содержания, пользователи в определенных пределах могут изменять оформление и структуру своих сайтов.

– Система поддерживает различные коммуникационные сервисы для общения пользователей между собой и получения обратной связи от них, к числу которых относятся обмен электронными сообщениями (в том числе чат в режиме реального времени) и списки рассылки, подписка на новости и информирование о происходящих в системе изменениях (например, об изменениях на сайтах каких-либо пользователей), тематические форумы и дискуссии, проведение опросов и голосований.

Система управления. Система управления сайтом — это программный комплекс, позволяющий автоматизировать процесс управления как сайтом в целом, так и сущностями в рамках сайта: макетами страниц, шаблонами вывода данных, структурой, информационным наполнением, пользователями и правами доступа, а также по возможности предоставляющий дополнительные сервисы: списки рассылки, ведение статистики, поиск, средства взаимодействия с пользователями и т. д.

Проектные решения в рамках разработки интеллектуального хранилища знаний и автоматизации процесса информационной поддержки принятия управленческих решений вместе с технологиями интеллектуального анализа данных, методами искусственного интеллекта, моделями представления информации, электронными базами по различным предметным областям послужат инструментом для повышения эффективности научной, инновационной и образовательной деятельности, а также в промышленности, в экономике, в медицине и т. д.

Практическая реализация проекта позволяет использовать его как самонастраивающуюся, адаптивную, открытую интеллектуальную информационную систему с интегрированными функциями экспертной системы и подсистемами интеллектуального анализа данных (извлечения знаний).



Концептуальная схема WEB системы интеллектуальной обработки данных

Предложенный математический аппарат и методика его использования в процессе анализа данных являются инструментом научной оценки и формального обоснования принимаемых инженерных решений в различных областях прогнозирования, классификации, принятия решений.

СПИСОК ЛИТЕРАТУРЫ

1. Интеллектуальные информационные системы: Интеллектуальная информационная технология. Экспертные системы: Учеб. пособие / Д. В. Гаскаров, Д. В. Сикулер, В. В. Фомин, И. К. Фомина. СПб.: СПГУВК, 2004. 362 с.

2. Искусственный интеллект: В 3 кн. Кн. 1. Системы общения и экспертные системы: Справочник / Под ред. Э. В. Попова. М.: Радио и связь, 1990. 464 с.

3. *Марманис Х., Бабенко Д.* Алгоритмы интеллектуального Интернета: Пер. с англ. СПб.: Символ-Плюс, 2011. 480 с.

4. *Фомин В. В.* Автоматизация логического моделирования программного обеспечения с применением формального аппарата семиотических систем. СПб.: Энергоатомиздат, Санкт-Петербургское отделение, 2000. 250 с.

REFERENCES

1. Intellektual'nye informacionnye sistemy: Intellektual'naja informacionnaja tehnologija. Jekspertnye sistemy: Ucheb. posobie / D. V. Gaskarov, D. V. Sikuler, V. V. Fomin, I. K. Fomina. SPb.: SPGUVK, 2004. 362 s.

2. Iskusstvennyj intellekt: V 3 kn. Kn. 1. Sistemy obwenija i jekspertnye sistemy: Spravochnik / Pod red. Je. V. Popova. M.: Radio i svjaz', 1990. 464 s.

3. *Marmanis X., Babenko D.* Algoritmy intellektual'nogo Interneta: Per. s angl. SPb.: Simvol-Pljus, 2011. 480 s.

4. *Fomin V. V.* Avtomatizacija logicheskogo modelirovanija programmnogo obespechenija s primeneniem formal'nogo apparata semioticheskih sistem. SPb.: Jenergoatomizdat, Sankt-Peterburgskoe otdelenie, 2000. 250 s.