

Зайцева Евгения Алексеевна
студент, ФГБОУ ВО
«Государственный университет
управления», г. Москва
e-mail: eneige@yandex.ru

ПОСТРОЕНИЕ СОЦИАЛЬНОГО ПРОФИЛЯ КЛИЕНТА ДЛЯ ПЕРСОНАЛИЗАЦИИ БАНКОВСКОГО ПРЕДЛОЖЕНИЯ НА ОСНОВЕ ОТКРЫТЫХ ИСТОЧНИКОВ ИНФОРМАЦИИ С ПРИМЕНЕНИЕМ ТЕХНОЛОГИЙ БОЛЬШИХ ДАННЫХ

Аннотация. Банковская сфера – одна из самых динамично развивающихся в области информационных технологий. Современные средства аналитики больших данных предоставляют широкий спектр возможностей по сбору и последующей обработке гетерогенных данных, в том числе и персональных. Потребность в построении платформы для персонализации клиентского предложения обусловлена экспоненциальным ростом информации и стремлением к диджитализации банков. Эта платформа позволит значительно повысить уровень понимания потребностей своего клиента и персонализации банковских продуктов и сервисов.

Ключевые слова: персонализация, большие данные, анализ данных, неструктурированные данные, социальный профиль клиента, IBM InfoSphere BigInsights, IBM i2.

Zaytseva Evgeniya
Student, State University
of Management, Moscow
e-mail: eneige@yandex.ru

PERSONALIZATION OF BANKING SERVICES BASED ON open source analysis via use of big data

Abstract. The banking sector is one of the most dynamically developing areas in the field of information technology. Modern big data analytics tools provide a wide range of opportunities for the collection and subsequent analysis of heterogeneous data, including personal one. The need to build a platform for personalization of the client's offer is due to the exponential growth of information and banks' desire to go digital. This platform will significantly increase the level of understanding of banks' customers' needs and personalize banking products and services properly.

Keywords: personalization, big data, data analysis, data structuring, social profile, IBM InfoSphere BigInsights, IBM i2.

Быстрый рост объемов информации в обществе ставит бизнес в новые условия функционирования. Самыми большими по генерируемому информационному потоку источниками становятся социальные сети, хранилища мультимедийного контента, системы контроля и мониторинга. Из-за объема и гетерогенности данных, сбор и анализ такого рода информации с помощью классических информационных систем весьма затруднителен: стандартные реляционные системы управления базами данных и сервера не справляются с экспоненциально растущим объемом [2]. В таких случаях на помощь призывается сравнительно новая концепция больших данных (англ. big data). Основным ее преимуществом является скорость обработки разнородной информации, независимо от ее объема и разнообразия. Если разложить большие данные на традиционные составляющие, можно выделить основные технологии [2]:

1) NoSQL – модель данных, представляющая возможность отойти от ограничений, заданных при создании реляционных баз данных. Это становится возможным благодаря согласованности и атомарности данных.

2) MapReduce – технология параллельных вычислений над большими массивами данных в распределенных кластерах.

3) Hadoop – представляет собой комплекс программного обеспечения, предоставляющий широкий набор возможностей по разработке и использованию распределенных программ на кластерах. Одним из популярных применений в настоящее время является его использования для реализации поисковых и контекстных механизмов высоконагруженных веб-сайтов.

В статье рассмотрена возможность применения аналитической платформы IBM InfoSphere BigInsights, реализующей вышеприведенные технологии, в сочетании с IBM i2, которая позволяет работать с нереляционными графовыми базами данных, с целью построения социального профиля клиента на основе открытых источников

информации. Актуальность работы обусловлена популяризацией «Интернета вещей» и необходимостью соответствия информационного и технического обеспечения, а также растущей потребностью бизнеса знать каждого клиента в лицо и персонализировать предложение или услугу.

Обсудим, что такое персонализация и зачем она нужна. Персонализация – это процесс, включающий поиск и обработку информации об интересах клиента с целью последующего личного обращения к клиенту в более личной форме, с учетом его потребностей и желаний. Индивидуальное взаимодействие располагает клиента к приобретению услуг или продуктов, заставляет его чувствовать себя уникальным и особенным. В банковской сфере это особенно важно. В эпоху высокой конкуренции знать своего клиента – это необходимое условие успешного функционирования банковского бизнеса. В настоящее время знать о потребностях каждого клиента достаточно затруднительно. Так, например, согласно данным Сбербанка на данный момент его розничными клиентами являются около 127 млн человек в России и более 10 млн за ее пределами. Помимо этого, также нельзя упускать возможность анализировать бывших и потенциальных клиентов.

Персонализация становится актуальной в силу следующих причин:

- возникновение проблем, связанных с ростом потоков данных. Банки имеют дело с постоянно растущими наборами данных, что означает, что они нуждаются в возможностях, которые могут помочь фильтровать сетевую информацию быстрее и эффективнее;
- необходимость лучшего понимания целевых сетей. Из-за динамичности целевых сетей банковскому сектору необходимо быстро определять потенциальных ключевых лиц/групп для лучшего распределения ограниченных ресурсов;
- необходимо смотреть за пределы структуры клиентской сети с учетом ее динамики. Определять характеристики сетей, которые не являются очевидными, а также анализировать, как эти сети изменяются с течением времени;
- в социальных сетях не все взаимодействия являются соответствующими действительности, и банки должны быть в состоянии использовать такие методы, как взвешивание отношений между субъектами, чтобы принять во внимание, как такие взаимоотношения могут повлиять на клиентскую сеть и взаимоотношения с клиентом в целом.

Таким образом, обычные средства аналитики не смогут обеспечить персонализацию банковского предложения для каждого клиента. Именно это и является ключевой идеей данной статьи. Чтобы решить вышеперечисленные проблемы предлагается методика построения социального профиля клиента.

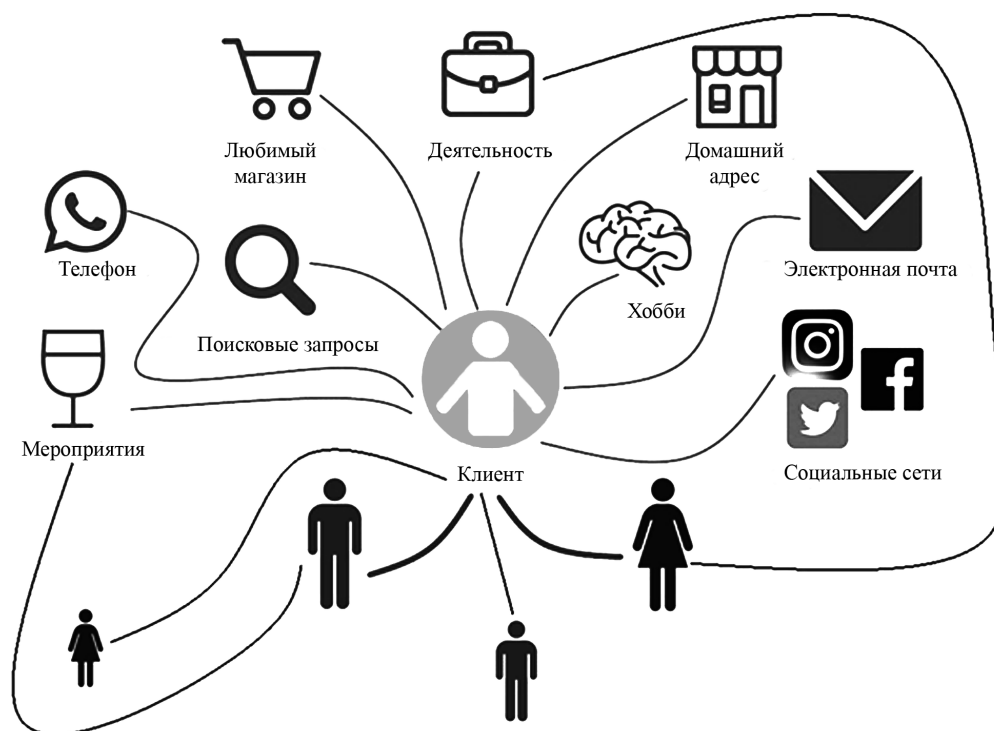
Социальный профиль клиента состоит из двух основных частей:

- профиль интересов;
- профиль окружения.

Профиль интересов помогает выявить информацию о клиенте на основе открытых источников. Социальные сети, поисковые запросы – все это может сказать о человеке гораздо больше, чем он сам. Профиль окружения выявляет круг лиц, с которыми общается клиент. При дополнительном анализе можно оценивать не только возможности, но и угрозы.

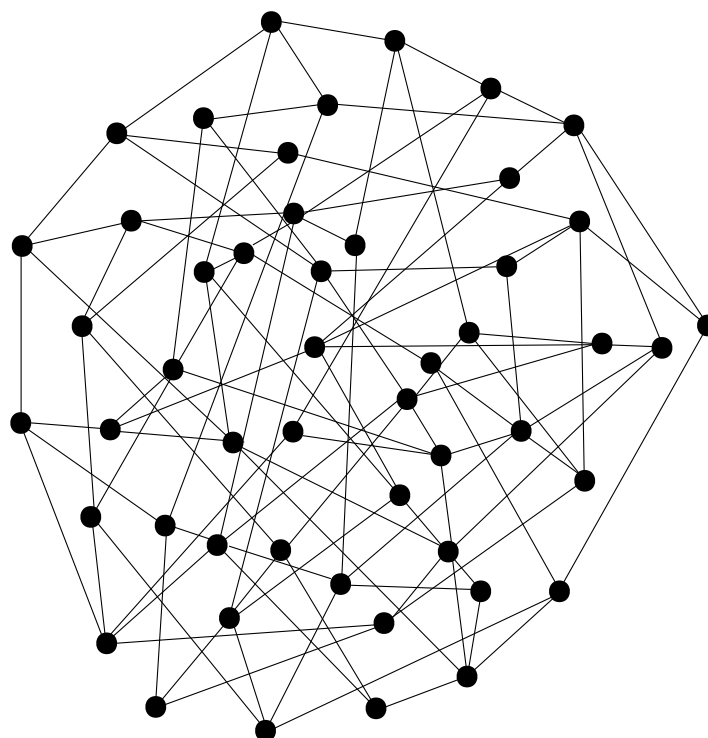
Использование программного продукта IBM i2 позволяет выявить скрытые взаимосвязи между данными, собрать их в цельные структуры и визуализировать результаты как в табличном представлении, так и в виде графов. Эти результаты можно интерпретировать путем выявления характеристик клиента, которые могли быть не указаны явно, имели противоречивые сведения, или были неизвестны. Такими характеристиками могут быть возраст, хобби, значимые события, связи с другими людьми, известность в какой-либо области деятельности и т. д. (см. рис. 1).

IBM i2 позволяет рассчитывать такие показатели, как взаимозависимость, близость, степень влияния и характеристические векторы (включая концентратор и авторитет), центральные меры, чтобы обеспечить различные точки зрения на социальные отношения в сети. Это позволяет анализировать роль клиента в его окружении, людей, которые влияют на него и его жизнь, тем самым понимая и узнавая о нем и о его потребностях больше. На рис. 2 представлен пример ненаправленной сети пользователей. Точки на рис. 2 – это связанные взаимоотношениями в сети клиенты, ребра – это взаимосвязи между ними. Отображение возможно менять и рассматривать в контексте определенных показателей, таких как степень близости, центральность, направление связи и другие. На рис. 2 сеть представлена схематично.



Составлено автором по материалам исследования

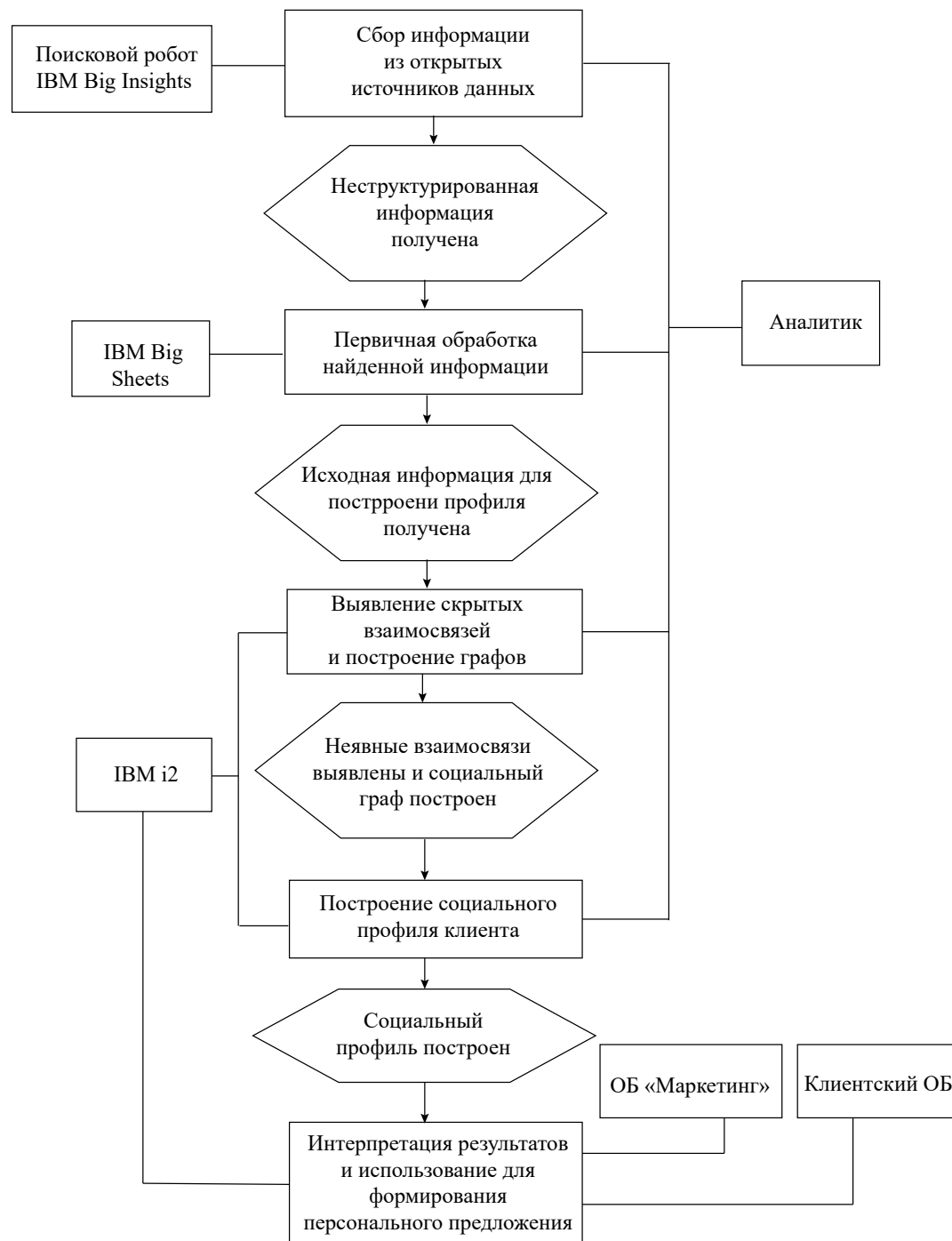
Рис. 1. Пример графа типов сущностей для социального профиля клиента



Составлено автором по материалам исследования

Рис. 2. Пример ненаправленной сети пользователей

Последовательность шагов проектирования социального профиля и элементы системы представлены на рисунке 3 с помощью программного продукта Aris Express.



Составлено автором по материалам исследования

Рис. 3. Модель процесса создания социального профиля клиента

Слева выделены элементы информационной платформы, отвечающие за тот или иной процесс. Справа выделяются ключевые лица, которые контролируют и выполняют этот процесс. ОБ «Маркетинг» – операционный блок «Маркетинг»; клиентский ОБ – операционный блок по работе с клиентами.

Блочная организационная структура характерна для многих российских и зарубежных банков (например, Альфа-Банк) [5]. Для первоначального сбора исходных данных можно воспользоваться двумя встроенными

приложениями платформы IBM BigInsights [4]: BoardReader осуществляет поиск в блогах и новостных лентах на основе ключевых слов и заданного временного диапазона, в течение которого появлялась искомая информация; WebCrawler является поисковым роботом, который был специально создан для сбора больших данных в распределенные базы данных по типу HBase. Используется для сбора информации внутри корпоративных баз данных.

В качестве входных параметров поиска необходимо задать критерии, с помощью которых можно с достаточной степенью точности идентифицировать человека (ФИО, никнейм, e-mail адрес, номер телефона, идентификаторы учетных записей). Полученная в ходе поиска информация представляется в виде списка с указанием критерия поиска и ссылки на конечный адрес, где была найдена информация, что по своей структуре напоминает работу обычных поисковых сервисов. После того, как список получено, можно добавить дополнительные критерии поиска и начать обработку неструктурированных данных.

Результаты поиска представляются во фрагментированном, необработанном виде. Для дальнейшего анализа требуется удалить повторяющуюся информацию и шумы (пустые абзацы, знаки табуляции в конце строк, части HTML-кода). Помимо этого, необходимо найти явные зависимости в текстах и собрать их в единое целое. Для этого используется платформа IBM BigSheets, с помощью которой отфильтровывается ненужная информация, прошедшая предварительную очистку, объединяются и представляются данные в виде книг – электронных таблиц, поддерживающих обработку с помощью BigSQL и AQL языков. Впоследствии их можно полностью или частично экспортировать в CSV/TSV файлы. Благодаря этому становится возможным проведение последующего детального анализа с извлечением структурированной информации и представлением ее в графическом виде для оценки.

IBM i2 состоит из нескольких программных продуктов, позволяющих выявлять скрытые зависимости между данными, собирать их в структуры и визуализировать результаты как в табличном представлении, так и в виде графов [3]. Таким образом, с его помощью, можно выявлять характеристики человека, которые либо не указаны явно, либо имеют противоречивые сведения, такие, как, например, возраст, увлечения, связи с другими людьми, известность в какой-либо области деятельности, любимые магазины, семейное положение и многое другое [1]. Смысл заключается в создании базы данных социального профиля и построении на ее основе социального графа.

В результате глубокого анализа, обеспеченного постоянно обновляющейся информацией, банк имеет максимально полную и актуальную информацию о своих текущих, бывших и потенциальных клиентах, способствуя построению максимально доверительных отношений с ними. Эта технология позволяет делать уникальные, персональные предложения в режиме real-time, анализировать клиентские сегменты и улучшать взаимодействие с клиентами через маркетинговые каналы, доступные банку.

Библиографический список

1. Анализ структурированных и неструктурированных данных с помощью ContentAnalytics / Центр компетенции по технологии IBM BigData. – М., 2017. – 66 с.
2. Бершадский, А. М. Разработка и моделирование гетерогенных инфраструктур для беспроводного информационного обеспечения процессов мониторинга / А. М. Бершадский, А. Г. Финогеев, А. С. Бождай // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2010. – № 1. – С. 36-46.
3. Центр компетенции технологии IBM BigData / Выявление скрытых связей на основе анализа текстов с помощью i2 – М., 2017. – 66 с.
4. Центр компетенции технологии IBM BigData / Исследование текстовой информации с помощью DataExplorer – М., 2017. – 23 с.
5. Официальный сайт Альфа-Банка: [Электронный ресурс]. – Режим доступа: https://alfabank.ru/about/corporate_governance/orgstructure/ (дата обращения: 18.03.2018).

References

1. Tsentr kompetentsii po tekhnologii IBM BigData Analiz strukturirovannykh i nestrukтуриrovannykh dannykh s pomoshh'yu ContentAnalytics [*Structured and unstructured data analyze using Content Analytics*]. Moscow, 2017. p. 66.
2. Bershadskij, A. M. Razrabotka i modelirovanie geterogennykh infrastruktur dlya besprovodnogo informatsionnogo obespecheniya protsessov monitoringa / A. M. Bershadskij, A. G. Finogeev, A. S. Bozhday [*Design and simulation of heterogeneous wireless*

- infrastructures for information support of processes of monitoring*]. Izvestiya vysshikh uchebnykh zavedenij. Povolzhskij region. Tekhnicheskie nauki. 2010. № 1. pp. 36-46.
3. Tsentr kompetentsii tekhnologii IBM BigData Vyyavlenie skrytykh svyazej na osnove analiza tekstov s pomoshh'yu i2 [*Identification of hidden relationships based on text analysis using i2*]. Moscow, 2017. pp. 66.
 4. Issledovanie tekstovoj informatsii s pomoshh'yu DataExplorer / Tsentr kompetentsii tekhnologii IBM BigData [*Textual information analyze using Data Explorer*]. Moscow, 2017. p. 23.
 5. Ofitsial'nyj sajt Al'fa-Banka [*Alfa Bank's official page*]. Available at: https://alfabank.ru/about/corporate_governance/orgstructure/ (accessed 18.03.2018).