

**Kamila N. Muratova**, Master of Tech. Sci., lecturer, «Information Systems» Department, International Information Technology University.

**Bakhtgerey Sinchev**, Doct. of Tech. Sci., Professor, «Information Systems» Department, International Information Technology University.

**Авторлар туралы мәлімет:**

**Ауезова Әнел Саттарқызы**, PhD-докторанты, Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының лекторы.

**Муратова Камила Нурланқызы**, магистр, Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының лекторы.

**Синчев Бахтгерей**, т.ғ.д., Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының профессоры.

УДК 530.1, 681.3.06

**Нәдіров Н.Қ.<sup>1</sup>, Дүйсебекова К.С.<sup>1</sup>**

<sup>1</sup>Международный университет информационных технологий, Алматы, Казахстан

## **РАЗРАБОТКА СИСТЕМЫ ФОРМИРОВАНИЯ ПРОФИЛЯ КЛИЕНТА НА ОСНОВЕ BIGDATA С ИСПОЛЬЗОВАНИЕМ СЕМАНТИЧЕСКОГО АНАЛИЗА**

***Аннотация.** По мере того, как бизнес-процессы становятся все более и более ориентированными на клиента, чрезвычайно важны маркетинговые и управленческие отношения с клиентами, что заслуживает особого внимания с учетом глобализации и растущей конкуренции на рынке. На данном этапе возникает много всевозможных вопросов «Кто такие клиенты?» и «Как их идентифицировать?». В настоящее время «Какую информацию необходимо собирать о потенциальных клиентах?» и «Каковы сейчас уровни обслуживания?».*

*Цель разработчика заглянуть в «голову» собственного клиента и осознать, чего он желает. Эти познания станут неоценимыми позднее, когда найденные данные и информация будут изучены, осознание всех задач и вопросов станет неотъемлемой частью успеха будущего проекта.*

*В данной статье представлен способ подготовительной обработки данных для умственного изучения данных, доктрина использования, способы изучения больших размеров данных и извлечения полезной информации, содержащейся в них полезной информации для формирования профиля клиента.*

***Ключевые слова:** Большие данные, программное обеспечение, информационная система, структурированные и неструктурированные данные, социальный профиль человека.*

### **Введение**

Актуальность темы анализа состоит в том, что сейчас случается повсеместное проникновение IT-продуктов в жизнь современного человека независимо от его сферы работы. Все это дает большое число преимуществ, которые, непременно, положительно влияют на работу человека.

Обычно каждый проект с приложением проходит несколько шагов либо актуальных циклов от постановки трудности до её визуализации либо представления готового решения. В статье мы рассмотрим любой из этих шагов больше тщательно на примере усовершенствования продаж, принятая схема процесса представлена на Рисунке 1.



Рисунок 1 - Стадия процесса

Чтобы начать работу с данными, необходимо выполнить ряд шагов, показанных на Рисунке 1. Поставить задачу, определить цель исследования, кто наш клиент, чего именно клиент ожидает от нас. И другие вопросы, определяющие проблемы и цели. Надо будет собрать необработанные данные, чтобы решить проблему. Чтобы подготовить и обработать данные, мы должны сначала очистить данные, чтобы превратить их в форму, которую мы можем продолжить анализировать в более удобной форме. Построение модели данных - самый важный этап этого проекта, на котором необходимо применить все технологии для анализа данных, показывает результат визуализации происходящего во время исследования.

Основные этапы и процесс представления больших данных рассматриваются для выбора информации, необходимой для создания социального профиля человека. Анализируются системы и сервисы для сбора данных, в том числе социальные сети, которые предлагают множество возможностей. Среди них есть довольно крупные системы и проекты, которые помогают собирать информацию легально. Наконец, из полученных исследований сделаны некоторые выводы.

**Целью** этого анализа является разработка системы формирования профиля клиента на базе больших данных с внедрением семантического изучения.

Для достижения поставленной цели нужно решить последующие научные задачи:

- Изучение методов машинного обучения;
- Исследование больших данных;
- Уточнение структуры профиля клиентов, типы и характеристики, которые их характеризуют.
- Исследование и создание данных о клиентах;
- Обобщение и классификация зарубежного опыта для увеличения профиля клиента;
- Проведение исследования доступных методов анализа профиля клиентов и определение наиболее эффективных методов для казахстанских компаний.
- Разработка системы профилей клиентов для производителей товаров и услуг с целью предоставления практических советов по их созданию.

Разработка системы профилей клиентов, повышающей продуктивность продаж предприятия, которая включает оценку клиентского капитала и прогнозирование действий клиентов.

Этот этап подразумевает использование ранее полученной информации о данных. Здесь применяются методы из области статистики, исследования операций машинного обучения и т.д. Процесс моделирования состоит из четырех шагов:

- планирование показателей и выбор модели;
- исследование модели и получение результатов;
- проверка адекватности созданной модели;
- применение отлаженной модели к исследуемым данным.

Такие модели предсказывают выбор между двумя или более дискретными альтернативами, в отличие от стандартных эконометрических моделей, оперирующих в основном непрерывными переменными.

При составлении социального профиля человека можно ограничиться минимальными параметрами таких как фамилия, имя, пол, дата рождения, адрес профиля в социальной сети, характеристики.

Рассмотрим архитектуру предлагаемого решения. В дополнение к приложению, серверная часть системы включает последующие составляющие, подобные как планировщик, поисковый робот, службу обработки данных, сервер реляционной информационной базы и сервер хранилища данных.

Главная цель, которая решается планировщиком, обновление и интеграция соц. профилей претендентов в уже существующие в базе данных по истечении данного промежутка времени. Данный механизм предназначен для поддержания актуальности данных, так как соц. профиль претендента может временно изменяться в зависимости от его активности на сайтах соцсетей.

### Исследование данных (Сбор данных)

После загрузки необходимых данных на их подготовку и очистку уходит много времени. На этом этапе происходит интеллектуальный анализ данных. Изучение данных должно помочь понять их. Для этих целей используется описательная статистика и простое моделирование. Здесь можно применять различные техники визуализации, от простых графиков и диаграмм до графиков и диаграмм, пример которых показан на Рисунке 2.

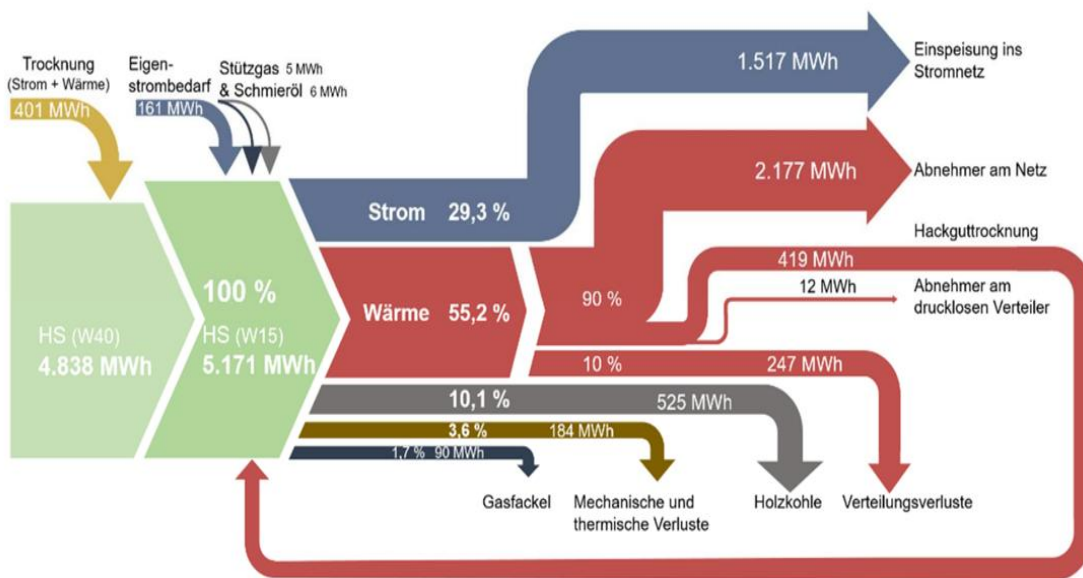


Рисунок 2 - Интеллектуальный анализ данных

Рассмотрим главные приемы визуализации данных. На Рисунке 3 представлен метод совмещения диаграмм, а на Рисунке 4 показан метод связывания диаграмм, при котором изменения одной диаграммы переносятся на другую.

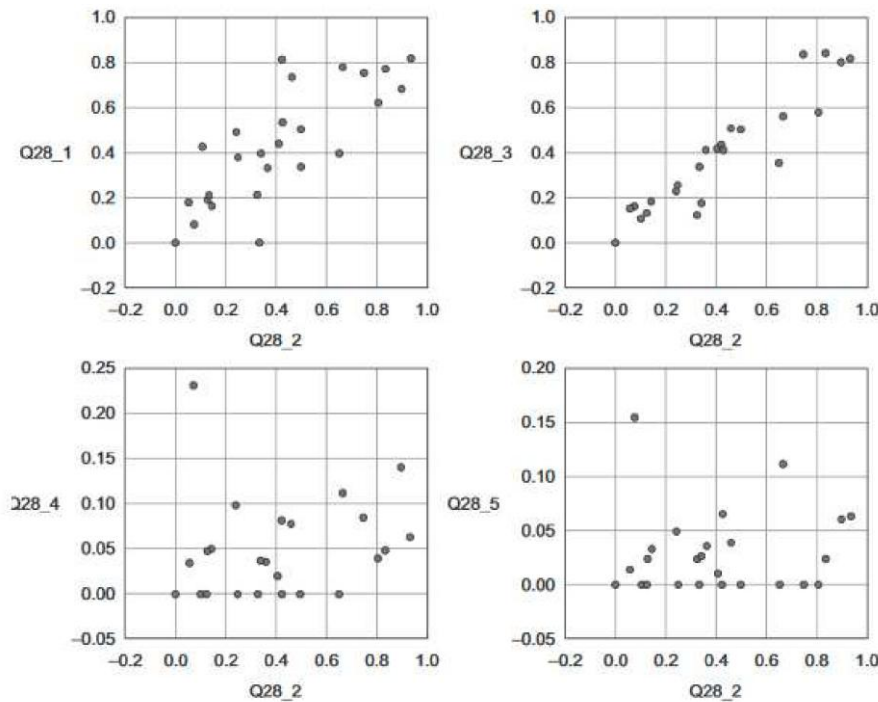


Рисунок 3 - Метод совмещения диаграмм

Все эти методы представлены в виде визуализации, но никто не запрещает использовать другие методы на этом этапе.

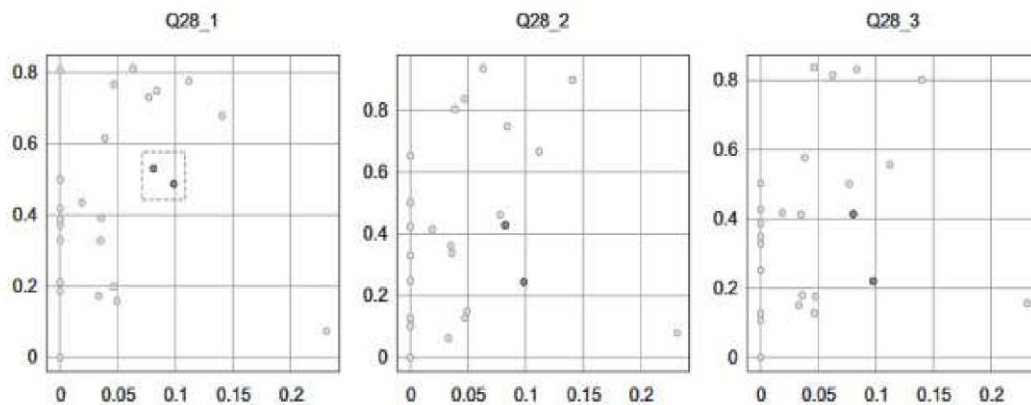


Рисунок 4 - Метод связывания диаграмм

**Моделирование данных**

Данный шаг подразумевает внедрение до этого приобретенной информации о данных. Он употребляет способы статистики, операционных исследовательских работ, машинного обучения и др.

Процесс моделирования состоит из 3-х шагов: выбор модели, исследование модели и получение итогов.

**Подготовка данных**

На этом этапе работа с данными не завершается, еще необходимо провести проверку и удостовериться, что они незапятнанные. Данный шаг обычно занимает огромную часть ежедневной работы работника. Любая допущенная вами ошибка может востребовать дополнительной работы.

Специалист анализирует данные в процессе их подготовки к ввозу и анализу.

Подготовка включает проверку типов данных на сходимость и выявление отсутствующих и покоробленных данных. Учитываются главные обнаруженные в данных ошибки, которые нужно поправить. Прежде всего это ошибки ввода и излишнее число пробелов, несоответствие знаков, некорректные значения, отличия от общих правил, различные единицы измерения и различные уровни агрегирования.

### **Математическая постановка задачи**

Формирование профиля клиента с использованием больших данных и имитационного обучения выглядит следующим образом: существует фактор-вектор  $\bar{X}$ , отражающий информацию, представленную в профиле клиента, которая влияет на зависимую переменную  $Y$ , которая, в свою очередь принимает всего два значения: 1 - если заказчик наш покупатель и 0 в противном случае. Тогда вероятность того, что  $Y = 1$ , должна быть выражена как функция факторов согласно формуле (1):

$$\text{Probability: } (Y=1|\bar{X}) = F(X) \tag{1}$$

Несмотря на то, что все профили клиентов не могут быть заполнены одинаково, из вышесказанного можно сделать вывод, что какая информация будет присутствовать в профиле клиента, почти наверняка.

- Пол клиента (1 - мужской, 2 - женский).
- Возраст (в годах).
- Работа (1 - да, 2 - нет).
- Средняя стоимость покупки (в месяц / год).
- В основном то, что он покупает (количество).
- Покупатель (0 - нет, 1 - да).

Как видно, модель должна включать в себя достаточно большой объем качественных данных, показывающих наличие или отсутствие той или иной особенности в рассматриваемом объекте, и может иметь как номинальную, так и ранговую [5].

Предполагая, что зависимая переменная  $Y$ , которая представляет собой возможность или невозможность взять на работу кандидата (или, в случае с рекрутинговым агентством – рекомендовать его клиенту), принимает только два значения:  $\{0; 1\}$  (см. формулу 2).

$$(Y = 1 | x) = F(x^T \beta) \tag{2}$$

Вероятность того, что она примет соответствующее значение можно выразить как функции некоторых факторов (3):

$$(Y = 0 | x) = 1 - F(x^T \beta) \tag{3}$$

Набор характеристик  $\beta$  отражает воздействие изменения каждого фактора на итоговую возможность. Задача заключается в том, чтобы найти пригодную функцию в правой части уравнения. Одна из возможностей – применять регулярную линейную регрессию (4), также мы можем построить регрессионную модель (5):

$$F(x, \beta) = x^T \beta \tag{4}$$

$$y = x^T \beta + \varepsilon \tag{5}$$

Широкую известность в эконометрике приобрела так называемая пробит-модель, использующая функцию (6) стандартного нормального распределения:

$$\Pr obability (Y = y | x) = \int_{-\infty}^{x^T \beta} \phi(t) dt = \Phi(x^T \beta) \quad (6)$$

Для обработки данных можно создать свой элемент системы, можно применять фреймворк Apache Spark.

Apache Spark - это программный пакет с открытым начальным кодом в целях выполнения распределенной обработки неструктурированных и полуструктурированных данных.

В качестве СУБД для реляционных БД можно применять SQL Server, приложений с базами данных можно применять ORM Entity.

Framework. Для хранения неструктурированных данных предлагается применять Apache Hadoop.

### Исследование семантики в рамках теории «Смысл $\Leftrightarrow$ Текст»

При разработке теории «Значение  $\Leftrightarrow$  Текст» было введено понятие лексической функции. Лексическая функция - определенное смысловое отношение, к примеру, «Равенство по значению», «противоположность значения», «обобщающее понятие» и т.д. Пусть будет несколько лексических единиц - слова и словосочетания. После эта лексическая функция связывает каждую из этих единиц с набором лексических единиц, которые находятся с начальной единицей в соответственном семантическом отношении.

Значения лексических функций от различных доводов могут быть всецело либо отчасти схожими. Значения различных функций аргумента также могут быть схожими. Кроме обычных лексических предикатов, их композиции - составные предикаты - могут употребляться для описания лексической сопоставимости.

Помимо простых лексических предикатов, для описания лексической сочетаемости могут использоваться и их комбинации – составные предикаты.

Необыкновенную роль в исследовании семантики играет валентность слов, другими словами способность слов вступать в связь с иными словами. Есть два типа валентностей слов: синтаксические и семантические. Однако это деление время от времени бывает достаточно условным. Ситуации, которые описываются словами на естественном языке, имеют, обычно, от 1-го до 4 семантических компонент, либо семантических актантов. При всем этом каждое слово соединено с глубокими синтаксическими актантами - зависимыми словами, которые соответствуют предмету и дополнениям. Семантическая валентность определяется лексическим изучением ситуации, данной определенным словом. Синтаксические валентности определяются числом синтаксических актантов, которые были представлены конкретно в тексте и данных контекстом.

С формальной точки зрения мы имеем конструкцию, описанную ниже (см. формулу 7). Чтобы не связывать с каждым глаголом (и другими словами) отдельный предикат, рассматривается предикат, размерность которого больше на 1:

$$P^{val}(y, x_1, x_2, \dots, x_n) \quad (7)$$

при всем этом  $y$  будет само слово, а  $x_1, x_2, \dots, x_n$  – его валентности. Различать синтаксические и семантические актанты (см. формулу 9) можно при помощи мультииндексов, которые указывают, какие конкретно актанты заданы в тексте (см. формулу 8), (см. формулу 10). Запись

$$P_{1\dots n}^{val}(y, x_1, x_2, \dots, x_n). \quad (8)$$

Формула значит, что заданы актанты

$$i_1, i_2, \dots, i_k. \tag{9}$$

А именно, если заданы все актаны, выходит

$$P_{i_1 i_2 \dots i_k}^{val}(y, x_{i_1}, x_{i_2}, \dots, x_{i_k}) \tag{10}$$

Некоторые варианты (мультииндексных наборов) (см. Формулу 11) могут быть непремлемыми в языке. Если огромное количество  $i_1, i_2, \dots, i_k$  приемлемо, импликация имеет место

$$\begin{aligned} & \forall y \forall x_1 \dots \forall x_n (P_{1\dots n}^{val}(y, x_1, x_2, \dots, x_n) \rightarrow \\ & \rightarrow P_{i_1 i_2 \dots i_k}^{val}(y, x_{i_1}, x_{i_2}, \dots, x_{i_k})). \end{aligned} \tag{11}$$

В данном случае набор слов в толково-комбинаторном словаре можно изучать как под-модель начальной модели, являющейся языком. Лексические предикаты, которые сейчас определены на более узком наборе, будут действовать аналогичным образом.

### Семантические роли и семантические ограничения

Семантические сети разрешают представить семантику некоторого слова в согласовании с его внутренней структурой. Если вместе с этой структурой учитываются грамматические индивидуальности, то значение утверждения может быть представлено в определениях семантических ролей и связанных семантических ограничений.

Кроме термина «семантические роли» в литературе употребляются также понятия, направленные на определенную тематику роли, тета-роли, глубочайшие кейсы. Дж. Грубер и К. Филлмор числятся основателями этого направления семантических исследовательских работ. На самом деле, эти понятия тесновато соединены с семантическими и глубинными синтаксическими факторами, исследователь которых И.А. Мельчук. Вот некоторые из семантических ролей, обсуждаемых в статьях.

По числу доводов и их семантическим свойствам почти все глагольные лексемы можно поделить на классы: глаголы физического действия (рубить, видел), глаголы восприятия (глядеть, слушать), глаголы о манере речи (орать, шептать) и т.д. В каждом классе есть более четкое деление. В числе глаголов физического сотрудничества глаголы глагольной формы (агент, инструмент, объект) имеют такую же семантическую структуру предиката-аргумента: разламывать - разламывать, разламывать - делить и т.д. Иная структура предиката-аргумента свойственна для глаголов глагольной формы (агент, инструмент, цель) и т.д.

### Профили поведения клиентов

Поведение покупателей - это процесс, средством которого люди решают, следует ли, что, когда, где, как и у кого получать товары и услуги. Однако Моуэн и Минор определяют поведение клиентов параличному.

Поведенческое профилирование основано на отношении покупателей к продукту, его использовании или реакции на него.

Рекламодатели считают, что поведенческие переменные, которые является наилучшими отправными точками для построения поведенческого профилирования, включают:

1. Случай: клиенты формируются по времени, когда они получают мысль о приобретении, совершении покупки либо использовании обретенного продукта. Компания может избрать один из видов маркетинговой стратегии перед праздниками.

2. Достоинства: это процесс, разделяющий клиентов на базе разных преимуществ, которые они могут найти в продукте.

3. Статус пользователя: профилирование в согласовании с непользователями, которые были пользователями, возможными пользователями, новенькими и неизменными пользователями продукта, компания может настроить и индивидуализировать собственный маркетинг для каждой группы.

4. Скорость использования: профиль коэффициента использования разделяет клиентов зависимо от того, сколько они потребляют продукт. Они разбиты на группы непользователей, пользователей легких, средних и тяжелых продуктов.

5. Этап готовности клиента: этап готовности клиента относится к осведомленности клиента и интереса к продукту.

6. Статус лояльности: клиент также быть может профилирован в согласовании с его лояльностью. Твердые лоялисты - это клиенты, которые не один раз приобретают один и тот же продукт; разбитые лоялы - это клиенты, которые привязаны к двум либо трем брендам и получают их случайным образом; сдвиговые лоялы - это клиенты, переключаемые с 1-го бренда на иной, придерживаясь этого бренда некое время, пока не переключатся на иной.

7. Отношение: клиентов можно поделить зависимо от того, относятся ли они к продукту с энтузиазмом, положительно, индифферентно, негативно либо воинственно. Изучая отношение покупателей к бренду либо продукту, компания получает широкий диапазон познаний о рынке и собственных клиентах.

### **Выводы**

Из всего вышесказанного можно сделать еще один очень важный вывод. Выражая свою мысль, человек ориентируется на необходимые ему вещи. В данной работе были рассмотрены главные способы и основы сегментного анализа маркетинга для формирования профиля клиента. В действительности есть значительно больше способов определения рыночного раздела. В конечном результате, сегментация - не самоцель. Ведущей ее задачей считается ответ на вопрос - имеют ли быть уделены устойчивые группы покупателей предоставленного продукта. В случае, если эти группы не отличаются, то работа компании имеет возможность быть направлена на всех клиентов предоставленной продукции (так именуемая стратегия глобального маркетинга).

В заключении можно с уверенностью сказать, что разработка системы профилирования клиентов на основе Bigdata посредством семантического анализа имеет право на существование и будет востребована на рынке программного обеспечения.

### **ЛИТЕРАТУРА**

1. А.В. Шмида. Новые методы работы с большими данными: победные стратегии управления в бизнес-аналитике// Научно-практический сборник. – М.: ПАЛЬМИРА, 2016. – 528 с.
2. Базенков Н.И., Губанов Д.А. Обзор информационных систем анализа социальных сетей: Управление большими системами, 2013. – 357-394.
3. Бодянский Е.В. Искусственные нейронные сети: архитектуры, обучение, применения: монография Е.В. Бодянский, О.Г. Руденко. – Харьков: Телетех, 2004. – 369 с.
4. Веснин В.Р. Управление человеческими ресурсами. Теория и практика: учебник / В.Р. Веснин. – М.: Проспект, 2014. – 688 с.
5. Вязигин А.В. Оценка персонала высшего и среднего звена. А.В. Вязигин. – М.: Вершина, 2006. – 256 с.



6. Дорофеев В.Д. Эффективность управления трудовыми ресурсами предприятия при внедрении системы менеджмента качества: монография В.Д. Дорофеев, А.Н. Шмелева. – Пенза: Информационно-издательский центр ПГУ, 2008. – 210 с.
7. Дрогобыцкий И.Н. Системный анализ в экономике: учебник для студентов вузов, обучающихся по специальностям «Математические методы в экономике», «Прикладная информатика» И. Н. Дрогобыцкий. - 2-е изд., перераб. и доп. – М.: ЮНИТИ-ДАНА, 2012. – 423 с.
8. Маслова В.М. Управление персоналом: толковый словарь. В.М. Маслова. – М.: Издательско-торговая корпорация «Дашков и К», 2013. – 120 с.
9. Зайцева Т.В., Зуб А.Т. Управление персоналом. – М.: ИД «ФОРУМ»: ИНФРА-М, 2013. – 336 с.
10. Губанов Д.А., Новиков Д.А., Чхартишвили А.Г. Социальные сети: модели информационного влияния, управления и противоборства. – М.: Изд-во физико-математической литературы, 2010. – 228 с.
11. Лахути Д.Г., Рубашкин В.Ш. Средства и процедура концептуальной интерпретации входных сообщений на естественном языке: Изв АН СССР: Сер. Технич. киберн. – 1987. – № 2. – С. 49–59.
12. Рубашкин В.Ш. Семантический компонент в системах понимания текста: КИИ-2006: тр. 10 Национ. конф. по искусствен. интеллекту с междунар. участ. 2006. [Электронный ресурс] URL: <http://www.raai.org/resurs/papers/kii-2006/#dokladi> (дата обращения: 22.10.2020).
13. Падучева Е.В. Семантика вида и точка отсчета: Изв. АН СССР: Сер. лит. и яз. – 1986. – Т. 45. – № 5. – С. 18–25.
14. Падучева Е.В. Отпредикатные имена в лексикографическом аспекте: Научно-техническая информация. – 1991. Сер. 2. – № 5. – С. 21–31.
15. Хабаров С.П. Представление знаний в информационных системах: конспекты лекций. [Электронный ресурс] URL: <http://www.habarov.spb.ru/bz/bz07.htm> (дата обращения: 23.10.2020).
16. Луценко Е.В. Представление знаний в информационных системах: электрон. учеб. пособие для студентов. – Краснодар: Изд-во КубГАУ, 2010. – 428 с.
17. Константинова И.С., Митрофанова О.А. Онтологии как системы хранения знаний: Информационно-телекоммуникационные системы: Всерос. конкурс. отбор статей, 2008. – 54 с.

#### REFERENCES

1. A.V. Shmida “*Novye metody raboty s bol'shimi dannymi: pobednye strategii upravlenija v biznes-analitike*” [New methods of working with big data: a winning strategy for the management of the business intelligence], Nauchno-prakticheskiĭ sbornik. М.: PAL"МIRA 2016. – 528P.
2. Bazenkov N.I., Gubanov D.A. *Obzor informacionnyh sistem analiza social'nyh setej: Upravlenie bol'shimi sistemami*, [Overview of information systems for social network analysis: Managing large systems] 2013. – 357-394 PP.
3. Bodjanskij, E. V. *Iskusstvennye nejronnye seti: arhitektury, obuchenie, primenenija: monografija E. V. Bodjanskij, O.G. Rudenko*. [Artificial neural networks: architectures, training, applications: monograph by E. V. Bodyansky, O. G. Rudenko] – Har'kov: Teleteh, 2004. – 369 PP.
4. Vesnin, V. R. *Upravlenie chelovecheskimi resursami. Teorija i praktika: uchebnik* [Human resource management. Theory and practice: textbook]/ V. R. Vesnin. – М.: Prospekt, 2014. – 688 P.
5. Vjazigin, A. V. *Ocenka personala vysshego i srednego zvena*. [Evaluation of senior and middle management personnel. A.V. Vyazigin] A. V. Vjazigin. – М.: Vershina, 2006. – 256 P.
6. Dorofeev, V. D. *Jeffektivnost' upravlenija trudovymi resursami predprijatija pri vnedrenii sistemy menedzhmenta kachestva* [Efficiency of the company's human resources management in the imple-

- mentation of the quality management system] monografiya V. D. Dorofeev, A. N. Shmeleva. – Penza: Informacionno-izdatel'skij centr PGU, 2008. – 210 P.
7. Drogobyckij, I. N. *Sistemnyj analiz v jekonomike: uchebnik dlja studentov vuzov, obuchajushhihsja po special'nostjam «Matematicheskie metody v jekonomike»*, [System analysis in economics: a textbook for university students studying in the specialties "Mathematical methods in Economics"] «Prikladnaja informatika» I. N. Drogobyckij. - 2-e izd., pererab. i dop. - M.: JUNITI-DANA, 2012. - 423 P.
  8. Maslova V. M. *Upravlenie personalom: tolkovyj slovar'*. [Personnel Management: explanatory dictionary] V. M. Maslova. – M.: Izdatel'sko-torgovaja korporacija «Dashkov i K», 2013. – 120 P.
  9. Zajceva T. V., Zub A. T. *Upravlenie personalom*. [Human Resources Management] – M.: ID «FORUM»: INFRA-M, 2013– 336 P.
  10. Gubanov D.A., Novikov D.A., Chhartishvili A.G. *Social'nye seti: modeli informacionnogo vlijaniya, upravlenija i protivoborstva*. [Social networks: models of information influence, management and confrontation] M.: Izd- vo fizikomatematicheskoi literatury, 2010. – 228 c.
  11. Lahuti D.G., Rubashkin V.Sh. *Sredstva i procedura konceptual'noj interpretacii vhodnyh soobshhenij na estestvennom jazyke*. [Tools and procedure for conceptual interpretation of input messages in natural language] Izv AN SSSR: Ser. Tehnich. kibern. 1987. № 2. 49–59 P.
  12. Rubashkin V.Sh. *Semanticheskij komponent v sistemah ponimaniya teksta*: [Semantic component in the system of understanding the text] KII-2006: tr. 10 Nacion. konf. po iskusstven. intellektu s mezhdunar. uchast. 2006. [Electronic resource] URL: <http://www.raai.org/resurs/papers/kii-2006/#dokladi> (accessed: 12.02.2021).
  13. Paducheva E.V. *Semantika vida i tochka otscheta*. [View semantics and reference point] Izv. AN SSSR: Ser. lit. i jaz. 1986. T. 45. № 5. 18–25 P.
  14. Paducheva E.V. *Otpredikatnye imena v leksikograficheskom aspekte*: [Predicate names in the lexicographic aspect] Nauchno-tehnicheskaja informacija. 1991. Ser. 2. № 5. 21–31 P.
  15. Habarov S.P. *Predstavlenie znaniy v informacionnyh sistemah: konspekty lekcij*. [Knowledge representation in information systems: lecture notes] [Electronic resource] URL: <http://www.habarov.spb.ru/bz/bz07.htm> (accessed: 12.02.2021).
  16. Lucenko E.V. *Predstavlenie znaniy v informacionnyh sistemah: jelektron. ucheb. posobie dlja studentov*. [Knowledge representation in information systems: electron. studies. manual for students] Krasnodar: Izd-vo KubGAU, 2010. – 428 P.
  17. Konstantinova I.S., Mitrofanova O.A. *Ontologii kak sistemy hranenija znaniy: Informacionno-telekommunikacionnye sistemy*: [Ontologies as Knowledge Storage Systems: Information and Telecommunications Systems] Vseros. konkurs. otbor statej, 2008. – 54 P.

Нәдіров Н.Қ.<sup>1</sup>, Дуйсебекова К.С.<sup>1</sup>

**Семантикалық талдауды қолдану арқылы Bigdata негізінде клиент профилін құру жүйесін қалыптастыру**

**Аңдатпа.** Бизнес-процестер клиенттерге көбірек бағытталған сайын, тұтынушылармен маркетингтік және басқарушылық қатынастар өте маңызды, бұл жаһандану мен нарықтағы өсіп келе жатқан бәсекелестікті ескере отырып, ерекше назар аударуға тұрарлық. Осы кезеңде "Клиенттер деген кім?" және "Оларды қалай анықтауға болады?". Қазіргі уақытта "Әлеуетті клиенттер туралы қандай ақпарат жинау керек?" және "Бүгінгі күні қызмет көрсету деңгейлері қандай?" деген мәселелер туындайды.

Әзірлеушінің мақсаты–клиенттің талғамына қарап, оның не қалайтынын білу. Кейінірек табылған мәліметтер мен ақпараттар талданған кезде алынған білімнің баға жетпес құндылығы байқалады, барлық мақсаттар мен мәселелерді түсіну болашақ жоба жетістіктерінің ажырамас бөлігі болып табылады. Аталған мақалада деректерді алдын ала өңдеу әдісі, пайдалану тұжырымдамасы, деректердің үлкен көлемін талдау әдістері және тұтынушы профилін қалыптастыру үшін олардағы пайдалы ақпаратты алу ұсынылған.

**Түйінді сөздер:** Үлкен деректер, бағдарламалық жасақтама, ақпараттық жүйе, құрылымдалған және құрылымдалмаған деректер, адамның әлеуметтік профилі.

**Nadirov N.K.<sup>1</sup>, Duisenbekova K. S.<sup>1</sup>**

**Development of a client profile formation system based on Big data using semantic analysis**

**Abstract.** As business processes become more and more customer-centric, marketing and management relationships with customers are becoming extremely important, which deserves special attention given the globalization and growing competition in the market. At this stage, many questions arise. Who are the clients? And how to identify them? Currently what information needs to be collected about potential customers? What are the service levels now? The developer's goal is to read the client's mind and understand what he wants. This knowledge will become invaluable later when the data and information found are analyzed. Understanding of all goals and problems is an integral part of the success of a future project. This paper presents a data preprocessing method for data mining based on the concept of using techniques to analyze huge amounts of data and extract the useful information it contains to form a customer profile.

**Key words:** Big data, software, information system, structured and unstructured data, social profile of a person.

**Авторлар туралы мәлімет:**

**Нәдіров Нұртас Қонысұлы**, Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының магистранты.

**Дүйсебекова Кулянда Сейтбековна** – PhD, Халықаралық ақпараттық технологиялар университетінің «Ақпараттық жүйелер» кафедрасының қауымдастырылған профессоры.

**About authors:**

**Nurtas K. Nadirov**, Master's student, «Information Systems» Department, International Information Technology University.

**Duisenbekova K. Saitbekova**, PhD, Assoc. Professor, Department of «Information Systems», International Information Technology University.

**Сведения об авторах**

**Нәдіров Нұртас Қонысұлы**, магистрант кафедры «Информационные системы» Международного университета информационных технологий.

**Дүйсебекова Кулянда Сейтбековна**, PhD, ассоц.профессор кафедры «Информационных систем» Международного университета информационных технологий.