

УДК 378.4

DOI: 10.15587/2313-8416.2019.

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ БОЛЬШИХ ДАННЫХ ДЛЯ МОНИТОРИНГА СОЦИАЛЬНЫХ ПРОЦЕССОВ В УНИВЕРСИТЕТЕ

Рагимова Назиля Али, Абдуллаев Вугар Гаджимахмуд, Халилов Матлаб Этибар

В данной статье рассматривается применение больших данных для проведения мониторинга социальных процессов. Для этого рассмотрены взаимосвязь больших данных и социологии. В качестве примера проведем мониторинг в университете. Для этого были рассмотрены некоторые основные методы и инструменты больших данных, на основе которых составили алгоритм для определения социальных процессов и проведения их мониторинга

Ключевые слова: большие данные, социология, социальные процессы, мониторинг социальных процессов, университет

Copyright © 2019, Рагимова Назиля Али, Абдуллаев Вугар Гаджимахмуд, Халилов Матлаб Этибар.
This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0>).

1. Введение

Появление первых ЭВМ является предпосылкой информационной эпохи. Началом самой информационной эпохи можно считать появление микропроцессоров и персональных компьютеров. Начиная с этого момента, главной продукцией в экономике становится информация и знания. Те, кто обладают ими, доминируют в экономике. Согласно отчету аналитической компании IDC «Digital Universe Study», объем электронных данных в 2006 году составил 0,18 зеттабайт, а в 2011 году объем составил 1,8 зеттабайт, до 2020 года весь объем данных нашей планеты достигнет 40 зеттабайт, где по статистике на каждого жителя Земли приходится около 5200 гигабайт данных [1].

Стремительно растущий объем данных провоцирует на поиск новых средств для их хранения и обработки. В связи с этим был введен термин «большие данные».

Большие данные – это структурированные и неструктурированные данные, неспособные обрабатываться традиционными методами обработки данных. К традиционным методам обработки данных можно отнести метод сравнения, метод применения абсолютных и относительных величин, метод применения средних величин, метод группировки, балансовый метод, графический метод, графоматематический метод. Эти методы могут реализовываться с помощью SQL в сочетании с одним из объ-

ектно-ориентированных языков программирования. Проблема больших данных заключается в том, что они представлены в формате, отличающимся от структурированного формата БД, ими могут быть веб-журналы, видеозапись, GPS данные, машинный код и т. д. [2].

Основными характеристиками больших данных считаются объем (volume), скорость накопления (velocity) и разнообразие (variety). К ним также принято относить достоверность (veracity), ценность (value), жизнеспособность (viability), изменчивость (variability), визуализация (visualization).

Из вышеперечисленных характеристик можно сформулировать следующие основные принципы для работы с большими данными:

– горизонтальная масштабируемость – это главный принцип анализа больших данных. Это означает, что при увеличении объема данных, необходимо увеличить количество вычислительных узлов, не теряя производительность.

– отказоустойчивость. Не исключено, что количество вычислительных узлов будет расти, а значит, растет вероятность выхода их из строя. Поэтому инструменты больших данных должны быть готовы к таким ситуациям и способны принять соответствующие меры.

– локальность данных. Необходимо хранить и обрабатывать данные в одном и том же физическом

сервере, в противном случае расходы на передачу данным между серверами могут быть колоссальными.

Применение больших данных нашли свое отражение во многих аспектах нашей жизни, включая в экономике [3], в менеджменте и бизнесе [4], антропологии [5], истории [6], а также в социологии.

По утверждению McKinsey&Company, в экономике существует пять основных подходов использования больших данных:

1. Организация «прозрачной» экономики;
2. Принятие математически обоснованных управленческих решений;
3. Узкое сегментирование клиентов с учетом персональных положений;
4. Увеличение скорости в принятии решений благодаря сложной аналитики;
5. Развитие товаров и услуг следующего поколения. [7]

По данным IDC, 90 % данных, хранимых на серверах компаний, являются практически полезными, но неиспользуемыми. Полезную информацию для бизнеса в компании в основном получают из CRM и телефонии (АТС). Системы CRM содержат информацию о продажах по территориям, временам года, сумме и количестве заказов. А АТС содержит данные о продолжительности ожидания на линии, продолжительности разговоров, алгоритмов распознавания входящих и исходящих звонков, номерах телефонов. [8].

Большие данные в менеджменте способны автоматизировать и обобщить функциональные подходы к поиску и отбору сотрудников, улучшению качества работы персонала и повышению производительности труда, решению задач в области образования команд путем соотношения качеств людей. Главной задачей больших данных в управлении персоналом является управление талантами и улучшении методов подбора персонала. [9]

2. Литературный обзор

Внедрение больших данных в социологию порождает два вида социологии: вычислительная социология (вычислительные социальные науки), направленная на сбор и анализ данных; социальную информатику (электронные социальные науки, цифровые социальные исследования), предназначенная для накопления и анализ информации. Оба этих видов были направлены на исследования, анализ и мониторинг социальных феноменов [10].

По К. Чэффи-Ревилья «вычислительные социальные науки» — это интегрированный междисциплинарный поиск в социальном исследовании посредством вычислений при увеличивающемся масштабе информационных процессов [10].

Д. Ватс считает термин «вычислительные социальные науки» ярлыком, который описывает симуляции агент-основанных моделей. Эти науки обозначают возникновение взаимное пересечения компьютерных и социальных наук, взаимодействия, которые включают в себя анализ web-масштабных данных наблюдения, виртуальных экспериментов в лабораторном стиле и вычислительное моделирование [10].

В корпорации «Microsoft» также создано подразделение вычислительных социальных наук. При определении «вычислительных социальных наук» добавляется статистика, считая, что это – междисциплинарная область, привлекающая экспертизу, масштабные статистические и машинообучаемые методы, охватывающая несколько самостоятельных социальных наук, включая социологию, экономику, психологию, политические науки, маркетинг, когда преобладают масштабные демографические, поведенческие и сетевые данные для исследования человеческой активности и взаимоотношений.

Следуя из вышеперечисленных определений этого термина, можно сделать следующий вывод:

Вычислительные социальные науки – это междисциплинарные взаимное пересечения компьютерных и социальных наук, направленные на междисциплинарный поиск в социальных исследованиях, посредством экспертизы, машинообучаемых методов, экономики, социологии, психологии, маркетинга, политических наук.

К. Чэффи-Ревилья предложив 5 методов исследований для «вычислительных социальных наук», надо отметить, что эти методы не являются эксклюзивными для «вычислительных социальных наук»:

1. Автоматический сбор информации. В «вычислительных социальных наук» применяется анализ текста и контент-анализ по мониторингу информации о событиях и изучении политических риторик;
2. Анализ социальных сетей. Это используется для обеспечения безопасности социальных сетей, изучении террористических сетей;
3. Геопространственный анализ. Применяя географические информационные системы, исследователи изучают пространственные слои распространения идей;
4. Моделирование сложности. Применяет математические средства для понимания взаимодействия между элементами в системе, а также определений интенсивных конфликтов;
5. Агент-основанное моделирование. Используя этот вид моделирования, исследователи изучают изменение сред и возникновений новых организаций [10].

Обратной стороной монеты «вычислительных социальных наук» является социальная информатика. Социальная информатика – это наука, применяющая современные цифровые технологии, направленные на изучение социальных проблем и возможностей применения социальных исследований.

Одними из главных объектов изучения социологии являются процессы, протекающие в обществе, называемые социальными процессами. В основе социальных процессов лежат разногласия между социальными группами людей, одни группы которых получают пользу в отличие с другими социальными группами людей. Такой расклад вещей является естественной эволюцией общества. Формулирующим критерием чего является временная составляющая. Эта составляющая придает объективированный характер, который дает возможность проследить все свойства объекта в зависимости от времени. Временная составляющая особенно интересна в изучении социально-экономических и политических процессов.

Социальный процесс – это изменение социального характера в обществе, которое вызвано желанием определенных групп воздействовать на текущие ситуации в социуме для удовлетворения своих интересов. Источниками этих процессов являются люди.

Неоднородность в позициях субъектов социального общества определяет вектор социальных процессов, которые стремятся достигнуть равновесия друг с другом. В результате такого взаимодействия интересов сообществ возникают действия неизвестных сил, вызванные непосредственно этим взаимодействием. Результатом этих действий устанавливает направление социальных процессов [7]. Подчинение социальных процессов субъектами общества своему вектору поведения и варианту вероятных действий является приоритетной их задачей.

Невзирая на большое разнообразие социальных процессов, социологи Чикагской школы социологии Роберт Парк и Эрнст Берджер смогли классифицировать их на шесть групп:

1. Кооперация – это процесс объединения группы лиц, ради общего интереса с целью получить выгоду. Для этого надо взаимопонимание и установление правил сотрудничества внутри группы. Социологи, проведя наблюдение, установили, что в основе кооперации лежат корыстные цели;

2. Конкуренция – это борьба субъектов за получения различных ресурсов (деньги, власть, любовь и др.). По своей сути конкуренция является борьба за вознаграждения, для чего необходимо опередить соперника с теми же целями;

3. Адаптация (приспособление) – это процесс, когда происходит принятие норм и ценностей новой среды индивидом при неудовлетворении норм и ценностей текущей среды потребностей индивида. В этом процессе выделяют подчинение, компромисс и терпимость;

4. Конфликт – это процесс, требующий полного подчинения или открытого противодействия изменения, происходящим в социуме;

5. Ассимиляция – это процесс, где часть общества утрачивает определенную степень своих культурных черт и замещает их заимствованиями из другой части общества;

6. Амальгамация – это процесс, который возникает при смешивании групп субъектов общества. Отличием от ассимиляции заключается в том, что после завершения процесса амальгамации границы между группами стираются [11].

К этим процессам также добавляют еще два процесса: поддержание границ и систематические связи. Границы между социальными группами являются одним из основных аспектов социальной жизни. Для их поддержания, установлении и модификации выделяется много времени и энергии. Пограничные линии социальных групп отделяют их членов от всего остального общества. [12] Процессы ассимиляции и амальгамации сопровождаются стиранием границ между социальными группами, уничтожением имеющегося разделения для создания общих черт группы.

Систематические связи – это процессы установления связи между социальными группами, кото-

рые находятся в установленных социальных границах. При отсутствии у какой-либо группы связи с другими группами приведет к ее изоляции.

Мониторинг социальных процессов – совокупность методов сбора, обработки и анализа социальных процессов для обнаружения закономерностей развития общества и каждого индивида в отдельности. Объектами социального мониторинга являются вся совокупность социальных явлений и процессов. [13].

На данный момент истории мониторинг социальных процессов проводится путем анализа социальных сетей и глобальной сети Internet, нежели прямым наблюдением объекта.

Социальные процессы способны влиять на экономическое развитие и политическую ситуацию региона или даже страны. Имея результаты мониторинга, эксперты способны воздействовать на эти процессы.

Мониторинг социальных процессов – это сложный процесс, определяющий не только конкретно свои цели, но и реализуется на основе определенных принципов:

- полнота социальной информации;
- оперативность сбора данных;
- сопоставимость получаемых данных;
- сочетание обобщенных и дифференциальных

оценок и выводов в процессе социального мониторинга.

Средства больших данных могут быть полезны для выполнения функций и задач мониторинга социальных процессов. Применяя эти средства, появляется возможность сбор информации об объекте из социальных сетей, неструктурированных и слабоструктурированных баз данных. Потом применение различных инструментов больших данных позволяет определить сами процессы в обществе. Далее провести сам мониторинг социальных процессов. Следом появляется возможность определить к положительным или отрицательным изменениям приведут социальные процессы и применить соответствующие меры.

Все это позволяет наблюдать за объектами и явлениями, выявлять тенденции развития, прогнозирования возможных последствий, производить поиск необходимых мер по предотвращению негативных тенденций и поддержания позитивных, ведущих социум к дальнейшему развитию.

3. Цель и задачи исследования

Цель исследования – определить уровень готовности студентов к социальной жизни с помощью анализа данных о преподавателях.

Для решения этой задачи необходимо:

1. Определить данные, необходимые для определения социальных процессов;
2. Определить методы и инструменты для проведения мониторинга социальных процессов в университете;
3. Составить алгоритм для проведения мониторинга социальных процессов в университете;
4. Построить прогнозы и принять соответствующие меры.

4. Данных для проведения мониторинга социальных процессов в университете

В университетах главными источниками социальных процессов являются студенты и преподаватели. Взаимоотношения студентов с преподавателями, студентов и преподавателей между собой являются объектами мониторинга социальных процессов в вузах.

Каждый семестр в университете накапливается огромное количество данных о студентах и преподавателях. Проанализировав эти данные, можно определить протекающие социальные процессы в вузе. Такими данными о студентах являются результаты промежуточных экзаменов, лабораторных работ, самостоятельных работ, данные о посещаемости, балл, набранный в течение семестра, результаты экзаменов, общий балл по предмету. Кроме данных о студентах, в университетах накапливаются данные о преподавателях: стаж, количество статей, выпущенных в зарубежных журналах, квалификация преподавателей, количество руководимыми студентами, учебная нагрузка, соотношение числа студентов к одному преподавателю, количество преподающим студентам, количество преподающим группам, проекты. Просмотрев эти данные, можно выявить социальные процессы, где главным источником является студент и преподаватель.

Примером процесса кооперации является объединение студентов в группы; объединение преподавателей в кафедры по какой-либо области в науке для ее развития и преподаванию студентам учебных программ; написание дипломных и диссертационных работ, где преподаватель и студент объединяют свои усилия; совместная работа нескольких преподавателей над одним проектом или же написании статьи.

Конкуренция проявляется, когда студенты одной специальности ведут борьбу за стипендию. Здесь единственным параметром является средний балл за семестр по всем предметам. Примером конкуренции среди преподавателей может быть конкурс на определенную должность (заведующий кафедры, декан).

Наиболее распространённым социальным процессом в вузе является конфликт между преподавателем и студентом. Для определения конфликтного преподавателя необходимы следующие данные: общее количество студентов, не сдавших экзамен, процент этих студентов от общего числа студентов, количество студентов, не допущенных к экзамену из-за посещаемости, результаты сессий по другим экзаменам у студентов. А для определения конфликтного студента необходимы – средний балл за семестр, количество академических задолженностей, результаты посещаемости. Обычно конфликт между преподавателем и студентом возникает из-за балла, где студенты ведут конкуренцию за стипендию.

Процесс адаптации наблюдается у первокурсников, где они приспособляются к новой среде. Мониторинг этого процесса осуществляется на основе данных о посещаемости и статистики успеваемости, если она улучшается, то процесс адаптации успешен. Конечным результатом этого процесса является окончание вуза, в противном случае студент исключается из вуза. Также к адаптации относится переезд студентом из региона в большие города для получение высшего образования. В университете адаптация

проявляется во время общения преподавателя со студентами, где студент адаптируется к условиям преподавателя, а преподаватель ищет нужный подход обучения. Также к адаптации относится учебная нагрузка преподавателя.

Примером процесса ассимиляции для преподавателей является повышение квалификации. У студентов ассимиляция проявляется при взятии на работу.

Неотъемлемой частью университета является объединения первокурсников в группы, где происходит процесс амальгамации, т. е. идет процесс обмена культурных качеств между студентами. Студенты объединяются в группы по выбранным специальностям. То же самое происходит, когда преподаватели объединяются в кафедры.

Социальные границы в университетах выделяются две группы людей: студенты и преподаватели. В свою очередь студенты делятся на группы, а преподаватели на кафедры.

Социальной связью здесь выступает урок, где устанавливается контакт между студентами и преподавателями. Еще сюда можно отнести проведения различных конференций, где устанавливается связь между преподавателями, учеными и студентами из различных университетов.

5. Результаты исследования

Для обработки больших данных имеются множество разнообразных методов и инструментов. Международная консалтинговая компания McKinsey&Company выделила несколько основных методов анализа больших данных.

Методы Data Mining (интеллектуальный анализ данных, ИАД) – набор методов для обнаружения в данных ранее неизвестных, нетривиальных, практически полезных знаний, которые необходимы для принятия решения. К методам ИАД относятся:

– Обучение ассоциативных правил. Совокупность методов для выявления ассоциативных правил в больших массивах данных;

– Классификация – это назначение объектов, наблюдений или событий к одному из ранее объявленных классов;

– Кластерный анализ – статистический метод отнесение объектов к группам благодаря выявлению общего признака;

– Регрессия – совокупность статистических методов для изучения воздействия одной или нескольких независимых переменных на зависимую и др.

Краудсорсинг – это метод позволяет одновременно различным пользователям производить сбор данных из различных источников, количество которых не ограничено.

Смешение и интеграция данных – совокупность методов для интегрирования неоднородных данных из различных источников для проведения над ними интеллектуального анализа. В качестве примера можно привести примеры методов цифровой обработки сигналов, обработки естественного языка, включая тональный анализ.

Машинное обучение – совокупность методов, преследующее цель создание алгоритмов самообуче-

ние на базе эмпирических данных. Отличительной чертой машинного обучения является не прямое решение поставленной задачи, а обучение во время применения решений сходных задач.

Искусственные нейронные сети – математическая модель, которая построена на основе биологических нейронных сетей.

Распознавание образов – совокупность методов классификации и идентификации образов, характеризующиеся определенным набором свойств и признаков.

Прогнозная аналитика – совокупность методов анализа данных, которые концентрируются на прогнозировании возможного поведения объектов для принятия оптимальных решений.

Имитационное моделирование – вид математического моделирования для построения модели, которые описывают реальную систему виртуально. Это используется в основном для прогнозирования и проведения экспериментов.

Пространственный анализ – совокупность методов анализа пространственных данных.

Статистический анализ – совокупность методов сбора, организации и интерпретации больших данных.

Визуализация данных – совокупность методов представления данных в виде графиков, диаграмм, анимированных изображений и анимации.

К инструментам больших данных можно отнести продукты 1010data (1010edge, 1010connect, 1010reveal, 1010equities), продукты Apache Software Foundation (Apache Hive, Apache Chukwa, Apache Hadoop, Apache Pig!), фреймворк MapReduce, язык R, библиотека Pandas, NoSQL и другие. [1]

Компания 1010data предлагает услуги анализа данных для получения более интеллектуального решения и принятия более оптимального решения для бизнеса. Для этого следует размещать свои данные на серверах компании. Продуктами этой компании является:

– 1010edge – интуитивная платформа корпоративного анализа данных, поддерживающие все потребности в данных и аналитике. Основными его возможностями являются предоставления сбора и управления данных, анализа и моделирования, создания отчетов и визуализации, разработки приложений, совместного использования и монетизации данных;

– 1010connect – детализированный портал, позволяющий совместно использовать и контролировать распределение данных за пределами предприятия, что позволяет создать основу для беспрецедентного сотрудничества с деловыми партнерами или превратить данные в дифференцированный высоко-рентабельный генератор регулярных доходов. Основные его возможности являются предоставление собственного облака, гибкого управления и управления разрешениями, многоуровневого доступа и контролируемого распространения, безопасности и высокой доступности;

– 1010reveal – набор интеллектуальных решений для потребителей высокой четкости;

– 1010equities – этот продукт предоставляет альтернативные решения для передачи данных на стороне покупателя.

MapReduce – это фреймворк распределенных вычислений, разработанная Google, применяемая для параллельных вычислений над наборами данных с размером до нескольких петабайт. Он состоит из трех этапов:

1. Этап Map. Здесь данные обрабатываются методами map (), определяемый пользователем. На данном этапе происходит фильтрация и предобработка данных. Метод map () выдает множество пар ключ-значение.

2. Этап Shuffle. Протекает незаметно для пользователей. Здесь вывод функции map () распределяется по одному ключу, полученному на выводе этапа map.

3. Этап Reduce. Вывод второго этапа является входов для метода reduce (). Этот метод, задаваемый пользователем, вычисляет конечный результат для отдельного ключа. Множество значений, возвращенные этим этапом, является выводом MapReduce.

Самым распространённым инструментом для работы с большими данными является Apache Hadoop, реализован на базе MapReduce. Apache Hadoop – это набор библиотек и утилит для разработки и выполнения распределенных программ, которые работают на кластерах, способных состоят из нескольких тысяч узлов. Hadoop состоит из 4 модулей:

– Hadoop Common – это набор библиотек управления файловыми системами Hadoop и сценарии создания нужной инфраструктуры и управления распределенной обработки;

– Hadoop Distributed File System – файловая система для хранения файлов больших объемов;

– Yet Another Resource Negotiator – модуль управления ресурсами кластеров и планирование заданий;

– Hadoop MapReduce – программный каркас для кодирования распределенных вычислений на основе MapReduce.

Apache выстроила целую экосистему вокруг Hadoop, которая содержит целый набор мощных инструментов, облегчающую работу с большими данными:

– Hive – инструмент для создания HiveSQL запросов над большими данными. Способен превращать обычные SQL запросы в серию MapReduce-задач;

– Pig – язык программирования запросов к большим слабоструктурированным данным, одна строчка которого способна превращаться в последовательность MapReduce-задач;

– Hbase – колоночная БД, которая реализует на основе NoSQL;

– Mahout – библиотека машинного обучения;

– Apache Spark – это движок распределенной обработки данных, использующий компоненты Hadoop (HDFS и YARN);

– Apache Tez – фреймворк, который работает поверх Hadoop YARN для обработки групповых данных, нуждающиеся интеграции с Hadoop YARN;

– Apache Solr – инструмент фасетного и полнотекстового поиска, интеграции с БД, динамической кластеризации и обработка документов со сложным форматом;

– Apache Sqoop и Flume – инструменты управления потоками данных;

– Zookeeper и Oozie – инструменты для координирования и планирования задач;

– Apache Storm – инструмент, предоставляющий безопасность потоковой обработки данных;

– Apache Kafka – инструмент для быстрой обработки программных сообщений между программами.

Установка Hadoop раньше представляла собой довольно трудной задачей: необходимо было настраивать каждую машину в кластере. В связи с увеличением популярности экосистемы Hadoop возникли компании, предоставляющие сборки экосистемы Hadoop, а также мощные инструменты для управления Hadoop-кластером. Forrester Research выделяют ряд дистрибутива Hadoop: Cloudera, Hortonworks и MapR.

Cloudera – это первая компания и ведущий поставщик Hadoop и владеет собственным дистрибутивом Hadoop, называемым Cloudera CDH. Она предоставляет ПО для получения доступа, хранения, анализа, защиты, управления и поиска информации. Cloudera имеет собственный модуль управления кластером Cloudera Manager и довольно высокой ценой технического сопровождения (\approx \$4 000 в год за один узел кластера), позволить себе такое могут только большие корпорации.

Дистрибутивом Hortonworks является Hortonworks Data Platform, включающий сервисы DataPlane для интеграции с решениями сторонних производителей ПО и собственный модуль управления кластера Hortonworks Management Center на базе Apache Ambari, входящая в состав каждого дистрибутива абсолютно бесплатно.

Дистрибутив MapR предпочитает свою распределенную файловую систему MapR-FS, собственную БД MapR-DB и использует уникальный распределенный брокер программных сообщений MapR Event Store вместо Apache Kafka. MapR обеспечивает баланс между стабильностью и быстродействием, сохраняя при этом максимально простое использование.

Также есть российский дистрибутив ArenaData, полностью локализован на русский язык и основан на открытых проектах Apache Software Foundation.

Библиотека Pandas делает Python мощным средством для анализа и визуализации данных. В качестве источника данных могут служить SQL, HTML, Excel файлы, текстовые файлы. Основными структурами хранения данных в Pandas являются:

– Series – это проиндексированный одномерный массив значений;

– DataFrame – это проиндексированный многомерный массив значений, столбец которого является структурой Series.

R – язык программирования для статистической обработки данных и работы с графикой. Отличительной особенностью данного языка является его большой спектр статистических и численных методов и расширяемостью с помощью пакетов, предоставляющие собой библиотеки для работы специальных функций или областей применения.

Четкого определения у NoSQL нету. В общем NoSQL – термин для обозначения подходов, реализующих СУБД, отличающиеся от реляционных СУБД. В отличие от традиционных СУБД, NoSQL имеет следующие свойства: базовая доступность, гибкое состояние, согласованность в конечном счете. Для работы с большими данными NoSQL использует модель «Семейство столбцов». Системы, использующие эту модель, хранят данные как разреженные матрицы, где строки и столбцы используют как ключи. В основном эта модель используется в Apache HBase, Apache Cassandra, ScyllaDB, Apache Accumulo, Hypertable.

6. Выводы

Методы и инструменты больших данных могут использоваться для проведения мониторинга социальных процессов в университете. Учитывая все это, алгоритм мониторинга социальных процессов в университете будет выглядеть следующим образом:

1. Необходимо определить источники информации. В данном случае в качестве источников информации может служить электронная система образования, группы студентов и страницы, имеющие отношение к университету;

2. Следует составить единую базу данных, куда входили бы вся информация, полученная из источников. Для составления базы можно использовать Hive, Hbase или же что-то из NoSQL;

3. Далее следует определить данные, которые необходимы для определения социального процесса. Для определения социальных процессов в университете данные описаны выше. Следом необходимо провести анализ данных для получения оценок социальных процессов, перечисленных выше. Для определения оценок социальных процессов можно использовать Hadoop и его дистрибуторы, Spark, язык R и другие средства обработки больших данных;

4. После получения оценок мониторинга можно строить прогнозы тех или иных процессов. Получив результаты прогнозов, следует принять соответствующие действия. Для прогнозирования можно использовать имитационное моделирование для построения виртуальной модели протекающих процессов.

Литература

1. Gantz, J., Reinsel, D. (2012). The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the Far East. IDC. Available at: <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>
2. Большие данные (Big Data) (2017). Available at: [http://www.tadviser.ru/index.php/Статья:%20Большие_данные_\(Big_Data\)](http://www.tadviser.ru/index.php/Статья:%20Большие_данные_(Big_Data))

3. Einav, L., Levin, J. (2014). The Data Revolution and Economic Analysis. *Innovation Policy and the Economy*, 14, 1–24. doi: <http://doi.org/10.1086/674019>
4. Frizzo-Barker, J., Chow-White, P. A., Mozafari, M., Ha, D. (2016). An empirical study of the rise of big data in business scholarship. *International Journal of Information Management*, 36 (3), 403–413. doi: <http://doi.org/10.1016/j.ijinfomgt.2016.01.006>
5. Сивков, Д. (2017). Большие данные в этнографии: вызовы и возможности. *Социология науки и технологий*, 8 (1), 56–67.
6. Bearman, P. (2015). Big Data and historical social science. *Big Data & Society*, 2 (2). doi: <http://doi.org/10.1177/2053951715612497>
7. Технология Big Data в экономике (2016). Available at: <http://ru.datasides.com/big-data-in-economics/>
8. Свилас, В. (2018). Big Data поможет увеличить прибыль вашей компании. Как это работает? Rusbase. Available at: <https://rb.ru/opinion/big-data-pomozhet/>
9. Кесаев, У. С., Алехно, В. В. (2017). Перспективы применения Big Data в управлении персоналом. *Nauka-Rastudent.ru*. Available at: <http://nauka-rastudent.ru/37/3942/>
10. Журавлева, Е. Ю. (2015). Социология в сетевой среде: к цифровым социальным исследованиям. *Социологические исследования*, 8, 25–33. Available at: http://socis.isras.ru/files/File/2015/2015_8/Zhuravleva.pdf
11. Понятие социального процесса. Available at: https://psyera.ru/ponyatie-socialnogo-processa_8350.htm
12. Виды социальных процессов. Available at: https://psyera.ru/vidy-socialnyh-processov_9845.htm
13. Мониторинг социальных процессов. Available at: https://spravochnik.ru/sociologiya/suschnostiprincipysocialnyh-processov/monitoring_socialnyh-processov/
14. Васильева, Л. Г. (2010). Социализация студентов и система управления образовательным процессом в филиале вуза (на примере Арсеньевского городского округа). *Молодой ученый*, 10, 301–303.
15. Big Data от А до Я. Часть 2: Hadoop (2015). Available at: <https://habr.com/ru/company/dca/blog/268277/>
16. Введение в анализ данных с помощью Pandas (2013). Available at: <https://habr.com/ru/post/196980/>
17. Big data. Методы и техники анализа больших данных. Available at: <https://www.it.ua/ru/knowledge-base/technology-innovation/big-data-bolshie-dannye>
18. Какой Hadoop лучше: сравнение 4 самых популярных дистрибутивов. Available at: <https://medium.com/@bigdataschool/какой-hadoop-лучше-сравнение-4-самых-популярных-дистрибутивов-48adbb74b84c>

Received date 12.08.2019

Accepted date 23.08.2019

Published date 30.08.2019

Рагимова Назилия Али, кандидат технических наук, доцент, кафедра “Компьютерная инженерия”, Азербайджанский государственный университет нефти и промышленности, пр. Азадлыг, 16/21, г. Баку, Азербайджан, AZ1010

Абдуллаев Вугар Гаджимахмуд, кандидат технических наук, доцент, кафедра “Компьютерная инженерия”, Азербайджанский государственный университет нефти и промышленности, пр. Азадлыг, 16/21, г. Баку, Азербайджан, AZ1010

Халилов Матлаб Этибар, кафедра “Компьютерная инженерия”, Азербайджанский государственный университет нефти и промышленности, пр. Азадлыг, 16/21, г. Баку, Азербайджан, AZ1010