

Министерство образования Республики Беларусь
Учреждение образования
«Белорусский государственный университет
информатики и радиоэлектроники»

Факультет компьютерного проектирования

Кафедра инженерной психологии и эргономики

Е. Б. Карпович, Д. А. Пархоменко

**ОСНОВЫ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ
ДЕЯТЕЛЬНОСТИ**

*Рекомендовано УМО по образованию в области информатики
и радиоэлектроники в качестве пособия для специальности 1-58 01 01
«Инженерно-психологическое обеспечение информационных технологий»,
направлений специальности 1-40 05 01-09 «Информационные системы
и технологии (в обеспечении промышленной безопасности)», 1-40 05 01-10
«Информационные системы и технологии (в бизнес-менеджменте)»*

Минск БГУИР 2017

УДК 001.895+002.6(076)
ББК 32.973.202я73+73я73
К26

Рецензенты:

кафедра интеллектуальных систем
Белорусского национального технического университета
(протокол №2 от 06.10.2015);

главный научный сотрудник государственного научного учреждения
«Объединенный институт проблем информатики Национальной академии наук
Беларуси», доктор технических наук, доцент Л. Д. Черемисинова

Карпович, Е. Б.

К26 Основы информационно-аналитической деятельности : пособие /
Е. Б. Карпович, Д. А. Пархоменко. – Минск : БГУИР, 2017. – 55 с. : ил.
ISBN 978-985-543-229-7.

Представлены вопросы, задания и тесты для проведения практических занятий.
Материал изложен в соответствии с рабочей программой дисциплины.

УДК 001.895+002.6(076)
ББК 32.973.202я73+73я73

ISBN 978-985-543-229-7

© Карпович Е. Б., Пархоменко Д. А., 2017
© УО «Белорусский государственный университет
информатики и радиоэлектроники», 2017

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
Классификация информации.....	5
Способы поиска и структурирования данных.....	7
Методы анализа и интерпретации данных.....	12
Выборочный метод сбора данных.....	16
Модели представления знаний	19
Современное представление об информационных технологиях, их развитии.....	28
Понятие цифровой трансформации.....	37
Интернет вещей (Internet of Things – IOT).....	42
Большие данные (Big Data).....	45
ЛИТЕРАТУРА.....	54

Библиотека БГУИР

ВВЕДЕНИЕ

В настоящее время в мире наблюдается дефицит специалистов, обладающих знаниями об информационно-коммуникационных технологиях, методах и способах их применения в различных прикладных областях. Постоянно появляются новые инструментарии (такие как облачные технологии, большие данные, Интернет вещей, социальные медиа и др.), оказывающие влияние на цифровую трансформацию всех сфер современного общества.

Учебная дисциплина «Основы информационно-аналитической деятельности» направлена на восполнение существующего пробела в системе подготовки IT-специалистов. Данная дисциплина предполагает изучение теории и практики информационно-аналитической деятельности, что способствует усвоению знаний и формированию навыков в области информационно-коммуникационных технологий, приобретению умений по их практическому применению в прикладных сферах.

Практические занятия нацелены на стимулирование учебно-познавательной деятельности обучающихся, формирование творческого подхода к решению актуальных проблем в сфере будущей профессиональной деятельности, содействие развитию инновационного мышления, активизации самообразования.

Классификация информации

Информация – это некоторая последовательность (упорядоченность) сведений, знаний, которые актуализируемы (получаемы, передаваемы, преобразуемы, сжимаемы или регистрируемы) с помощью некоторых знаков (символьного, образного, жестового, звукового, сенсомоторного типа). Это приращение, развитие, актуализация знаний, возникающие в процессе целеполагающей интеллектуальной деятельности человека.

Информация может существовать в пассивной (неактуализированной) и активной (актуализированной) форме.

Пример. Информация актуализируется сообщениями, при этом формы облачения информации в сообщения различны, например, для живых существ – сигналы, жесты, для технических устройств – сигналы. Информация, передаваемая от одного человека другому, может передаваться символами (письмо), жестами (сигнальщик на боевом корабле), звуками (диктор), геометрическими фигурами (чертежник), художественными образами (балерина). Информация, передающаяся животными, может быть передана звуками (лай, вой, писк), ситуационным поведением (образами). Информация в технических устройствах, автоматах может быть передана электрическими, магнитными, световыми импульсами, как это происходит в ЭВМ.

Информация по отношению к окружающей среде (или к использующей ее среде) бывает трех типов: входная, выходная и внутренняя.

Входная информация (по отношению к окружающей среде) – информация, которую система воспринимает от окружающей среды.

Выходная информация (по отношению к окружающей среде) – информация, которую система выдает в окружающую среду.

Внутренняя, внутрисистемная информация (по отношению к системе) – информация, которая хранится, перерабатывается, используется только внутри системы, т. е. актуализируемая лишь только подсистемами системы. Это несколько идеализированное (особенно с точки зрения физики открытых систем) понятие.

Пример. Человек воспринимает, обрабатывает входную информацию, например данные о погоде на улице, формирует выходную реакцию – ту или иную форму одежды. При этом используется внутренняя информация, например, это генетически заложенная (или приобретенная) физиологическая информация о реакции, например о «морозостойкости» человека.

Пример. Генетически заложенная в молекулах ДНК информация и приобретенная информация (в памяти) влияют на поведение, на адаптацию человека в окружающей среде. В машинах первого поколения внутренняя структура определялась тысячами ламп, причем каждая из них отдельно была невысокой надежности, т. е. вся система была ненадежной в работе. Это влияло на входную информацию, например, такие ЭВМ не были способны на работу в

многозадачном режиме, в режиме реального времени (обработки сообщений по мере получения входных данных).

Информация по отношению к конечному результату проблемы бывает:

- исходная (на начало актуализации этой информации);
- промежуточная (от начала до завершения актуализации информации);
- результирующая (после завершения ее актуализации).

Пример. При решении системы линейных алгебраических уравнений информация о методах решения, среде реализации, входных данных (источники, точность и т. д.), размерности системы и т. д. является исходной информацией; информация о совместности системы уравнений, численных значениях корня и т. д. – результирующая; информация о текущих состояниях коэффициентов уравнений реализации схемы Гаусса – промежуточная.

Информация по изменчивости при ее актуализации бывает:

- постоянная (никогда не изменяемая при ее актуализации);
- переменная (изменяемая при актуализации);
- смешанная – условно-постоянная (или условно-переменная).

Пример. В известной физической задаче определения дальности полета снаряда артиллерийского орудия информация об угле наклона орудия может быть переменной, информация о начальной скорости вылета снаряда – постоянной, а информация о координатах цели – условно-постоянной.

По отношению к цели системы выделяют информацию синтаксическую, семантическую, прагматическую.

Синтаксическая информация состоит из совокупности знаков, объединяемых в тот или иной алфавит (т. е. полный ассортимент определенных знаков, например буквенных, цифровых). В сообщении эти знаки расставляются не произвольно, не в случайном порядке. Последовательность и место каждого знака обусловлены установленными правилами. Эти правила очень важны: их отсутствие привело бы к невозможности прочесть сообщение (чертеж, схему, текст).

Семантическая информация основной задачей ставит вопрос о нахождении оптимальных для фиксации значений видов знаковых выражений и оптимальных для формирования смысла построения сообщений. Таким образом, семантическая информация – это информация, предполагающая актуальную или потенциальную возможность раскрытия потребителем закодированного теми или иными знаковыми средствами ее содержания, смысла сообщения.

Прагматическая информация инициирует определенный образ мыслей, определенное поведение, то или иное состояние.

Возможна также классификация информации и по другим признакам:

- по стадии использования (первичная, вторичная);
- по полноте (избыточная, достаточная, недостаточная);
- по отношению к элементам системы (статическая, динамическая);
- по отношению к структуре системы (структурная, относительная);

- по отношению к управлению системой (управляющая, советующая, преобразующая, смешанная);
- по отношению к территории (региональная, местная, относящаяся к юридическому или к физическому лицу, смешанная);
- по доступу (открытая или общедоступная, закрытая или конфиденциальная, смешанная);
- по предметной области, по характеру использования (статистическая, коммерческая, нормативная, справочная, научная, учебная, методическая и т. д., смешанная) и др.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАНИЯ

1. Дайте определение понятию «информация».
2. Чем отличается информация от знания?
3. По каким критериям производится классификация информации?
4. Приведите собственные примеры входной, выходной и внутренней информации; исходной, промежуточной и результирующей; постоянной, переменной и смешанной; первичной и вторичной; синтаксической, семантической и прагматической.
5. По какому признаку выделяют научную, методическую, учебную информацию?

Способы поиска и структурирования данных

Формирование информации происходит в процессе сбора и передачи, т. е. обработки данных. В широком смысле данные представляют собой факты, текст, графики, картинки, звуки, аналоговые или цифровые видеосегменты.

Данные могут быть получены в результате измерений, экспериментов, арифметических и логических операций.

Данные должны быть представлены в форме, пригодной для хранения, передачи и обработки.

Данные – это необработанный материал, предоставляемый поставщиками данных и используемый потребителями для формирования информации на основе данных.

Понятия «информация» и «знания» с философской точки зрения являются понятиями более высокого уровня, чем понятие «данные», которое возникло относительно недавно. Существует также понятие «метаданные». В состав метаданных могут входить каталоги, справочники, реестры. *Метаданные* содержат сведения о составе *данных*, содержании, статусе, происхождении, местонахождении, качестве, форматах и формах представления, условиях доступа, приобретения и использования, авторских, имущественных и смежных с ними правах на данные и др.

Сама по себе информация в чистом виде бессмысленна. Информация, не представляющая интерес, является шумом. Отсюда следует вывод, что информация – это чье-то тактическое знание, передаваемое в виде символов и при помощи каких-либо прикладных средств.

Понятие информации непосредственно связано с сущностью процессов внутри информационной системы, тогда как понятие «знание» скорее ориентировано на качество процессов. Понятие «знание» тесно связано с процессом принятия решений. По мере продвижения вверх по информационной пирамиде объемы данных переходят в ценность решений. Несмотря на различия, рассмотренные понятия не являются разрозненными и несвязанными. Они есть часть одного потока: у истока его находятся данные, в процессе передачи которых возникает информация и в результате использования информации при определенных условиях возникают знания.

По определению Денхема Грэя, «знания – это абсолютное использование информации и данных совместно с потенциалом практического опыта людей, способностями, идеями, интуицией, убежденностью и мотивациями».

Знания отличают от информации следующие свойства: структурированность (наличие ясных названий и заголовков, удобного представления структуры); удобство доступа и усвоения; лаконичность, позволяющая быстро осваивать и перерабатывать знания и повышать «коэффициент полезного содержания»; непротиворечивость.

Одно из главных применений знаний – возможность передачи знаний другим и способность делать выводы на их основе. Для этого должны существовать процедуры обработки знаний. Способность делать выводы означает для машины наличие процедур обработки и вывода, а также подготовленность структур данных для такой обработки, т. е. наличие специальных форматов знаний.

Граница между информацией (данными) и знаниями нечеткая и зависит от воспринимающего субъекта. Слух или намек может для умного человека стать знанием или источником знаний, а для не любящего думать и ворох информации останется просто информацией, без превращения в знания. Довольно часто можно встретить людей с хорошей памятью, но не имеющих привычки думать. Это показывает, что существует зависимость качества знаний от наличия и мощности процедур обработки.

Наилучший (вырожденный) пример информации, не превращающейся в знание, – иностранный язык. Наличие информации в иностранном тексте очевидно, однако в знание ее превратить невозможно, если не знать языка, или очень трудоемко, если пользоваться словарем.

Знакомый многим машинный переводчик – типичный пример автоматизированной процедуры извлечения знаний, дающей их существенный прирост, но при этом все равно крайне несовершенной. Переводчик-человек – пример более совершенной процедуры извлечения знаний.

Отличия процедур извлечения знаний человека и машины в настоящее время определяются не только большим числом используемых правил продукции в процессе обобщения и логического вывода, более эффективной реализацией этих процедур, но также и в наличии у человека функций самообучения, т. е. автоматического формирования новых правил логического вывода.

Зачастую текст и на родном языке может быть настолько же трудным для извлечения смысла, как и текст на чужом языке, например, в силу смысловой загруженности, специальных терминов или большого объема.

Именно рост объемов компьютерной информации создает огромную потребность в извлечении знаний.

Поиск

Поиск – это простейший способ доступа к текстовым данным, скорее подбор информации, чем извлечение знаний. Поисковые машины выполняют следующие функции: индексирование текстов и поиск по ключевым словам (по индексу); морфологический поиск – разбор и отождествление различных грамматических форм слов; логический язык запросов, позволяющий задавать условия на совместное вхождение ключевых слов в искомый документ; ранжирование по степени соответствия документа запросу.

Этой простой возможности при росте объемов текстовых баз становится совершенно недостаточно, и в последнее время поисковые машины начинают оснащаться средствами извлечения знаний. В первую очередь новшества появляются в поисковых машинах сети Интернет, а затем постепенно проникают в средства поиска, входящие в корпоративные системы документооборота. Рассмотрим некоторые из этих новшеств.

Итеративный поиск: функция «найти похожие». Данная возможность позволяет постепенно уточнить запрос: указать на один или несколько найденных документов и попросить найти документы, повествующие «о том же». Выполняется такой поиск путем превращения документа в поисковый запрос (естественно, с определенной степенью «сжатия», так как запрос слишком большим быть не может). В Интернете данная функция представлена, например, «Яндексом» (www.yandex.ru) и «Рамблером» (www.rambler.ru).

Поиск по выборке

Если по первому запросу поисковая машина нашла слишком много документов, то второй запрос с заданием дополнительных терминов можно провести только по ним и тем самым уточнить поиск. Эта функция очень проста в исполнении (нужно только запомнить предыдущую выборку) и реализована в большинстве популярных машин сети Интернет, включая «Яндекс» и «Рамблер».

Запрос на естественном языке

В отличие от формальных языков запросов с логическими операторами, запрос на естественном языке позволяет пользователю просто задать поисковой машине вопрос в свободной форме, как если бы он задавался человеку: «Как

выйти замуж за миллионера?», «Скажи, где купить привод CD-ROM?» и т. д. Реализуется данная функция путем отбрасывания шумовых слов и выражений («как», «скажи», «где»), выделения грамматической структуры запроса, подстановки синонимов и т. п.

Большинство поисковых машин декларирует подобную возможность, и пользователь действительно может ввести в строке запроса «естественный» набор слов, не разделяя их логическими операторами, однако зачастую эта функция фактически не работает. Например, «Рамблер» даже не отбрасывает «шумовых» слов и показывает все вхождения слов «как», «за» и подобных им, что, конечно, не имеет смысла и не нужно пользователю.

Тезаурусы

Тезаурусы (словари) служат для так называемого расширения запроса и включают синонимы, антонимы, родственные слова, «вышестоящие» и «нижестоящие» категории и понятия (военный – офицер – капитан).

Хотя использование тезауруса позволяет сделать поиск по-настоящему смысловым, пока большинство поисковых машин тезауруса не имеют, в том числе и упоминавшиеся российские машины. Трудно сказать, чем это вызвано, видимо, высокой стоимостью и сложностью лексикографической (словарной) работы.

Поиск типа «где»

Как видно из сказанного выше, поисковые машины отыскивают все упоминания нужного понятия в документах. В последнее время стало ясно, что пользователя часто интересует не только поиск «чего-нибудь» где угодно и в любом количестве, но и наоборот – поиск одного конкретного места, где находится уже известная ему вещь: интернет-сервер или домашняя страница. Чтобы удовлетворить эту потребность, начали появляться особые поисковые машины – службы имен, позволяющие по «обычному» имени ресурса Интернета найти точное место, где он расположен. Например, калифорнийская служба имен RealNames (www.realnames.com) и Российская национальная служба имен (www.names.ru).

Обе эти службы позволяют вводить естественно-языковые имена в адресную строку программы просмотра (браузера) и сразу попадать на нужный сервер, а также допускают использование различных синонимов для одного имени и задание неполных имен. Делается это путем поддержания базы имен и их соответствий обычным формальным адресам Интернета, а также рассылки пользователям драйверов для распознавания естественных имен в браузере. Впрочем, данные службы дают возможность не только адресоваться к глобальной сети, но и создавать локальные службы естественно-языковых имен для внутренних сетей.

Все вышеперечисленные функции основаны на статистике и морфологическом анализе текста. Реализация принципиально новых функциональных возможностей поисковых систем и повышение их интеллектуальности требует использования полного лингвистического анализа

текстов (запроса пользователя и индексируемых ресурсов). В качестве примера существующих поисковых систем, в которых реализован синтаксический и семантический анализ, можно привести AskNet. В данной системе реализован полный лингвистический анализ русско- и англоязычных текстов, что позволило реализовать некоторые принципиально новые функции.

Каталоги

Каталоги, т. е. рубрикаторы или классификаторы, организующие множества документов посредством деревьев категорий, наполненных ссылками на документы и прочие ресурсы.

Таким же образом строятся каталоги и классификаторы вне Internet, в частности в правовых справочных базах данных, в корпоративных информационных системах и т. д.

Каталоги предоставляют более удобный вид доступа к данным, так как они больше, чем поисковые машины, похожи на знания за счет своей структурированности. Однако для эффективного использования каталога пользователю нужно правильно подобрать принцип структуризации, который был применен при его создании и который зачастую существует только в головах авторов каталога. Поэтому для компенсации неудобств (если уж пользователь не знает, на какой ветке дерева «по логике» данного каталога лежит нужная ссылка, то не найдет ее никогда) каталоги в последнее время стали включать и средства поиска. И наоборот – поисковые машины стали снабжаться каталогами.

Видимо, в дальнейшем эти два вида доступа окончательно переплетутся; вопрос только в том, как автоматизировать составление каталогов.

Аннотирование

Раньше в книгах каждая глава начиналась с краткого изложения: «Робинзон видит человеческие следы, находит Пятницу, приучает его солить пищу, собирает первый урожай киви». Эти изложения (аннотации) выносились в оглавление, что было очень удобно.

Многие из существующих каталогов содержат аннотации для внесенных в них ссылок. Например, Yahoo содержит аннотированные ссылки на 500 000 интернет-страниц, разбитых на 25 000 категорий. Аннотации составляются вручную, например, для сервера по экологическим проблемам – специалистом по экологии, а для медицинского сервера – специалистом по медицине и т. д. Однако даже для каталогов по узким темам или географически ограниченными областями Интернета ручная классификация и аннотирование начинают представлять собой проблему из-за увеличения объема информации в Интернете (увеличивается сложность системы, количество связей, дублей и т. д.) Возникает потребность в автоматизации каталожной работы, в частности в автоматическом составлении аннотаций.

Гипертекст

Давно применявшиеся ссылки в книгах с возникновением компьютерных методов хранения текстов преобразовались в гипертекст. Гипертекст

формируется в результате представлений текста как ассоциативно связанных блоков информации. Ассоциативная связь – это соединение, сближение представлений: смежных, противоположных, аналогичных и т. д.

Гипертекст значительно отличается от обычного текста. Обычные (линейные) тексты имеют последовательную структуру и предусматривают их чтение слева направо и сверху вниз. Использование гипертекста позволяет фиксировать отдельные идеи, мысли, факты, а затем связывать их друг с другом, двигаясь в любых направлениях, определяемых ассоциативными связями. В тех случаях, когда к блокам текста добавляются большое число изображений и запись звука, гипертекст превращается в гиперсреду.

Однако способ организации информации через ссылки, кроме удобства доступа (увидел ссылку – щелкнул мышью), имеет несколько существенных недостатков, среди которых низкая наглядность, неочевидность структуры, плохая запоминаемость пользователем пройденного пути и т. д. Человек с трудом запоминает вложенные структуры с уровнем вложенности более трех, а этим количеством уровней никогда не удастся обойтись при создании крупных структур данных.

Тем не менее способ организации гипертекстовых связей сам по себе как дополнительное средство доступа достаточно удобен и также допускает автоматизацию при создании больших массивов знаний.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАНИЯ

1. Как соотносятся понятия «шум», «данные», «информация», «знание»?
2. В каком виде могут быть представлены данные?
3. Дайте определение понятию «знание».
4. Определите достоинства и недостатки основных способов поиска данных.
5. Приведите собственные примеры «машинных» и неавтоматизированных способов поиска данных.
6. Изобразите графически информационную систему.

Методы анализа и интерпретации данных

Основной формой знания, получаемого на эмпирическом этапе, является научный факт и совокупность эмпирических обобщений. На теоретическом уровне получаемое знание фиксируется в форме законов, принципов и научных теорий. Основными методами, используемыми на эмпирическом этапе, являются наблюдение, эксперимент, индуктивное обобщение. На теоретическом этапе познания используются такие методы, как анализ и синтез, идеализация, индукция и дедукция, аналогия, гипотеза и др.

Граница между эмпирическим и теоретическим уровнями весьма условна, их самостоятельность является относительной. Эмпирическое переходит в тео-

ретическое, а то, что когда-то было теоретическим, на другом, более высоком этапе развития, становится эмпирически доступным. В любой сфере научного познания, на всех уровнях наблюдается диалектическое единство теоретического и эмпирического. Ведущая роль в этом единстве зависимости от предмета, условий и уже имеющихся, полученных научных результатов принадлежит то эмпирическому, то теоретическому этапу познания. Основой единства эмпирического и теоретического уровней научного познания выступает единство научной теории и научно-исследовательской практики.

В процессе обучения выполняются следующие этапы теоретического обобщения:

- 1) факты;
- 2) модель;
- 3) следствия;
- 4) применение.

1 этап. Накопление и анализ фактов и их связей в процессе предметно-материальной деятельности человека.

2 этап. Абстрагирование – отвлечение от конкретных явлений и формулировка обобщения с использованием той или иной идеальной его формы.

3 этап. Получение и обсуждение всевозможных конкретных выводов и следствий из главной закономерности – абстрактной формулы, закона, принципа.

4 этап. Применение полученных знаний к конкретным объектам и явлениям.

Так, например, *научными методами эмпирического исследования* являются наблюдения, описания, измерения, эксперименты.

Наблюдение – целенаправленное восприятие явлений объективной действительности.

Основные функции наблюдения:

- фиксация и регистрация фактов;
- предварительная классификация фактов, уже зафиксированных на основе определенных принципов, сформулированных на основе существующих теорий;
- сравнение зафиксированных фактов.

Описание – фиксация средствами естественного или искусственного языка сведений об объекте. Это указание признаков предмета или явления как существенных, так и несущественных. Описание, как правило, применяется в отношении единичных, индивидуальных объектов для более полного ознакомления с ними. Его метод – дать наиболее полные сведения об объекте.

Измерение – сравнение объекта по каким-либо сходным свойствам или сторонам. Это определенная система фиксации и регистрации количественных характеристик исследуемого объекта с помощью различных измерительных приборов и аппаратов. Путем измерения определяется отношение одной количественной характеристики объекта к другой, однородной с ней, принятой за единицу измерения.

Основными функциями метода измерения является, во-первых, фиксация количественных характеристик объекта, во-вторых, классификация и сравнение результатов измерения.

Эксперимент – наблюдение в специально создаваемых и контролируемых условиях, что позволяет восстановить ход явления при повторении условий.

Эксперимент имеет свои специфические особенности по сравнению с другими методами:

- позволяет исследовать объекты в так называемом «чистом» виде;
- позволяет исследовать свойства объектов в экстремальных условиях, что способствует более глубокому проникновению в их сущность;
- важным преимуществом эксперимента является его повторяемость, благодаря чему в научном познании этот метод приобретает особое значение и ценность.

Существуют различные виды эксперимента: исследовательский; проверочный; воспроизводящий и др.

Методы теоретического уровня познания

Анализ – это расчленение предмета на его составляющие (стороны, признаки, свойства, отношения) с целью их всестороннего изучения.

Синтез – это объединение ранее выделенных частей (сторон, признаков, свойств, отношений) предмета в единое целое.

Анализ и синтез диалектически противоречивые и взаимообусловленные методы познания. Познание предмета в его конкретной целостности предполагает предварительное расчленение его на составляющие и рассмотрение каждой из них. Это задачу выполняет анализ. Он дает возможность выделить существенное, то, что составляет основу связи всех сторон изучаемого объекта. Диалектический анализ является средством проникновения в сущность вещей. Но играя важную роль в познании, анализ не дает знания конкретного, знания объекта как единства многообразного, единства различных определений. Эту задачу выполняет синтез. Следовательно, анализ и синтез органично взаимодействуют и взаимообуславливают друг друга на каждом этапе процесса теоретического познания.

Абстрагирование – это метод отвлечения от некоторых свойств и отношений объекта и одновременно сосредоточение основного внимания на тех, которые являются непосредственным предметом научного исследования. Абстрагирование способствует проникновению познания в сущность явлений, движению познания от явления к сущности. Понятно, что абстрагирование расчленяет, огрубляет, схематизирует целостную подвижную действительность. Однако именно это и позволяет более глубоко изучить отдельные стороны предмета «в чистом виде».

Важную роль в исследовании играют и приемы обобщения. На первый взгляд они кажутся простыми, но в действительности требуют глубокого понимания исследуемых явлений. Иногда они связаны с возникновением и вве-

дением в обиход, в практику деятельности новых понятий. *Обобщение* – это метод научного познания, который фиксирует общие признаки и свойства определенной группы объектов, осуществляет переход от единичного до общего и общего.

В процессе познания нередко приходится, опираясь на уже существующие знания, делать выводы, которые являются новым знанием о неизвестном. Это осуществляется с помощью таких методов, как индукция и дедукция.

Индукция – это такой метод научного познания, когда на основании знания об отдельном делается вывод об общем. Это способ рассуждения, посредством которого устанавливается обоснованность выдвинутого предположения или гипотезы. В реальном познании индукция всегда выступает в единстве с дедукцией, органически связана с нею.

Дедукция – это метод познания, когда на основе общего принципа логическим путем из одних положений как истинных выводится новое истинное знание об отдельном. С помощью этого метода отдельное познается на основе знания общих закономерностей.

Классификация – это разделение явлений, а следовательно, и понятий, характеризующих их, на определенные классы, позволяющие увидеть специфику явлений, их разнообразие, свойства, связи и зависимости, общее и специфическое, и посредством этого проникнуть в их сущность.

Разновидностями классификации являются декомпозиция, стратификация.

Декомпозиция – это особый вид классификации, не допускающий произвольного критерия. Декомпозиция предназначена для установления связанных между собой содержательных элементов некоторой объективной целостности.

Стратификация – это определение слоев (страт) в многослойном явлении, т. е. зависимостей особого вида. В исследовании управления такими стратами могут быть внешняя и внутренняя среда, технические средства и человеческие ресурсы, стратегия и тактика управления и т. д.

Особым вариантом классификации является типология. *Типология* – это группировка объектов на основе их подобия некоторому образцу, который именуется типом, эталоном или идеальным образом. Здесь каждое явление в большей или меньшей степени приближается к одному из эталонов.

Отличие типологии от классификации в том, что типология допускает существование таких явлений, которые не соответствуют ни одному из выделенных типов.

Типология превосходит классификацию своей универсальностью. Она является первоначальной операцией любых систематизаций.

Моделирование – это изучение объекта (оригинала) путем создания и исследования его копии (модели), которая по своим свойствам в определенной степени воспроизводит свойства исследуемого объекта.

Моделирование используется тогда, когда непосредственное изучение объектов по некоторым причинам невозможно, затруднено или нецелесообразно. Аналитик создает модель как подобие изучаемого объекта. Модели могут

быть записаны в виде различных изображений, схем, математических формул и т. д. Преимуществом использования моделей при исследованиях является простота модели в сравнении с исследуемым объектом. При этом модели позволяют выделить в объекте наиболее существенные факторы с точки зрения цели исследования и не отвлекаться на маловажные детали. Из этого следует, что модель обладает свойством неполноты, поскольку является по своему определению абстрактной.

Идеализация – это способ логического моделирования, благодаря которому создаются идеализированные объекты. Идеализация направлена на процессы мысленного построения возможных объектов. Результаты идеализации – произвольные. В предельном случае они соответствуют отдельным реальным свойствам объектов или допускают интерпретацию их, исходя из данных эмпирического уровня научного познания. Идеализация связана с «мысленным экспериментом», в результате которого из гипотетического минимума некоторых признаков поведения объектов открываются или обобщаются законы их функционирования. Границы эффективности идеализации определяются практикой.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАНИЯ

1. Как соотносятся эмпирический и теоретический уровни познания?
2. Перечислите этапы теоретического обобщения.
3. Приведите собственные примеры дедуктивных и индуктивных умозаключений.
4. Составьте собственную классификацию.
5. Сформулируйте цель возможного исследования посредством наблюдения. Определите критерии и показатели изучаемого явления.
6. Составьте описание модели исследуемого явления.

Выборочный метод сбора данных

На практике по тем или иным причинам не всегда возможно или целесообразно исследовать все интересующие нас объекты. Тогда применяют выборочный метод – ограничиваются изучением лишь некоторой части объектов.

Выборочный метод можно сформулировать как получение социальной информации путем строго научного обоснования выбора части социальных объектов как метода исследования всей их совокупности. По разным признакам выделяют определенную совокупность выборочных методов. Так, выделяются сплошные и несплошные методы. В результате сплошного исследования образуется генеральная совокупность – множество социальных объектов, которые являются предметом изучения в пределах очерченных программой исследования и территориально-временными границами. В случае применения несплошных методов исследования создается выборочная совокупность, отражающая часть объектов генеральной совокупности. Наиболее часто используются три

метода несплошного исследования: монографический, метод основного массива и выборочный.

На практике часто применяются упрощенные варианты выборочного метода, в результате чего выборочная совокупность конструируется таким образом, чтобы при минимуме исследуемых объектов удавалось с необходимой степенью гарантии представить всю генеральную совокупность. В качестве упрощенных методов выборки используются следующие – простой, случайный, систематический, серийный, стратифицированный. Выборочная совокупность может создаваться на основе многоступенчатой и комбинированной методологии.

Выборочный метод в условиях недостатка информации о генеральной совокупности предполагает применение неслучайных методов отбора, наиболее известными из которых являются стихийная выборка и квотная выборка.

В *стихийной выборке* не учитывается соответствие структуры выборочной совокупности испытуемых их генеральной совокупности. Иными словами, это выборка первых встречных, в процессе которой исследователь нередко бессознательно руководствуется личной симпатией или антипатией, тем самым снижая объективность результатов исследования.

Квотная выборка – это выборка, при которой соблюдаются характерные для генеральной совокупности соотношения лиц с определенными наборами социальных параметров, которые признаны значимыми для исследования.

Число единиц наблюдения в составе выборочной совокупности называется объемом выборки.

Свойство выборочной совокупности отражать исследуемые характеристики генеральной совокупности называется *репрезентативностью*. Очевидно, что абсолютно репрезентативную (отражающую по всем параметрам) выборку сформировать невозможно, но необходимо обеспечивать репрезентативность по основным направлениям исследования.

Основными требованиями к выборке являются:

- репрезентативность;
- случайность формирования (каждый объект генеральной совокупности должен иметь равную вероятность быть отобранным);
- достаточность объема для получения статистически значимых результатов.

Выборочный метод имеет ряд преимуществ:

- снижение затрат на проведение исследования;
- ускорение сбора и последующей обработки информации;
- широкая область применения (при некоторых видах обследований для сбора данных необходимо привлечь высококвалифицированный персонал или воспользоваться специальным оборудованием; как правило, и то и другое ограничено, в этих случаях удобнее применить выборочные обследования);
- увеличение достоверности (за счет использования опытного и квалифицированного персонала для обследования).

Взаимосвязь объема выборки и репрезентативности

Репрезентативность не зависит от объема выборки. Репрезентативность достигается только тогда, когда в выборку отобраны объекты из разных групп, при условии, что их доли в генеральной и выборочной совокупности равны. Репрезентативность выборки зависит только от методики отбора единиц из генеральной совокупности в выборочную совокупность и не зависит от объема последней. Очевидно, что чем больше объем выборки, тем выше ее точность, однако неверно распределенная выборка в 2000 единиц даст большее искажение, чем хорошо распределенная выборка в 200 единиц.

Чем более однородна генеральная совокупность, тем меньший объем выборочной совокупности потребуется для получения точных результатов.

Понятие ошибки репрезентативности, виды ошибок репрезентативности

При проведении выборочного наблюдения нельзя даже теоретически получить абсолютно точные данные, как при сплошном обследовании. Обусловлено это тем, что наблюдению подвергается не вся совокупность, а только ее часть, поэтому при проведении выборочного наблюдения неизбежна некоторая свойственная ему погрешность (ошибки). Принято различать среднюю и предельную ошибки выборки.

Ошибка репрезентативности – расхождение между выборочной характеристикой и характеристикой генеральной совокупности.

Выделяют следующие ошибки репрезентативности:

- систематические (возникают в результате нарушения научных принципов отбора единиц совокупности);
- случайные (возникают в результате несплошного наблюдения).

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАНИЯ

1. По каким критериям опрашиваемых можно разделить генеральную совокупность?
2. По какому критерию выделяют группы, представляющие наибольший интерес для исследования?
3. Приведите примеры случаев, в которых целесообразно использовать выборочный метод.
4. Какие источники можно использовать для определения соотношения между классами для квотной выборки?
5. Верно ли, что чем больше выборка, тем достовернее полученные данные?
6. Что такое репрезентативность выборки?
7. Дополните предложение. Квотной выборкой называется такой вид выборочной совокупности, при котором

Модели представления знаний

Важное место в информационных системах занимает проблема представления знаний, являющаяся ключевой. Можно выделить вопросы, актуальные для любых информационных систем:

- приобретение новых знаний и их взаимодействие с уже существующими;
- организация ассоциативных связей;
- явность и доступность знаний;
- выбор соотношения декларативной и процедурной составляющих представления, что влияет на экономичность системы, полноту, легкость кодировки и понимания.

Модели представления знаний можно условно разделить на *декларативные* и *процедурные*.

Декларативная модель представления знаний основывается на предположении, что проблема представления какой-либо предметной области решается независимо от того, как эти знания потом будут использоваться. Поэтому можно сказать, что модель состоит из двух частей: статических описательных структур знаний и механизма вывода, оперирующего этими структурами и практически независимого от их содержательного наполнения. При этом оказываются раздельными синтаксические и семантические аспекты знания, что является достоинством указанных форм представления знаний из-за возможности достижения их определенной универсальности.

В декларативных моделях не содержатся в явном виде описания выполняемых процедур. Эти модели представляют собой множество утверждений. Предметная область представляется в виде синтаксического описания ее состояния. Вывод решений основывается в основном на процедурах поиска в пространстве состояний.

В процедурном представлении знания содержатся в небольших программах, которые определяют, как выполнять специфичные действия (как поступать в специфичных ситуациях).

При этом можно не описывать все возможные состояния среды или объекта для реализации вывода. Достаточно хранить некоторые начальные состояния и процедуры, генерирующие необходимые описания ситуаций и действий. При процедурном представлении знаний семантика непосредственно заложена в описание элементов базы знаний, за счет чего повышается эффективность поиска решений.

«Семантическая» – означает смысловая, а сама семантика – это наука, устанавливающая отношения между символами и объектами, которые они обозначают, т. е. наука, определяющая смысл знаков.

Понятие *семантической сети* основано на идее о том, что память формируется через ассоциации между понятиями. Понятие «ассоциативная память» появилось еще во времена Аристотеля. В информатику оно вошло в связи с работами по использованию простых ассоциаций для представления значений

слов в базе данных. Разработка семантических сетей относится к 1960 г., когда они использовались для моделирования обработки естественного языка, для представления смысла (семантики) выражения. Отсюда и происходит их название. Квиллиан предположил, что наша способность понимать язык может быть охарактеризована некоторым множеством базовых понятий (концептов) и правил. Модель Р. Квиллиана и А. Коллинза предполагает, что семантическая память состоит из обширной сети понятий, которые составлены из единиц и свойств и соединены рядом ассоциативных связей. Несмотря на то что отдельные стороны модели подверглись критике (например, то, что сила ассоциативных связей в пределах сети варьируется), модификации этой модели явились основой для создания последующих моделей. Теперь они используются в качестве структуры, пригодной для представления информации общего вида.

Базовыми функциональными элементами семантической сети служит структура из двух компонентов – узлов и связывающих их дуг. Таким образом, семантической сетью называется ориентированный граф с конечными вершинами (рис. 1). Каждый его узел представляет собой некоторое понятие, а дуга – отношение между парой понятий. Можно считать, что каждая из таких пар отношений представляет простой факт. Узлы в семантической сети соответствуют объектам, понятиям или событиям. Они обладают определенной маркировкой, позволяющей идентифицировать этот узел.

Основной принцип семантической сети: знания, которые семантически связаны между собой (связаны по смыслу) должны храниться рядом. В семантической сети имеется два типа дуг:

- 1) является (is);
- 2) имеет частью (has part).

Дуги обладают свойством транзитивности – устанавливают отношения иерархии наследования в сети (элементы низкого уровня наследуют свойства высокого).

В качестве простого примера семантической сети рассмотрим предположения: «Студент Иванов является мужчиной»; «Мужчина является человеком». Студент Иванов *является* мужчиной *является* человеком.

Очевидно, что отношение «является» транзитивно, т. е. из этой сети мы можем вывести третье утверждение, хотя оно и не было сформулировано в чистом виде «Студент Иванов является человеком». Свойство транзитивности позволяет экономить память, поскольку информация о сходных узлах может не повторяться в каждом узле сети, а храниться в одном центральном узле. Это свойство модели памяти получило наименование *когнитивной экономии*.

Квиллиан также ввел разделение между видами узлов. Один вид узлов он назвал *узлами-типами*. Такой узел представляет концепт, связанный с конфигурацией других узлов – *узел-лексем*. Это в определенной степени напоминает толковый словарь, в котором каждое понятие определяется другими понятиями, также присутствующими в этом словаре, причем и их смысл толкуется с помощью еще каких-либо понятий в этом словаре.

Например, можно определить смысл слова «машина» как конструкцию из связанных компонентов, которые передают усилия для выполнения определенной работы. Это потребует присоединения узла-типа для слова «машина» к узлам-лексемам «конструкция», «компонент» и т. д. Однако в дополнение к связям, сформированным для определения смысла, могут существовать связи к другим узлам-лексемам, например «телетайп», «офис». Эти связи представляют знания о том, что телетайпы являются одним из видов машин, которые используются в офисе.

В качестве понятий обычно выступают абстрактные или конкретные объекты (огурец, машина, любовь, Маша). В качестве отношений наиболее часто используются следующие:

- связи типа «часть – целое» («экземпляр – класс», «элемент – множество» и т.п.);
- функциональные (определяемые обычно глаголами «производит», «влияет» и т. п.);
- количественные (больше, меньше, равно и т. п.);
- пространственные (далеко от, близко от, за, под, над и т. п.);
- временные (раньше, позже, в течение и т. п.);
- атрибутивные (иметь свойство, иметь значение);
- логические (И, ИЛИ, НЕ);
- казуальные (причинно-следственные).

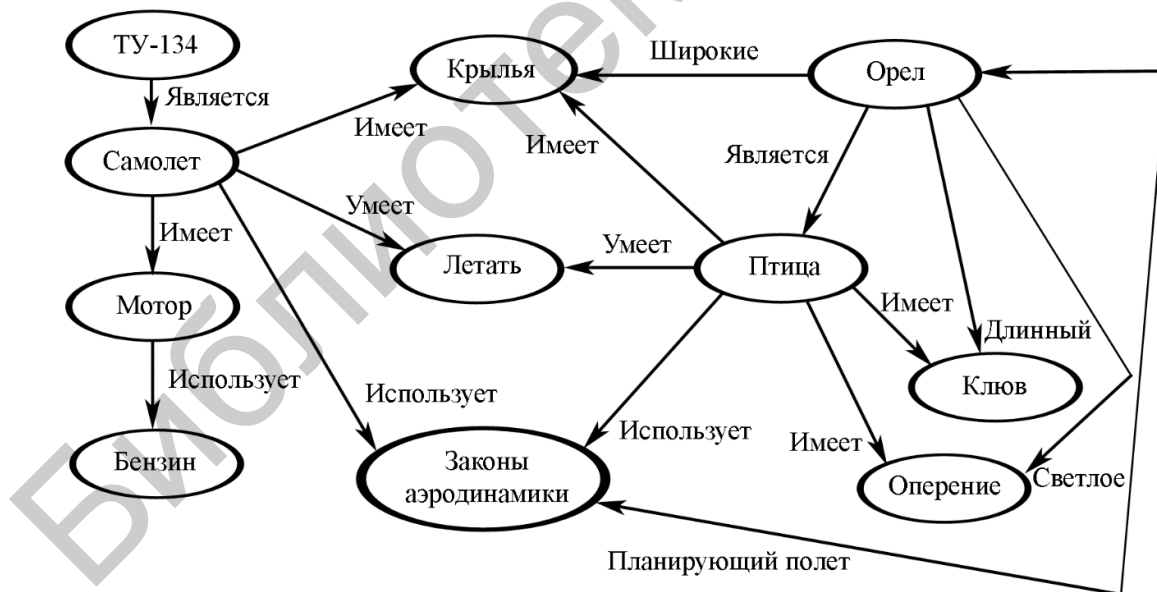


Рис. 1. Семантическая сеть

Семантические сети широко используются в экспертной системе в качестве языка представления знаний, в системах распознавания речи и понимания естественного языка. Непосредственное отношение к сетевым моделям имеют исследования по реляционным, сетевым и иерархическим базам данных (БД).

Проблема поиска решений в базе знаний типа семантической сети сводится к задаче поиска фрагмента сети, соответствующего некоторой подсети, отражающей поставленный запрос к базе.

Существует довольно обширный перечень проблем, при решении которых представление, базирующееся на формализме семантических сетей, оказывается весьма полезным. И использование узлов и связей в сети для представления понятий и отношений может показаться само собой разумеющимся.

Достоинства семантических сетей:

1. Универсальность, достигаемая за счет выбора соответствующего набора отношений. В принципе с помощью семантической сети можно описать сколь угодно сложную ситуацию, факт или предметную область.

2. Наглядность системы знаний, представленной графически.

3. Близость структуры сети, представляющей систему знаний, семантической структуре фраз на естественном языке.

4. Соответствие современным представлениям об организации долговременной памяти человека.

Недостатки семантических сетей:

1. Сетевая модель не дает (точнее не содержит) ясного представления о структуре предметной области, поэтому формирование и модификация такой модели затруднительны.

2. Сетевые модели представляют собой пассивные структуры, для обработки которых необходим специальный аппарат формального вывода.

3. Проблема поиска решения в семантической сети сводится к задаче поиска фрагмента сети, соответствующего подсети, отражающей поставленный запрос. Это, в свою очередь, обуславливает сложность поиска решения в семантических сетях.

4. Представление, использование и модификация знаний при описании систем реального уровня сложности оказывается трудоемкой процедурой, особенно при наличии множественных отношений между ее понятиями.

Таким образом, структура сетей, как правило, не дает четкий ответ на два важных вопроса: что же действительно стоит за узлами и связями в сети? как можно эффективно обрабатывать информацию, хранящуюся в такой структуре?

Одним из способов ликвидировать эти недостатки является переход к специальному представлению узлов в сети и унификация связей между узлами (фреймами).

По определению М. Минского, фрейм есть «структура данных для представления стереотипных ситуаций». Идея состояла в том, чтобы сконцентрировать знания в единой структуре данных, а не распределять их между множеством более мелких структур. Такие знания либо сосредоточены в самой структуре данных, либо доступны из этой структуры (например, хранятся в другой структуре, связанной с фреймом). Доступ к связанным структурам осуществляется посредством присоединенных процедур. По существу фрейм оказался тем средством, которое помогло связать декларативные и процедурные знания. По

своей структуре фреймы похожи на семантические сети. Можно считать, что фрейм – это сложный узел сети. Часть специалистов по системам искусственного интеллекта даже считает, что нет необходимости выделять фреймовые модели представления знаний, так как в них объединены все основные особенности моделей остальных типов. Таким образом, можно дать определение фрейму как минимально возможному описанию сущности какого-то явления, события, ситуации, процесса или объекта. Минимально возможное означает, что при дальнейшем упрощении описания теряется его полнота, оно перестает определять ту единицу знаний, которой предназначено.

В системе фреймов предпринимается попытка судить о классе объектов, используя представление знаний о *прототипах*, которые хорошо представляют различные объекты данного класса. Например, «прототипическая» птица, пусть это будет воробей, может летать, а потому есть основания полагать, что это свойство *всех* птиц, хотя существуют редкие виды птиц, которые этим свойством не обладают, например пингвины. Именно в этом смысле воробей является лучшим экземпляром категории «птицы», чем пингвин, поскольку он представляет более типичные свойства объектов своего класса.

При решении практических задач подобных исключений весьма много, а границы между классами оказываются очень размыты. Система фреймов оказывается полезной по той причине, что она дает средства структурирования знаний, связанных с приложением правил и классификации объектов.

В качестве идентификатора фрейму присваивается имя фрейма. Это имя должно быть единственным во всей фреймовой системе.

Фрейм имеет определенную внутреннюю структуру, состоящую из множества элементов, называемых *слотами*, которым также присваиваются имена. Каждый слот, в свою очередь, представляется определенной структурой данных. В значение слота подставляется конкретная информация, относящаяся к объекту, описываемому этим фреймом. Таким образом, каждый фрейм имеет специальный слот, заполненный наименованиями сущностей, и слоты, заполненные значениями различных атрибутов, ассоциирующихся с объектом.

Как правило, фреймы организованы в виде «ослабленной иерархии» (или «гетерархии»), в которой фреймы, расположенные ниже в иерархии, могут наследовать значения слотов разных фреймов, расположенных выше.

Фундаментальная идея состоит в том, что свойства и процедуры, расположенные выше, являются более или менее фиксированными, поскольку они представляют те вещи или понятия, которые в большинстве случаев являются истинными для интересующей нас сущности. В то же время фреймы более нижних уровней имеют слоты, которые должны быть заполнены наиболее динамической информацией, подверженной частым изменениям. Если такая динамическая информация отсутствует из-за неполноты наших знаний о предмете, то слоты фреймов более нижних уровней заполняются данными, унаследованными от фреймов верхних уровней.

Значением слота может быть практически что угодно: числа, формулы, тексты на естественном языке или программы, правила вывода или ссылки на другие слоты данного фрейма или других фреймов. Кроме того, со слотами может ассоциироваться любая дополнительная информация (например, процедура вычисления значения этого слота в случае отсутствия его явного заполнения, процедура обновления значения слота при изменении значения другого слота). В качестве значения слота даже может выступать набор слотов более низкого уровня, что позволяет реализовывать во фреймовых представлениях «принцип матрешки». Связи между фреймами задаются значениями специального слота с именем «Связь».

Фреймы с заполненными слотами называются *экземплярами фрейма*.

Для описания объекта, ситуации или события строится сеть фреймов. Фрейм представляет собой сложный узел такой сети. Как уже отмечалось выше, слоты фрейма обладают *присоединенными процедурами*, позволяющими осуществлять связь в структуре фрейма. С каждым фреймом связана информация о том, как его использовать, как поступать в чрезвычайных ситуациях.

Поиск решения во фреймовой модели представления знаний означает поиск по всем фреймам (сети фреймов) модели, которая наилучшим образом описывает ситуацию. В процессе поиска слоты фреймов заполняются информацией. Полученные экземпляры фреймов проверяются на адекватность рассматриваемой ситуации. Если выявлено хотя бы одно несовпадение, рассматривается следующий фрейм.

Для автоматизации процесса использования и представления знаний предпочтительной является унифицированная, однородная форма представления информации. Поэтому в большинстве последних исследований, касающихся представления знаний, предпочтение отдается фреймам. Узлы в сети представлены как иерархия фреймов, связи организованы связями между фреймами. Эффективность обработки обеспечивается подключением к узлам специфических процедур, на которые возлагается вычисление значений переменных в ответ на запросы или при обновлении значений других свойств узлами фрейма.

Модель фрейма является достаточно универсальной, поскольку позволяет отобразить все многообразие знаний о мире через:

- фреймы-структуры, для обозначения объектов и понятий (заем, залог, вексель);
- фреймы-роли (менеджер, кассир, клиент);
- фреймы-сценарии (банкротство, собрание акционеров, празднование именин);
- фреймы-ситуации (тревога, авария, рабочий режим устройства) и др.

Важнейшим свойством теории фреймов является заимствованное из теории семантических сетей наследование свойств. И во фреймах, и в семантических сетях наследование происходит по *АКО-связям* (*A-Kind-Of = это*). Слот АКО указывает на фрейм более высокого уровня иерархии, откуда неявно наследуются, т. е. переносятся значения аналогичных слотов.

Например, в сети фреймов понятие «ученик» наследует свойства фреймов «ребенок» и «человек», которые находятся на более высоком уровне иерархии. Так, на вопрос: «Любят ли ученики сладкое?», следует ответ: «Да», так как этим свойством обладают все дети, что указано во фрейме «ребенок». Наследование свойств может быть частичным. Так, возраст для учеников не наследуется из фрейма «ребенок», поскольку указан явно в своем собственном фрейме.

Основным преимуществом фреймов как модели представления знаний является способность отражать концептуальную основу организации памяти человека, а также ее гибкость и наглядность.

Продукционные модели или модель, основанная на правилах, позволяет представить знания в виде предложений типа: *если* (условие), *то* (действие).

Под *условием* понимают некоторое предложение-образец, по которому осуществляется поиск в базе знаний, а под *действием* – действия, выполняемые при успешном исходе поиска (они могут быть промежуточными, выступающими далее как условие, и терминальными или целевыми, завершающими работу системы).

При использовании продукционной модели база знаний состоит из набора правил. Программа, управляющая перебором правил, называется машиной вывода. Чаще всего вывод бывает *прямой* (от данных к поиску цели) или *обратный* (от цели для ее подтверждения к данным). Данные – это исходные факты, на основании которых запускается машина вывода – программа, перебирающая правила из базы.

Продукционные модели являются наиболее популярным способом представления знаний. При организации знаний с использованием продукционных моделей в базе знаний (БЗ) содержатся правила продукций, а в базе данных (БД) содержится информация, которая отображает текущее состояние решаемой задачи. Инициализацию необходимого правила осуществляет блок управления.

Трудности возникают при создании моделей нечетких знаний. Формализация таких знаний осуществляется на основе теории нечетких множеств. Развиваются также модели на основе искусственных нейронных сетей (ИНС), многоагентных систем, генетических алгоритмов и другие модели представления и обработки знаний.

Нейронные сети

Нейронные сети – это адаптивные системы для обработки и анализа данных, которые представляют собой математическую структуру, имитирующую некоторые аспекты работы человеческого мозга и демонстрирующую такие его возможности, как способность к неформальному обучению, способность к обобщению и кластеризации неклассифицированной информации, способность самостоятельно строить прогнозы на основе уже предъявленных временных рядов. Главным их отличием от других методов, например таких, как экспертные системы, является то, что нейросети в принципе не нуждаются в заранее известной модели, а строят ее сами только на

основе предъявляемой информации. Именно поэтому нейронные сети и генетические алгоритмы вошли в практику всюду, где нужно решать задачи прогнозирования, классификации, управления – иными словами, в области человеческой деятельности, где есть плохо алгоритмизуемые задачи, для решения которых необходимы либо постоянная работа группы квалифицированных экспертов, либо адаптивные системы автоматизации, каковыми и являются нейронные сети.

Нейронная сеть принимает входную информацию и анализирует ее способом, аналогичным тому, что использует наш мозг. Во время анализа сеть обучается (приобретает опыт и знания) и выдает выходную информацию на основе приобретенного ранее опыта.

Основная задача аналитика, использующего нейронные сети для решения какой-либо проблемы, – создать наиболее эффективную архитектуру нейронной сети, т. е. правильно выбрать вид нейронной сети, алгоритм ее обучения, количество нейронов и виды связей между ними. Эта работа не имеет формализованных процедур, она требует глубокого понимания различных видов архитектур нейронных сетей, включает в себя много исследовательской и аналитической работы и может занять достаточно много времени.

Для неформализованных задач нейросетевые модели могут на порядок превосходить традиционные методы решения. Но применение нейронных сетей целесообразно, если:

- накоплены достаточные объемы данных о предыдущем поведении системы;
- не существует традиционных методов или алгоритмов, которые удовлетворительно решают проблему;
- данные частично искажены, частично противоречивы или неполны и поэтому традиционные методы выдают неудовлетворительный результат.

Нейронные сети наилучшим образом проявляют себя там, где имеется большое количество входных данных, между которыми существуют неявные взаимосвязи и закономерности. В этом случае нейросети помогут автоматически учесть различные нелинейные зависимости, скрытые в данных. Это особенно важно в системах поддержки принятия решений и системах прогнозирования.

Преимущества. Нейросети являются незаменимыми при анализе данных, в частности для предварительного анализа или отбора, выявления «выпадающих фактов» или грубых ошибок человека, принимающего решения. Целесообразно использовать нейросетевые методы в задачах с неполной или «зашумленной» информацией, особенно в задачах, где решение можно найти интуитивно и при этом традиционные математические модели не дают желаемого результата.

Методы нейронных сетей могут использоваться независимо или же служить хорошим дополнением к традиционным методам статистического анализа, большинство из которых связаны с построением моделей, основанных

на тех или иных предположениях и теоретических выводах (например, что искомая зависимость является линейной или что некоторая переменная имеет нормальное распределение). Нейросетевой подход не связан с такими предположениями – он одинаково пригоден для линейных и сложных нелинейных зависимостей, особенно же эффективен в разведочном анализе данных, когда ставится цель выяснить, имеются ли зависимости между переменными. При этом данные могут быть неполными, противоречивыми и даже заведомо искаженными. Если между входными и выходными данными существует какая-то связь, даже не обнаруживаемая традиционными корреляционными методами, то нейронная сеть способна автоматически настроиться на нее с заданной степенью точности. Кроме того, современные нейронные сети обладают дополнительными возможностями: они позволяют оценивать сравнительную важность различных видов входной информации, уменьшать ее объем без потери существенных данных, распознавать симптомы приближения критических ситуаций и т. д.

Области, в которых эффективность применения нейронных сетей доказана на практике:

- выявление тенденций, корреляций, типовых образцов и исключений в больших объемах данных;
- сравнительный анализ;
- системы слежения за состоянием оборудования;
- проектирование и оптимизация;
- прогнозирование (например, потребления энергии);
- распознавание рукописных символов, в том числе автоматическое распознавание и аутентификация подписи;
- распознавание и обработка видео- и аудиосигналов.

Нейронные сети могут быть использованы и в других задачах. Основными предопределяющими условиями их использования являются наличие «исторических данных», используя которые нейронная сеть сможет обучиться, а также невозможность или неэффективность использования других, более формальных, методов.

Нейронные сети, используемые в сфере искусственного интеллекта, традиционно рассматриваются как упрощенные модели нейронных сетей в головном мозге, хотя вопрос о том, в какой мере искусственные нейронные сети отражают реальную структуру головного мозга, по-прежнему остается открытым.

Предметом исследований в теоретической нейробиологии является вопрос об уровне сложности и свойствах, которыми должны обладать отдельные нейроны, для того чтобы воспроизвести нечто, похожее на разум животных.

Исторически сложилось, что развитие компьютерной техники шло от архитектуры фон Неймана, которая основана на **последовательной** обработке и исполнении явных инструкций. С другой стороны, разработка искусственных

нейронных сетей основывалась на моделях обработки информации в биологических системах, которые предполагают параллельную обработку информации, а также использование неявных инструкций на основе распознавания «сенсорных» входов из внешних источников.

Другими словами, нейронная сеть представляет собой сложный статистический процессор (в отличие от систем, основанных на последовательной обработке и исполнении команд).

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАНИЯ

1. Каковы критерии выбора модели представления знаний?
2. В чем заключаются сходства и различия таких моделей представления знаний, как семантическая сеть и фрейм?
3. Перечислите достоинства и недостатки данных моделей.
4. Приведите собственный пример семантической сети, используя все перечисленные в тексте виды отношений. Изобразите ее графически.
5. Создайте фрейм.
6. Нейросети в искусственном интеллекте традиционно сравнивают с нейронной сетью человека. Проведите аналогию между операциями по приему и переработке информации человеком и машиной.

Современное представление об информационных технологиях, их развитии

Информационные технологии

Способности и возможности людей обрабатывать информацию ограничены, особенно в условиях всевозрастающих массивов информации. Поэтому появилась необходимость использовать способы хранения, обработки и передачи информации (информационные технологии), удаленные от одушевленного носителя – человека.

Термин **технология** (от греч. «techne» – искусство, мастерство, умение и «logos» – знания, наука) обозначает совокупность методов обработки, изготовления, изменения состояния, свойств, формы, осуществляемых в процессе производства продукции. Любая технология связана с выполнением определенных операций и процессов, изменением качества, формы, состояния и содержания материала, объекта и т. п. Например, простейшим видом технологии, практически не использующим какие-либо технические средства, является доставка почтальоном почтовых отправлений (писем, телеграмм, газет и журналов) по указанным адресам.

Технологии, предназначенные для решения информационных задач с помощью различных методов и программно-технических средств, например, связанных с приемом и хранением информации, ее обработкой и преобразованием в форму, удобную для человека, называют *информационными*, а иногда – *компьютерными*. Компьютерными их называют потому, что компьютеры составляют основу технических средств информационных технологий (ТС ИТ).

Информационные технологии – это методы и способы, использующие компьютерные программно-технические средства, отдельные или совокупные информационные процессы и операции для достижения поставленных целей.

Информационные технологии используют при решении различных (социальных, экономических, производственных, культурных) и иных проблем, связанных с деятельностью людей и окружающей их природой.

Под термином *информационные технологии* понимают:

– совокупность программно-технических средств вычислительной техники (СВТ), приемов, способов и методов их применения, предназначенных для сбора, хранения, обработки, передачи и использования информации в конкретных предметных областях;

– совокупность методов, производственных и программно-технологических средств, объединенных для обеспечения сбора, хранения, обработки, вывода и распространения информации.

Следует помнить, что свойства информации определяют свойства информационных технологий. Информационные технологии предназначены для снижения трудоемкости процессов использования информационных ресурсов. Практически любой технологический процесс может быть частью сложного процесса. Он также может включать в себя набор простых (менее сложных) технологических процессов и операций.

Основные этапы развития информационных технологий

ИТ (информационные технологии) являются наиболее важной составляющей процесса использования информационных ресурсов общества. К настоящему времени они прошли несколько эволюционных этапов, смена которых определялась главным образом развитием научно-технического прогресса, появлением новых технических средств переработки информации.

Существует несколько точек зрения на развитие информационных технологий с использованием компьютеров, которые определяются различными признаками деления. Общим для всех изложенных ниже подходов является то, что с появлением ПК (персональных компьютеров) начался новый этап развития ИТ. Основной целью становится удовлетворение персональных информационных потребностей человека как для профессиональной, так и для бытовой сферы. В процессе своего развития информационные технологии прошли через ряд этапов, начало которых связывается с появлением ЭВМ.

1. Первый этап охватывает период с конца 60-х до начала 70-х годов, когда с появлением ЭВМ первого поколения встала задача ускорения процесса кодирования программ по заранее формализованным алгоритмам. Общие за-

траты на программирование составляли в тот период лишь несколько процентов от довольно высокой стоимости аренды ЭВМ, поэтому важнейшей задачей, стоящей перед программистами, была экономия машинных ресурсов. В связи с этим основные затраты на обработку данных находились тогда в почти прямой зависимости от затраченного на них машинного времени.

С появлением ЭВМ второго поколения началась разработка алгоритмических языков программирования. В наибольшей степени решению поставленных задач на этом этапе развития ИТ способствовало применение операционных систем, ориентированных на пакетный режим обработки данных.

2. Второй этап развития ИТ охватывает период с начала 70-х до начала 80-х годов, который характеризуется появлением моделей единой системы ЭВМ третьего поколения, отличающихся друг от друга только быстродействием и объемом оперативной памяти. С развитием средств автоматизации относительный вес машинных ресурсов в общих расходах на обработку данных начал неуклонно снижаться. Успехи в развитии электроники привели к быстрому снижению стоимости работы ЭВМ, тогда как расходы на разработку и сопровождение программ почти не снижались, а в ряде случаев имели тенденцию к росту. В это время наблюдается переход от критерия эффективного исполнения программ к критерию эффективного программирования, а главной задачей становится экономия уже не машинных, а человеческих ресурсов. Для этой цели были разработаны пакеты прикладных программ для автоматизации решения различных экономических задач и системы управления базами данных.

3. Третий этап развития информационных технологий охватывает период с начала 80-х до начала 90-х годов. В этот период появилась тенденция замены программистов на конечных пользователей, т. е. специалистов в конкретной предметной области, но не имеющих профессиональной подготовки в области вычислительной техники и программирования, благодаря появлению на рынке компьютерных средств настольных микроЭВМ, ориентированных на персональный режим работы и получивших название персональных компьютеров (ПК). ПК в отличие от прежних используемых средств вычислительной техники имеет уже такие программно-аппаратные ресурсы, которые обеспечивают дружественное взаимодействие машины и пользователя. Если на ранних этапах развития информационных технологий программист-профессионал должен был вмешиваться в содержательную часть работы пользователя-непрограммиста, то в настоящее время программист включается в процесс формализации знаний только на инструментальном (программном) уровне, оставляя наиболее трудную для его понимания содержательную часть задачи специалисту в данной предметной области. В это время широкое распространение получили диалоговые информационные технологии, автоматизированные рабочие места пользователей, табличные и графические процессоры, а также базы данных и локальные вычислительные сети, основанные на распределенной обработке данных. Если раньше для обработки каждого вида информации (текст, таблицы, графики, базы данных) существовали отдельные информаци-

онные технологии, то сейчас они объединяются в интегрированные пакеты прикладных программ.

4. Четвертый этап, начавшийся с начала 90-х годов, характеризуется разработкой информационных технологий для автоформализации знаний. Основной задачей настоящего этапа развития информационных технологий является разработка инструментальных средств, облегчающих непрограммирующим профессионалам процесс самостоятельной формализации их индивидуальных знаний. Причем на смену технологии, основанной на обработке данных по формализованным алгоритмам, приходит технология, основанная на интеллектуализации работы ЭВМ. Такая технология связана с обработкой неформализованной информации, требующей применения алгоритмов, по своим функциональным особенностям все более приближающимся к человеческому сознанию. Начинает осуществляться интеграция ИТ. Сетевые, гипертекстовые и мультимедийные технологии включаются практически во все предметные ИТ, повышая эффективность их использования. С конца 90-х годов и по настоящее время широко внедряется всемирная паутина Интернет и локальная корпоративная сеть Интранет, а также появляются технологии информационных хранилищ, электронного документооборота и поддержки принятия решений.

5. Пятый этап. Дальнейшее развитие информационных технологий специалисты связывают с использованием в XXI веке нанотехнологий и суперкомпьютеров для выполнения различных информационных процессов с помощью объединенных вычислительных мощностей этих компьютеров, расположенных в любых местах нашей планеты и связанных между собой с помощью телекоммуникаций (Интернета).

С точки зрения используемых *видов инструментария информационных технологий* выделяют шесть этапов:

1-й этап (до второй половины XIX века) связан с использованием «ручных» информационных технологий. Их инструментом в основном являлись канцелярские принадлежности и средства почтовой связи, обеспечивавшие пересылку писем, пакетов и бандеролей.

2-й этап (с конца XIX века) называют периодом «механических» технологий. В этот период к названному инструментарию добавляются средства оргтехники (пишущие машинки, телеграф, телефон, магнитофоны и диктофоны). Информационные коммуникации поддерживаются с помощью более совершенных средств доставки почты.

3-й этап (1940–1960-е годы) относят к «электрическим» технологиям, инструмент которых составляют: большие ЭВМ и программное обеспечение к ним, электрические пишущие машинки, настольные копиры, портативные диктофоны и т. п. В этот период развиваются и совершенствуются существующие информационные коммуникации, появляются телевидение, системы передачи данных по воздушным и безвоздушным линиям связи.

4-й этап (с начала 1970-х годов) характеризуют «электронные» технологии. Их основной инструментарий – большие ЭВМ с создаваемыми на их базе

автоматизированными системами управления (АСУ) и информационно-поисковыми системами (ИПС). Появляются факсимильные средства передачи данных, компьютерные вычислительные и информационные коммуникации: локальные и междугородные вычислительные сети.

5-й этап (с середины 1980-х годов) характеризуется использованием новых компьютерных технологий. Основным инструментом в этот период становится персональный компьютер. Для него создается множество различных программных продуктов и периферийных устройств. Появляются автоматизированные рабочие места (АРМ), в том числе локальные (на одном персональном компьютере) и системы поддержки принятия решений. Информационные коммуникации называют телекоммуникациями. Они включают локальные, региональные глобальные (международные) и иные компьютерные сети. Рост сложности информационных систем (ИС) вызывает разобщенность и разнородность разработчиков, пользователей, аппаратных средств и т. п., необходимость их интеграции.

6-й этап (с начала XXI века) определяют как период формирования информационных обществ. Он характеризуется глобализацией информационных технологий и связанным с ними применением суперкомпьютеров, квантовых и нанокompьютеров и технологий. В области телекоммуникаций все чаще используются оптические проводные и беспроводные системы, а также иные беспроводные коммуникации.

Инструментарий информационных технологий порой называют базой или платформой информационных технологий.

Платформа информационных технологий

Данный термин не имеет однозначного определения. Платформой называют функциональный блок, интерфейс и сервис которого определяется некоторым стандартом. К **платформе** (от англ. «platform»), или базе, **информационных технологий** относят аппаратные средства, устройства и комплексы (компьютеры и периферийные устройства к ним, оргтехника), телекоммуникации, программные продукты и математическое обеспечение, позволяющие пользователям практически в любых предметных областях достигать поставленных целей. С точки зрения информационных технологий считается, что «платформа» соответствует «опорной» их части. **Опорная технология** – это совокупность программно-технических средств, на основе которых реализуются информационные системы и подсистемы.

Платформа – это аппаратно-программный комплекс, обеспечивающий базовый набор сервисов, необходимых пользователям для выполнения определенных задач.

Платформы могут создаваться для выполнения локальных задач, а могут быть универсальными. Они могут модернизироваться, расширяться, полностью заменяться или обновляться. Характеристики универсальной платформы позволяют использовать ее при решении большого круга задач. Выделяют аппарат-

ную, операционную (программную), административную, транспортную, прикладную и коммуникативную платформы.

Аппаратная платформа – это техническое обеспечение вычислительной системы (IBM PC, Macintosh и т. д.), включающее и тип процессора.

Операционная платформа обеспечивает интерфейс между прикладными программами и группой операционных систем (MS DOS, Windows, OS/2, UNIX и т. д.). Она устанавливается на соответствующие компьютеры и позволяет работать с различными программными продуктами. Пользователь приобретает программный продукт и информационную технологию, ориентированные на имеющуюся у него платформу.

Платформа управления сетью (административная платформа) – это комплекс программ, предназначенных для управления сетью и входящими в нее системами. Такая платформа обеспечивает:

- контроль работы устройств и состояния кабелей;
- контроль деловых процедур;
- контроль других аспектов функционирования сети.

Транспортная платформа обеспечивает передачу данных через коммуникационную сеть.

Прикладная платформа связана с прикладными и обслуживающими процессами. Она не зависит от типов коммуникационных сетей.

Коммуникативная платформа – это комплекс информационных материалов (методик, практических рекомендаций), обеспечивающий эффективную совместную работу людей, например в организации.

Таким образом, «платформа» является важной составляющей структуры информационных технологий.

Структура информационной технологии – это внутренняя организация ИТ, представляющая взаимосвязь входящих в нее компонентов.

Другой ее составляющей являются базы знаний, состоящие из баз и банков данных, а также пользовательского интерфейса (рис. 2).



Рис. 2. Структура информационной технологии

Роль информационных технологий в развитии экономики и общества

Развитие экономики тесно связано с развитием любого общества, потому что невозможно рассматривать какие-либо экономические задачи и проблемы вне общества. В любом обществе одновременно создается и используется много различных технологий. При этом общественные процессы включают такие технологии, как экономическая, социальная, политическая, духовная, экологическая, демографическая, информационная и др.

Информационные технологии могут существовать самостоятельно. В большинстве случаев они связаны с различными, осуществляемыми в обществе процессами. В этих процессах информационным технологиям отводится определенная роль. Так, например, информационным технологиям в экономике отводится роль, связанная с управлением государством и бизнесом.

Информационные технологии используются в электронной коммерции, обеспечивают доступ к финансовым рынкам; способствуют решению проблем, связанных с увеличением занятости, притоком инвестиций, особенно в малый и средний бизнес, с подъемом производительности; расширяют возможности всех слоев общества; находят применение в дистанционном образовании и телемедицине, в управлении окружающей средой и ее мониторинге для предотвращения и ликвидации катастроф и др.

Информационные технологии являются стратегически важной отраслью, влияющей на все стороны жизнедеятельности любого современного общества. Специалисты отмечают, что главная их цель заключается в том, чтобы людям в любом уголке планеты стало лучше жить.

В управлении государством использование информационных технологий прежде всего помогает государственным органам контролировать сбор налогов и расходы, собирать статистику и выполнять другие функции, направленные на укрепления государства.

В бизнесе информационные технологии являются главным инструментом управления компанией, контроля за издержками, способом увеличения производительности труда и доходов. Они предоставляют средства анализа финансовой и производственной деятельности, оценки эффективности бизнеса, маркетинга, управления производством и взаимоотношениями с клиентами, хранения информации, обучения и контроля знаний, сбора и анализа различных статистических данных.

ООН разрабатывает проекты, которые позволяют ускорить экономический рост и подъем уровня жизни населения в разных странах с помощью информационных и коммуникационных технологий (ИКТ). Представители ряда государств, члены международного сообщества заявляют, что информационные и коммуникационные технологии становятся основой (базой) для создания глобальной экономики, основанной на знаниях. По их мнению, ИКТ способны внести важный вклад в ускорение экономического роста и обеспечение устойчивого развития различных стран, способствовать искоренению нищеты и эффективной интеграции государств в глобальную экономику.

Реализация этих проектов позволит эффективно развивать международное сотрудничество, совместно решать экономические, экологические, военные, правовые и иные проблемы, в том числе в образовании, туризме и культуре. В результате произойдет интеграция государственных и банковских, общественных структур, производств и других образований.

Ключевые элементы мировой экономики (финансовая глобализация, глобализация рынков товаров и услуг) опираются на информационные ресурсы, размещенные в глобальных сетях типа Интернета. Их интеграция осуществляется в результате широкого использования мировых телекоммуникаций, например космических средств связи (спутники связи).

Специалисты утверждают, что современный бизнес немыслим без его информационного обеспечения с помощью Интернета. При этом глобальная мировая экономика работает как единое целое в реальном масштабе времени, образуя единое мировое информационное пространство. Такая технология способствует резкому росту объемов информационных продуктов и услуг, использованию электронной торговли в бизнесе.

Использование информационно-коммуникационных технологий открывает широкие возможности для экономического роста и социального развития государств, но одновременно создает проблемы и риски, порождает углубление межгосударственного и внутригосударственного неравенства. В частности, речь идет о неравных возможностях людей создавать и использовать имеющиеся электронные информационные ресурсы, особенно в Интернете.

Если доступ к использованию этих технологий не будет расширен, то значительная часть населения развивающихся стран не получит пользы от них. Громадный потенциал информационно-коммуникационных технологий недостаточно применяется. Это привело к появлению «цифрового разрыва» («цифровой пропасти», «цифрового водораздела»), «виртуального барьера» на пути торговли. Такой барьер способен изолировать от рынков производителей, организации и государства, которые не имеют доступ к новым технологиям.

Жизненный цикл информации. Информационная сфера

Информация может существовать кратковременно (например, в памяти калькулятора в процессе проводимых на нем вычислений), в течение некоторого времени (например, при подготовке какой-либо справки) или очень долго (например, при хранении важных личных, коммерческих, общественных или государственных данных). Эти периоды времени определяют *жизненный цикл информации*, состоящий из следующих стадий: появление, существование и исчезновение («смерть»).

Поскольку информация имеет цену и является товаром, ее зачастую воспринимают как услугу, продукт или изделие. *Жизненный цикл изделия* затрагивает два основных его состояния.

Первое состояние связано с процессами его производства, осуществляемыми от момента подготовки проекта до выпуска конкретного изделия. В основе его лежит *концепция управления жизненным циклом изделия* (от англ.

«Product Lifecycle Management», PLM), объединяющая существующие разработки в единое интегрированное решение. Она затрагивает конструкторский, технологический, производственный этапы, завершением которых является коммерческий этап. Такое решение включает: систему управления инженерными данными (от англ. «Product Data Management», PDM), связывающую все компоненты и обеспечивающую взаимодействие с системами, предназначенными для управления ресурсами предприятия (ERP), взаимоотношениями с клиентами (CRM) и поставщиками (SCM). Данная концепция распространяется на предприятия как с дискретным, так и с непрерывным производством. Реализация PDM систем способствует повышению эффективности разработки изделий, снижению расходов и времени на проектирование, повышению качества и себестоимости выпускаемой продукции, сокращению ошибок и более простому учету требований клиентов. Однако предприятиям при этом приходится решать технологические, финансовые, организационные и психологические проблемы. При этом наибольшей из них является проблема внутренней неорганизованности на предприятии, когда отсутствует общая идеология и различные структуры пытаются решать свои локальные задачи, как правило, дешевыми программными и техническими средствами.

Второе состояние определяет период существования изделия с момента его выпуска, эксплуатации, когда изделие становится продуктом или услугой, и до окончания использования (утилизации).

Жизненный цикл свойственен большинству живых и неживых объектов, например человеку, животным или растениям. В информационных технологиях в этом случае говорят о жизненном цикле технических средств, компьютерных программ, сайта или портала, линии связи, соединяющей, например, провайдера интернет-услуг и его пользователя.

Развитие информационных технологий осуществляется за счет научно-технического прогресса (НТП), способствующего созданию новых средств производства, совершенствованию различных служб обслуживания и т. п. В результате создаются огромные распространяемые в обществе массивы (объемы) информации, которые формируют информационную среду (сферу).

Под **информационной сферой** понимают любую деятельность, направленную на:

- создание и распространение информации;
- формирование информационных ресурсов, подготовку и предоставление информационных продуктов и услуг;
- потребление информации.

Негативные последствия внедрения информационных технологий

Наравне с «цифровым разрывом» и «виртуальным барьером» изменения информационных технологий выполняемых работ часто могут оказывать негативное воздействие на людей (информационный шум и др.), участвующих в этих процессах, вызывая у них различные отрицательные реакции (информационный, психологический барьеры и др.).

Информационный шум означает, что в общем объеме полученных полезных данных есть посторонние сигналы (шумы). В ИПС (информационно-поисковая система) он свидетельствует о том, что в результате поиска по запросу пользователь получил не соответствующую его запросу (нерелевантную) информацию.

Информационный барьер – один из факторов, препятствующих получению нужной информации, затрудняющий использование документов как источников информации. Во многом он вызван законами развития потоков информации: постоянным ростом количества публикаций, рассеянием их в различных изданиях, старением публикаций и, наоборот, их актуализацией. Информационный барьер влияет на расслоение как информации, так и общества. Его появлению и углублению способствуют такие явления, как информационный шум, психологический барьер и др.

Психологический барьер обычно возникает, как защитная реакция человека на попытки изменить налаженную последовательность его действий. Он связан с необходимостью выполнять новые сложные виды работ, с перегрузками, появляющимися при поиске данных, их выборе в большом массиве полученных сведений и изучении отобранных материалов, составляющих порой несколько сотен и даже тысяч документов.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАНИЯ

1. Дайте определение и краткую характеристику понятию «информационные технологии».
2. Назовите основные этапы развития информационных технологий.
3. Дайте краткую характеристику каждому этапу развития информационных технологий.
4. Что означает термин «Платформа ИТ»?
5. Какова роль информационных технологий в развитии экономики и общества?
6. Дайте пояснение понятиям «Жизненный цикл информации» и «Информационная сфера».
7. Назовите негативные последствия внедрения информационных технологий.
8. Перечислите причины возникновения негативных последствий внедрения информационных технологий.

Понятие цифровой трансформации

Цифровая трансформация относится к изменениям, связанным с применением цифровых технологий во всех аспектах человеческого общества. Цифровую трансформацию можно рассматривать как третий этап использования цифровых технологий: цифровая компетенция → цифровая грамотность → цифровая трансформация.

Ключевым элементом любой **трансформации** являются *изменения*. Но если речь заходит о **цифровой трансформации**, то эти изменения совершенно нового рода. Это изменение самого подхода к ведению бизнеса.

Информационные технологии все глубже проникают в нашу жизнь: не только в рабочий контекст, но и в контекст отдыха. С каждым днем число устройств, подключенных к Интернету, становится все больше. По оценкам аналитиков уже в 2014 году число подключенных к Интернету устройств (Internet of Things) составит 2 млрд штук, а к 2020 году увеличится в 10 раз.

Цифровая трансформация определяется не тем, какие технологии используют компании для оказания своих услуг, а изменениями, которые происходят непосредственно с поведением потребителей и «средой их обитания».

Еще совсем недавно основой всего являлся *бренд* и предлагаемый рынку *товар* или *услуга*, которыми стремились обладать *потребители*. В настоящее время *потребитель* ставит приоритеты и выбирает, а *бренды* вынуждены вращаться вокруг него, чтобы привлечь к себе внимание.

Например, розничный банк. Долгие годы банки контролировали каналы взаимодействия с клиентами – клиент должен был приходить в офис для выполнения большинства операций. Банк стабильно и безопасно чувствовал себя, «окруженный стройным заборчиком». С ростом числа digital-каналов банки уже не могут контролировать все способы взаимодействия с клиентом – они лишь могут успевать реагировать на действия, выполняемые их клиентами, и всячески стараться их предугадать. Исчезли барьеры вокруг банка, и клиент помещается в центр всей стратегии дальнейшего развития.

Важно понимать, что диджитализация сервисов и услуг еще не является цифровой трансформацией.

Цифровая трансформация – это предоставление *новых* услуг *новым* способом. Необходимы новые сервисы в цифровом формате.

Для того чтобы быть конкурентноспособными, компаниям необходимо успевать предлагать услуги, максимально опирающиеся на интересы и намерения клиента в момент контакта с ним по любому из существующих каналов взаимодействия – необходимо научиться понимать клиента.

Цифровая трансформация открывает новые возможности для развития и перехода к digital-бизнесу. С ростом числа услуг в цифровом формате, digital-каналов взаимодействия с клиентом можно получить всю историю взаимоотношений с клиентом в «цифровом формате». Появляются данные, из которых компания может извлечь новые знания. И возникают новые модели работы, когда можно воспользоваться данными внешних организаций для обогащения своих собственных.

Монетизация данных и диджитализация формируют два новых подхода в развитии организации:

1. Bimodal IT

Поиск новых знаний – это инновационные процессы с использованием

agile-методологий и подходов, отличных от стандартных ИТ процессов в организациях. При этом текущие задачи и процессы, обеспечивающие жизнедеятельность организации, не остаются без внимания. Так, все в больших организациях появляется так называемое – Bimodal ИТ – разделения ИТ на две совершенно разные, часто независимые друг от друга, команды – традиционное ИТ, отвечающее за стабильность, и инновационное ИТ, характеризующее скорость и гибкость.

2. Open API

С появлением массивов информации, позволяющих тестировать различные идеи, неизбежно столкновение с проблемой нехватки либо собственных ресурсов, либо собственных идей. Появляется новая модель работы Data-As-a-Service. На российском и белорусском рынке примеров пока крайне мало, а в западных компаниях таких кейсов становится все больше – организации открывают доступ к своим данным и сервисам внешним командам и партнерам.

Цифровая трансформация является реальным способом повышения продуктивности. Так, правительство Великобритании приводит очень любопытные данные по своим цифровым проектам. В «Докладе о цифровой эффективности» говорится, что онлайн-операции могут быть в 20 раз дешевле телефонных, в 30 раз дешевле почтовых и в 50 раз менее затратными, чем выполненные «очно». При переходе полностью на цифровой формат госаппарат мог бы сэкономить от 1,7 до 1,8 млрд ф. ст. ежегодно. Однако речь идет не только о деньгах. Растут и ожидания людей. Им нужен быстрый и удобный доступ к госуслугам в удобное для них время и наиболее комфортным способом. Британское правительство стремится оправдать ожидания общества и намерено использует цифровые технологии для улучшения качества услуг и снижения бюрократических расходов.

Традиционный подход к ведению бизнеса не дает полностью использовать цифровые технологии. Инновации в сфере ИТ способны изменить способы взаимодействия с клиентами, оптимизировать рабочие процессы и улучшить бизнес-модель.

Компания «Financial Times» добилась значительного успеха с помощью цифрового преобразования бизнеса. Бумажная газета превратилась в *уникальный бренд с контентом высочайшего качества*, воспроизводимым на любых устройствах. Увеличился не только общий тираж Financial Times, включая печатную версию. Количество читателей, купивших онлайн-подписку, растет на 23 % ежегодно, что составляет более двух третей от общей аудитории. И все это происходит в сфере, где считается, что люди больше не будут платить за содержимое.

Многие предприятия используют потенциал аналитических систем, чтобы иметь более подробные представления о клиентах. Например, некоторые страховые компании совершенствуют свои продуктовые портфели и структуру затрат благодаря аналитическим выкладкам по андеррайтингу и

ценообразованию. Другие компании с помощью аналитики проводят экспериментальное изучение поведенческих моделей клиентов. Так, компания, владеющая ресторанами, исследует ценообразование и продвижение через франчайзинговую сеть. Эксперимент позволяет динамически регулировать цены на продукцию в зависимости от спроса, погодных условий, уровня запасов и близости времени закрытия заведений.

Лучшие компании – так называемая «цифроэлита» (Digirati) – сочетают активность в сфере цифровых технологий и сильное руководство, совершая переход от просто использования ИТ к трансформации бизнеса. Это то, что называют «цифровая зрелость». Компании по этому признаку существенно различаются, и те, которые являются более зрелыми с точки зрения цифровых технологий, превосходят своих конкурентов в бизнесе.

В Слоуновском центре цифрового бизнеса при Массачусетском технологическом институте исследователь Джордж Вестерман совместно с Cargemini разработал модель **цифровой зрелости** компаний и связал корпоративные финансовые показатели с уровнем цифровой зрелости [7].

Цифровая зрелость представлена как сочетание двух отдельных, но взаимосвязанных параметров.

Первый – цифровая активность. К цифровой активности относятся вложения в технологические проекты, которые должны менять методы работы компании: способы взаимодействия с клиентами, внутренние процессы и бизнес-модели. Компании в самых разных отраслях инвестируют в интересные им цифровые проекты. Однако во многих организациях эти инвестиции не скоординированы, а иногда дублируют друг друга.

Второй параметр – активность управления трансформацией. Этот параметр связан с созданием управленческих возможностей, необходимых для стимуляции изменений. Сюда можно отнести наличие видения желаемого результата, моделей управления и вовлечения пользователей для реализации заданного курса, отлаженных процессов взаимодействия ИТ и бизнес-подразделений для внедрения технологических изменений. Отдельные элементы трансформационной активности работают сообща – при помощи управления, осуществляемого сверху вниз, и инноваций, идущих во встречном направлении. Так поддерживается режим непрерывной трансформации.

Бывает, что эти элементы работают чересчур медленно или слишком консервативны по своей сути. В итоге они не позволяют компании инвестировать достаточное количество средств и ресурсов в необходимые инновации.

Два параметра (два измерения) в итоге дают нам четыре типа цифровой зрелости (рис. 3).



Рис. 3. Типы цифровой зрелости

В левом нижнем квадранте графика – так называемые «Новички» цифрового мира. Новички – компании, в которых не используют продвинутые цифровые технологии. В то же время они могут активно применять более традиционные инструменты – ERP и электронную коммерцию. Компании могут оставаться новичками вполне осознанно и добровольно, однако чаще всего оказываются в этом квадранте случайно. Сюда их приводят недостаточные знания о существующих возможностях или весьма скромные инвестиции в новые технологии при отсутствии эффективного управления трансформацией.

В левом верхнем квадранте расположились «Модники». Модники уже внедряли или экспериментировали с трендовыми цифровыми приложениями. Какие-то проекты приносили определенную пользу, другие – давали нулевой результат. Возможно вместе эти проекты смотрятся довольно неплохо, однако внедрялись они без какой-либо концепции, позволяющей получить синергетический эффект. У модников есть мотивация к привнесению изменений на основе цифровых технологий. Как правило, их проблемы связаны с отсутствием понимания, как максимизировать бизнес-выгоды с помощью продуманной стратегии цифровой трансформации. В квадрант модников могут попасть организации, у которых недостаточно проработана на корпоративном уровне система управления изменениями. При этом в отдельных подразделениях конкретные цифровые проекты могут находиться на более серьезных стадиях зрелости.

Правый нижний квадрант – угол «Консерваторов». Они ценят осмотрительность превыше любых инноваций. Консерваторы осознают необходимость создания единой концепции, системы управления и корпоративной культуры как средств для получения весомой отдачи от

инвестиций в современные разработки. Тем не менее они скептически относятся к новым трендам в области цифровых технологий. Иногда себе же во вред. Консерваторы стремятся тратить разумно, но излишняя осторожность заставляет отказываться от возникающих возможностей, которыми с удовольствием пользуются конкуренты.

Правый верхний квадрант – это «Элита» цифрового мира. Элита действительно понимает, как и какие выгоды можно извлечь из цифровой трансформации. В их случае четкое видение трансформации сочетается с продуманной системой управления, вовлечением пользователей и серьезными вложениями в открывающиеся возможности. Следуя своему видению, искусно пользуясь механизмами вовлечения, элита формирует особую цифровую культуру, которая определяет дальнейшие изменения и способствует их успешному воплощению. Инвестируя в новые технологии и аккуратно координируя цифровые проекты, элита постоянно наращивает свое преимущество в области цифровой трансформации.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАНИЯ

1. Дайте определение понятию «Цифровая трансформация».
2. Какова роль цифровой трансформации в развитии различных сфер жизни общества?
3. Дайте характеристику двум новым подходам в развитии организаций (bimodal IT и open API).
4. Что подразумевает параметр «цифровая активность»?
5. Что подразумевает параметр «активность управления трансформацией»?
6. Дайте характеристику четырем типам цифровой зрелости.

Интернет вещей (Internet of Things – IOT)

Интернет вещей – концепция вычислительной сети физических объектов («вещей»), оснащенных встроенными технологиями для взаимодействия друг с другом или с внешней средой, рассматривающая организацию таких сетей, как явление, способное перестроить экономические и общественные процессы, исключаящее из части действий и операций необходимость участия человека.

Концепция сформулирована в 1999 году как осмысление перспектив широкого применения средств радиочастотной идентификации для взаимодействия физических объектов между собой и с внешним окружением. Наполнение концепции Интернета вещей многообразным технологическим содержанием и внедрение практических решений для ее реализации начиная с 2010-х годов считается восходящим трендом в информационных технологиях, прежде всего благодаря повсеместному распространению беспроводных сетей, появлению облачных вычислений, развитию технологий межмашинного взаимодействия, началу активного перехода на IPv6 и освоению программно-конфигурируемых сетей.

Исторические факты

В 1926 году Никола Тесла в интервью для журнала «Collier's» сказал, что в будущем радио будет преобразовано в «большой мозг», все вещи станут частью единого целого, а инструменты, благодаря которым это станет возможным, будут легко помещаться в кармане.

В 1990 году выпускник MIT, один из отцов протокола TCP/IP, Джон Ромки создал первую в мире интернет-вещь. Он подключил к сети свой тостер.

Сам термин «Интернет вещей» (Internet of Things) был предложен Кевинном Эштоном в 1999 году. В этом же году был создан Центр автоматической идентификации (Auto-ID Center), занимающийся радиочастотной идентификацией (RFID) и сенсорными технологиями, благодаря которому эта концепция и получила широкое распространение.

В 2008–2009 годах произошел переход от Интернета людей к Интернету вещей, т. е. количество подключенных к сети предметов превысило количество людей.

Определение понятия

Определений Интернета вещей очень много. Для примера можно привести определение профессора Роба Ван Краненбурга: *IOT – концепция пространства, в котором все из аналогового и цифрового миров может быть совмещено – это переопределяет наши отношения с объектами, а также свойства и суть самих объектов.*

Интернет вещей – это не просто множество различных приборов и датчиков, объединенных между собой проводными и беспроводными каналами связи и подключенных к сети Интернет, это еще и более тесная интеграция реального и виртуального миров, в котором общение производится между людьми и устройствами.

Предполагается, что в будущем «вещи» станут активными участниками бизнеса, информационных и социальных процессов, где они смогут взаимодействовать и общаться между собой, обмениваясь информацией об окружающей среде, реагируя и влияя на процессы, происходящие в окружающем мире, без вмешательства человека.

По мнению Роба Ван Краненбурга, Интернет вещей представляет из себя «четырёхслойный пирог»:

- 1-й уровень связан с идентификацией каждого объекта;
- 2-й уровень предоставляет сервис по обслуживанию потребностей потребителя (можно рассматривать как сеть собственных «вещей», частный пример – «умный дом»);
- 3-й уровень связан с урбанизацией городской жизни (это концепция «умного города», где вся информация, которая касается жителей этого города, стягивается в конкретный жилой квартал, в ваш дом и соседние дома);
- 4-й уровень – сенсорная планета.

Интернет вещей можно рассматривать как сеть сетей, в которой небольшие малосвязанные сети образуют более крупные.

Понимание «вещей» в концепции ИВ

Каждая «вещь», будь то «умный» холодильник или подключенный к Интернету интеллектуальный индикатор, имеет три основных атрибута, которые будут говорить о ее принадлежности к ИВ:

– **сенсоры:** ИВ-устройства и системы имеют сенсоры, отслеживающие и измеряющие активность в реальном мире;

– **подключение:** интернет-подключение либо доступно самому устройству, либо осуществляется посредством хаба, смартфона или базовой станции, в то время как вся ИВ-система обычно подключена к Интернету или «облаку»;

– **процессоры:** как и всякое вычислительное устройство, ИВ-устройство будет обладать некоторой вычислительной мощностью, позволяющей анализировать входящую информацию и передавать ее.

Примеры практического применения концепции ИВ

Подключенная реклама и маркетинг. Cisco полагает, что эта категория (подразумеваются подключенные к Интернету рекламные щиты) будет одной из лучших категорий ИВ, наряду с умными фабриками и системами поддержки телекоммуникационных услуг.

Интеллектуальные системы организации дорожного движения. Исследование Machina в газете, подготовленной для Ассоциации GSM, прогнозирует к 2020 году доход в 100 млрд дол. США благодаря применению ИВ в системе дорожных выплат и штрафов. Также источником дохода будет смарт-управление местами для парковки, которое, как ожидают, принесет около 30 млрд дол. США.

Системы утилизации отходов. Одним из новшеств стало оснащение баков мусора и урн RFID-датчиками, которые позволяют увидеть, когда мусор был выброшен. Опираясь на это, город Кливленд смог исключить 10 маршрутов погрузки и сократить эксплуатационные расходы на 13 %. В Цинциннати объем бытовых отходов упал на 17 %, а объем переработки вырос на 49 % посредством использования системы «платите, если вы мусорите» – программа, которая штрафует тех, кто превышает заданный лимит отходов.

Смарт-электросети, которые могут отслеживать и координировать уровень потребления энергии. Они обеспечат экономию от 200 до 500 млрд дол. США в год к 2025 году согласно McKinsey Global Institute.

Смарт-аквасистемы и счетчики. Города Доха, Сан-Паулу и Пекин уменьшили утечки на 40–50 %, помещая датчики в насосы и другую водную инфраструктуру.

В качестве примера использования концепции ИВ в сегменте потребительской электроники, несомненно, стоит выделить **Google Glass:** устройство, снабженное прозрачным дисплеем над правым глазом (в оправе наподобие очков), способное снимать фото- и видеоматериалы, выводить различную информацию в формате дополненной реальности и т. п.

Также в последнее время стали очень популярными **фитнес-трекеры**, такие как Jawbone UP, Fitbit, Nike + Fuelband и т. п., позволяющие измерять физические нагрузки пользователя, время его сна и следить за рационом питания.

Не стоит забывать о различных **термостатах, датчиках** дыма и освещенности, которые в совокупности с другими технологическими решениями могут объединяться в так называемый «умный дом».

Проблемы и недостатки

Самой главной проблемой на сегодняшний день является отсутствие стандартов в данной области, что затрудняет возможность интеграции предлагаемых на рынке решений и во многом сдерживает появление новых. Также для полноценного функционирования такой сети необходима автономность всех «вещей», т. е. датчики должны научиться получать энергию из окружающей среды, а не работать от батареек, как это происходит сейчас.

Наличие огромной сети, контролирующей весь окружающий мир, глобальная открытость данных и прочие особенности могут иметь и негативные последствия. Каждый сам может составить себе список возможных угроз и проблем, которые несет в себе эта технология.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАНИЯ

1. Как появилось понятие «Интернет вещей»?
2. Дайте определение понятию «Интернет вещей».
3. Назовите четыре уровня развития Интернета вещей.
4. Что представляет собой «вещь» в рамках концепции Интернета вещей?
5. Приведите примеры практического применения концепции Интернета вещей.
6. Какие проблемы и опасности может создать Интернет вещей в жизни человека?
7. Каковы перспективы развития Интернета вещей?
8. Где Интернет вещей может принести максимум пользы для человека?

Большие данные (Big Data)

К категории **Большие данные (Big Data)** относится информация, которую уже невозможно обрабатывать традиционными способами, в том числе структурированные данные, медиа и случайные объекты. Для работы с ними на смену традиционным монолитным системам должны прийти новые массивно-параллельные решения, которые состоят из нескольких независимых блоков, число которых можно неограниченно увеличивать и отказ одного из которых не приводит к разрушению всей системы.

Из названия можно предположить, что термин «большие данные» относится просто к управлению и анализу больших объемов данных. Согласно

отчету McKinsey Institute «Большие данные – новый рубеж для инноваций, конкуренции и производительности». Термин «Большие данные» относится к наборам данных, размер которых превосходит возможности типичных баз данных (БД) по занесению, хранению, управлению и анализу информации. И мировые репозитории данных, безусловно, продолжают расти. В представленном в середине 2011 года отчете аналитической компании IDC «Исследование цифровой вселенной» (Digital Universe Study), подготовку которого спонсировала компания EMC, предсказывалось, что общий мировой объем созданных и реплицированных данных в 2011 году может составить около 1,8 Збайта (1,8 трлн Гбайт) – примерно в 9 раз больше того, что было создано в 2006 году.

«Большие данные» предполагают нечто большее, чем просто анализ огромных объемов информации. Проблема не в том, что организации создают огромные объемы данных, а в том, что большая их часть представлена в формате, плохо соответствующем традиционному структурированному формату БД, – это веб-журналы, видеозаписи, текстовые документы, машинный код или, например, геопространственные данные. Все это хранится во множестве разнообразных хранилищ, иногда даже за пределами организации. В результате корпорации могут иметь доступ к огромному объему своих данных и не иметь необходимых инструментов, чтобы установить взаимосвязи между этими данными и сделать на их основе значимые выводы. К тому же данные сейчас обновляются все чаще и чаще, и традиционные методы анализа информации не могут угнаться за огромными объемами постоянно обновляемых данных, что в итоге и открывает дорогу технологиям больших данных.

Понятие больших данных подразумевает работу с информацией огромного объема и разнообразного состава, весьма часто обновляемой и находящейся в разных источниках в целях увеличения эффективности работы, создания новых продуктов и повышения конкурентоспособности. Консалтинговая компания Forrester дает краткую формулировку: «Большие данные объединяют техники и технологии, которые извлекают смысл из данных на экстремальном пределе практичности».

Сфера больших данных характеризуется пятью V:

– **Volume** – объем. Накопленная база данных представляет собой большой объем информации, который трудоемко обрабатывать и хранить традиционными способами, для них требуются новый подход и усовершенствованные инструменты.

– **Velocity** – скорость. Данный признак указывает как на увеличивающуюся скорость накопления данных (90 % информации было собрано за последние два года), так и на скорость обработки данных. В последнее время стали более востребованы технологии обработки данных в реальном времени.

– **Variety** – многообразие, т. е. возможность одновременной обработки структурированной и неструктурированной разноформатной информации. Главное отличие структурированной информации – это то, что она может быть

классифицирована. Примером такой информации может служить информация о клиентских транзакциях. Неструктурированная информация включает в себя видео-, аудиофайлы, свободный текст, информацию, поступающую из социальных сетей. На сегодняшний день 80 % информации входит в группу неструктурированной. Данная информация нуждается в комплексном анализе, чтобы сделать ее полезной для дальнейшей обработки.

– **Veracity** – достоверность данных. Все большее значение пользователи стали придавать достоверности имеющихся данных. Так, у интернет-компаний есть проблема по разделению действий, проводимых роботом и человеком на сайте компании, что приводит в конечном счете к затруднению анализа данных.

– **Value** – ценность накопленной информации. Большие данные должны быть полезны компании и приносить определенную ценность для нее. К примеру, помогать в усовершенствовании бизнес-процессов, составлении отчетности или оптимизации расходов. Если накопленные объемы данных характеризуются указанными выше пятью признаками, то их можно относить к числу больших.

Источники больших данных

В качестве источников больших данных могут выступать непрерывно поступающие данные с измерительных устройств, события от радиочастотных идентификаторов, потоки сообщений из социальных сетей, метеорологические данные, данные дистанционного зондирования земли, потоки данных о местонахождении абонентов сетей сотовой связи, устройств аудио- и видеорегистрации. Массовое распространение перечисленных выше технологий и принципиально новых моделей использования различного рода устройств и интернет-сервисов послужило отправной точкой для проникновения больших данных едва ли не во все сферы деятельности человека. В первую очередь в научно-исследовательскую деятельность, коммерческий сектор и государственное управление.

В ходе исследования (весна 2013 года) под названием Cisco Connected World Technology Report, проведенного в 18 странах независимой аналитической компанией InsightExpress, были опрошены 1 800 студентов колледжей и такое же количество молодых специалистов в возрасте от 18 до 30 лет. Опрос проводился, чтобы выяснить уровень готовности ИТ-отделов к реализации проектов Big Data и получить представление о связанных с этим проблемах, технологических изъянах и стратегической ценности таких проектов.

Большинство компаний собирает, записывает и анализирует данные. Тем не менее, говорится в отчете, многие компании в связи с Big Data сталкиваются с целым рядом сложных деловых и информационно-технологических проблем. К примеру, 60 % опрошенных признают, что решения Big Data могут усовершенствовать процессы принятия решений и повысить конкурентоспособность, но лишь 28 % заявили о том, что уже получают реальные стратегические преимущества от накопленной информации.

Сферы применения больших данных

Сфера использования технологий больших данных разнообразна и широка. Можно узнать о предпочтениях клиентов, об эффективности маркетинговых кампаний или провести анализ рисков. Согласно результатам опроса, проведенного IBM Institute, о направлениях использования Big Data в компаниях 53 % используют большие данные в сфере клиентского сервиса, 40 % – в сфере операционной эффективности, 7 % – в сфере риск-менеджмента.

Big Data являются одной из самых быстрорастущих сфер информационных технологий. Общий объем получаемых и хранимых данных удваивается каждые 1,2 года. За период с 2012 по 2014 год количество данных, ежемесячно передаваемых мобильными сетями, выросло на 81 %. По оценкам Cisco, в 2014 году объем мобильного трафика составил 2,5 Эбайта (единица измерения количества информации, равная 10^{18} стандартным байтам) в месяц, а уже в 2019 году он будет равен 24,3 Эбайтам.

Применение больших данных в различных отраслях

Большие данные используют в здравоохранении, телекоммуникациях, торговле, логистике, финансовых компаниях, а также в государственном управлении. Ниже приводится несколько примеров.

Розничная торговля. В базах данных розничных магазинов может быть накоплено множество информации о клиентах, системе управления запасами, поставками товарной продукции. Данная информация может быть полезна во всех сферах деятельности магазинов. С помощью накопленной информации можно управлять поставками товара, его хранением и продажей. На основании накопленной информации можно прогнозировать спрос и поставки товара. Также система обработки и анализа данных может решить и другие проблемы ритейлера, например оптимизировать затраты или подготовить отчетность.

Финансовые услуги. Большие данные дают возможность проанализировать кредитоспособность заемщика, также они полезны для кредитного скоринга и андеррайтинга. Внедрение технологий больших данных позволит сократить время рассмотрения кредитных заявок. С помощью больших данных можно проанализировать операции конкретного клиента и предложить подходящие именно ему банковские услуги.

Телекоммуникационная отрасль. Операторы сотовой связи наравне с финансовыми организациями имеют одни из самых объемных баз данных, что позволяет им проводить наиболее глубокий анализ накопленной информации. Главной целью анализа данных является удержание существующих клиентов и привлечение новых. Для этого компании проводят сегментацию клиентов, анализируют их трафики, определяют социальную принадлежность абонента.

Помимо использования Big Data в маркетинговых целях, технологии применяются для предотвращения мошеннических финансовых операций.

Горнодобывающая и нефтяная промышленности. Большие данные используются как при добыче полезных ископаемых, так и при их переработке и сбыте. Предприятия могут на основании поступившей информации делать

выводы об эффективности разработки месторождения, отслеживать график капитального ремонта и состояния оборудования, прогнозировать спрос на продукцию и цены.

По данным опроса Tech Pro Research, наибольшее распространение большие данные получили в телекоммуникационной отрасли, а также в инжиниринге, ИТ, на финансовых и государственных предприятиях. По результатам данного опроса менее популярны большие данные в образовании и здравоохранении.

Компании, использующие большие данные (несколько примеров)

Сегодня во многих зарубежных компаниях активно внедряются большие данные. Уже используют ресурсы Big Data такие известные компании, как Nasdaq, Facebook, Google, IBM, VISA, Master Card, Bank of America, HSBC, AT&T, Coca Cola, Starbucks и Netflix.

HSBC использует технологии больших данных для противодействия мошенническим операциям с пластиковыми картами. С помощью Big Data компания увеличила эффективность службы безопасности в 3 раза, распознавание мошеннических инцидентов – в 10 раз. Экономический эффект от внедрения данных технологий превысил 10 млн дол. США.

Антифрод **VISA** позволяет в автоматическом режиме вычислить операции мошеннического характера, система на данный момент помогает предотвратить мошеннические платежи на сумму 2 млрд дол. США ежегодно.

Суперкомпьютер Watson компании **IBM** анализирует в реальном времени поток данных по денежным транзакциям. По данным IBM, Watson на 15 % увеличил количество выявленных мошеннических операций, на 50 % сократил ложные срабатывания системы и на 60 % увеличил сумму денежных средств, защищенных от транзакций такого характера.

Procter & Gamble с помощью больших данных проектируют новые продукты и составляют глобальные маркетинговые кампании. P&G создал специализированные офисы Business Spheres, где можно просматривать информацию в реальном времени. Таким образом, у менеджмента компании есть возможность мгновенно проверять гипотезы и проводить эксперименты. По мнению P&G, большие данные помогают в прогнозировании деятельности компании.

По мнению **Caterpillar**, ее дистрибьюторы ежегодно выпускают от 9 до 18 млрд дол. США прибыли только из-за того, что не внедряют технологии обработки больших данных. Big Data позволили бы клиентам более эффективно управлять парком машин за счет анализа информации, поступающей с датчиков, установленных на машинах. Уже существует возможность анализировать состояние ключевых узлов, их степени износа, управлять затратами на топливо и техническое обслуживание.

Luxottica group является производителем спортивных очков таких марок, как Ray-Ban, Persol и Oakley. Технологии больших данных компания применяет для анализа поведения потенциальных клиентов и «умного» смс-маркетинга. В

результате Big Data Luxottica group выделила более 100 млн наиболее ценных клиентов и повысила эффективность маркетинговой кампании на 10 %.

С помощью Yandex Data Factory разработчики игры **World of Tanks** анализируют поведение игроков. Технологии больших данных позволили проанализировать поведение 100 тыс. игроков World of Tanks с использованием более 100 параметров (информация о покупках, играх, опыт и др.). В результате анализа был получен прогноз оттока пользователей. Данная информация позволяет уменьшить уход пользователей и работать с участниками игры адресно. Разработанная модель оказалась на 20–30 % эффективнее стандартных инструментов анализа игровой индустрии.

Министерство труда Германии использует большие данные в работе, связанной с анализом поступающих заявок на выдачу пособий по безработице. Так, проанализировав информацию, стало понятно, что 20 % пособий выплачивалось незаслуженно. С помощью Big Data министерство труда сократило расходы на 10 млрд евро.

Детская больница Торонто внедрила проект Project Artemis. Это информационная система, которая собирает и анализирует данные по младенцам в реальном времени. Система ежесекундно отслеживает 1260 показателей состояния каждого ребенка. Project Artemis позволяет прогнозировать нестабильное состояние ребенка и начать профилактику заболеваний у детей.

Методики анализа больших данных

Существует множество разнообразных методик анализа массивов данных, в основе которых лежит инструментарий, заимствованный из статистики и информатики (например, машинное обучение). Список не является полным, однако в нем отражены наиболее востребованные в различных отраслях подходы. При этом следует понимать, что исследователи продолжают работать над созданием новых методик и совершенствованием существующих. Кроме того, некоторые из перечисленных методик вовсе не обязательно применимы исключительно к большим данным и могут с успехом использоваться для меньших по объему массивов (например, A/B-тестирование, регрессионный анализ). Безусловно, чем более объемный и диверсифицируемый массив подвергается анализу, тем более точные и релевантные данные удастся получить на выходе.

A/B testing – методика, в которой контрольная выборка поочередно сравнивается с другими. Тем самым удается выявить оптимальную комбинацию показателей для достижения, например, наилучшей ответной реакции потребителей на маркетинговое предложение. Большие данные позволяют провести огромное количество итераций и таким образом получить статистически достоверный результат.

Association rule learning – набор методик для выявления взаимосвязей, т. е. ассоциативных правил, между переменными величинами в больших массивах данных. Используется в **data mining**.

Classification – набор методик, которые позволяют предсказать поведение потребителей в определенном сегменте рынка (принятие решений о покупке, отток, объем потребления и др.). Используется в **data mining**.

Cluster analysis – статистический метод классификации объектов по группам за счет выявления заранее не известных общих признаков. Используется в **data mining**.

Crowdsourcing – методика сбора данных из большого количества источников.

Data fusion and data integration – набор методик, которые позволяют анализировать комментарии пользователей социальных сетей и сопоставлять с результатами продаж в режиме реального времени.

Data mining – набор методик, который позволяет определить наиболее восприимчивые для продвигаемого продукта или услуги категории потребителей, выявить особенности наиболее успешных работников, предсказать поведенческую модель потребителей.

Ensemble learning. В этом методе задействуется множество предикативных моделей за счет чего повышается качество сделанных прогнозов.

Genetic algorithms. В этой методике возможные решения представляют в виде «хромосом», которые могут комбинироваться и мутировать. Как и в процессе естественной эволюции, выживает наиболее приспособленная особь.

Machine learning – направление в информатике (исторически за ним закрепилось название «искусственный интеллект»), которое преследует цель создания алгоритмов самообучения на основе анализа эмпирических данных.

Natural language processing (NLP) – набор заимствованных из информатики и лингвистики методик распознавания естественного языка человека.

Network analysis – набор методик анализа связей между узлами в сетях. Применительно к социальным сетям позволяет анализировать взаимосвязи между отдельными пользователями, компаниями, сообществами и т. п.

Optimization – набор численных методов для редизайна сложных систем и процессов для улучшения одного или нескольких показателей. Помогает в принятии стратегических решений, например состава выводимой на рынок продуктовой линейки, проведении инвестиционного анализа и др.

Pattern recognition – набор методик с элементами самообучения для предсказания поведенческой модели потребителей.

Predictive modeling – набор методик, которые позволяют создать математическую модель перед заданным вероятным сценарием развития событий. Например, анализ базы данных CRM-системы на предмет возможных условий, которые подтолкнут абоненты сменить провайдера.

Regression – набор статистических методов для выявления закономерности между изменением зависимой переменной и одной или несколькими независимыми. Часто применяется для прогнозирования и предсказаний. Используется в **data mining**.

Sentiment analysis. В основе методик оценки настроений потребителей лежат технологии распознавания естественного языка человека. Они позволяют вычлениить из общего информационного потока сообщения, связанные с интересующим предметом (например, потребительским продуктом). Далее оценить полярность суждения (позитивное или негативное), степень эмоциональности и др.

Signal processing – заимствованный из радиотехники набор методик, который преследует цель распознавания сигнала на фоне шума и его дальнейшего анализа.

Spatial analysis – набор отчасти заимствованных из статистики методик анализа пространственных данных – топологии местности, географических координат, геометрии объектов. Источником больших данных в этом случае часто выступают геоинформационные системы (ГИС).

Statistics – наука о сборе, организации и интерпретации данных, включая разработку опросников и проведение экспериментов. Статистические методы часто применяются для оценочных суждений о взаимосвязях между теми или иными событиями.

Supervised learning – набор основанных на технологиях машинного обучения методик, которые позволяют выявить функциональные взаимосвязи в анализируемых массивах данных.

Simulation. Моделирование поведения сложных систем часто используется для прогнозирования, предсказания и проработки различных сценариев при планировании.

Time series analysis – набор заимствованных из статистики и цифровой обработки сигналов методов анализа повторяющихся с течением времени последовательностей данных. Одно из очевидных применений – отслеживание рынка ценных бумаг или заболеваемости пациентов.

Unsupervised learning – набор основанных на технологиях машинного обучения методик, которые позволяют выявить скрытые функциональные взаимосвязи в анализируемых массивах данных. Имеет общие черты с **Cluster Analysis**.

Visualization – методы графического представления результатов анализа больших данных в виде диаграмм или анимированных изображений для упрощения интерпретации облегчения понимания полученных результатов.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАНИЯ

1. Дайте определение понятию «большие данные».
2. Назовите пять признаков, на основании которых объемы данных можно назвать большими данными.
3. Дайте подробную характеристику каждому из пяти признаков больших данных.
4. Назовите основные источники больших данных.
5. Назовите сферы наиболее широкого применения больших данных.

6. В каких отраслях и как используются большие данные?
7. Проиллюстрируйте ответ на вопрос №6 конкретными примерами (компаниями).
8. Назовите и дайте краткую характеристику основным методикам работы с большими данными.

Библиотека БГУИР

ЛИТЕРАТУРА

Основная

1. Гайдышев, И. Анализ и обработка данных. Специальный справочник / И. Гайдышев. – СПб. : Питер, 2001.
2. О'Кифф, Джон. Нешаблонное мышление / Джон О'Кифф. – М. : Манн, Иванов и Фербер, 2013.
3. Осипович, В. С. Сборник задач к практическим работам по освоению прикладного программного обеспечения: пособие для студентов / В. С. Осипович, О. С. Потоцкая, К. Д. Яшин. – Минск, 2012.
4. Industrial Internet Insights Report for 2015. – Accenture, 2014.
5. Ten IT-enabled business trends for the decade ahead. Updated research. – McKinsey Global Institute, 2013.
6. The Internet of things. Business index. A quiet revolution gathers pace. – A report from The Economist Intelligence Unit, 2013.
7. The Digital Advantage: How digital leaders outperform their peers in every industry. MIT Center for Digital Business, 2014

Дополнительная

1. Гансвиндт, Т. Инновации – будущее информационного общества / Т. Гансвиндт, А. Гоголь. – М. : Энергомашиностроение, 2005.
2. Загоруйко, Н. Г. Прикладные методы анализа данных и знаний / Н. Г. Загоруйко. – Новосибирск : ИМ СО РАН, 1999.
3. Загоруйко, Н. Г. Когнитивный анализ данных / Н. Г. Загоруйко. – Новосибирск : Академическое издательство «Гео», 2012.
4. Кристенсен, К. М. Дилемма инноватора. Как из-за новых технологий погибают сильные компании / К. М. Кристенсен. – М. : Альпина Бизнес Букс, 2004.
5. Солсо, Р. Когнитивная психология / Р. Солсо. – СПб. : Питер, 2006.
6. Саати, Т. Аналитическое планирование. Организация систем / Т. Саати, К. Кернс. – М. : Радио и связь, 1991.
7. Саати, Т. Принятие решений при зависимостях и обратных связях: аналитические сети / Т. Саати – М. : Издательство ЛКИ, 2008.
8. Талеб, Нассим Николас. Одураченные случайностью / Нассим Николас Талеб. – М. : Манн, Иванов и Фербер, 2012.
9. Татарова, Г. Г. Методология анализа данных в социологии (введение) : учебник для вузов / Г. Г. Татарова. – М. : NOTA BENE, 1999.
10. Шей, Тони. Доставляя счастье / Тони Шей. – М. : Манн, Иванов и Фербер, 2013.

Учебное издание

Карпович Екатерина Борисовна
Пархоменко Дарья Александровна

**ОСНОВЫ
ИНФОРМАЦИОННО-АНАЛИТИЧЕСКОЙ
ДЕЯТЕЛЬНОСТИ**

ПОСОБИЕ

Редактор *Е. И. Герман*
Корректор *Е. Н. Батурчик*
Компьютерная правка, оригинал-макет *В. М. Задоя*

Подписано в печать 13.02.2017. Формат 60x84 1/16. Бумага офсетная. Гарнитура «Таймс».
Отпечатано на ризографе. Усл. печ. л. 3,37. Уч.-изд. л. 4,0. Тираж 100 экз. Заказ 25.

Издатель и полиграфическое исполнение: учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий №1/238 от 24.03.2014,
№2/113 от 07.04.2014, №3/615 от 07.04.2014.
ЛП №02330/264 от 14.04.2014.
220013, Минск, П. Бровки, 6