

ОБЩИЕ ВОПРОСЫ GENERAL QUESTIONS

УДК 004.9

П. Ю. Богатырев, аспирант, e-mail: PFight77@gmail.com,

А. В. Коськин, д-р техн. наук, проф., e-mail: kav1959@rambler.ru,

Федеральное государственное бюджетное образовательное учреждение высшего образования
Орловский государственный университет имени И. С. Тургенева, г. Орел

Структура информационного пространства организации с применением теоретико-множественного подхода к систематизации информации

Рассматривается способ построения системы управления информационными ресурсами организации с точки зрения структурирования информации. Вводится обобщенная модель систематизации, интегрирующая подходы, используемые в современных системах управления содержанием, на основе теоретико-множественного подхода.

Ключевые слова: информация, управление знаниями, управление содержанием, модель, систематизация, теория множеств, структурирование, иерархия, метки

Введение

В процессе управления знаниями в организации накапливается значительное количество информации в различных представлениях и формах. Под знаниями здесь понимается "любое слово, факт, пример, событие, правило, гипотеза или модель, которые усиливают понимание или исполнение в определенной области деятельности или дисциплины" [1], а также "все виды информации, включая неструктурированный контент (письма, эскизы, фото) и данные (в базах данных и хранилищах данных)" [2]. Создание общедоступных баз данных, способных накапливать и предоставлять доступ к данным из различных географических точек, в настоящее время является задачей, имеющей много готовых технических решений [1–3]. Однако систематизация всего объема накопленных сведений для обеспечения фактической доступности с учетом временных затрат на поиск нужной информации является на настоящий момент слабо разработанной научной проблемой.

Постановка задачи

Организацию можно рассматривать как единый организм, в котором все сотрудники работают над достижением общих целей. Базу знаний организации в этом случае можно представить как память этого единого существа, аналогичную памяти человека. Проводя аналогию с человеческой памятью, мы можем сформулировать требования к системе управления базой знаний:

- размещение информации в базе знаний должно быть простым, эффективным и быстрым;

- информация в базе должна храниться в таком виде, который обеспечивает эффективную работу с ней;
- получение необходимых сведений должно быть таким же простым, как припоминание ранее известного факта человеком.

Очевидно, что построить систему, которая в полной мере могла бы сравниться с человеческой памятью, затруднительно, однако к этому следует стремиться. Чем более эффективными и простыми будут процессы размещения и получения информации в данной системе, тем ближе мы будем к идеалу. В практическом же смысле приближение к идеалу памяти будет означать повышение эффективности организации за счет накопления и использования опыта сотрудников, обмена информацией между участниками процесса и, как следствие, повышение конкурентоспособности, прибыльности и социальной значимости всего предприятия.

Архитектура информационной системы

Одним из основных направлений в данной области является разработка систем искусственного интеллекта, способных осуществлять высокорелевантный поиск по базе знаний за счет семантического анализа запроса и самих знаний [2, 3]. Такой подход, хотя и не лишен преимуществ, обладает существенным недостатком. Он подразумевает, что в процессе накопления данные слабо структурируются, образуя большие массивы малодоступной информации. И только в конце жизненной цепочки, на этапе доступа к собранным знаниям осуществляется некоторый анализ и обработка информации. На наш взгляд, для достижения наиболь-

шей эффективности систематизация должна осуществляться на всех этапах работы с информацией, в том числе уже на этапе ее сбора. Полученная таким образом база знаний будет доступна не только автоматическим системам с искусственным интеллектом, но в ней также сможет найти нужные сведения сам человек, используя те сведения, которые у него есть в памяти для поиска нужных данных. Ко всему прочему, задача построения такой базы знаний видится более просто реализуемой с технической точки зрения, чем построение сложных систем, способных в полной мере выявлять семантическое и прагматическое содержание текста.

Если учитывать многообразие информационного пространства организации, то создание единой системы, которая хранила бы внутри себя все возможные виды информационных объектов, видится непрактичным. В любой организации есть набор устоявшихся инструментов для выполнения тех или иных задач. Например, почта отправляется через почтовый клиент, файлы хранятся на FTP-серверах, различного рода статьи — на внутреннем интранет-портале, задачи — в некотором менеджере задач (например, Actionspace) и т.д. Все эти решения являются узкоспециализированными и в наилучшей степени подходят для решения поставленных перед ними конкретных задач. Однако, каждое из этих решений, как правило, существует отдельно от остальных. Каждое из них образует небольшую базу знаний, содержащую часть информационного пространства организации. Более того, поскольку все эти системы ориентированы на выполнение специальных задач, часто они не включают достаточных средств для систематизации информации.

Одним из способов решения данной проблемы является разработка информационной системы, которая стала бы интегратором, надстройкой над всеми остальными программными системами, функционирующими в организации. Единственной зада-

чей данной системы будет систематизация информации, в то время как сами данные будут храниться там же, где и обычно. Письма будут храниться на почтовом сервере, файлы на FTP-серверах, но в системе-интеграторе будут находиться ссылки на все эти ресурсы, организованные наиболее оптимальным образом. Это позволит ввести единую модель систематизации для всех информационных объектов, определить единый интерфейс, через который можно получить любую необходимую информацию, и использовать при этом мощные специально разработанные для этих целей инструменты систематизации, сущность которых и является предметом данного исследования.

Технически данная система может состоять из Web-интерфейса, предоставляющего быстрый доступ к базе знаний с любого устройства (в том числе мобильного), а также набора шлюзов для доступа к различным хранилищам информации (например, для осуществления полнотекстового поиска по содержимому документов), набора специализированных клиентов (плагинов), которые будут внедряться в существующие программные системы организации (почтовые клиенты, браузеры, файловые менеджеры и т.д.) и предоставлять пользователям простой интерфейс для добавления обрабатываемых в них информационных объектов в интегрированную базу знаний. Общая архитектура системы представлена на рис. 1. Штриховыми линиями на рисунке обозначены элементы, которые являются внешними для разрабатываемой системы.

Проблема выбора структуры базы систематизации

Ключевой проблемой при построении данной системы является логическая структура базы систематизации. От того, насколько она будет гибка в описании структурируемого информационного пространства, зависит эффективность всей системы. Анализ существующих способов систематизации информации показал, что одним из основных способов упорядочивания сведений является разбивка множества информационных объектов на именованные подмножества [6]. Данные подмножества могут иметь плоскую структуру и пересекаться произвольным образом (такие подмножества обычно называют "метки", или "тэги"), либо выстраиваются в иерархическую структуру без пересечений ("директории", "папки") [6]. Стоит отметить, что при большом числе информационных объектов иерархическая организация — один из самых эффективных способов борьбы со сложностью [4, 5]. Таким образом, структура базы систематизации должна включать в себя выделение в массиве данных некоторых групп или подмножеств, формируемых по некоторым объективным или субъективным критериям, значимым с точки зрения пользователей базы знаний, и образующих иерархическую структуру.

Иерархическая организация множеств в данном вопросе не может просто ограничиться соотноше-

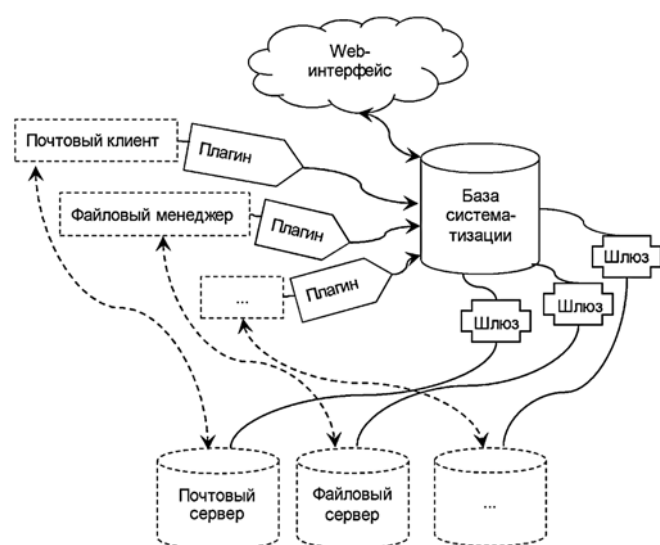


Рис. 1. Архитектура интегрирующей информационной системы



Рис. 2. Вариант иерархической организации множеств с непересекающимися подмножествами



Рис. 3. Вариант иерархической организации множеств с пересекающимися подмножествами

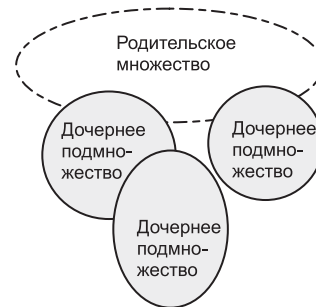


Рис. 4. Вариант иерархической организации множеств с несвязанными подмножествами

нием "надмножество включает в себя подмножество", так как это не позволяет смоделировать существующие, устоявшиеся способы систематизации [6]. Рассмотрим возможные способы выделения подмножеств информационных объектов с теоретико-множественной точки зрения, учитывая иерархический характер взаимосвязей.

А. Дети являются непересекающимися подмножествами родителя (рис. 2).

В. Дети являются пересекающимися подмножествами родителя, могут охватывать не все множество (рис. 3).

С. Родительское множество никак не связано с дочерними подмножествами (рис. 4).

Д. Набор независимых иерархий, существующих одновременно, т. е. множества каждой иерархии могут иметь произвольные пересечения на разных уровнях, которые никак не ограничены.

Варианты построения одной иерархии А, В, С получаются путем сочетания ответов на два вопроса, представленных в табл. 1. Вариант D объединяет несколько иерархий, каждая из которых может строиться различными способами.

Наиболее распространенной является модель А. Эта модель используется в иерархии директорий файловой системы, а также реализована во множествах систем для организации информации.

Модель В в упрощенном, одноуровневом виде известна как способ организации с помощью "тэгов", или "меток". Информационные системы, использующие данную модель, как правило, не вводят никакой иерархии, собирая все метки на одном уровне. Это можно рассматривать как одноуровневую иерархию, т. е. есть один родительский узел, который содержит в себе все дочерние подмножества.

Модель С встречается как расширение модели В в программе для ведения заметок Evernote. Данная

система позволяет строить иерархию меток, однако родительские метки никак не связаны с дочерними. То есть, выбрав родительскую метку, можно увидеть только те объекты, которые связаны непосредственно с ней. Чтобы увидеть объекты, связанные с дочерними метками, нужно выбрать каждую из них в отдельности. В таком случае теряется какой-либо смысл в отношении родитель — потомок как отношении между множествами. Поэтому в варианте С более рационально узлами дерева считать не множества информационных объектов как таковые, а группы, объединяющие другие узлы и листья дерева.

Модель D слабо распространена, однако в ограниченном виде присутствует в некоторых системах. Например, часто метки и директории сосуществуют одновременно. Так, в программе Evernote заметки можно располагать в блокнотах (аналоги директорий) и в то же время прикреплять к ним произвольные метки. Таким образом, мы имеем две независимые модели, которые существуют одновременно.

Предлагаемая структура базы систематизации

Все описанные типы взаимосвязи между множествами можно объединить в одну обобщенную модель, основанную на использовании элементов систематизации трех различных типов, представленных в табл. 2.

Таблица 2

Элементы систематизации обобщенной модели

Элемент систематизации	Свойства элемента
Директория	Содержимое директорий на одном уровне иерархии не пересекается. Все дочерние элементы ограничены содержимым директории
Метка	Содержимое меток на одном уровне иерархии может пересекаться. Все дочерние элементы ограничены содержимым метки
Группа	Не содержит собственных элементов, но образует уровень иерархии. Дочерние элементы ограничены содержимым родителя группы (если родитель — другая группа, то ограничены содержимым ближайшего родителя, имеющего собственные элементы, если таковой существует)

Принцип формирования вариантов иерархической организации множеств

Таблица 1

Вариант	Потомки могут пересекаться?	Потомки ограничены родителем?
А	—	+
В	+	+
С	±	—

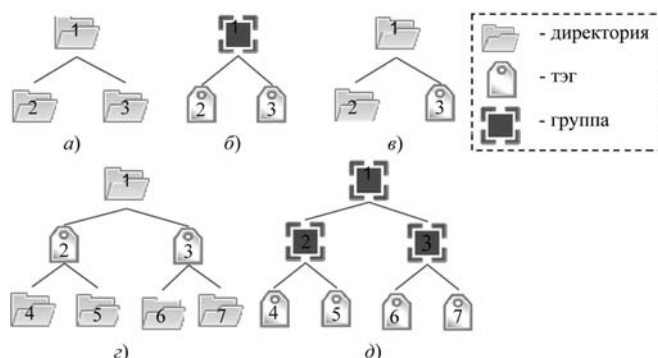


Рис. 5. Некоторые варианты построения иерархий

Строя иерархию из данных элементов систематизации, можно моделировать все описанные способы отношения множеств. С помощью директорий можно построить модель А, меткам соответствует модель В, группы образуют модель С и D. Рассмотрим некоторые варианты построения иерархий на их основе (рис. 5).

Иерархия, построенная с помощью директорий (под литерой "а" на рисунке) будет соответствовать классической иерархии директорий в файловой системе. Структуру общеизвестного способа систематизации в виде облака тэгов, или меток, можно представить с помощью группы и содержащихся в ней меток (литера "б" на рисунке).

При комбинировании всех трех элементов можно получить более сложные варианты структур. Под литерой "в" на рисунке на одном уровне иерархии расположена метка и директория. При этом оба этих элемента находятся в одной роди-

тельской директории. Таким образом, элементы 2 и 3 будут содержать подмножества элемента 1 и смогут пересекаться. Другой интересный случай изображен под литерой "г", в котором содержимое директорий 4, 5 может пересекаться с содержимым директорий 6, 7, так как могут пересекаться множества их родительских меток 2 и 3. Это не нарушает целостности модели, поскольку директории ограничены только тем, что их содержимое не пересекается с другими директориями на одном уровне иерархии. Пересечения с другими уровнями иерархии зависят уже от типа родительских элементов.

Вариант использования групп изображен под литерой "д". Группы позволяют группировать другие элементы систематизации, не накладывая на них никаких ограничений. Группы не содержат собственных элементов, поэтому их потомки ограничены только содержимым родителя самой группы, если таковой имеется. Подобный способ систематизации, в частности, используется в программе Evernote для иерархической организации меток, о чем уже упоминалось выше.

Рассмотрим операции по работе с множествами в контексте поддержания целостности модели, т.е. выполнение ограничений, указанных в табл. 2. Операции представлены в табл. 3.

Как видно из таблицы, ограничение накладывается только на операции добавления и удаления объекта. Все остальные операции легко реализуются через эти две базовые операции и поэтому не требуют особой логики для поддержания целостности.

Рассмотренная обобщенная модель, с одной стороны, может использоваться как самодостаточный инструмент для систематизации информационных

Таблица 3

Поддержание целостности в обобщенной модели

Операция	Ограничение	Действие
Добавление объекта в директорию или тэг	Дочерние элементы ограничены содержимым родителя	Добавить объект также во все родительские множества, которые имеют собственные элементы (тэги и директории)
Добавление объекта в директорию	Содержимое директорий на одном уровне иерархии не пересекается	Если объект принадлежит другой директории, то 1) подниматься по иерархии обеих директорий вверх до тех пор, пока не встретятся элементы, находящиеся на одном уровне; 2) если оба этих элемента являются директориями, удалить объект из директории либо запретить операцию
Удаление объекта из директории или метки	Дочерние элементы ограничены содержимым родителя	Удалить объект из всех дочерних элементов либо отменить операцию.
Перенос элемента в другое место иерархии	—	1. Создание копии элемента и его дочерних элементов в целевом месте иерархии. 2. Удаление оригинального элемента и его дочерних элементов*. 3. Добавление объектов в копии элементов с проверкой всех ограничений
Объединение двух элементов	—	1. Удаление объединяемых элементов*. 2. Создание объединенного элемента. 3. Добавление объектов в объединенный элемент с проверкой всех ограничений
Смена типа элемента	—	1. Удаление исходного элемента*. 2. Создание целевого элемента. 3. Добавление в целевой элемент объектов с проверкой всех ограничений

* При удалении необходимо выбрать, следует или нет удалять объекты также из родительских элементов. Этот выбор не может быть сделан автоматически, так как имеет прагматическую составляющую.

объектов. С другой стороны, данная модель может служить базой для построения других, более сложных способов систематизации. Например, представляется перспективным осуществлять привязку элементов данной модели к элементам модели процессов предметной области или структурной иерархии компании. Такая привязка позволит четче сформулировать семантику и содержание элементов систематизации, а также упростит поиск нужных информационных объектов, так как модель будет оперировать с хорошо известными сотрудникам сущностями и процессами.

Заключение

Основным преимуществом рассмотренной модели систематизации является ее обобщающий характер. Данная модель, во-первых, включает в себя все основные возможные варианты иерархической взаимосвязи множеств, а во-вторых, позволяет смоделировать уже существующие способы систематизации, такие как директории и метки вместе, а не по отдельности. При использовании только элементов вида "директория" мы получаем семантику, аналогичную семантике файловой системы. При использовании только элементов типа "метка" без иерархических взаимосвязей формируется классическая форма использования меток в современных информационных системах. Объединение же меток, директорий и групп позволяет повысить гибкость модели, делая возможным описание более

сложных свойств информационных объектов, подлежащих систематизации.

Предполагается, что данная модель будет реализована в виде программной системы управления знаниями для автоматизации процессов экстернализации, интернализации и комбинации (согласно модели, введенной Нонака и Такеучи [7]). Дальнейшее направление исследований в данной области заключается в изучении алгоритмов взаимодействия с пользователями системы, а также в построении других моделей систематизации более высокого уровня, основанных на использовании описанной обобщенной модели.

Список литературы

1. Мильнер Б. З. Концепция управления знаниями в современных организациях // Российский журнал менеджмента. Институт экономики РАН, Государственный университет управления. 2003. № 1. С. 57–76.
2. Гаврилова Т. А., Кудрявцев Д. В. Информационные технологии управления знаниями. URL: <http://bigc.ru/publications/bigspb/km/itkm/> (дата обращения 20.02.2016).
3. Тузовский А. Ф., Чириков С. В., Ямпольский В. З. Системы управления знаниями (методы и технологии) / Под общ. ред. В. З. Ямпольского. Томск: Изд-во НТЛ, 2005. 260 с.
4. Буч Г. Объектно-ориентированный анализ и проектирование. 2-е изд. М.: Вильямс, 2010. 720 с.
5. Марка Д., Клемент М. Методология структурного анализа и проектирования SADT. М.: Метатехнология, 1993. 284 с.
6. Богатырев П. Ю., Коскин А. В. Систематизация информации при реализации инновационных проектов IT-компаниями // Информационные системы и технологии. 2016. № 2. С. 51–58.
7. Нонака И., Такеучи Х. Компания — создатель знания. Зарождение и развитие инноваций в японских фирмах: Пер. с англ. А. Трактинского. М.: ЗАО "Олимп-Бизнес", 2011. 384 с.

P. Yu. Bogatyrev, Postgraduate, e-mail: PFight77@gmail.com,

A. V. Koskin, Professor, e-mail: kav1959@rambler.ru

Orel State University named after I. S. Turgenev

Structure of Integrated Information Space of a Company, Based on Set-Theoretic Approach

Approach of developing a content management system of organization in information structure aspect is presented in the article. Generalized model of information systematization that combine known methods, used in modern content management systems is introduced. The core of the model is a set-theoretic approach for structuring a data.

Keywords: information, knowledge management, content management systems, organization, model, systematization, structuring, set theory, hierarchy, tags

References

1. Mil'ner B. Z. Konceptija upravlenija znanijami v sovremennyh organizacijah (Concept of knowledge management in modern organizations). *Rossiiskij zhurnal menedzhmenta*, Institut ekonomiki RAN, Gosudarstvennyj universitet upravlenija, 2003, no. 1, pp. 57–76.
2. Gavrilova T. A., Kudrjavcev D. V. Informacionnye tehnologii upravlenija znanijami (Information systems for knowledge management). URL: <http://bigc.ru/publications/bigspb/km/itkm/> (access date 20.02.2016).
3. Tuzovskij A. F., Chirikov S. V., Jampol'skij V. Z. Pod obshh. red. V. Z. Jampol'skogo. *Sistemy upravlenija znanijami — metody i tehnologii* (Systems of knowledge management — methods and technologies), Tomsk: Izd-vo NTL, 2005. 260 p.
4. Buch G. *Ob'ektno-orientirovannyj analiz i proektirovanie* (Object-Oriented Analysis and Design with Applications). 2-e izd, Moscow: Vil'jams, 2010, 720 p.
5. Marka D., Klement M. *Metodologija strukturnogo analiza i proektirovanija SADT* (Structured Analysis & Design Technique), Moscow, Meta Tehnologija, 1993. 284 p.
6. Bogatyrev P. U., Koskin A. V. Sistematizacija informacii pri realizacii innovacionnyh proektov IT-kompanijami (Systematization of information in the implementation of innovative projects it-companies), *Informacionnye sistemy i tehnologii*, 2016, no. 2, pp. 51–58.
7. Nonaka I., Takeuchi H. *Kompanija — sozdatel' znanija. Zarozhdenie i razvitie innovacij v japonskih firmah* (The Knowledge — Creating Company: How Japanese Companies Create the Dynamics of Innovation). [Per. s angl. A. Traktinskogo], Moscow: ZAO "Olimp-Biznes", 2011. 384 p.

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ INTELLIGENT SYSTEMS AND TECHNOLOGIES

УДК 621.391:004.934

О. Я. Родькина, канд. техн. наук, доц., olgarodkina04@gmail.com,

В. А. Никольская, канд. филол. наук, доц., nivt08@rambler.ru

Нижегородский государственный лингвистический университет им. Н. А. Добролюбова,
Нижний Новгород

К проблеме распознавания психоэмоционального состояния человека по речи с использованием автоматизированных систем

Проведено исследование возможности использования программного комплекса "Фонетический детектор лжи", основанного на методе оценки качества речи по общесистемному шенноновскому критерию минимума требуемой избыточности, для экспресс-оценки психоэмоционального состояния человека по его речи. Полученные результаты открывают новые перспективные направления использования ФДЛ как эффективного инструмента в задачах психолингвистики и гендерного анализа.

Ключевые слова: информатика, человеко-машинное взаимодействие, технологии распознавания речи, информационная теория качества речи, фонетический детектор лжи, психолингвистика

Введение

Речь служит человеку, прежде всего, для передачи смысла сообщения. Но это лишь одна из ее функций, поскольку она является еще и инструментом реализации и идентификации эмоционального состояния человека. Ведь в речи передаются самые тонкие эмоции, уловить которые не всегда возможно даже человеческому уху или специалисту. Именно поэтому проблема автоматического распознавания психоэмоционального состояния человека по его голосу представляет значительный интерес как в теоретическом плане, так и для решения различных прикладных задач: в педагогике при обучении иностранным языкам; в психологии для определения личности по голосу, для изучения коммуникативных процессов; в теории массового обслуживания для улучшения качества обслуживания в call-центрах; в медицине и других сферах.

Анализ работ по данной проблематике позволил выделить в целом несколько групп признаков и методов, позволяющих на современном этапе с тем или иным успехом идентифицировать различные речевые образцы: амплитудно-частотные признаки, кепстральные, спектрально-временные и на основе нелинейной динамики [1]. Однако все указанные подходы имеют некоторые недостатки. Так, в источнике [1] показано, что распознавать и идентифицировать различные эмоции с помощью прос-

тых спектральных характеристик звукового сигнала некорректно. Амплитудно-частотные признаки используют многие исследователи, поскольку они информативны. Например, в качестве акустических коррелятов эмоциональных и функциональных состояний рассматривается ряд частотных, временных и мощностных характеристик голоса. Показано, в частности, что стенические состояния повышают основной тон, а астенические ведут к понижению показателей основного тона. Обнаружена взаимосвязь между акустическими параметрами речи в эмоциональном состоянии, обусловленная индивидуальными особенностями говорящих, но использование этих признаков в качестве инструмента идентификации эмоционально окрашенной речи в полной мере, по мнению исследователей, не представляется возможным [2, 3]. Спектрально-временные признаки позволяют отражать своеобразие спектра голосовых импульсов у разных лиц, характеризуют особенности речи, связанные со скоростью перестройки артикуляционных органов речи говорящего, характеризуют синхронность их работы. В то же время в этой группе признаков были выделены параметры, остающиеся неизменными при повышении уровня сигнала [4]. Группа признаков эмоционально окрашенной речи по кепстральным коэффициентам позволяет выделить сигнал возбуждения из сигнала речи. Для группы признаков не-