

УДК 004.421.2

ИСПОЛЬЗОВАНИЕ ОНТОЛОГИЧЕСКОГО ПОДХОДА ДЛЯ ЗАЩИТЫ ДАННЫХ ПРИ ИХ ПЕРЕСЫЛКЕ И АРХИВАЦИИ

А.А. Муромский¹, Н.П. Тучкова²

*Вычислительный центр им. А.А. Дородницына Федерального исследовательского центра
«Информатика и управление» РАН, Москва, Россия*

¹murom@ccas.ru, ²natalia_tuchkova@mail.ru

Аннотация

Проблема передачи текстов в информационных системах непосредственно связана с разработкой алгоритмов сжатия информации, управлением потоками информации, транспортной задачей, контекстным анализом информации для ее адресной доставки. Одним из ключевых вопросов является однозначность трактовки передаваемой информации. В работе предлагается один из подходов к рассмотрению проблемы управления текстами: сжатия и передачи информации в локальной информационной системе, предназначенной для управления персоналом и управления частными текстами. Рассматриваются проблемы сжатия текстов для их передачи по каналам связи и архивации с сохранением смыслового содержания. Для структуризации текстов предлагается использовать метод онтологического анализа. Сжатие текстов предлагается выполнить с применением теории цепных дробей. Предлагаемый подход позволяет выполнить сжатие текста без потерь и может быть использован как вариант защиты текстов при сохранении и пересылке.

Ключевые слова: *цепные дроби, передача текста без потерь, алгоритм сжатия текста без потерь, онтологический анализ текста.*

Цитирование: *Муромский, А.А. Использование онтологического подхода для защиты данных при их пересылке и архивации / А.А. Муромский, Н.П. Тучкова // Онтология проектирования. – 2016. – Т. 6, №2(20). - С. 136-148. – DOI: 10.18287/2223-9537-2016-6-2-136-148.*

Введение

Тема распространения информации во все времена была важнейшей для всех областей человеческой деятельности и общественного развития. С появлением информационных технологий возникли новые проблемы, связанные с кибербезопасностью, которые обычно обсуждаются в глобальном государственном аспекте. Тем не менее, практически каждый руководитель вынужден на своем уровне так или иначе решать проблемы передачи информации в процессе управления без искажения, а обычный пользователь сталкивается с проблемами защиты частных архивов, как от несанкционированного использования, так и от потерь при передаче по каналам связи или копировании.

Задача сжатия информации без потерь не теряет своей актуальности на протяжении всего «цифрового» периода развития цивилизации, и особенно, при обработке «больших данных». Для хранения и передачи текстовой информации очень важно уменьшить её объём, не искажая смыслового содержания. Собственно, проблема передачи текста является причиной возникновения предметной области, называемой кодированием.

Известны определения этого понятия. *Кодирование* - присвоение числовых кодов позициям в социологической анкете [1, с.141]. *Кодирование* (codification, coding, encoding) - процесс отображения состояния одной физической системы через состояние некоторой другой системы, производимой с целью передачи информации [2, с.125].

В первом определении: анкета - носитель информации при её движении. Важнейшее изобретение в обсуждаемой области - азбука Морзе. Код Морзе использует звуковые или световые сигналы. Для кодирования на письме используются символы тире «-» и точки «.».

Одной из главных проблем всякого кодирования является защита закодированной информации в процессе передачи. В области кодирования применяются средства для автоматизации процессов кодирования и декодирования, т.е. шифрования и дешифрования.

Шифр (code, cipher) - система условных знаков (символов) и их комбинация, которым присваиваются определённые значения и которые используются при передаче сообщений в зашифрованном виде. Шифр является видом кода, значения элементов которого и правила кодирования известны ограниченному числу лиц [2, с.127].

Способы кодирования весьма разнообразны: от элементарных до использования теории кодирования (см., например [3]). В век информационных технологий оцифровка текстов стала обычной процедурой при передаче информации. Для кодировки используются стандарты, разрабатываемые международными комитетами и ассоциациями. С 1968 года используется код ASCII (American Standard Code of Information Interchange).

Распространение телекоммуникационных технологий естественным образом привело к росту объёмов оцифрованных данных и развитию алгоритмов сжатия информации для её передачи и архивирования. Теория кодирования считается одним из наиболее важных разделов прикладной математики [4].

В настоящей работе для передачи текста предлагается воспользоваться одним из разделов теории чисел, а именно, использовать цепные дроби (непрерывные дроби) с целыми положительными элементами - натуральными числами. Использование цепных дробей для алфавитных текстов некоторой длины имеет определённое практическое значение для сжатия текста. Объектом передачи является рациональное число, которое однозначно идентифицирует текст в силу свойств цепных дробей.

1 Сжатие информации и онтологический анализ текста

1.1 Сжатие текста без потерь

Рассмотрим процесс передачи текста по каналам связи. Естественно, что для передачи данных используются международные цифровые стандарты двоичной системы исчисления, но на предварительном этапе для ускорения процесса передачи больших текстов их желательно «сжать» с помощью каких-нибудь алгоритмов кодирования. Кодирование отдельных слов реализуется различными весьма эффективными способами, которые не всегда успешно применимы в случае больших текстов [5]. При кодировании для пересылки (передачи) текстов больших объёмов возникают вопросы эффективности, связанные с временным фактором и возможными потерями. А именно, насколько процесс кодирования ускорит (или замедлит) процесс передачи текста и насколько используемый метод кодирования позволяет сохранить смысловое содержание текста, то есть осуществить сжатие текста без потерь.

Действительно, если алгоритм кодирования-декодирования текста требует временных затрат, сопоставимых с временем передачи текста без кодирования, то возникает вопрос о целесообразности его применения. Если же сжатие текста привело к искажению смыслового содержания, то кодирование также бессмысленно. То есть возникает некоторое противоречие между эффективностью самого процесса сжатия текста, скоростью и надёжностью передачи текстов. Сравнению технических свойств алгоритмов, связанных со степенью сжатия, скоростью и качеством сжатия текстов посвящено много работ, в частности, аналитический обзор [6].

В качестве развития методов сжатия текстов предлагается разбить его на два этапа: семантического анализа текста, с целью создания его структуры, и собственно кодирования отдельных частей в соответствии с выделенной структурой. Это принятый подход, варианты которого составляют методы структурирования текстов и методы кодирования. Применение онтологического подхода к анализу текстов формируется в некоторое направление интеллектуального анализа данных.

В настоящей работе используется онтологический анализ текста для выявления его структуры и теория цепных дробей для кодирования текстовых фрагментов.

1.2 Терминология и онтологический анализ текста

Один из подходов к рассмотрению проблемы структуризации текста - это использование технологии онтологического моделирования для представления информации. С точки зрения сжатия информации можно выделить несколько этапов в работе с онтологиями в информационных системах:

- формирование словарей;
- выявление ключевых слов;
- выявление связей (иерархических, ассоциативных, семантических и др.);
- формирование модели данных (онтологии);
- подготовка запросов для поиска по ключевым словам (архивация форм запросов, архивация поисковых образов);
- разметка данных по ключевым словам (архивация связей);
- архивация самих данных в соответствии с выбранной структурой (моделью).

Эти этапы информационной обработки необходимы для организации поиска и хранения данных в любой информационно-поисковой системе. Когда речь идет о хранении или передаче по каналам связи текстовых данных, то задачу сжатия для обеспечения скорости и сохранения содержания приходится решать всегда.

Использование словарей ключевых слов, помещённых в онтологию предметной области (ПрО) текста - один из способов структурирования текста, позволяющий отразить семантические связи, формальное представления тематики и тем самым сохранить смысловое содержание текста [7].

Терминология в передаче информации также имеет первостепенное значение. Применение технологии контролируемой лексики на основе словарей и тезаурусов служит, в том числе, и защите передаваемой информации от искажения. Терминологический и онтологический анализы - фактически связанные процессы, в результате которых необходимо получить словарь со связями, в котором однозначно определяются понятия некоторой ПрО или тематики структурируемого текста.

На рисунке 1 представлен один из вариантов предварительной обработки текста с целью выявления ключевых слов для последующего кодирования и передачи по каналам связи. Для пересылки и хранения текстов (особенно текстов больших объёмом) логично сначала выявить структуру текста (первый этап сжатия), а потом кодировать отдельные его части (второй этап сжатия). Это поможет сохранить смысловое содержание текста при его сжатии. На рисунке 1 показан вариант анализа текста, когда сначала составляется словарь, далее в нём выявляются термины (главные и связанные), что позволяет сформировать тезаурус текста. Далее на его основе можно обратиться к онтологическому анализу и в результате, опираясь на структуру текста (из онтологии) и ключевые слова (из тезауруса), сформировать список слов для кодирования, которые будут передаваться по частям, соответственно структуре. Если текст сравнительно небольшой, то можно пересылать его весь. Для контроля качества полученного сообщения пересылка сопровождается служебной информацией в виде сжатого по

общему алгоритму списка ключевых слов из тезауруса и темы (содержания, идентификатора ПрО и т.д.) из онтологии.

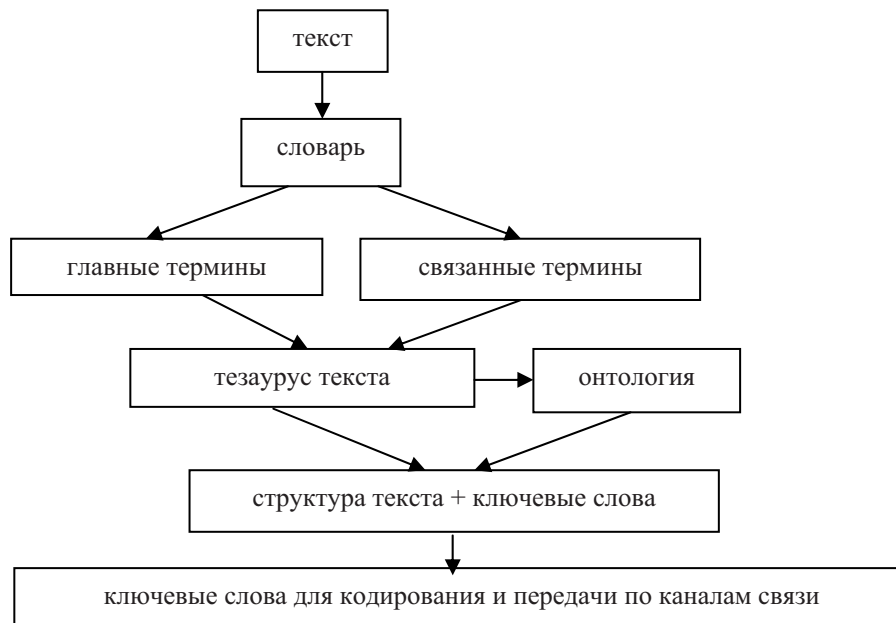


Рисунок 1 - Схема (вариант) терминологического и онтологического анализа текста

Рассмотрим вариант такой обработки текста на следующем примере:

«Текст» - состоит из «Р» - предложений и «S» - слов. Выявим наиболее часто употребляемые в «Тексте» термины и выясним, являются ли они главными терминами, на основе которых можно создать словарь со связями (локальный тезаурус для фрагмента текста).

Например, проанализируем подраздел 1.1. настоящей статьи. Этот текст состоит из 256 слов и 12 предложений. Следующие слова (для удобства запишем их как существительные в именительном падеже единственного числа) встречаются несколько раз: текст - 19, кодирование - 12, передача - 7, процесс - 5, анализ - 4, метод - 4, смысл - 3, алгоритм - 3. Все остальные, такие как, «семантический», «потери» и т.д., встречаются три и меньше раз. Можно продолжить этот список и учесть, например, синонимы и антонимы: кодирование и сжатие (2); передача и пересылка (1); текст и данные (2); кодирование - декодирование (1); смысл - бессмысленный (3).

Можно считать, что «текст» и «кодирование» - главные термины, а остальные с ними связаны. Сами существительные «текст» и «кодирование» также связаны между собой в словосочетания, а также употребляются с определёнными глаголами. Если взять эти словосочетания, то получим связи: «кодирование-декодирование - текста» (3 раза); «анализ - текста» (3 раза); «сжатие - текста» (5 раз); «передача - текста» (7 раз с учётом синонимов), «метод - кодирования» и др.

По этому списку слов и связей можем определить «тематику» текста как текст «о сжатии и передаче текста». Далее, используя метод сжатия текста с помощью цепных дробей, записать двенадцать предложений в виде двенадцати рациональных дробей и переслать эти числа. Затем добавим ещё одну дробь с сообщением о «тематике» текста, которую уже определили из терминологического анализа словаря. В итоге, передаётся не 256 слов со знаками препинания, а тринадцать чисел.

Заметим, что для формального выяснения «тематики» текста необходимо обратиться к онтологии, в которой содержатся наиболее часто встречающиеся в тексте термины (главные

термины). Если таковая есть, то ПрО текста, его тематика, определяется однозначно (см. связь «тезаурус текста – онтология» на рисунке 1). Если текст небольшой, то достаточно терминологического анализа словаря. Для больших объёмов текстов уже выгодно построить онтологию [8] и в этом случае фрагменты информации будут однозначно тематически идентифицироваться.

Разумеется, существуют и другие способы разбиения текста и сжатия текстовых фрагментов. Способ, предлагаемый в данном случае, отличается тем, что сам текст не передаётся, а передаётся число. Заметим, что можно передать вообще одну дробь, соответствующую тексту, то есть сжать весь текст до одного числа (рациональной дроби), причём это будет сделано с помощью цепной дроби без потерь.

2 Защита данных при хранении и пересылке

2.1 Терминология и защита передаваемой информации

Защита передаваемой информации, текста, представляет собой необходимое условие процесса движения информации. Способы, методы, оборудование, в частности кодеры и декодеры, как правило, входят в функции государственной сферы и не составляют предмета данной работы. Сама тема защиты информации весьма распространённая и составляет одну из проблем передачи информации. В этом отношении можно отметить сборник комитета научно-технической информации [9], который содержит 106 терминов на русском и английском языках. Представляет большой интерес издание института языкознания [10], в котором рассмотрены:

- информационно-языковые проблемы современного общества;
- теоретические и методологические вопросы обработки информации на естественном языке;
- вопросы обработки номинативных единиц естественных и искусственных языков при решении информационных задач.

Целая глава работы [10] посвящена информационным тезаурусам.

Отметим, что использование цепных дробей в какой-то мере уже является защитой, так как объектом передачи является несократимая дробь, но при условии отсутствия сведений о её образовании, происхождении. При этом реализуется сжатие без потерь.

На практике в большинстве случаев применяют методы сжатия информации без потерь в виде различных текстов с определённым ключом. Кодирование, как способ преобразования текста для передачи, и метод защиты информации встречаются в цифровую эпоху в том или ином виде в различных областях жизнедеятельности.

Очевидные аспекты для обсуждения в этом контексте: информация и деньги, цена, стоимость информации, стоимость защиты информации, как быть в нужном месте в нужное время (временной фактор составляет часть защищаемой информации и не менее важен). В них изначально заложены противоречия и основы противоборства в современном мире. Информация - именно та область, где эти противоречия выражаются в концентрированном виде.

2.2 Защита частных данных

Сохранение частной жизни и скопление данных, поддающееся проверке, обсуждались на недавней конференции по кибербезопасности [11]. Действительно, в современном мире частные архивы данных стали в большей степени подвергаться опасности по причине их представления в цифровом виде, в том числе и в открытом доступе. Их можно уничтожить

(«стереть»), потерять при пересылке, сделать доступными для окружающих случайно или преднамеренно. Сжатие информации можно применить как вариант защиты частных данных.

Системы кодирования и шифрования данных при пересылке используются практически всеми почтовыми программами. Но есть ещё проблемы защиты и поиска данных в собственных архивах, а также при общении через социальные и информационные сети.

В общем случае можно различать ситуации сжатия с использованием цепных дробей для архивирования собственных текстов и текстов некоторого сообщества.

Для личных текстов можно составить словари для обозначения их тематики и далее хранить не тексты, а соответствующие числа. Это позволит экономить место в памяти компьютера или мобильного устройства, а также быстро осуществлять поиск необходимой информации по словарю. Например, в отношении частных данных можно действовать по следующей схеме.

- 1) выбрать из своих текстов ключевые слова, которыми можно однозначно идентифицировать имеющиеся (или поступающие) тексты;
- 2) закодировать тексты, индексированные ключевыми словами, применяя алгоритм сжатия с помощью цепных дробей,
- 3) записать для хранения словарь и числа (дробь);
- 4) осуществлять поиск по ключевым словам, далее для получения текста декодировать дробь.

Несколько отличная схема может быть использована для сжатия текстов информационного сетевого сообщества. Поскольку сообщества формируются по интересам, то можно опираться на язык сообщества, составить для него словари, использовать онтологии ПрО [12], которые представляют интерес для участников. Далее, используя технологию сжатия текстов, например, теорию цепных дробей, пересылать числа, а не тексты. Вариант общения с использованием данных, сжатых согласно схеме, показанной на рисунке 2, может быть таким.

- 1) выбор онтологии ПрО, если таковая есть, или составление «интересов» сообщества и создание «онтологии сообщества»;
- 2) составление «языка сообщества» на уровне словаря сообщества на основе терминов онтологии сообщества;
- 3) кодирование словаря с помощью алгоритмов сжатия (например, цепных дробей);
- 4) пересылка кодов (рациональных дробей), а не текстов внутри сообщества.

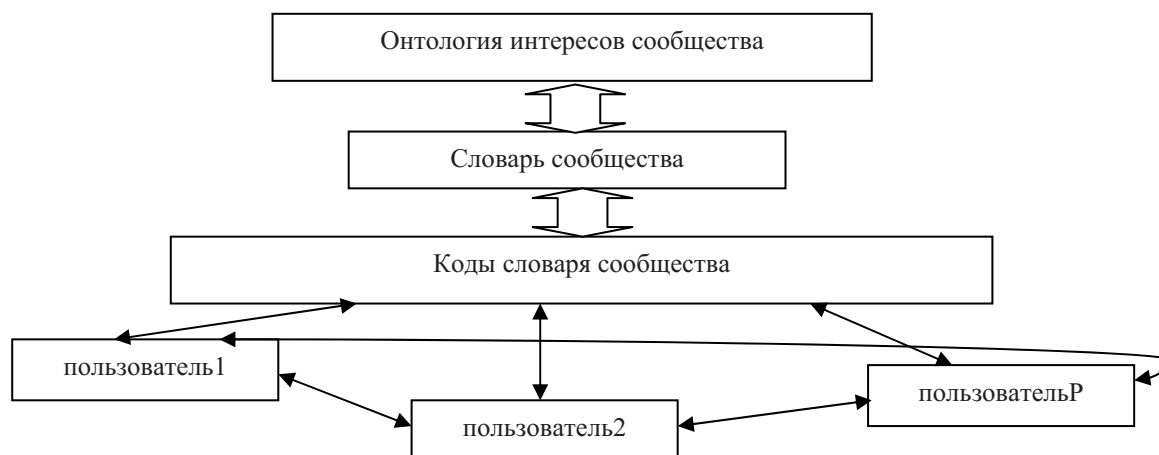


Рисунок 2 - Схема (вариант) сжатия текстов архива информационного сообщества

Рассмотренные примеры сжатия информации (текстов) в личных архивах и для использования в сетевом общении могут быть другими. Известны, например, «словари хиппи», популярные в середине 20 века [13]. Это было время, когда не было сетевого общения, но языки сообществ формировались, что продолжается на современном «цифровом» уровне.

2.3 Защита данных в управляющей системе

История знает немало событий, когда неверно понятая информация приводила к ошибкам на уровне управления. Известная фраза «сделайте что-нибудь», адресованная подчинённым разного уровня, зачастую приводила к вариантам в трактовках, а иногда и к трагическим последствиям¹. Особая роль передачи смысла (содержания) информации в управлении связана не только с передачей данных от управления к управляемому, но и с обратной связью, от управляемого объекта.

Однозначность описания ситуации позволяет принять правильное решение в управлении. В этом контексте экономика и наука оказались, в одной ситуации, поскольку они требуют инвестиций в условиях конкурентной борьбы. Современная торговля, как и реализация научных проектов, сопряжены с маркетинговыми исследованиями и рекламными продвижениями продукта. Проблемы передачи смысла и восприятие текста особенно ярко отражены в рекламных продуктах [14]. Использование графики и специальных обозначений, характерных для демонстрации научных и практических результатов, весьма эффективно, но в зависимости от культурных традиций и уровня подготовки могут восприниматься по-разному [15]. Язык был и остаётся основным средством коммуникации, а проблема передачи информации с помощью текста не теряет своей актуальности. Передача информации в системах ситуационного управления имеет целью управление людьми. В свою очередь, в плане управления людьми основные процессы - это употребление силы и передача информации [16].

Особенность современного управления составляет широкое применение мобильных устройств для быстрого оповещения населения и персонала о ситуации. Управление бизнес-процессами и персоналом требует чёткости и определённости в передаче информации. Текстовая информация, как правило, допускает варианты в трактовке, поэтому важно обеспечить недопущение двусмысленностей.

Первую и необходимую часть работы по организации передачи данных составляет создание словарей для конкретной ситуации. Основа описания ситуации - это использование контролируемой лексики [17].

Использование алгоритмов сжатия текстов - следующий этап в подготовке данных к их передаче в информационной сети. Сжатие информации выполняется различными путями с применением алгоритмов кодирования [18].

В качестве примера можем рассмотреть схему рисунка 3, где предлагается ввести элемент «словарь» в процесс передачи информации от руководства персоналу. Анализируя бизнес-процессы предприятия на информационном уровне, можно использовать технологию онтологического моделирования и сформировать «словари распоряжений» для различного уровня руководства. Словари кодируются и для передачи используются, например, цепные дроби. Пользуясь словарями и мобильными устройствами персонала, руководители могут быть уверены, что текст распоряжения будет доставлен без потерь и искажений, что гарантируется особенностями алгоритма.

¹Например, история убийства архиепископа Томаса Бекета, после фразы Генриха II: "Who will rid me of this troublesome priest?" (https://en.wikipedia.org/wiki/Thomas_Becket#cite_note-Warren508-13). Один из вариантов перевода из словаря цитат: «Неужели никто не избавит меня от этого мятежного попа?». Oxford Dictionary of Quotations. Edited by Elizabeth Knowles. Oxford University Press (1999, 2004, 2009) (<http://www.oxfordreference.com/>). Независимо от того, что король сказал, это интерпретировалось как королевская команда, и Томас Бекет был убит.

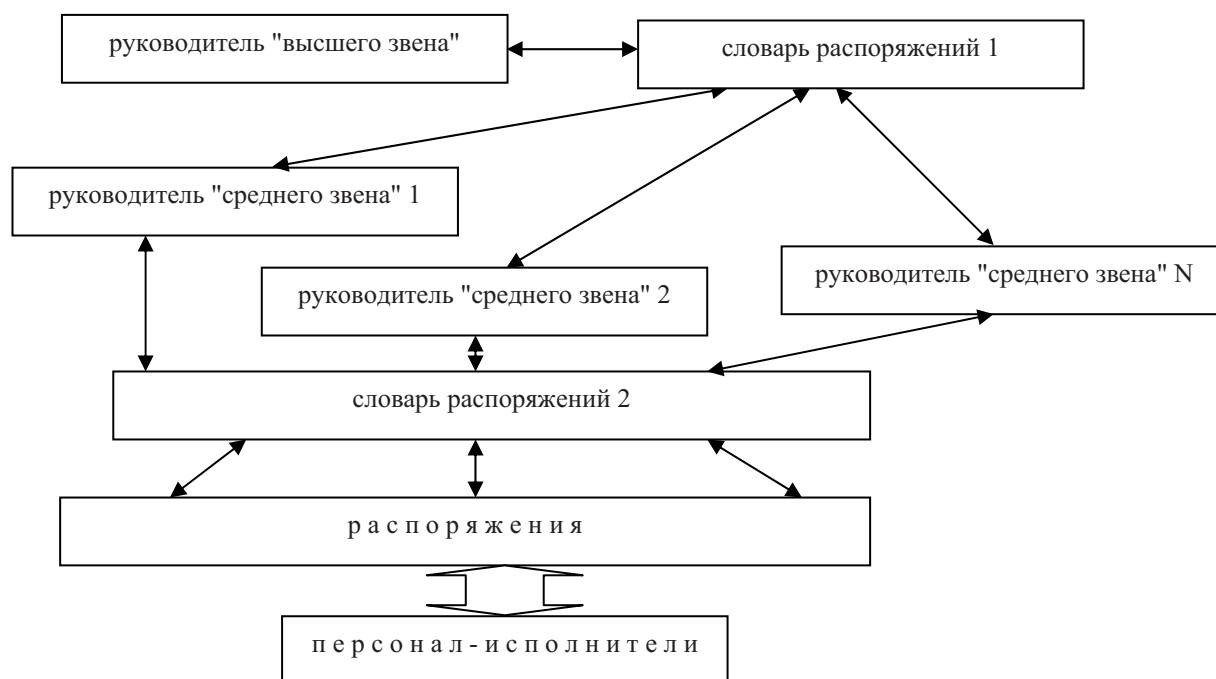


Рисунок 3 - Схема (вариант) сжатия текстов распоряжений для однозначного их толкования

Отличительные особенности предложенных алгоритмов составляют передача смысла текста без потерь, минимальный объём занимаемой памяти передающего и приёмного устройств (например, мобильного телефона), возможность восстановления при утере текста и скорость².

2.4 Кодирование без потери данных

Для кодирования текста, предназначенного для передачи по каналам связи, в работе используются цепные дроби.

Каждой букве алфавита соответствует номер, например, как в таблице 1.

Таблица 1 – Соответствие номеров и букв русского алфавита

Буква	№	Буква	№	Буква	№	Буква	№	Буква	№	Буква	№	Буква	№
А	2	Д	6	З	10	Л	14	П	18	У	22	Ч	26
Б	3	Е	7	И	11	М	15	Р	19	Ф	23	Ш	27
В	4	Ё	8	Й	12	Н	16	С	20	Х	24	Щ	28
Г	5	Ж	9	К	13	О	17	Т	21	Ц	25	Ъ	29
												Я	34

В таблице 2 также в качестве примера указаны необходимые в тексте знаки препинания и соответствующие им номера.

Таблица 2 – Соответствие номеров и знаков препинания

Знак	№	Знак	№	Знак	№	Знак	№	Знак	№	Знак	№
«.»	35	«.»	36	«:»	37	«?»	38	«!»	39	«-»	40

² Авторы в этой статье оставили в стороне принципы формирования словарей распоряжений, т.е. сами формализмы семантического сжатия без потерь. Редакция приглашает исследователей высказаться по этой актуальной теме на страницах нашего журнала. *Прим.ред.*

В дальнейшем используем известные свойства цепных дробей с целыми элементами (см., например [19-22]).

Напомним одно из основных утверждений (см. [19, с. 25-30]):

«Теорема 14. Каждому вещественному числу α соответствует единственная цепная дробь, имеющая это число своим значением. Эта дробь конечна, если число α рационально, и бесконечна, если оно иррационально».

Рассматриваем здесь только конечные цепные дроби, так что в нашем случае упомянутое в теореме 14 число α - рационально.

На основе изложенного и таблиц 1 и 2, констатируем взаимно-однозначное соответствие между записями в виде цепных дробей, элементами которых являются буквы русского алфавита и номерами этих букв, соответственно.

Поясним сказанное на примере слова «АББА». Запишем выражения для этого слова в виде цепной дроби:

$$(1) \quad t = \frac{1}{A + \frac{1}{B + \frac{1}{B + \frac{1}{A}}}}, \quad c = \frac{1}{2 + \frac{1}{3 + \frac{1}{3 + \frac{1}{2}}}}.$$

Вычисляя значение дроби, получим, что $c = 23/53$, причём $t \leftrightarrow c$, соответственно, при назначенной нумерации букв из таблиц 1 и 2.

Приведённый пример со словом «АББА» распространяется на любой текст, которому взаимно-однозначно соответствует рациональное число, несократимая дробь. Для больших текстов будет очень большая дробь. Чтобы избежать этого, можно сжимать текст по частям, разделяя его на фрагменты, например, из 10 предложений и т.п.

Запись цепной дроби представляют в следующем виде (см. [19, с.8]):

$[a_0; a_1, a_2, \dots, a_n]$ - конечная цепная дробь ($1 \leq n < \infty$). В нашем случае:

$$(2) \quad АББА \leftrightarrow [0; 2, 3, 3, 2] \leftrightarrow 23/53$$

Приведём ещё характерные примеры.

Пример 1. Текст: «объект в 1-ом квадрате».

В рамках изложенного, пример должен быть следующим: «Объект в первом квадрате». Это правило распространяется на все случаи. Можно охватить и приведённый выше пример с «1-ым», но это приводит к некоторому усложнению. Поэтому в такого рода случаях употребляется только текст.

Пример 2. Транспортная задача (ТрЗ) о передаче текста.

Пусть O - отправитель (поставщик) в ТрЗ, $П$ - получатель в ТрЗ. $Об$ - объект передачи в ТрЗ. Рассмотрим общую схему для ТрЗ: $O \rightarrow Об \rightarrow П$.

На основании изложенного под «Об» можем понимать значение цепной дроби. Далее, перед « $O \rightarrow$ » должен быть реализован процесс кодирования текста (см. выше), а после « $\rightarrow П$ » процесс декодирования и получения в итоге исходного текста. Наглядно весь процесс в данной ТрЗ проиллюстрируем с использованием реального текста «АББА» (3).

$$O \rightarrow ABBA \rightarrow \frac{1}{A + \frac{1}{B + \frac{1}{A}}} \rightarrow \frac{1}{2 + \frac{1}{3 + \frac{1}{3 + \frac{1}{2}}}} \rightarrow \frac{23}{53} \rightarrow$$

(3)

$$\rightarrow \frac{1}{2 + \frac{1}{3 + \frac{1}{3 + \frac{1}{2}}}} \rightarrow \frac{1}{A + \frac{1}{B + \frac{1}{B + \frac{1}{A}}}} \rightarrow ABBA \rightarrow \Pi$$

Заключение

Изложенные подходы могут быть использованы для автоматического построения аннотаций и архивирования их в сжатом виде, а также в мобильных системах для распространения внутренних распоряжений при управлении [23].

В применении к информационным системам необходимо учитывать следующие проблемы:

- адекватного поиска информации (удовлетворение условию релевантности);
- соответствия информационной потребности (удовлетворение условию пертинентности);
- формального представления темы.

Эти проблемы также можно рассматривать с позиции использования кодирования словарей, что позволит ускорить и уточнить поиск.

Продолжение исследований по этой тематике предполагается в развитии алгоритма сжатия текста с использованием алфавитно-цифровых текстовых элементов для передачи информации.

Список источников

- [1] *Кравченко, А.И.* Краткий социологический словарь / А.И. Кравченко – М.: Проспект, 2013. - 352 с.
- [2] *Жданова, Г.С.* Словарь терминов по информатике на русском и английском языках / Г.С. Жданова, Е.С. Колобродова, В.А. Полушкин, А.И. Черный – М.: Наука, 1971. - 259 с.
- [3] *Романенко, А.Е.* Заметки по теории кодирования / А.Е. Романенко, А.Ю. Румянцев, А. Шень – М.: МЦНМО, 2011. - 80 с.
- [4] *Аршинов, М.Н.*, Коды и математика. Рассказы о кодировании / М.Н. Аршинов, М.Е. Мадовский – М.: Наука, 1983. - 144 с.
- [5] Big Data Technology and Applications. First National Conference, BDTA 2015, Harbin, China, December 25-26, 2015. Proceedings. Eds Chen, W., Yin, G., Zhao, G., Han, Q., Jing, W., Sun, G., Lu, Z. - Singapore: Springer 2016. 324 p.
- [6] *Kodituwakku, S.R.* Comparison of lossless data compression algorithms for text data S.R. Kodituwakku, U. S.Amarasinghe / Indian Journal of Computer Science and Engineering V. 1 N. 4 P. 416-425.
- [7] *Englmeier, K.* Domain ontology: automatically extracting and structuring community language from texts / K. Englmeier, F. Murragh, J. Mothe (<https://www.researchgate.net> Valid on 06 April 2016).
- [8] *Aussenac-Gilles, N* Supervised text analysis for ontology and terminology engineering/ N. Aussenac-Gilles (<https://www.researchgate.net> Valid on 06 April 2016).
- [9] Теория передачи информации. Терминология Под. ред. В.И. Сифорова. – М.: Наука, 1979. - 24 с. (Серия: «Сборники рекомендованных терминов». Вып. 94.)

- [10] Лингвистические вопросы алгоритмической обработки сообщений. Ответ. редакторы: д.т.н. Р.Г. Котов, К.И. Курбаков. – М.: Наука, 1983. - 246 с.
- [11] **Tran, H.N.** Privacy-Preserving and Verifiable Data Aggregation / H.N. Tran, R.H. Deng, H.H. Pang Proceedings of the Singapore Cyber-Security Conference (SG-CRC) January 14-15. A Mathur and A. Roychoudhury (Eds.) - Singapore, 2016, P. 115-122.
- [12] **Englmeier, K.** Domain ontology: automatically extracting and structuring community language from texts / K. Englmeier, F. Murragh, J. Mothe (<https://www.researchgate.net> Valid on 06 April 2016).
- [13] **Рожанский, Ф.И.** Сленг хиппи: Материалы к словарю / Ф.И. Рожанский - СПб. - Париж. Европейский Дом 1992. 64 с.
- [14] **Карманова, Т.И.** Проблемы перевода в международном маркетинге/ Т.И. Карманова // Молодой ученый. 2014. №11. С. 207-209.
- [15] **Moiseev, E.I.** Problems of information support of collective expert activity / E.I. Moiseev, A.A. Muromskiy, N.P. Tuchkova // Proceedings and Program of the International Workshop «Contingency management, intelligent, agent-based computing and cyber security in energy sector» CV/IAC/CS/ES -2015 (March, 5-10, 2015, Khankh-Irkutsk). Irkutsk: ESI SB RAS, 2015. P. 24-25.
- [16] **Доценко, Е.Л.** Психология манипуляции: феномены, механизмы и защита / Е.Л. Доценко – М.: ЧеРо, Издательство МГУ, 1997. 344 с.
- [17] <http://www.controlledvocabulary.com> (доступ получен: 31.01.2016).
- [18] **Ватолин, Д.** Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео / Д. Ватолин, А. Ратушняк, М. Смирнов, В. Юкин – М.: ДИАЛОГ-МИФИ, 2002. - 384 с.
- [19] **Арнольд, В.И.** Цепные дроби / В.И. Арнольд – М.: МЦНМО, 2001. - 40 с.
- [20] **Хинчин, А.Я.** Цепные дроби / А.Я. Хинчин – М.: Гос. изд. ФМ, 1961. - 112 с.
- [21] **Хованский, А.Н.** Приложения цепных дробей и их обобщений к вопросам приближенного анализа / А.Н. Хованский – М.: ГИТТЛ, 1951. - 203 с.
- [22] **Воробьев, Н.Н.** Числа Фибоначчи / Н.Н. Воробьев – М.: Наука, 1969. - 110 с.
- [23] **Muromskiy, A.A.** Texts management into the applications expert systems / A.A. Muromskiy, N.P. Tuchkova // Proceedings and Program of the International Workshop «Contingency management, intelligent, agent-based computing and cyber security in energy sector» CV/IAC/CS/ES -2016 (March, 11-16, 2016, Lisvjanka-Irkutsk). Irkutsk: ESI SB RAS, 2016. P. 35-37.
-

ONTOLOGICAL APPROACH TO THE DATA PROTECTION FOR THEIR TRANSFER AND ARCHIVING

A.A. Muromskii¹, N.P. Tuchkova²

Dorodnicyn Computing Centre, Federal Research Centre of Computer Science and Control RAS, Moscow, Russia

¹ murom@ccas.ru, ² natalia_tuchkova@mail.ru

Abstract

The text's transmission problem in the information systems is directly connected with development of the information compression algorithms, information management, transport task, and the contextual analysis of information for its targeted delivery. One of key questions is the unambiguous interpretation information transmitted. One of approaches to consideration of the problem of control of the text: the compression and transmission of information in the local information system for personnel managing and control of the private texts. Problems of compression of texts for their transfer on communication channels and archiving with saving the semantic contents are considered. For structurization of texts it is offered to use a method of the ontological analysis. Compression of texts is offered to be executed with application of the theory of chain fractions. The offered approach allows to execute compression of the text without loss and can be used as option of protection of texts at preservation and transfer.

Key words: chain fractions, lossless text transmission, lossless data compression algorithm, ontological analysis of the text.

Citation: Muromskii AA, Tuchkova NP. Ontological approach to the data protection for their transfer and archiving. *Ontology of designing*. 2016; 2(20): 136-148. DOI: 10.18287/2223-9537-2016-6-2-136-148.

References

- [1] Kravchenko AI. Concise Sociological Dictionary [In Russian]. - M.: Prospect, 2013. - 352 p.
- [2] Jdanova GS, Kolobrodova ES, Polushkin VA, Chernuy AI. Glossary of terms in computer science at the Russian and English languages [In Russian]. - M.: Science, 1971. - 259 p.
- [3] Romanenko AE, Rumjancev AY, Shen A. Notes on Coding Theory [In Russian]. - M.: MCNMO, 2011. - 80 p.
- [4] Arshinov MN, Madovskij ME. Codes and mathematics. Stories about coding [In Russian]. - M.: Science, 1983. - 144 p.
- [5] Big Data Technology and Applications. First National Conference, BDTA 2015, Harbin, China, December 25-26, 2015. Proceedings. Eds Chen, W., Yin, G., Zhao, G., Han, Q., Jing, W., Sun, G., Lu, Z. - Singapore: Springer 2016. 324 p.
- [6] Kodituwakku SR, Amarasinghe US, Kodituwakku SR. Comparison of lossless data compression algorithms for text data / Indian Journal of Computer Science and Engineering V.1, N.4, P.416-425.
- [7] Englmeier K, Murragh F, Mothe J. Domain ontology: automatically extracting and structuring community language from texts / (<https://www.researchgate.net> Valid on 06 April 2016).
- [8] Aussenac-Gilles N. Supervised text analysis for ontology and terminology engineering / (<https://www.researchgate.net> Valid on 06 April 2016).
- [9] The theory of information transmission. Terminology [In Russian]. Ed. VI Siforov - M.: Science, 1979. - 24 p.
- [10] Linguistic issues of algorithmic processing messages [In Russian]. Ed. RG Kotov, KI Kurbakov. - M.: Science, 1983. - 246 p.
- [11] Tran HN, Deng RH, Pang HH. Privacy-Preserving and Verifiable Data Aggregation / Proceedings of the Singapore Cyber-Security Conference (SG-CRC) January 14-15. A. Mathur and A. Roychoudhury (Eds.) - Singapore, 2016, P. 115-122.
- [12] Englmeier K, Murragh F, Mothe J. Domain ontology: automatically extracting and structuring community language from texts / K. Englmeier, (<https://www.researchgate.net> Valid on 06 April 2016).
- [13] Rozanskij FI. Hippie slang: Materials for the dictionary [In Russian]. - SPb.-Paris. European House. 1992. - 64 p.
- [14] Karmanova TI. Translation problems in international marketing [In Russian]. Young scientist. 2014. N11. P.207-209.

- [15] *Moiseev EI, Muromskiy AA, Tuchkova NP.* Problems of information support of collective expert activity / E.I. Moiseev, // Proceedings and Program of the International Workshop «Contingency management, intelligent, agent-based computing and cyber security in energy sector» CV/IAC/CS/ES -2015 (March, 5-10, 2015, Khankh-Irkutsk). Irkutsk: ESI SB RAS, 2015. P. 24-25.
- [16] *Docenko EL.* Psychology manipulation: phenomena, mechanisms and protection [In Russian]. – М.: Moscow State University Press, 1997. 344 p.
- [17] <http://www.controlledvocabulary.com> (Valid on 06 April.2016).
- [18] *Vatolin D, Ratushnjak A, Smirnov M, Yurkin V.* Data compression methods. The device archiver, compression of images and video [In Russian]. – М.: DIALOG-MIFI, 2002. - 384 p.
- [19] *Arnold VI.* Continued fractions [In Russian]. – М.: MCNMO, 2001. - 40 p.
- [20] *Hincin AY.* Continued fractions [In Russian]. – М.: Gos. izd. FM, 1961. - 112 p.
- [21] *Hovanskij AN.* Applications continued fractions and their generalizations to the issues of the approximate analysis [In Russian]. – М.: GITTL, 1951. - 203 p.
- [22] *Vorobev NN.* Fibonacci Numbers [In Russian]. – М.: Science, 1969. - 110 p.
- [23] *Muromskiy AA, Tuchkova NP.* Texts management into the applications expert systems // Proceedings and Program of the International Workshop «Contingency management, intelligent, agent-based computing and cyber security in energy sector» CV/IAC/CS/ES -2016 (March, 11-16, 2016, Lisvjanka-Irkutsk). Irkutsk: ESI SB RAS, 2016. P. 35-37.

Сведения об авторах



Муромский Александр Александрович, 1926 г. рождения, с.н.с., Вычислительный центр Федерального исследовательского центра «Информатика и управление» РАН (ФИЦ ИУ РАН), к.ф.-м.н., окончил механико-математический факультет МГУ им. М.В.Ломоносова и университет им. Н.Э. Баумана, долгие годы работал в ВИНТИ. Специалист в области математического анализа и информационных технологий. Автор более 50 работ в области функционального анализа и информационных технологий.

Muromskiy Alexander Alexandrovich (b. 1926) senior researcher of CCAS (Federal Research Centre of Computer Science and Control RAS), PhD, graduated from mechanics and mathematics faculty of Lomonosov MSU and the university of N.E. Bauman, for many years worked in VINITI. He is a specialist in the field of the mathematical analysis and information technologies. He has published over 50 papers in the field of functional analysis and information technology.



Тучкова Наталья Павловна, 1955 г. рождения, с.н.с., Вычислительный центр Федерального исследовательского центра «Информатика и управление» РАН (ФИЦ ИУ РАН), к.ф.-м.н., окончила факультет вычислительной математики и кибернетики МГУ им. М.В.Ломоносова. Специалист в области алгоритмических языков и информационных технологий. Автор более 40 работ в области вычислительной математике и информационных технологиях.

Tuchkova Natalia Pavlovna (b. 1955) senior researcher of CCAS (Federal Research Centre of Computer Science and Control RAS), PhD, graduated from CS faculty of Lomonosov MSU. She is a specialist in the field of programming languages and information technologies. She published more than 40 papers in the field of computational mathematics and information technology.