

представление. Преподавание истории может и должно заполнять собой пустующие ниши и даже вторгаться на, казалось бы, чужие территории. Только таким образом оно станет интересной, полезной и неотъемлемой частью общего образовательного пространства технического вуза и только тогда сможет достичь своей основной цели – воспитания гармонично развитой личности, сочетающей гуманитарный и технократический стиль мышления.

Литература

1. Грико Т.И. Сохранение истории (реферативная работа студентов) // Материалы (CD) Международной научно-технической конференции ААИ «Автомобиле- и тракторостроение в России: приоритеты развития и подготовка кадров», 27-28 марта 2012 г. М., МГТУ «МАМИ», 2012. с. 266.
2. Примером исследования культурно-исторического пространства Университета может служить статья: Грико Т.И., Круглова Т.А. Из истории преподавания гуманитарных дисциплин в техническом вузе (на примере Константиновского межевого института) // Актуальные проблемы гуманитарных, социально-экономических наук и высшего образования / Материалы межвузовской научно-практической конференции кафедры «История и политология». М.: Университет машиностроения, 2013. с. 12–19.
3. См., например: Петрунин Е.Н. Создание, становление и развитие университета. Часть первая // Московский государственный технический университет «МАМИ». К 145-летию со дня основания / Под ред. А.В. Николаенко и В.В. Серебрякова. М.: МГТУ «МАМИ», 2010.
4. Панин Е.В., Пигарев С.П. Военные страницы гвардии сержанта А.Н. Соколова // Актуальные проблемы гуманитарных, социально-экономических наук и высшего образования / Материалы межвузовской научно-практической конференции кафедры «История и политология». М.: Университет машиностроения, 2013. с. 84–86.
5. Комплекс разнопланового материала для самостоятельной работы студентов представлен в учебном пособии кафедры «История и политология» Университета машиностроения. См.: История России в 4-х частях / Харламова Т.И., Грико Т.И., Баранова А.Н., Колесова М.А., Пономарев М.В., Рогожкин В.А., Минухина Н.В., Максимовских Л.В., Дмитриев А.Е. М.: МГТУ «МАМИ», 2011.
6. Рыбина М.В. Интерактивные методы обучения в процессе формирования навыков работы со специальной литературой гуманитарного цикла // Материалы (CD) Международной научно-технической конференции ААИ «Автомобиле- и тракторостроение в России: приоритеты развития и подготовка кадров», 27-28 марта 2012 г. М., МГТУ «МАМИ», 2012. с. 335–337.
7. Харламова Т.И. Преподавание гуманитарных дисциплин с учетом профиля вуза // Известия МГТУ «МАМИ», 2013, № 1 (15). т. 6, с. 201–209.

Дистанционные технологии в образовании. Алгоритмы сжатия информации и форматы данных для передачи текстовой, звуковой и видеоинформации

Акимов В.А.

Университет машиностроения
v_akimov5@mail.ru

Аннотация. В статье рассмотрены особенности применения алгоритмов компрессии информации в мультимедийных технологиях обучения. Приведены примеры алгоритмов сжатия с потерями и без потерь и основных форматов сжатия текстовой, графической и аудиовизуальной информации. Приведен анализ эффективности тех или иных алгоритмов и даны рекомендации по их применению для разных видов информации. Приведен пример кодирования по методу Хаффмана.

Ключевые слова: дистанционные технологии обучения, сжатие информа-

Характерной чертой современного общества является потребность в постоянном совершенствовании, повышении квалификации, получении дополнительного образования и новых компетенций. С появлением компьютеров появилась возможность удовлетворять эту потребность дистанционно, не выходя из дома или непосредственно на рабочем месте в удобное для слушателя время. Такие технологии дистанционного обучения как телеконференции, вебинары, мультимедийные курсы стали мощным средством обучения, которые позволяют существенно ускорить и повысить эффективность взаимодействия слушателей с университетом. Однако такое количество информации потребовало применения мощных алгоритмов ее сжатия и декомпрессии при хранении, передаче по каналам связи и выдаче пользователю. Выбор нужного алгоритма сжатия, формата файла и коэффициента сжатия информации непосредственно влияет на качество мультимедийного контента, рисунков и текстовых данных и соответственно влияет на восприятие слушателем учебных материалов. В данной статье мы рассмотрим возможные варианты представления данных в процессе обучения дистанционным методом.

Все данные в электронных вычислительных машинах представляют собой последовательность двоичных символов. При этом каждой букве, цифре или знаку препинания соответствует 8-битный или 16-битный код из таблицы кодировки.

8-битная таблица кодировки называется ASCII (American Standard Code for Information Interchange) – американская стандартная кодировочная таблица для печатных символов и некоторых специальных кодов. 16-битная называется Unicode и позволяет кодировать символы практически из всех имеющих хождение языков.

Все данные, с которыми имеет дело человек в повседневной жизни (возможно кроме ценников в супермаркете), являются избыточными. Это связано с психологической особенностью человека: избыточность делает данные более понятными и улучшает качество восприятия информации. При этом степень избыточности зависит от типа данных и способа их кодировки. Например, избыточность текстовых данных в формате ASCII ниже, чем в формате Unicode. Еще выше степень избыточности аудиоданных и графических данных, и самыми избыточными в настоящий момент являются видеоданные (в этом нетрудно убедиться, посмотрев на размер файла, содержащего 10 минутный видеоролик в разрешении HD). Улучшая восприятие информации, избыточность играет отрицательную роль при обработке, передаче и хранении информации, поскольку она увеличивает нагрузку на оборудование, каналы связи и увеличивает требуемое место на диске, а значит, и стоимость хранения.

В связи с этим проблема уменьшения избыточности данных является актуальной и постоянно идет процесс создания новых и совершенствования старых методов и алгоритмов сжатия и архивирования информации, графики аудио и видео контента. Если метод сжатия применяется к текстовым файлам, то применяют термин архивация данных, а программные средства, реализующие этот метод, называются архиваторами.

Методы сжатия данных делятся на сжатие с потерями и сжатие без потерь.

Сжатие данных без потерь (Lossless data compression) – метод сжатия данных (видео, аудио, графики, документов), при использовании которого закодированные данные могут быть полностью восстановлены с точностью до бита. Этот тип сжатия принципиально отличается от сжатия данных с потерями. Такой метод сжатия всегда используется в файловых архиваторах, а также иногда при сжатии графической, аудио и видео информации, при этом для каждого из типов, как правило, существуют свои оптимальные алгоритмы.

Наиболее известные алгоритмы сжатия без потерь: преобразование Барроуза-Уилера; преобразование Шиндлера; алгоритм DEFLATE; дельта-кодирование; энтропийное кодирование; инкрементное кодирование; алгоритмы Лемпеля-Зива; алгоритм Шеннона-Фано; алгоритм Хаффмана; адаптивное кодирование Хаффмана; усечённое двоичное кодирование; арифметическое кодирование; адаптивное арифметическое кодирование; кодирование растояний; энтропийное кодирование; унарное кодирование; кодирование Фибоначчи; кодиро-

вание Голомба; кодирование Райса; кодирование Элиаса.

Примерами форматов сжатия без потерь информации могут быть:

- PNG, GIF, TIFF – для графических данных;
- AVI – для видеоданных;
- ZIP, ARJ, RAR, и т.д. – для произвольных типов данных.

Сжатие данных с потерями – метод сжатия (компрессии) данных, при использовании которого распакованные данные отличаются от исходных, но степень отличия не является существенной с точки зрения их дальнейшего использования. Чаще всего метод сжатия с потерями используется для сжатия аналоговых (аудио, видео) и графических данных, IP-телефонии, передаче графики и видео в WEB, но он никогда не используется при архивации. Преимущество этого метода заключается в том, что он дает существенно больший коэффициент компрессии при умеренном снижении качества контента. Однако при использовании сжатия с потерями необходимо учитывать, что повторное сжатие с потерями снижает качество контента и при этом увеличивает размер файла. Поэтому для редактирования данных и перекодирования в другой формат необходимо иметь оригинал.

Наиболее известные алгоритмы сжатия: JPEG; линейное предсказывающее кодирование; А-закон; Мю-закон; фрактальное сжатие; трансформирующее кодирование; векторная квантизация; вейвлетное сжатие.

Примерами форматов сжатия с потерями информации могут быть:

- JPEG – для графических данных;
- MPG – для для видеоданных;
- MP3 – для аудиоданных.

Эффективность работы алгоритма сжатия характеризуется коэффициентом сжатия. Коэффициент сжатия информации – это отношение объема информации подвергнутой алгоритму сжатия к исходному объему информации.

$$K_{\text{сжат}} = \frac{N_{\text{comp}}}{N_{\text{исх}}} \cdot 100\%,$$

где: N_{comp} – размер файла после применения алгоритма сжатия;

$N_{\text{исх}}$ – размер исходного файла.

Рассмотрим метод сжатия информации без потерь как наиболее часто применяемый к текстовой информации более подробно.

Существует много разных практических методов сжатия без потери информации, которые, как правило, имеют разную эффективность для разных типов данных и разных объемов. Однако в основе этих методов лежат три теоретических алгоритма:

- алгоритм RLE (Run Length Encoding);
- алгоритмы группы KWE(KeyWord Encoding);
- алгоритм Хаффмана.

Кодирование повторов – простой алгоритм сжатия данных, который оперирует сериями данных, то есть последовательностями, в которых один и тот же символ встречается несколько раз подряд. При кодировании строка одинаковых символов, составляющих серию, заменяется строкой, которая содержит сам повторяющийся символ и количество его повторов.

Сжатие способом кодирования серий (RLE)

Наиболее известный и простой алгоритм сжатия информации обратимым путем – это кодирование серий последовательностей (Run Length Encoding - RLE). Суть данного подхода состоит в замене цепочек или серий повторяющихся байтов или их последовательностей на один кодирующий байт и счетчик числа их повторений. Проблема всех аналогичных методов заключается лишь в определении способа, при помощи которого распаковывающий алгоритм мог бы отличить в результирующем потоке байтов закодированную серию от других, не закодированных последовательностей. Решение проблемы достигается обычно простановкой меток вначале закодированных цепочек. Такими метками могут быть, например, характерные значения битов в первом байте закодированной серии, значения первого байта закодирован-

Серия 7. Теоретические и прикладные аспекты высшего профессионального образования
ной серии и т.п. Лучший, средний и худший коэффициенты сжатия – 1/32, 1/2, 2/1.

Данные методы, как правило, весьма эффективны для сжатия растровых графических изображений (BMP, PCX, TIFF), т.к. последние содержат достаточно длинных серий повторяющихся последовательностей байтов.

Недостатком метода RLE является низкая степень сжатия или стоимость кодирования файлов с малым числом серий и, что еще хуже – с малым числом повторяющихся байтов в сериях. К положительным сторонам алгоритма, пожалуй, можно отнести только то, что он не требует дополнительной памяти при работе и быстро выполняется. Интересная особенность группового кодирования в формате PCX заключается в том, что степень архивации для некоторых изображений может быть существенно повышена всего лишь за счет изменения порядка цветов в палитре изображений.

Сжатие по методу Хаффмана

Для текстовых файлов чаще других употребляется кодировка Хаффмана, заключающаяся в том, что повторяющиеся цепочки символов текста заменяются кодовыми комбинациями разной длины. При этом чем чаще данная цепочка встречается в тексте, тем короче кодовая комбинация, используемая для ее обозначения. Методика Хаффмана гарантирует кодирование с наибольшим для данного распределения коэффициентом сжатия.

Пример: Для кодирования слова: “Длинношеее” сначала производится анализ относительной частоты появления повторяющихся цепочек:

Цепочка символов	Относительная вероятность	Кодовая комбинация
eee	0,3	00
н	0,2	01
д	0,1	100
л	0,1	101
и	0,1	110
о	0,1	111
ш	0,1	1000

Кодовая комбинация для кодирования слова будет выглядеть так: 100.101.110.01.01.111.1000.00.00.00 (разделитель “.” использован для наглядности).

Вместе с этой кодовой комбинацией должна быть передана и таблица кодировки.

Применительно к сжатию изображений в основе такого метода лежит учет частоты появления одинаковых байт в изображении. При этом пикселям исходного изображения, которые встречаются большее число раз, сопоставляется код меньшей длины, а встречающимся редко – код большей длины (т.е. формируется префиксный код переменной длины). Для сбора статистики требуется два прохода по файлу: один – для просмотра и сбора статистической информации, второй – для кодирования. Коэффициенты сжатия: 1/8, 2/3, 1.

При использовании такого метода требуется запись в файл и таблицы соответствия кодируемых пикселей и кодирующих цепочек. Такое кодирование применяется в качестве последнего этапа архивации в JPEG. Методы Хаффмана дают достаточно высокую скорость и умеренно хорошее качество сжатия.

Основным недостатком данного метода является зависимость степени сжатия от близости вероятностей символов к величине 2^n , поскольку каждый символ кодируется целым числом бит. Так, при кодировании данных с двухсимвольным алфавитом сжатие всегда отсутствует, т.к., несмотря на различные вероятности появления символов во входном потоке алгоритм, фактически сводит их до 1/2. Такой алгоритм реализован в формате TIFF.

Литература

1. Ватолин Д. Ратушняк А. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео. - М., 2003. 381 с
2. Ватолин Д. и др. Методы сжатия данных. - М., 2003. 176 с
3. Сэлмон Д. Сжатие данных, изображений и звука.
4. Миано Дж. Форматы и алгоритмы сжатия изображений в действии. - М., Издательство «Триумф», 2003. 336 с.
5. Сойфер В.А. Методы компьютерной обработки изображений. Учебник.