

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования

**«Дальневосточный федеральный университет»**

**(ДВФУ)**

|  |
| --- |
| **ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ**  **Департамент математического и компьютерного моделирования** |

**ДОКЛАД**

**о практическом задание по дисциплине АИСД**

«Алгоритм Лемпеля-Зива-Велча»

направление подготовки 09.03.03 «Прикладная информатика»

профиль «Прикладная информатика в компьютерном дизайне»

|  |  |  |
| --- | --- | --- |
|  |  | Выполнил студент  гр. Б9121-09.03.03пикд  Виноходова А.А  \_\_\_\_\_\_\_\_\_\_\_\_\_ |
| Доклад защищен:  С оценкой \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  | *(подпись)*  Руководитель практики  Доцент ИМКТ А.С Кленин  *(должность, уч. звание)*  \_\_\_\_\_\_\_\_\_\_\_\_\_  *(подпись)*  «\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_ 2022г. |
| Рег. № \_\_\_\_\_\_  «\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2022 г. |  |  |

г. Владивосток

2022

Содержание

Введение

Суть и назначение

Алгоритм Лемпеля-Зива-Велча – это универсальный алгоритм сжатия данных без потерь.

Авторство (кто, год, зачем)

Алгоритм был опубликован Терри Велчем в 1984 году в качестве улучшенного алгоритма LZ78, опубликованного Абрахамом Лемпелем и Якобом Зивом в 1978 г., и получил название LZW.

История развития.

Наиболее распространенными модификациями алгоритма LZW являются: LZC (Lempel-Ziv Compress, 1985 г.), LZT (Lempel-Ziv-Tischer, 1985 г.), LZMW (Lempel-Ziv-Miller-Wegman, 1985 г.) и LZAP (Lempel-Ziv All Prefixes, 1988 г.).

Состояние, реализация

На момент своего появления алгоритм LZW стал первым широко используемым на компьютерах методом сжатия данных.

Перспектива использования

В настоящее время используется в файлах формата TIFF, PDF, GIF, PostScript и других, а также отчасти во многих популярных программах сжатия данных (ZIP, ARJ, LHA).

Описание метода

Формальное

При сжатии сообщения стандартный алгоритм LZW создает словарь строк. Каждой строке присваивается уникальных 12-битный код. Сначала словарь заполняется всеми односимвольными строками, содержащимися в сообщении. Максимальный размер словаря составляет 4096 строк с кодами. Алгоритм считывает текст сообщения посимвольно слева направо и ищет максимальную строку, которой нет в словаре – WK, где W – строка, имеющаяся в словаре, а K – символ, следующий за ней в сообщении. Найденная строка WK вносится в словарь и ей присваивается уникальный код, программа выводит код строки W, а следующая рассматриваемая строка начинается символа K. В случае переполнения словаря он продолжает использоваться без добавления строк.

Математическое

1. Инициализировать словарь со всеми односимвольными строками входного сообщения.
2. Инициализировать строку W и присвоить ей первый символ входного сообщения.
3. Если КОНЕЦ\_СООБЩЕНИЯ, то вывести код для W и завершить алгоритм.
4. Считать очередной символ K из входного сообщения.
5. Если фраза WK уже есть в словаре, то присвоить входной фразе W значение WK и перейти к шагу 3.
6. Иначе выдать код W, добавить WK в словарь, присвоить входной фразе W значение K и перейти к шагу 3.

Пример.

Входное сообщение, которое необходимо закодировать, имеет вид:

ababcbababaaaaaaa

Программа инициализирует словарь, содержащий 3 односимвольных строки (a; b; c) и присваивает им уникальные 12-битные коды (1;2;3).

|  |  |  |
| --- | --- | --- |
| Фраза | Десятичный код | Двоичный код |
| a | 1 | 000000000001 |
| b | 2 | 000000000010 |
| c | 3 | 000000000011 |

В процессе кодирования словарь пополняется и приобретает следующий вид:

|  |  |  |
| --- | --- | --- |
| Фраза | Десятичный код | Двоичный код |
| a | 1 | 000000000001 |
| b | 2 | 000000000010 |
| c | 3 | 000000000011 |
| ab | 4 | 000000000100 |
| ba | 5 | 000000000101 |
| abc | 6 | 000000000110 |
| cb | 7 | 000000000111 |
| bab | 8 | 000000001000 |
| baba | 9 | 000000001001 |
| aa | 10 | 000000001010 |
| aaa | 11 | 000000001011 |
| aaaa | 12 | 000000001100 |

Процесс кодирования:

|  |  |
| --- | --- |
| Новая фраза | Вывод |
| ab | 1 |
| ba | 2 |
| abc | 4 |
| cb | 3 |
| bab | 5 |
| baba | 8 |
| aa | 1 |
| aaa | 10 |
| aaaa | 11 |
| - | 1 |

Тогда закодированное сообщение будет выглядеть следующим образом:

124358110111

Формальная постановка задачи

Написать программу, осуществляющую кодирование входного сообщения по алгоритму Лемпеля-Зива-Велча. Программа должна инициализировать динамический словарь всех односимвольных фраз входного сообщения и в процессе кодировки пополнять его новыми фразами. Каждой фразе должен присваиваться уникальный 12-битный код. Считывание текста сообщения происходит посимвольно слева направо.

Формат входного файла:

Входной файл содержит текст в кодировке UTF-8 без разделения на абзацы.

Формат выходного файла:

Выходной файл содержит строку целых чисел без разделителей.

Заключение

В ходе работы удалось реализовать алгоритм Лемпеля-Зива-Велча.

Список литературы

1. Welch T. A. A technique for high-performance data compression // Computer. — 1984. — Т. 6, № 17. — С. 8–19. — doi:10.1109/MC.1984.1659158.
2. Dinsky [Электронный ресурс] https://youtu.be/XsllPSupzy4
3. Arnold R., Bell T.[en]. A corpus for the evaluation of lossless compression algorithms // IEEE Data Compression Conference. — 1997. — С. 201–210. — doi:10.1109/DCC.1997.582019.
4. Bell T.[en], Witten I. H.[en], Cleary J. G.[en]. Modeling for text compression // ACM Computing Surveys[en]. — 1989. — Т. 21, № 4. — С. 557–591. — doi:10.1145/76894.76896.
5. Charikar M., Lehman E., Lehman A., Liu D., Panigrahy R., Prabhakaran M., Sahai A., shelat a. The smallest grammar problem // IEEE Transactions on Information Theory[en]. — 2005. — Т. 51, № 7. — С. 2554—2576. — doi:10.1109/TIT.2005.850116.
6. De Agostino S., Silvestri R. A worst-case analysis of the LZ2 compression algorithm // Information and Computation[en]. — 1997. — Т. 139, № 2. — С. 258–268. — doi:10.1006/inco.1997.2668.
7. De Agostino S., Storer J. A. On-line versus off-line computation in dynamic text compression // Information Processing Letters[en]. — 1996. — Т. 59, № 3. — С. 169–174. — doi:10.1016/0020-0190(96)00068-3.
8. Hucke D., Lohrey M., Reh C. P. The smallest grammar problem revisited // String Processing and Information Retrieval (SPIRE). — 2016. — Т. 9954. — С. 35–49. — doi:10.1007/978-3-319-46049-9\_4.
9. Lempel A., Ziv J. Compression of individual sequences via variable-rate coding // IEEE Transactions on Information Theory[en]. — 1978. — Т. 24, № 5. — С. 530–536. — doi:10.1109/TIT.1978.1055934.
10. Miller V. S[en], Wegman M. N.[en]. Variations on a theme by Ziv and Lempel // Combinatorial algorithms on words. — 1985. — Т. 12. — С. 131–140. — doi:10.1007/978-3-642-82456-2\_9.
11. Sheinwald D. On the Ziv-Lempel proof and related topics // Proceedings of the IEEE[en]. — 1994. — Т. 82, № 6. — С. 866–871. — doi:10.1109/5.286190.
12. Storer J. A. Data Compression: Methods and Theory. — New York, USA: Computer Science Press, 1988. — 413 с. — ISBN 0-7167-8156-5.
13. Ziv J. A constrained-dictionary version of LZ78 asymptotically achieves the finite-state compressibility with a distortion measure // IEEE Information Theory Workshop. — 2015. — С. 1–4. — doi:10.1109/ITW.2015.7133077.
14. Adobe Systems Incorporated. Document management — Portable document format — Part 1: PDF 1.7 (англ.). PDF 1.7 specification. Adobe (1 июля 2008). Дата
15. Wikipedia — Lempel — Ziv — Welch
16. Семенюк В.В. — Экономное кодирование дискретной информации
17. Метод LZW — сжатия данных — алгоритмы и методы
18. Алгоритмы сжатия и компрессии
19. Алгоритм LZW — Понятие алгоритма
20. Вирт Н. Алгоритмы и структуры данных/Н. Вирт. М.: Мир,1989.
21. Сибуя М. Алгоритмы обработки данных/М. Сибуя, Т. Ямамото. М.: Мир,1986.
22. Костин А.Е. Организация и обработка структур данных в вычислительных системах: учеб.пособ. для вузов/А.Е. Костин, В.Ф. Шаньгин . М.: Высш.шк., 1987.
23. Кнут Д. Искусство программирования для ЭВМ.Т.1: Основные алгоритмы:пер. с англ./Д. Кнут. М.:Мир,1978.
24. Кнут Д. Искусство программирования для ЭВМ.Т.3: Сортировка и поиск.: пер. с англ./Д.Кнут. М.:Мир,1978.
25. Кормен Т. Алгоритмы: построение и анализ./Т. Кормен, Ч. Лейзерсон, Р.Ривест. М.: МЦНМО, 2000
26. Кричевский Р.Е. Сжатие и поиск информации/Р.Е. Кричевский. М.: Радио и связь, 1989
27. Интернет ресурс. https://habr.com/ru/post/132683/