

Алгоритм Lempel- Ziv-Welch

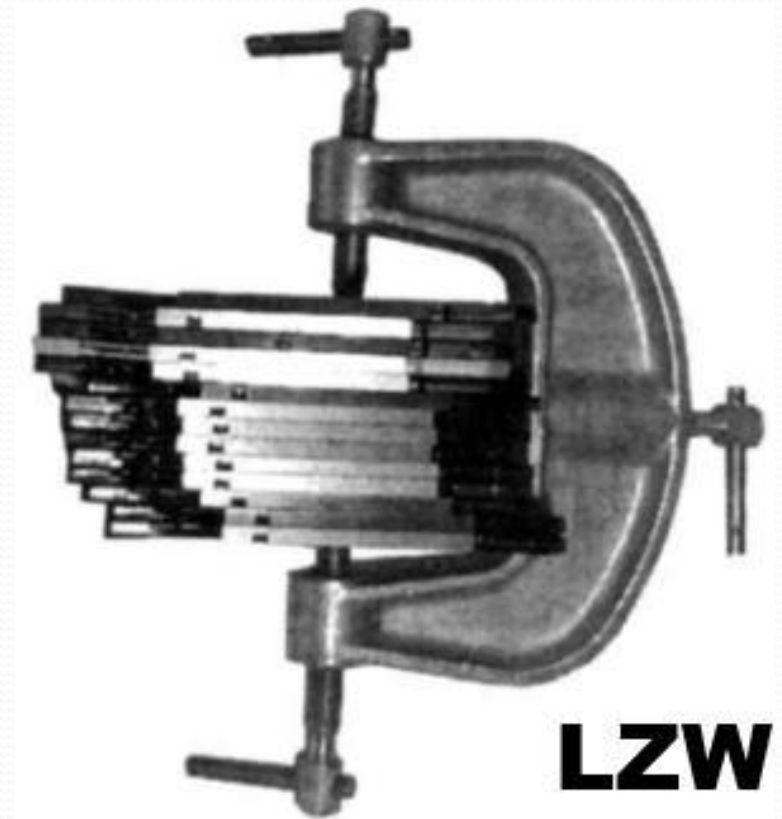
- **Алгоритм Лемпеля — Зива — Велча** (**Lempel-Ziv-Welch, LZW**) — это универсальный алгоритм сжатия данных без потерь, созданный Абрахамом Лемпелем ([англ.](#) *Abraham Lempel*), Якобом Зивом (англ. *Jacob Ziv*) и Терри Велчем (англ. *Terry Welch*). Он был опубликован Велчем в 1984 году, в качестве улучшенной реализации алгоритма LZ78, опубликованного Лемпелем и Зивом в 1978 году.

- Акроним «LZW» указывает на фамилии изобретателей алгоритма: Лемпель, Зив и Велч, но многие утверждают, что, поскольку патент принадлежал Зиву, то метод должен называться *алгоритмом Зива — Лемпеля — Велча*.

The image shows the letters 'LZW' in a large, bold, green, sans-serif font. The letters are centered within a white rectangular area, which is itself set against a light gray background.

Применение

- На момент своего появления алгоритм LZW давал лучший коэффициент сжатия, для большинства приложений, чем любой другой хорошо известный метод того времени. Он стал первым широко используемым на компьютерах методом сжатия данных.



- В 1987 году алгоритм стал частью стандарта на формат изображений GIF. Он также может (опционально) использоваться в формате TIFF.
- В настоящее время алгоритм содержится в стандарте PDF.



Пример

- Данный пример показывает алгоритм LZW в действии, показывая состояние выходных данных и словаря на каждой стадии, как при кодировании, так и при декодировании сообщения. С тем чтобы сделать изложение проще, мы ограничимся простым алфавитом — только заглавные буквы, без знаков препинания и пробелов. Сообщение, которое нужно сжать, выглядит следующим образом:
- TOBEORNOTTOBEORTOBEORNOT#

Маркер # используется для обозначения конца сообщения.

Кодирование

- Без использования алгоритма LZW, при передаче сообщения, как оно есть — 25 символов по 5 бит на каждый — оно займёт 125 бит. Сравним это с тем, что получается при использовании LZW:



- Таким образом, используя LZW, мы сократили сообщение на 29 бит из 125 — это почти 22 %.

Символ: Битовый код: Новая запись словаря:
(на выходе)

T	20 =	10100	
O	15 =	01111	27: TO
B	2 =	00010	28: OB
E	5 =	00101	29: BE
O	15 =	01111	30: EO
R	18 =	10010	31: OR <--

-со следующего символа начинаем использовать 6-битные группы

N	14 =	001110	32: RN
O	15 =	001111	33: NO
T	20 =	010100	34: OT
TO	27 =	011011	35: TT
BE	29 =	011101	36: TOB
OR	31 =	011111	37: BEO
TOB	36 =	100100	38: ORT
EO	30 =	011110	39: TOBE
RN	32 =	100000	40: EOR
OT	34 =	100010	41: RNO
#	0 =	000000	42: OT#

Общая длина = $6 \cdot 5 + 11 \cdot 6 = 96$ бит.

Патенты

PATENTED

PATENTED



PATENTED

- На алгоритм LZW и его вариации был выдан ряд патентов, как в США, так и в других странах. На LZ78 был выдан американский патент U.S. Patent 4 464 650, принадлежащий Sperry Corporation, позднее ставшей частью Unisys Corporation. На LZW в США были выданы два патента: U.S. Patent 4 814 746, принадлежащий IBM, и патент Велча U.S. Patent 4 558 302 (выдан 20 июня 1983 года), принадлежащий Sperry Corporation, позднее перешедший к Unisys Corporation. К настоящему времени сроки всех патентов истекли.

PATENTED

Unisys, GIF и PNG

При разработке формата GIF в CompuServe не знали о существовании патента U.S. Patent 4 558 302 . В декабре 1994 года, когда в Unisys стало известно об использовании LZW в широко используемом графическом формате, эта компания распространила информацию о своих планах по взысканию лицензионных отчислений с коммерческих программ, имеющих возможность по созданию GIF-файлов. В то время формат был уже настолько широко распространён, что большинство компаний-производителей ПО не имели другого выхода, кроме как заплатить. Эта ситуация стала одной из причин разработки графического формата PNG (неофициальная расшифровка: «PNG's Not GIF»), ставшего третьим по распространённости в WWW, после GIF и JPEG.



- В конце августа 1999 года Unisys прервала действие безвозмездных лицензий на LZW для бесплатного и некоммерческого ПО, а также для пользователей нелицензированных программ, призвав *League for Programming Freedom* развернуть кампанию «сожжём все GIF'ы» и информировать публику об имеющихся альтернативах.

- 20 июня 2003 года истёк срок оригинального американского патента, что означает, что Unisys не может больше собирать по нему лицензионные отчисления. Аналогичные патенты в Европе, Японии и Канаде истекли в 2004 году.





Конец