

Dictionary-based Encoding

*Computer Science Department
CS4481b/9628b: Image Compression
Winter 2017
Instructor: Mahmoud R. El-Sakka
Office: MC-419
Email: elsakka@csd.uwo.ca
Phone: 519-661-2111 x86996*

1

Topic 07: Dictionary-based Encoding

Dictionary-based Compressions

- How many words do we have in a dictionary?
 - Pocket dictionary: about 25,000 words
 - Full-size dictionary: about 60,000 words
- If we will give a sequential number for each of these words, how many bits do we need?
 - Pocket dictionary: we only need 15 bits (0 to 32,767)
 - Full-size dictionary: we only need 16 bits (0 to 64,535)

Dictionary-based Compressions

- Assume that an English word consists of 6 letters (on average)
- Consider having an English dictionary containing half a million words (without their definition)
 - This dictionary is searched for each word needs to be encoded
 - If a match is found,
 - this word is encoded by a pointer to that word (needs 20 bits)
 - 1 bit to say that match was found
 - 19 bits as a dictionary index to that word
 - Otherwise,
 - the word is encoded without any compression, i.e., a raw word (needs 50 bits on average),
 - 1 bit to say that match was not found
 - 7 bits to identify the length of the word
 - 6 characters per word (on average) \times 7 bits per characters
 - Assume that P is the probability of finding a word in the dictionary

Dictionary-based Compressions

- After reading and compressing N words, the size of the compressed file will be (on average)

$$N \times (20 \times P + 50 \times (1 - P)) \text{ bits},$$

- Without compression, we need

$$N \times 49 \text{ bits}$$

to encode these N words

- To achieve compression, the following relation must be hold

$$N \times (20 \times P + 50 \times (1 - P)) < N \times 49$$

$$(20 \times P + 50 \times (1 - P)) < 49$$

$$- 30 \times P + 50 < 49$$

or in other word, P must be greater than $1/30$!!

Statistical-based Versus Dictionary-based Compressions

- Statistical-based compression schemes
 - Use a statistical model of the data
 - The quality of the achieved compression depends on how good that model is
- Dictionary-based compression schemes
 - A data string is partitioned into sub-strings of symbols
 - Each sub-string is encoded as a dictionary index
 - The dictionary holds sub-strings of symbols and it may be
 - Static dictionary
 - Holds *predetermined* strings (*image independent*)
 - Sometimes allows strings addition
 - Does not allow strings deletion
 - Dynamic (adaptive) dictionary
 - Holds strings *previously found* in the input stream (*image dependent*)
 - Allows strings addition and deletion
 - Perhaps the most famous dictionary-based compression schemes are LZ77 and LZ78

LZ77

- LZ77 is a dictionary-based compression scheme described by Ziv and Lempel in 1977, and hence the name LZ77
- The main idea of LZ77 is to use *part* of the previously seen input stream as a dictionary (*searching buffer*)
 - The searching buffer is a sliding window on the input characters/pixels that have recently encoded
- The encoder scans the search buffer looking for a match to the current data symbol to be encoded
- If a match is found, it then tries to match as many following symbols as possible
- The encoder then continues the scan, trying to find longer matches
- The longest match is selected to generate the token
- LZ77 token has three parts
 - *an offset*,
 - *a match length*, and
 - *the code of the next mismatched symbol (from the input stream)* **Why?**

[illegible]

CS4481/9628: Image Compression

[illegible]

CS4481/9628: Image Compression

Encoding of "sir sid eastman easily teases sea sick seal" using LZ77

No.	Searching buffer	Data to be encoded	token
1		sir sid eastman easily teases sea sick seal	
2	s	ir sid eastman easily teases sea sick seal	(0, 0, "s")
3	si	r sid eastman easily teases sea sick seal	(0, 0, "i")
4	sir	sid eastman easily teases sea sick seal	(0, 0, "r")
5	sir	sid eastman easily teases sea sick seal	(0, 0, " ")
6	sir sid	eastman easily teases sea sick seal	(4, 2, "d")
7	sir sid e	astman easily teases sea sick seal	(4, 1, "e")
8	sir sid ea	stman easily teases sea sick seal	(0, 0, "a")
9	sir sid east	man easily teases sea sick seal	(6, 1, "t")
10	sir sid eastm	an easily teases sea sick seal	(0, 0, "n")
11	sir sid eastman	easily teases sea sick seal	(4, 1, "n")
12	sir sid eastman easi	ly teases sea sick seal	(8, 4, "i")
13	sir sid eastman easil	y teases sea sick seal	(0, 0, "l")

Encoding of "sir sid eastman easily teases sea sick seal" using LZ77

No.	Searching buffer	Data to be encoded	token
1		sir sid eastman easily teases sea sick seal	
2	s	ir sid eastman easily teases sea sick seal	(0, 0, "s")
3	si	r sid eastman easily teases sea sick seal	(0, 0, "i")
4	sir	sid eastman easily teases sea sick seal	(0, 0, "r")
5	sir	sid eastman easily teases sea sick seal	(0, 0, " ")
6	sir sid	eastman easily teases sea sick seal	(4, 2, "d")
7	sir sid e	astman easily teases sea sick seal	(4, 1, "e")
8	sir sid ea	stman easily teases sea sick seal	(0, 0, "a")
9	sir sid east	man easily teases sea sick seal	(6, 1, "t")
10	sir sid eastm	an easily teases sea sick seal	(0, 0, "n")
11	sir sid eastman	easily teases sea sick seal	(4, 1, "n")
12	sir sid eastman easi	ly teases sea sick seal	(8, 4, "i")
13	sir sid eastman easil	y teases sea sick seal	(0, 0, "l")
14	sir sid eastman easily	teases sea sick seal	(0, 0, "y")

Encoding of "sir sid eastman easily teases sea sick seal" using LZ77

No.	Searching buffer	Data to be encoded	token
1		sir sid eastman easily teases sea sick seal	
2	s	ir sid eastman easily teases sea sick seal	(0, 0, "s")
3	si	r sid eastman easily teases sea sick seal	(0, 0, "i")
4	sir	sid eastman easily teases sea sick seal	(0, 0, "r")
5	sir	sid eastman easily teases sea sick seal	(0, 0, " ")
6	sir sid	eastman easily teases sea sick seal	(4, 2, "d")
7	sir sid e	astman easily teases sea sick seal	(4, 1, "e")
8	sir sid ea	stman easily teases sea sick seal	(0, 0, "a")
9	sir sid east	man easily teases sea sick seal	(6, 1, "t")
10	sir sid eastm	an easily teases sea sick seal	(0, 0, "m")
11	sir sid eastman	easily teases sea sick seal	(4, 1, "n")
12	sir sid eastman easi	ly teases sea sick seal	(8, 4, "i")
13	sir sid eastman easil	y teases sea sick seal	(0, 0, "l")
14	sir sid eastman easily	teases sea sick seal	(0, 0, "y")
15	sir sid eastman easily t	eases sea sick seal	(7, 1, "t")

Encoding of "sir sid eastman easily teases sea sick seal" using LZ77

No.	Searching buffer	Data to be encoded	token
1		sir sid eastman easily teases sea sick seal	
2	s	ir sid eastman easily teases sea sick seal	(0, 0, "s")
3	si	r sid eastman easily teases sea sick seal	(0, 0, "i")
4	sir	sid eastman easily teases sea sick seal	(0, 0, "r")
5	sir	sid eastman easily teases sea sick seal	(0, 0, " ")
6	sir sid	eastman easily teases sea sick seal	(4, 2, "d")
7	sir sid e	astman easily teases sea sick seal	(4, 1, "e")
8	sir sid ea	stman easily teases sea sick seal	(0, 0, "a")
9	sir sid east	man easily teases sea sick seal	(6, 1, "t")
10	sir sid eastm	an easily teases sea sick seal	(0, 0, "m")
11	sir sid eastman	easily teases sea sick seal	(4, 1, "n")
12	sir sid eastman easi	ly teases sea sick seal	(8, 4, "i")
13	sir sid eastman easil	y teases sea sick seal	(0, 0, "l")
14	sir sid eastman easily	teases sea sick seal	(0, 0, "y")
15	sir sid eastman easily t	eases sea sick seal	(7, 1, "t")
16	sir sid eastman easily teases	s sea sick seal	(8, 3, "e")

Encoding of "sir sid eastman easily teases sea sick seal" using LZ77

No.	Searching buffer	Data to be encoded	token
1		sir sid eastman easily teases sea sick seal	
2	s	ir sid eastman easily teases sea sick seal	(0, 0, "s")
3	si	r sid eastman easily teases sea sick seal	(0, 0, "i")
4	sir	sid eastman easily teases sea sick seal	(0, 0, "r")
5	sir	sid eastman easily teases sea sick seal	(0, 0, " ")
6	sir sid	eastman easily teases sea sick seal	(4, 2, "d")
7	sir sid e	astman easily teases sea sick seal	(4, 1, "e")
8	sir sid ea	stman easily teases sea sick seal	(0, 0, "a")
9	sir sid east	man easily teases sea sick seal	(6, 1, "t")
10	sir sid eastm	an easily teases sea sick seal	(0, 0, "m")
11	sir sid eastman	easily teases sea sick seal	(4, 1, "n")
12	sir sid eastman easi	ly teases sea sick seal	(8, 4, "i")
13	sir sid eastman easil	y teases sea sick seal	(0, 0, "l")
14	sir sid eastman easily	teases sea sick seal	(0, 0, "y")
15	sir sid eastman easily t	eases sea sick seal	(7, 1, "t")
16	sir sid eastman easily tease	s sea sick seal	(8, 3, "e")
17	sir sid eastman easily teases	sea sick seal	(2, 1, " ")

Encoding of "sir sid eastman easily teases sea sick seal" using LZ77

No.	Searching buffer	Data to be encoded	token
1		sir sid eastman easily teases sea sick seal	
2	s	ir sid eastman easily teases sea sick seal	(0, 0, "s")
3	si	r sid eastman easily teases sea sick seal	(0, 0, "i")
4	sir	sid eastman easily teases sea sick seal	(0, 0, "r")
5	sir	sid eastman easily teases sea sick seal	(0, 0, " ")
6	sir sid	eastman easily teases sea sick seal	(4, 2, "d")
7	sir sid e	astman easily teases sea sick seal	(4, 1, "e")
8	sir sid ea	stman easily teases sea sick seal	(0, 0, "a")
9	sir sid east	man easily teases sea sick seal	(6, 1, "t")
10	sir sid eastm	an easily teases sea sick seal	(0, 0, "m")
11	sir sid eastman	easily teases sea sick seal	(4, 1, "n")
12	sir sid eastman easi	ly teases sea sick seal	(8, 4, "i")
13	sir sid eastman easil	y teases sea sick seal	(0, 0, "l")
14	sir sid eastman easily	teases sea sick seal	(0, 0, "y")
15	sir sid eastman easily t	eases sea sick seal	(7, 1, "t")
16	sir sid eastman easily tease	s sea sick seal	(8, 3, "e")
17	sir sid eastman easily teases	sea sick seal	(2, 1, " ")
18	sir sid eastman easily teases sea	sick seal	(4, 2, "a")

Encoding of "sir sid eastman easily teases sea sick seal" using LZ77

No.	Searching buffer	Data to be encoded	token
1		sir sid eastman easily teases sea sick seal	
2	s	ir sid eastman easily teases sea sick seal	(0, 0, "s")
3	si	r sid eastman easily teases sea sick seal	(0, 0, "i")
4	sir	sid eastman easily teases sea sick seal	(0, 0, "r")
5	sir	sid eastman easily teases sea sick seal	(0, 0, " ")
6	sir sid	eastman easily teases sea sick seal	(4, 2, "d")
7	sir sid e	astman easily teases sea sick seal	(4, 1, "e")
8	sir sid ea	stman easily teases sea sick seal	(0, 0, "a")
9	sir sid east	man easily teases sea sick seal	(6, 1, "t")
10	sir sid eastm	an easily teases sea sick seal	(0, 0, "m")
11	sir sid eastman	easily teases sea sick seal	(4, 1, "n")
12	sir sid eastman easi	ly teases sea sick seal	(8, 4, "i")
13	sir sid eastman easil	y teases sea sick seal	(0, 0, "l")
14	sir sid eastman easily	teases sea sick seal	(0, 0, "y")
15	sir sid eastman easily t	eases sea sick seal	(7, 1, "t")
16	sir sid eastman easily tease	s sea sick seal	(8, 3, "e")
17	sir sid eastman easily teases	sea sick seal	(2, 1, " ")
18	sir sid eastman easily teases sea	sick seal	(4, 2, "a")
19	sir sid eastman easily teases sea sick	k seal	(30, 3, "c")

Encoding of "sir sid eastman easily teases sea sick seal" using LZ77

No.	Searching buffer	Data to be encoded	token
1		sir sid eastman easily teases sea sick seal	
2	s	ir sid eastman easily teases sea sick seal	(0, 0, "s")
3	si	r sid eastman easily teases sea sick seal	(0, 0, "i")
4	sir	sid eastman easily teases sea sick seal	(0, 0, "r")
5	sir	sid eastman easily teases sea sick seal	(0, 0, " ")
6	sir sid	eastman easily teases sea sick seal	(4, 2, "d")
7	sir sid e	astman easily teases sea sick seal	(4, 1, "e")
8	sir sid ea	stman easily teases sea sick seal	(0, 0, "a")
9	sir sid east	man easily teases sea sick seal	(6, 1, "t")
10	sir sid eastm	an easily teases sea sick seal	(0, 0, "m")
11	sir sid eastman	easily teases sea sick seal	(4, 1, "n")
12	sir sid eastman easi	ly teases sea sick seal	(8, 4, "i")
13	sir sid eastman easil	y teases sea sick seal	(0, 0, "l")
14	sir sid eastman easily	teases sea sick seal	(0, 0, "y")
15	sir sid eastman easily t	eases sea sick seal	(7, 1, "t")
16	sir sid eastman easily tease	s sea sick seal	(8, 3, "e")
17	sir sid eastman easily teases	sea sick seal	(2, 1, " ")
18	sir sid eastman easily teases sea	sick seal	(4, 2, "a")
19	sir sid eastman easily teases sea sic	k seal	(30, 3, "c")
20	sir sid eastman easily teases sea sick	seal	(0, 0, "k")

Encoding of "sir sid eastman easily teases sea sick seal" using LZ77

No.	Searching buffer	Data to be encoded	token
1		sir sid eastman easily teases sea sick seal	
2	s	ir sid eastman easily teases sea sick seal	(0, 0, "s")
3	si	r sid eastman easily teases sea sick seal	(0, 0, "i")
4	sir	sid eastman easily teases sea sick seal	(0, 0, "r")
5	sir	sid eastman easily teases sea sick seal	(0, 0, " ")
6	sir sid	eastman easily teases sea sick seal	(4, 2, "d")
7	sir sid e	astman easily teases sea sick seal	(4, 1, "e")
8	sir sid ea	stman easily teases sea sick seal	(0, 0, "a")
9	sir sid east	man easily teases sea sick seal	(6, 1, "t")
10	sir sid eastm	an easily teases sea sick seal	(0, 0, "m")
11	sir sid eastman	easily teases sea sick seal	(4, 1, "n")
12	sir sid eastman easi	ly teases sea sick seal	(8, 4, "i")
13	sir sid eastman easil	y teases sea sick seal	(0, 0, "l")
14	sir sid eastman easily	teases sea sick seal	(0, 0, "y")
15	sir sid eastman easily t	eases sea sick seal	(7, 1, "t")
16	sir sid eastman easily tease	s sea sick seal	(8, 3, "e")
17	sir sid eastman easily teases	sea sick seal	(2, 1, " ")
18	sir sid eastman easily teases sea	sick seal	(4, 2, "a")
19	sir sid eastman easily teases sea sic	k seal	(30, 3, "c")
20	sir sid eastman easily teases sea sick	seal	(0, 0, "k")
21	sir sid eastman easily teases sea sick seal		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2		(0, 0, "s")
3		(0, 0, "i")
4		(0, 0, "r")
5		(0, 0, " ")
6		(4, 2, "d")
7		(4, 1, "e")
8		(0, 0, "a")
9		(6, 1, "t")
10		(0, 0, "m")
11		(4, 1, "n")
12		(8, 4, "i")
13		(0, 0, "l")
14		(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2		s (0, 0, "s")
3		(0, 0, "i")
4		(0, 0, "r")
5		(0, 0, " ")
6		(4, 2, "d")
7		(4, 1, "e")
8		(0, 0, "a")
9		(6, 1, "t")
10		(0, 0, "m")
11		(4, 1, "n")
12		(8, 4, "i")
13		(0, 0, "l")
14		(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2		s (0, 0, "s")
3		si (0, 0, "i")
4		(0, 0, "r")
5		(0, 0, " ")
6		(4, 2, "d")
7		(4, 1, "e")
8		(0, 0, "a")
9		(6, 1, "t")
10		(0, 0, "m")
11		(4, 1, "n")
12		(8, 4, "i")
13		(0, 0, "l")
14		(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5		(0, 0, " ")
6		(4, 2, "d")
7		(4, 1, "e")
8		(0, 0, "a")
9		(6, 1, "t")
10		(0, 0, "m")
11		(4, 1, "n")
12		(8, 4, "i")
13		(0, 0, "l")
14		(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6		(4, 2, "d")
7		(4, 1, "e")
8		(0, 0, "a")
9		(6, 1, "t")
10		(0, 0, "m")
11		(4, 1, "n")
12		(8, 4, "i")
13		(0, 0, "l")
14		(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	si ^r	(0, 0, " ")
6	si ^r si ^d	(4, 2, "d")
7		(4, 1, "e")
8		(0, 0, "a")
9		(6, 1, "t")
10		(0, 0, "m")
11		(4, 1, "n")
12		(8, 4, "i")
13		(0, 0, "l")
14		(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	si ^r	(0, 0, " ")
6	si ^r si ^d	(4, 2, "d")
7	si ^r si ^d e	(4, 1, "e")
8		(0, 0, "a")
9		(6, 1, "t")
10		(0, 0, "m")
11		(4, 1, "n")
12		(8, 4, "i")
13		(0, 0, "l")
14		(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9		(6, 1, "t")
10		(0, 0, "m")
11		(4, 1, "n")
12		(8, 4, "i")
13		(0, 0, "l")
14		(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9	sir sid east	(6, 1, "t")
10		(0, 0, "m")
11		(4, 1, "n")
12		(8, 4, "i")
13		(0, 0, "l")
14		(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9	sir sid east	(6, 1, "t")
10	sir sid eastm	(0, 0, "m")
11		(4, 1, "n")
12		(8, 4, "i")
13		(0, 0, "l")
14		(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9	sir sid east	(6, 1, "t")
10	sir sid eastm	(0, 0, "m")
11	sir sid eastman	(4, 1, "n")
12		(8, 4, "i")
13		(0, 0, "l")
14		(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9	sir sid east	(6, 1, "t")
10	sir sid eastm	(0, 0, "m")
11	sir sid eastman	(4, 1, "n")
12	sir sid eastman easi	(8, 4, "i")
13		(0, 0, "l")
14		(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9	sir sid east	(6, 1, "t")
10	sir sid eastm	(0, 0, "m")
11	sir sid eastman	(4, 1, "n")
12	sir sid eastman easi	(8, 4, "i")
13	sir sid eastman easil	(0, 0, "l")
14		(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9	sir sid east	(6, 1, "t")
10	sir sid eastm	(0, 0, "m")
11	sir sid eastman	(4, 1, "n")
12	sir sid eastman easi	(8, 4, "i")
13	sir sid eastman easil	(0, 0, "l")
14	sir sid eastman easily	(0, 0, "y")
15		(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9	sir sid east	(6, 1, "t")
10	sir sid eastm	(0, 0, "m")
11	sir sid eastman	(4, 1, "n")
12	sir sid eastman easi	(8, 4, "i")
13	sir sid eastman easil	(0, 0, "l")
14	sir sid eastman easily	(0, 0, "y")
15	sir sid eastman easily t	(7, 1, "t")
16		(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9	sir sid east	(6, 1, "t")
10	sir sid eastm	(0, 0, "m")
11	sir sid eastman	(4, 1, "n")
12	sir sid eastman easi	(8, 4, "i")
13	sir sid eastman easil	(0, 0, "l")
14	sir sid eastman easily	(0, 0, "y")
15	sir sid eastman easily t	(7, 1, "t")
16	sir sid eastman easily tease	(8, 3, "e")
17		(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9	sir sid east	(6, 1, "t")
10	sir sid eastm	(0, 0, "m")
11	sir sid eastman	(4, 1, "n")
12	sir sid eastman easi	(8, 4, "i")
13	sir sid eastman easil	(0, 0, "l")
14	sir sid eastman easily	(0, 0, "y")
15	sir sid eastman easily t	(7, 1, "t")
16	sir sid eastman easily tease	(8, 3, "e")
17	sir sid eastman easily teases	(2, 1, " ")
18		(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9	sir sid east	(6, 1, "t")
10	sir sid eastm	(0, 0, "m")
11	sir sid eastman	(4, 1, "n")
12	sir sid eastman easi	(8, 4, "i")
13	sir sid eastman easil	(0, 0, "l")
14	sir sid eastman easily	(0, 0, "y")
15	sir sid eastman easily t	(7, 1, "t")
16	sir sid eastman easily tease	(8, 3, "e")
17	sir sid eastman easily teases	(2, 1, " ")
18	sir sid eastman easily teases sea	(4, 2, "a")
19		(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9	sir sid east	(6, 1, "t")
10	sir sid eastm	(0, 0, "m")
11	sir sid eastman	(4, 1, "n")
12	sir sid eastman easi	(8, 4, "i")
13	sir sid eastman easil	(0, 0, "l")
14	sir sid eastman easily	(0, 0, "y")
15	sir sid eastman easily t	(7, 1, "t")
16	sir sid eastman easily tease	(8, 3, "e")
17	sir sid eastman easily teases	(2, 1, " ")
18	sir sid eastman easily teases sea	(4, 2, "a")
19	sir sid eastman easily teases sea sic	(30, 3, "c")
20		(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9	sir sid east	(6, 1, "t")
10	sir sid eastm	(0, 0, "m")
11	sir sid eastman	(4, 1, "n")
12	sir sid eastman easi	(8, 4, "i")
13	sir sid eastman easil	(0, 0, "l")
14	sir sid eastman easily	(0, 0, "y")
15	sir sid eastman easily t	(7, 1, "t")
16	sir sid eastman easily tease	(8, 3, "e")
17	sir sid eastman easily teases	(2, 1, " ")
18	sir sid eastman easily teases sea	(4, 2, "a")
19	sir sid eastman easily teases sea sic	(30, 3, "c")
20	sir sid eastman easily teases sea sick	(0, 0, "k")
21		(9, 4, "l")

Decoding using LZ77

No.	Searching buffer	token
1		
2	s	(0, 0, "s")
3	si	(0, 0, "i")
4	sir	(0, 0, "r")
5	sir	(0, 0, " ")
6	sir sid	(4, 2, "d")
7	sir sid e	(4, 1, "e")
8	sir sid ea	(0, 0, "a")
9	sir sid east	(6, 1, "t")
10	sir sid eastm	(0, 0, "m")
11	sir sid eastman	(4, 1, "n")
12	sir sid eastman easi	(8, 4, "i")
13	sir sid eastman easil	(0, 0, "l")
14	sir sid eastman easily	(0, 0, "y")
15	sir sid eastman easily t	(7, 1, "t")
16	sir sid eastman easily tease	(8, 3, "e")
17	sir sid eastman easily teases	(2, 1, " ")
18	sir sid eastman easily teases sea	(4, 2, "a")
19	sir sid eastman easily teases sea sic	(30, 3, "c")
20	sir sid eastman easily teases sea sick	(0, 0, "k")
21	sir sid eastman easily teases sea sick seal	(9, 4, "l")

LZ77

- the *length* field may go beyond the searching window

□ Example 1:

searching buffer	Data to be encoded	token
he grows alf	alfa in his garden	

LZ77

- the *length* field may go beyond the search window

□ Example 1:

searching buffer	Data to be encoded	token
he grows alf	alfa in his garden	
he grows alfalfa	in his garden	(3, 4, " ")

LZ77

- the *length* field may go beyond the search window

□ Example 1:

searching buffer	Data to be encoded	token
he grows alf	alfa in his garden	
he grows alfalfa	in his garden	(3, 4, " ")
he grows alfalfa i	n his garden	(0, 0, "i")

LZ77

- the *length* field may go beyond the search window

□ Example 1:

searching buffer	Data to be encoded	token
he grows alf	alfa in his garden	
he grows alfalfa	in his garden	(3, 4, " ")
he grows alfalfa i	n his garden	(0, 0, "i")

□ Example 2:

searching buffer	Data to be encoded	token
I said sh	shshshshsh please	

LZ77

- the *length* field may go beyond the search window

□ Example 1:

searching buffer	Data to be encoded	token
he grows alf	alfa in his garden	
he grows alfalfa	in his garden	(3, 4, " ")
he grows alfalfa i	n his garden	(0, 0, "i")

□ Example 2:

searching buffer	Data to be encoded	token
I said sh	shshshshsh please	
I said shshshshshsh	please	(2, 10, " ")

LZ77

- the *length* field may go beyond the search window

□ Example 1:

searching buffer	Data to be encoded	token
he grows alf	alfa in his garden	
he grows alfalfa	in his garden	(3, 4, " ")
he grows alfalfa i	n his garden	(0, 0, "i")

□ Example 2:

searching buffer	Data to be encoded	token
I said sh	shshshshsh please	
I said shshshshshsh	please	(2, 10, " ")
I said shshshshshsh p	lease	(0, 0, "p")

LZ77

- To keep track of the occurrence of more patterns, the length of the searching buffer should be long
- The length of the searching buffer represents a trade-off between
 - Capturing more similarities within the file being compressed
 - Number of bits needed to represent *offset* values
- The LZ77 approach implicitly assumes that similar patterns occur close to each other
- The LZ77 algorithm, or a variant of it, followed by a variable-length encoding is used in:
 - Portable Network Graphics (**PNG**) image format
 - Popular compression packages, e.g., **PKZip**, **Zip**, and **gzip**

LZ77

- Using triple values to encode a single character can be inefficient
- Instead of a single LZ77's token format, two different kinds of token formats can be used:
 - when a sufficiently long match is found
 - a pointer token (*flag*, *match offset* and *match length*) is sent
 - otherwise
 - a character token (*flag* and *character*) is sent
- This method is called **LZSS**