



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN TP.HCM
KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN CÔNG NGHỆ PHẦN MỀM
MÔN: **CẤU TRÚC DỮ LIỆU VÀ GIẢI THUẬT**

BÁO CÁO ĐỒ ÁN CUỐI KỲ LỚP 16CNTN

LÊ THÀNH CÔNG
MSSV: 1612842

TP.HCM, ngày 08 tháng 12 năm 2017

1. Bảng tần số và bảng mã bit của thuật toán nén Huffman tĩnh

a. Bảng tần số (trích ra từ file data)

Kí tự	Tần số	Kí tự	Tần số	Kí tự	Tần số	Kí tự	Tần số
'\n'	3512	-	30	7	4	E	162
"	3512	.	4376	8	3	F	2
' '	27161	/	6709	9	2	G	2
!	1	0	9	:	4315	H	1
"	1	1	19	;	902	I	2
'	1359	2	7	=	6	J	4
(480	3	4	A	86	K	6
)	480	4	7	B	84	L	2
+	6	5	7	C	4	M	14
,	5145	6	1	D	4	N	86

Kí tự	Tần số	Kí tự	Tần số	Kí tự	Tần số	Kí tự	Tần số
O	2	b	1771	l	3850	v	2504
P	2	c	4779	m	2825	w	506
Q	1	d	3409	n	12127	x	440
S	11	e	4194	o	3063	y	1458
T	12	f	856	p	2541	z	232
U	4	g	4185	q	325	'	602
V	4	h	6602	r	4070	À	89
W	2	i	8080	s	3988	Ô	1
Y	2	j	1004	t	7710	Ò	1
a	5045	k	1946	u	3328	à	935

Kí tự	Tần số	Kí tự	Tần số	Kí tự	Tần số	Kí tự	Tần số
á	992	ò	159	ĩ	73	ə	1539
â	344	ó	465	ŋ	35	ε	83
ã	114	ô	469	ũ	43	з	22
æ	360	õ	22	ơ	280	ı	286
è	25	ù	111	ư	909	ŋ	69
é	122	ú	260	Õ	5	ƒ	307
ê	467	ý	128	ə	2	ư	54
ì	241	ă	221	ɑ	90	Λ	283
í	383	Đ	1	ɒ	50	3	96
ồ	51	đ	1613	ɔ	382	đ	63

Kí tự	Tần số	Kí tự	Tần số	Kí tự	Tần số	Kí tự	Tần số
'	294	3	67	ă	53	é	582
:	32	æ	1	â	342	è	349
:	9	ə	57	ă	325	ẽ	247
Δ	13	3	3	ă	63	ề	82
Λ	7	Λ	2	ă	25	ệ	435
α	3	ạ	596	ă	6	ỉ	97
ε	1	ả	518	ặ	150	ị	302
η	90	á	460	ẹ	42	ơ	231
θ	74	à	255	è	62	ò	148
3	11	ã	68	ẽ	34	ó	458

Kí tự	Tần số	Kí tự	Tần số	Kí tự	Tần số
ồ	202	ủ	196	'	5
ỗ	157	ứ	236	...	6
õ	69	ừ	99]	59
ộ	510	ử	119		1
ớ	303	ữ	109		
ờ	414	ự	1060		
ở	208	ỳ	25		
ỡ	48	ỷ	14		
ợ	299	ỹ	27		
ụ	167		1		

b. Bảng mã bit (sau khi xây dựng cây Huffman)

Kí tự	Mã bit	Kí tự	Mã bit	Kí tự	Mã bit
'\n'	111011	-	000101010010	7	010010111100110
"	111010	.	01011	8	1111001000000011
' '	110	/	10111	9	0100101111010100
!	10100000100100101	0	01001011110111	:	01010
"	01101110110001011	1	0110111011001	;	11110011
'	1000101	2	111101010001100	=	101010110100001
(00010000	3	010010111100101	A	10001000110
)	00010001	4	111101010001110	B	10001000101
+	101010110100010	5	111101010001111	C	010010111100001
,	10000	6	01101110110001001	D	010010111100000

Kí tự	Mã bit	Kí tự	Mã bit	Kí tự	Mã bit
E	1000100000	O	0100101111010001	b	1111000
F	0100101111010110	P	0100101111010011	c	01100
G	0100101111010010	Q	01001011110110101	d	111001
H	01001011110110100	S	10101010100010	e	01000
I	0100101111011000	T	10101010100011	f	10101111
J	010010111100011	U	010010111100010	g	00111

K	101010110100000	V	010010111100100	h	10110
L	0100101111001111	W	0100101111010111	i	0010
M	11110101000100	Y	0100101111010101	j	0001011
N	10001000111	a	01110	k	1111101

Kí tự	Mã bit	Kí tự	Mã bit	Kí tự	Mã bit
l	111111	v	011110	á	0001001
m	100011	w	01001000	â	100010011
n	1001	x	111100101	ã	11110101010
o	101001	y	1010001	æ	101000010
p	011111	z	1111011011	è	1010101101101
q	011011111	,	01101100	é	0001010001
r	00110	;	10100000101	ê	111101111
s	00011	À	01001011110110111	ì	1111100111
t	0000	Ô	10100000100100100	í	101010100
u	111000	à	11111000	õ	101010110111

Kí tự	Mã bit	Kí tự	Mã bit	Kí tự	Mã bit
ò	110111010	ĩ	01101001100	ə	1010100
ó	111101110	ŋ	010011100111	ε	10001000100
ô	111110010	ũ	101000001000	з	1010101010000
õ	1010000010011	σ	010011101	ı	010011111
ù	11110101001	ʌ	11110100	ŋ	01001110010
ú	010010110	Ŭ	011011101100011	ƒ	011011011
ý	0100101110	ə	0100101111001110	υ	111101010000
ă	1111001001	ɑ	10100000110	Λ	010011110
Đ	01001011110110011	ɒ	101010110101	З	10101010101
đ	1010110	ɔ	101000011	đ	00010101011

Kí tự	Mã bit	Kí tự	Mã bit	Kí tự	Mã bit
,	011010010	з	010010111111	ă	111100100001
.	000101010011	æ	01001011110110110	â	100010010
:	01101110110000	ə	111101010110	ă	011011110
Δ	11110010000001	З	1111001000000010	ă	00010101010
Λ	111101010001101	Λ	0100101111010000	ă	1010101101001
α	1010000010010011	ạ	01101010	ă	101010110100011
ε	01101110110001010	à	01001010	ă	0110111000
η	10100000111	ă	111101100	ẹ	011011101101
θ	01101001101	à	000101011	è	00010101000
3	10100000100101	ă	01001110000	ē	010011100110

Kí tự	Mã bit	Kí tự	Mã bit	Kí tự	Mã bit
é	01101000	ồ	1010101110	ủ	1010101011
è	101000000	ỗ	0110111001	ứ	1111100110
ẽ	000101001	õ	01001110001	ừ	10101011001

ê	01101110111	ộ	01001001	ử	0001010000
ệ	101011101	ớ	011011010	ữ	11110010001
ỉ	10101011000	ờ	101011100	ự	0100110
ì	011010111	ở	1010101111	ỳ	1010101101100
ọ	1111011010	ỡ	101010101001	ỷ	11110101000101
ồ	0110100111	ợ	011010110	ỹ	1111001000001
ố	111101011	ụ	1000100001		01101110110001000

Kí tự	Mã bit
,	101000001001000
...	111100100000000
∫	111101010111
	01001011110110010

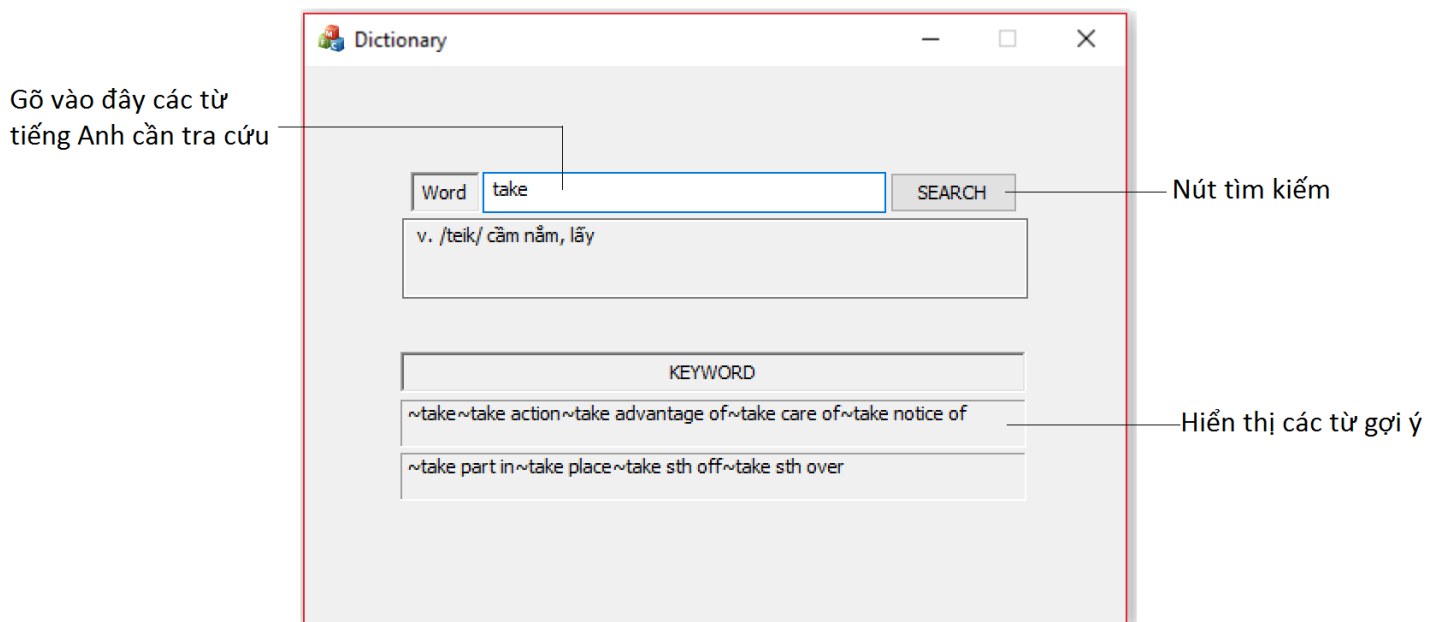
2. Hiệu suất nén

Kích thước từ điển trước khi nén: $N = 207221$ bytes

Kích thước từ điển sau khi nén: $D = 174553$ bytes

Tỷ lệ nén: $D = (N - M)/N * 100 = (207221 - 174553)/207221 * 100 = 15.76\%$

3. Giao diện phần mềm từ điển Anh-Việt



4. Các nguồn tham khảo

<http://www.geeksforgeeks.org/b-tree-set-1-introduction-2/>

<http://www.geeksforgeeks.org/b-tree-set-1-insert-2/>

<http://www.geeksforgeeks.org/b-tree-set-3delete/>

<http://www.geeksforgeeks.org/greedy-algorithms-set-3-huffman-coding/>

<https://trachanhso.net/giai-thuat-nen-huffman-phan-1/>

<https://vndoc.com/3000-tu-vung-tieng-anh-thong-dung-nhat/download>

https://www.siggraph.org/education/materials/HyperGraph/video/mpeg/mpegfaq/huffman_tutorial.html

http://btechsmartclass.com/DS/U5_T3.html