

程序代写代做 CS编程辅导



# COMT-3319 Web Data Compression and Search

WeChat: cstutorcs

Assignment Project Exam Help

BWT, MTE and Pattern Matching

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

# 程序代写代做CS编程辅导



- Burrows–Wheeler Transform (BWT) is an algorithm used to prepare data for use with data compression techniques such as bzip2.
- It was invented by Michael Burrows and David Wheeler in 1994 at DEC SRC, Palo Alto, California.
- It is based on a previously unpublished transformation discovered by Wheeler in 1983.

WeChat: cstutorcs

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

# Recall: 程序代写代做CS编程辅导=BWT

A

A

A

B

N

N

[

]

First



B

N

N

L

A

A

]

A

WeChat: cstutorcs

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

Last

# 程序代写代做 CS编程辅导



A  
A  
A  
B  
N  
N  
[  
]

B  
N  
N  
A  
A  
A  
A  
A

WeChat: cstutorcs

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

# 程序代写代做 CS编程辅导



-A  
=A  
A  
B  
N  
N  
[  
]

B  
N  
N  
[  
A  
A  
A  
]

WeChat: cstutorcs

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>



# 程序代写代做CS编程辅导



A  
A  
A  
B  
N  
N  
[  
]

B  
N  
N  
[  
A  
A  
]

WeChat: cstutorcs

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

# 程序代写代做 CS编程辅导



A  
A  
A  
B  
N  
N  
[  
]

B  
N  
N  
[  
A  
A  
]  
A

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>



# BANANA]

程序代写代做CS编程辅导



A  
A  
A  
B  
N  
N  
[  
]

B  
N  
N  
[  
A  
A  
]

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

# [BANANA]

程序代写代做 CS编程辅导



A  
A  
A  
B  
N  
N  
[  
]

B  
N  
N  
[  
A  
A  
A  
A  
A

WeChat: cstutorcs

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

# Example using C[ ] & Occ[ ]



Position	Symbol	Count
0	B	0
1	N	1
2	N	1
3	[	0
4	A	0
5	A	1
6	]	0
7	A	2

Symbol	# Less Than
A	0
B	3
N	4
[	6
]	7

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

# 程序代写代做CS编程辅导



Position	Symbol	Count
0	B	0
1	N	1
2	N	1
3	[	0
4	A	0
5	A	1
6	]	0
7	A	2

Symbol	# Less Than
A	0
B	3
N	4
[	6
]	7

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

# 程序代写代做CS编程辅导



Position	Symbol	Count
0	B	0
1	N	1
2	N	1
3	[	0
4	A	0
5	A	1
6	]	0
7	A	2

Symbol	# Less Than
A	0
B	3
N	4
[	6
]	7

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

# 程序代写代做[CS编程辅导]



Position	Symbol	Count
0	B	0
1	N	1
2	N	0
3	[	0
4	A	0
5	A	1
6	]	0
7	A	2

Symbol	# Less Than
A	0
B	3
N	4
[	6
]	7

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

# 程序代写代做 CS编程辅导



Position	Symbol	Count
0	B	0
1	N	1
2	N	1
3	[	0
4	A	0
5	A	1
6	]	0
7	A	2

Symbol	# Less Than
A	0
B	3
N	4
[	6
]	7

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

# 程序代写代做 CS编程辅导



Position	Symbol	Count
0	B	0
1	N	1
2	N	1
3	[	0
4	A	0
5	A	1
6	]	0
7	A	2

Symbol	# Less Than
A	0
B	3
N	4
[	6
]	7

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>



# 程序代做CS编程辅导



Position	Symbol	Count
0	B	0
1	N	1
2	N	1
3	[	0
4	A	0
5	A	1
6	]	0
7	A	2

Symbol	# Less Than
A	0
B	3
N	4
[	6
]	7

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

# 程序代写代做 CS编程辅导



Position	Symbol	Count
0	B	0
1	N	1
2	N	1
3	[	0
4	A	0
5	A	1
6	]	0
7	A	2

Symbol	# Less Than
A	0
B	3
N	4
[	6
]	7

WeChat: cstutorcs  
 Assignment Project Exam Help  
 Email: tutorcs@163.com  
 QQ: 749389476  
<https://tutorcs.com>

# 程序代写代做 CS编程辅导



Position	Symbol	Count
0	B	
1	N	0
2	N	1
3	[	0
4	A	0
5	A	1
6	]	0
7	A	2

Symbol	# Less Than
A	0
B	3
N	4
[	6
]	7

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

# 程序代写代做 CS编程辅导



Position	Symbol	Occ / Rank
0	B	
1	N	0
2	N	1
3	[	0
4	A	1
5	A	1
6	]	0
7	A	2

Occ / Rank

Symbol	# LessThan
A	0
B	3
N	4
[	6
]	7

C[]

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

# Move to Front (MTF)



Reduce entropy based on local frequency correlation

Usually used for BWT before an entropy-encoding step

Author and detail:

Original paper at webcms3

[http://www.arturocampos.com/ac\\_mtf.html](http://www.arturocampos.com/ac_mtf.html)

<https://tutorcs.com>

# Example: abbaabacdad



Symbol	List	
a	1	abcde.....
b	1	bacde.....
a	1	abcde.....
a	0	abcde.....
b	1	bacde.....
a	1	abcde.....
c	2	cabde.....
a	1	acbde.....
d	3	dacbe.....

To transform a general file, the list has 256 ASCII symbols.

WeChat: cstutorcs

Assignment Project Exam Help

Email: tut@cs.163.com

QQ: 749389476

https://tutorcs.com

# Example: 程序代写代做 CS编程辅导 abaaabbbccdddddcc

Symbols: abaaabbbccd

Codes (in ASCII binary)

01100100, 01100011, 0

Codes (in ASCII dec): 97, 98, 97, 97, 97, 98, 98, 98, 99, 99, 100, 100, 100,  
100, 99, 99



WeChat: cstutorcs

Assignment Project Exam Help

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

# Example: 程序代写代做 CS编程辅导 abaaabbbccdddddcc

Symbols: abaaabbbccdd

Codes (in ASCII binary)

01100100, 01100011, 0

Codes (in ASCII dec): 97, 98, 97, 97, 97, 98, 98, 98, 99, 99, 100, 100, 100, 100, 99, 99



WeChat: cstutorcs

Recall that Shannon's Entropy is maximized when there is max uncertainty, i.e., equal probability, like the example above (4 "97"s, 4 "98"s, 4 "99"s, 4 "100"s).

QQ: 749389476

e.g., Entropy  $H = 2.00$

<https://tutorcs.com>



# Example: 程序代写代做CS编程辅导 abaaabbbccdddddcc

Symbols: abaaabbbccdd

Codes (in ASCII binary)

01100100, 01100011, 0

Codes (in ASCII dec): 97, 98, 97, 97, 97, 98, 98, 98, 99, 99, 100, 100, 100, 100, 99, 99

WeChat: cstutorcs

List

Index: 0 1 2 3 4 97 98 99 100 101 102 ... 255

Value: .. .. .. .. .. a b c d e f ... ..

Assignment Project Exam Help

Codes (in ASCII): 97, 98, 97, 97, 97, 98, 98, 98, 99, 99, 100, 100, 100, 100, 99, 99

Email: tutorcs@163.com

Codes ( in MTF ): 97

QQ: 749389476

https://tutorcs.com

# Example: 程序代写代做 CS编程辅导 abaaabbbccdddddcc

Symbols: abaaabbbccdd

Codes (in ASCII binary)

01100100, 01100011, 0

Codes (in ASCII dec): 97, 98, 97, 97, 97, 98, 98, 98, 99, 99, 100, 100, 100, 100, 99, 99

WeChat: cstutorcs

List

Index: 0 1 2 3 4 97 98 99 100 101 102 ... 255

Value: .. .. .. .. .. a b c d e f ... ..

Assignment Project Exam Help

Codes (in ASCII): 97, 98, 97, 97, 97, 98, 98, 98, 99, 99, 100, 100, 100, 100, 99, 99

Email: tutorcs@163.com

Codes ( in MTF ): 97

QQ: 749389476

List

Index: 0 1 2 3 4 97 98 99 100 101 102 ... 255

Value: .. .. .. .. .. a b c d e f ... ..

https://tutorcs.com

move to front

# Example: 程序代写代做 CS编程辅导 abaaabbbccddddcc

## List

Index: 0 1 2 3 4 97 98 99 100 101 102 ... 255  
Value: a .. .. .. .. d e f ... ..



Codes (in ASCII): 97, 98, 97, 97, 97, 98, 98, 98, 99, 99, 100, 100, 100, 100, 99, 99

WeChat: cstutorcs

Codes ( in MTF ): 97, 98

Assignment Project Exam Help

Email: tutorcs@163.com

## List

Index: 0 1 2 3 4 97 98 99 100 101 102 ... 255  
Value: b a .. .. .. c d e f ... ..

QQ: 749389476

https://tutorcs.com

# Example: 程序代写代做 CS编程辅导 abaaabbbccddddcc

## List

Index: 0 1 2 3 4 ... 100 101 102 ... 255  
Value: b a .. .. .. d e f ... ..



Codes (in ASCII): 97, 98, 97, 97, 97, 98, 98, 98, 99, 99, 100, 100, 100, 100, 99, 99

WeChat: cstutorcs

Codes ( in MTF ): 97, 98, 1

Assignment Project Exam Help

Email: tutorcs@163.com

## List

Index: 0 1 2 3 4 97 98 99 100 101 102 ... 255  
Value: b a .. .. .. c d e f ... ..

QQ: 749389476

https://tutorcs.com

# Example: 程序代写代做 CS编程辅导 abaaabbbccddddcc

## List

Index: 0 1 2 3 4 ... 100 101 102 ... 255  
Value: a b .. .. .. d e f ... ..



Codes (in ASCII): 97, 98, 97, 97, 97, 98, 98, 98, 99, 99, 100, 100, 100, 100, 99, 99

WeChat: cstutorcs

Codes ( in MTF ): 97, 98, 1, 0,

Assignment Project Exam Help

Email: tutorcs@163.com

## List

Index: 0 1 2 3 4 97 98 99 100 101 102 ... 255  
Value: a b .. .. .. c d e f ... ..

QQ: 749389476

https://tutorcs.com

# Example: 程序代写代做 CS编程辅导 abaaabbbccddddcc

## List

Index: 0 1 2 3 4 ... 100 101 102 ... 255  
Value: a b .. .. .. d e f ... ..



Codes (in ASCII): 97, 98, 97, 97, 97, 98, 98, 98, 99, 99, 100, 100, 100, 100, 99, 99

WeChat: cstutorcs

Codes ( in MTF ): 97, 98, 1, 0, 0, 1, 0, 0, 99, 0, 100, 0, 0, 0, 1, 0

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

https://tutorcs.com

# Example: MTF decoding



List

Index: 0 1 2 3 4 ... 100 101 102 ... 255  
Value: .. .. .. .. .. d e f ... ..

Codes ( in MTF ): 97, 98, 1, 0, 0, 1, 0, 0, 99, 0, 100, 0, 0, 0,  
1, 0

Symbols: a, b

WeChat: cstutorcs

Assignment Project Exam Help

List

Index: 0 1 2 3 4 97 98 99 100 101 102 ... 255  
Value: b a .. .. .. c d e f ... ..

Email: [tutorcs@163.com](mailto:tutorcs@163.com)

QQ: 749389476

<https://tutorcs.com>

# Example: MTF decoding



List

Index: 0 1 2 3 4 ... 100 101 102 ... 255  
 Value: b a .. .. .. d e f ... ..

Codes ( in MTF ): 97, 98, 1, 0, 0, 1, 0, 0, 99, 0, 100, 0, 0, 0, 1, 0

Symbols: a, b, a

WeChat: cstutorcs

Assignment Project Exam Help

List

Index: 0 1 2 3 4 97 98 99 100 101 102 ... 255  
 Value: a b .. .. .. c d e f ... ..

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>



# Example: MTF decoding



List

Index: 0 1 2 3 4 ... 100 101 102 ... 255  
Value: a b .. .. .. d e f ... ..

Codes ( in MTF ): 97, 98, 1, 0, 0, 1, 0, 0, 99, 0, 100, 0, 0, 0,  
1, 0

Symbols: a, b, a, a

List

Index: 0 1 2 3 4 97 98 99 100 101 102 ... 255  
Value: a b .. .. .. c d e f ... ..

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

# Example: MTF decoding



## List

Index: 0 1 2 3 4 ... 100 101 102 ... 255  
Value: a b .. .. .. d e f ... ..

Codes ( in MTF ): 97, 98, 1, 0, 0, 1, 0, 0, 99, 0, 100, 0, 0, 0,  
1, 0

Symbols: a, b, a, a, a, b, b, b, c, c, d, d, d, d,  
c, c

**WeChat: cstutorcs**

**Assignment Project Exam Help**

**Email: tutorcs@163.com**

**QQ: 749389476**

**<https://tutorcs.com>**

# Example: MTF decoding



## List

Index: 0 1 2 3 4 ... 100 101 102 ... 255  
 Value: a b .. .. .. d e f ... ..

Codes ( in MTF ): 97, 98, 1, 0, 0, 1, 0, 0, 99, 0, 100, 0, 0, 0,  
 1, 0

Symbols: a, b, a, a, a, b, b, b, c, c, d, d, d, d,  
 c, c

The distribution of symbols is changed, with more  
*local* references (1 “97”, 1 “98”, 1 “99”, 1 “100”, 9  
 “0”s, 3 “1”s).  $\Rightarrow$  Reduced entropy

$H = 1.92$  <https://tutorcs.com>

# BWT compressor vs ZIP

程序代写代做 CS编程辅导

ZIP (i.e., LZW based)



BWT+RLE+MTF+AC

File Name	Raw Size	PKZIP Bits/Byte	BWT Size	BWT Bits/Byte
bib	111,261	25,821	29,567	2.13
book1	768,771	315,999	275,831	2.87
book2	610,856	209,061	186,592	2.44
geo	102,400	117,117	64,120	4.85
news	377,109	146,010	134,174	2.85
obj1	21,504	10,311	10,857	4.04
obj2	246,814	81,846	81,948	2.66

WeChat: cstutorcs

Assignment Project Exam Help

Email: tutorcs@163.com

QQ: 749389476

<https://tutorcs.com>

From <http://marknelson.us/1996/09/01/bwt/>