Extensives Beispiel zur Bestimmung der Levenshtein-Distanz von Zeichenketten über nicht-speicher-optimierte Matrix

Dr. Ricardo Usbeck https://github.com/RicardoUsbeck/NLU

16.07.2020

1/48

Dr. R. Usbeck Extensives Beispiel 16.07.2020

Exercise

• Compute Levenshtein distance matrix for oslo – snow



2 / 48

Dr. R. Usbeck Extensives Beispiel 16.07.2020

Each cell of Levenshtein matrix

cost of getting here from	cost of getting here from		
my upper left neighbor	my upper neighbor (dele-		
(copy or replace)	te)		
	the minimum of the		
cost of getting here from	three possible "move-		
my left neighbor (insert)	ments"; the cheapest		
	way of getting here		

3 / 48

		S	n	0	W	
	0	1 1	2 2	3 3	4 4	
o	1 1					
S	2 2					
I	3 3					
0	4					

		S	n	О	w	
	0	1 1	2 2	3 3	4 4	
o	1 1	1 2 ?				
S	2 2					
ı	3 3					
0	4					

		S	n	0	w	
	0	1 1	2 2	3 3	4 4	
o	1 1	1 2 2 1				
S	2 2					
I	3 3					
o	4					

		S	n	О	w
	0	1 1	2 2	3 3	4 4
o	1 1	1 2 2 1	2 3 2 ?		
S	2 2				
I	3 3				
o	4				

Dr. R. Usbeck Extensives Beispiel 16.07.2020 7/48

		S	n	0	W	
	0	1 1	2 2	3 3	4 4	
o	1 1	1 2 2 1	2 3 2 2			
S	2 2					
ı	3 3					
o	4					

		S	n	О	W	
	0	1 1	2 2	3 3	4 4	
0	1 1	1 2 2 1	2 3 2 2	2 4 3 ?		
S	2 2					
ı	3 3					
o	4					

		S	n	О	w
	0	1 1	2 2	3 3	4 4
0	1 1	1 2 2 1	2 3 2 2	2 4 3 2	
S	2 2				
I	3 3				
0	4				

		S	n	О	w
	0	1 1	2 2	3 3	4 4
0	1 1	1 2 2 1	2 3 2 2	2 4 3 2	4 5 3 ?
S	2 2				
I	3 3				
o	4				

		S	n	О	w
	0	1 1	2 2	3 3	4 4
0	1 1	1 2 2 1	2 3 2 2	2 4 3 2	4 5 3 3
S	2 2				
I	3 3				
o	4				

		9	S		n		0		w	
	0	1	1	2	2	3	3	4	4	
	1	1	2	2	3	2	4	4	5	
0	1	2	1	2	2	3	2	3	3	
	2	1	2							
S	2	3	?							
I	3									
o	4									

Dr. R. Usbeck Extensives Beispiel 16.07.2020

		S	n	О	w
	0	1 1	2 2	3 3	4 4
0	1 1	1 2 2 1	2 3 2	2 4 3 2	4 5 3 3
s	2 2	1 2 3 1			
I	3 3				
o	4				

Dr. R. Usbeck Extensives Beispiel

			9	S		ı	0		v	v
	0	_	1	1	2	2	3	3	4	4
	1		1	2	2	3	2	4	4	5
0	1		2	1	2	2	3	2	3	3
	2		1	2	2	3				
S	2		3	1	2	?				
I	3	-								
0	4	-								

			S		า	()	V	v
	0	1	1	2	2	3	3	4	4
0	_ 1	1	2	2	3	2	4	4	5
	1	2	1	2	2	3	2	3	3
	2	1	2	2	3				
S	2	3	1	2	2				
ı	3 3								
o	4								

		9	S		ı	()	v	v
	 0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3		
S	2	3	1	2	2	3	?		
ı	3								
o	4								

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 17/48

			5	5	r	1	()	v	v
	0	-	1	1	2	2	3	3	4	4
0	1		1	2	2	3	2	4	4	5
	1		2	1	2	2	3	2	3	3
	2		1	2	2	3	3	3		
S	2		3	1	2	2	3	3		
ı	3 3	-								
o	4	-								

			S		r	ı	()	v	v
	0	_	1	1	2	2	3	3	4	4
	1		1	2	2	3	2	4	4	5
0	1		2	1	2	2	3	2	3	3
	2		1	2	2	3	3	3	3	4
S	2		3	1	2	2	3	3	4	?
ı	3	-								
o	4	_								

Dr. R. Usbeck Extensives Beispiel 16.07.2020

		9	S		ı	()	w	
	 0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3	3	4
S	2	3	1	2	2	3	3	4	3
	3								
'	3								
o	4								

		9	5	r	ı	()	v	v
	 0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3	3	4
S	2	3	1	2	2	3	3	4	3
	3	3	2						
	3	4	?						
o	4								

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 21/48

		9	5	r	า	()	w	
	 0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3	3	4
S	2	3	1	2	2	3	3	4	3
	3	3	2						
ı	3	4	2						
o	4								

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 22 / 48

			S	ı	า	О		V	v
	0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3	3	4
S	2	3	1	2	2	3	3	4	3
	3	3	2	2	3				
'	3	4	2	3	?				
0	4								

		9	S		ı	0		W	
	0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3	3	4
S	2	3	1	2	2	3	3	4	3
	3	3	2	2	3				
'	3	4	2	3	2				
0	4								

			9	5	n		()	w	
	0	-	1	1	2	2	3	3	4	4
	1		1	2	2	3	2	4	4	5
0	1		2	1	2	2	3	2	3	3
	2		1	2	2	3	3	3	3	4
S	2		3	1	2	2	3	3	4	3
	3		3	2	2	3	3	4		
I	3		4	2	3	2	3	?		
0	4	-								

			S		r	ı	()	W	
	0)	1	1	2	2	3	3	4	4
	1	.	1	2	2	3	2	4	4	5
0	1	_	2	1	2	2	3	2	3	3
	2	2	1	2	2	3	3	3	3	4
S	2	2	3	1	2	2	3	3	4	3
	3	3	3	2	2	3	3	4		
'	3	Γ	4	2	3	2	3	3		
0	4									

			S		r	า	0		w	
	0	_	1	1	2	2	3	3	4	4
	1		1	2	2	3	2	4	4	5
0	1		2	1	2	2	3	2	3	3
	2		1	2	2	3	3	3	3	4
S	2		3	1	2	2	3	3	4	3
	3		3	2	2	3	3	4	4	4
ı	3		4	2	3	2	3	3	4	?
0	4	_								

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 27 / 48

		S		r	า	()	v	v
	0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3	3	4
S	2	3	1	2	2	3	3	4	3
	3	3	2	2	3	3	4	4	4
	3	4	2	3	2	3	3	4	4
0	4								

		9	5	r	ı	()	v	v
	 0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3	3	4
S	2	3	1	2	2	3	3	4	3
	3	3	2	2	3	3	4	4	4
'	3	4	2	3	2	3	3	4	4
	4	4	3						
0	1	-5	7						

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 29/48

				5	r	ı	()	w		
	_	0	1	1	2	2	3	3	4	4	
		1	1	2	2	3	2	4	4	5	
0		1	2	1	2	2	3	2	3	3	
		2	1	2	2	3	3	3	3	4	
S		2	3	1	2	2	3	3	4	3	
		3	3	2	2	3	3	4	4	4	
1		3	4	2	3	2	3	3	4	4	
		4	4	3							
0		1	5	3							

		9	5	r	1	()	v	v
	 0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3	3	4
S	2	3	1	2	2	3	3	4	3
	3	3	2	2	3	3	4	4	4
	3	4	2	3	2	3	3	4	4
	4	4	3	3	3				
0	1		2	1	2				

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 31/48

		9	5	r	ı	()	v	v
	 0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3	3	4
S	2	3	1	2	2	3	3	4	3
	3	3	2	2	3	3	4	4	4
1	3	4	2	3	2	3	3	4	4
	4	4	3	3	3				
0	1		2		2				

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 32/48

			5	r	า	()	W	
	 0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3	3	4
S	2	3	1	2	2	3	3	4	3
	3	3	2	2	3	3	4	4	4
1	3	4	2	3	2	3	3	4	4
	4	4	3	3	3	2	4		
0	1	5	3	1	3	1	7		

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 33/48

		9	5	r	ı	()	v	v
	 0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3	3	4
S	2	3	1	2	2	3	3	4	3
	3	3	2	2	3	3	4	4	4
'	3	4	2	3	2	3	3	4	4
	4	4	3	3	3	2	4		
0	1		2	1	3	1	2		

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 34/48

			5	r	ı	()	v	v
	 0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3	3	4
S	2	3	1	2	2	3	3	4	3
	3	3	2	2	3	3	4	4	4
	3	4	2	3	2	3	3	4	4
	4	4	3	3	3	2	4	4	5
0	1	5	3		3	1	2	3	7

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 35/48

			5	r	า	()	V	v
	 0	1	1	2	2	3	3	4	4
	1	1	2	2	3	2	4	4	5
0	1	2	1	2	2	3	2	3	3
	2	1	2	2	3	3	3	3	4
S	2	3	1	2	2	3	3	4	3
	3	3	2	2	3	3	4	4	4
	3	4	2	3	2	3	3	4	4
	4	4	3	3	3	2	4	4	5
0	1	5	3	1	3	1	2	3	3

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 36/48

			S		r	n		0		v
		0	1	1	2	2	3	3	4	4
		1	1	2	2	3	2	4	4	5
0		1	2	1	2	2	3	2	3	3
		2	1	2	2	3	3	3	3	4
S		2	3	1	2	2	3	3	4	3
		3	3	2	2	3	3	4	4	4
ı		3	4	2	3	2	3	3	4	4
		4	4	3	3	3	2	4	4	5
0		4	5	3	4	3	4	2	3	3

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 37/48

			S		r	n		0		v
	_	0	1	1	2	2	3	3	4	4
		1	1	2	2	3	2	4	4	5
0		1	2	1	2	2	3	2	3	3
		2	1	2	2	3	3	3	3	4
S		2	3	1	2	2	3	3	4	3
		3	3	2	2	3	3	4	4	4
1		3	4	2	3	2	3	3	4	4
		4	4	3	3	3	2	4	4	5
0		4	5	3	4	3	4	2	3	3

How do I read out the editing operations that transform oslo into snow?

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 38 / 48

			S		r	n		0		٧
		0	1	1	2	2	3	3	4	4
		1	1	2	2	3	2	4	4	5
0		1	2	1	2	2	3	2	3	3
		2	1	2	2	3	3	3	3	4
S		2	3	1	2	2	3	3	4	3
		3	3	2	2	3	3	4	4	4
ľ		3	4	2	3	2	3	3	4	4
		4	4	3	3	3	2	4	4	5
0		4	5	3	4	3	4	2	3	3

cost	operation	input	output
1	insert	*	w

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 39/48

			S		r	n		0		W
	_	0	1	1	2	2	3	3	4	4
		1	1	2	2	3	2	4	4	5
0		1	2	1	2	2	3	2	3	3
		2	1	2	2	3	3	3	3	4
S		2	3	1	2	2	3	3	4	3
		3	3	2	2	3	3	4	4	4
ı		3	4	2	3	2	3	3	4	4
		4	4	3	3	3	2	4	4	5
0		4	5	3	4	3	4	2	3	3

cost	operation	input	output
0	(copy)	0	0
1	insert	*	W

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 40 / 48

			9	5	r	า	()	V	v
	_	0	1	1	2	2	3	3	4	4
		1	1	2	2	3	2	4	4	5
0		1	2	1	2	2	3	2	3	3
		2	1	2	2	3	3	3	3	4
S		2	3	1	2	2	3	3	4	3
		3	3	2	2	3	3	4	4	4
<u>'</u>		3	4	2	3	2	3	3	4	4
		4	4	3	3	3	2	4	4	5
0		4	5	3	4	3	4	2	3	3

cost	operation	input	output
1	replace	1	n
0	(copy)	0	0
1	insert	*	W

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 41/48

				5	ı	า	()	V	v
	_	0	1	1	2	2	3	3	4	4
		1	1	2	2	3	2	4	4	5
0		1	2	1	2	2	3	2	3	3
		2	1	2	2	3	3	3	3	4
S		2	3	1	2	2	3	3	4	3
		3	3	2	2	3	3	4	4	4
		3	4	2	3	2	3	3	4	4
		4	4	3	3	3	2	4	4	5
0		4	5	3	4	3	4	2	3	3

cost	operation	input	output
0	(copy)	S	S
1	replace	I	n
0	(copy)	0	0
1	insert	*	W

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 42 / 48

			S		n		0		W	
		0	1	1		2	3	3	4	4
		1	1	2	2	3	2	4	4	5
0		1	2	1	2	2	3	2	3	3
		2	1	2	2	3	3	3	3	4
S		2	3	1	2	2	3	3	4	3
		3	3	2	2	3	3	4	4	4
		3	4	2	3	2	3	3	4	4
		4	4	3	3	3	2	4	4	5
0		4	5	3	4	3	4	2	3	3

cost	operation	input	output
1	delete	0	*
0	(copy)	S	S
1	replace	I	n
0	(copy)	0	0
1	insert	*	W

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 43 / 48

			(c	á	a	1	t	(:	á	ì	1	
		0	1	1	2	2	3	3	4	4	5	5	6	6
С	_	1	0 2	0	2 1	3 1	3 2	2	3	5 3	5 4	6 4	6 5	7 5
а	_	2 2	3	1	0 2	2 0	2 1	3 1	3 2	4 2	3	5 3	5 4	6 4
t		3	3 4	2	3	1 1	0 2	0	$\frac{2}{1}$	3 1	3 2	4 2	3	5 3

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 44/48

				С		a		t	(С		a	1	t
	0	_	1	1	2	2	3	3	4	4	5	5	6	6
С	- 1 1	_	0 2	0	2 1	3 1	3 2	2	3	5 3	5 4	6 4	6 5	7 5
а	2 2	_	3	1	0 2	0	2 1	3 1	3 2	2	3	5 3	5 4	6 4
t	3	_	3 4	2	$\frac{2}{3}$	1 1	$\frac{0}{2}$	2 0	$\frac{2}{1}$	3 1	$\frac{3}{2}$	<u>4</u> 2	3	5 3

cost	operation	input	output
1	insert	*	С
1	insert	*	а
1	insert	*	t
0	(copy)	С	С
0	(copy)	а	a
0	(copy)	t	t

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 45/48

		(С		а	1	t	(2		a		t
	 0	1	1	2	2	3	3	4	4	5	5	6	6
	1	0	2	2	3	3	4	3	5	5	6	6	7
С	1	2	0	1	1	2	2	3	3	4	4	5	5
	2	2	1	0	2	2	3	3	4	3	5	5	6
а	2	3	1	2	0	1	1	2	2	3	3	4	4
_	3	3	2	2	1	0	2	2	3	3	4	3	5
۱ ۱	3	4	2	3	1	2	0	1	1	2	2	3	3

cost	operation	input	output
0	(copy)	С	С
1	insert	*	а
1	insert	*	t
1	insert	*	С
0	(copy)	а	а
0	(copy)	t	t

			(С		a	1	t	(С	í	a	1	t
	_	0	1	1	2	2	3	3	4	4	5	5	6	6
С		1	0 2	2 0	2 1	3 1	3 2	2	3	5 3	5 4	6 4	6 5	7 5
а		2 2	3	1	0 2	2 0	2 1	3 1	3 2	4 2	3	5 3	5 4	6 4
t		3	3 4	2 2	2 3	1 1	0 2	2 0	2 1	3 1	3 2	4 2	3	5 3

cost	operation	input	output
0	(copy)	С	С
0	(copy)	а	а
1	insert	*	t
1	insert	*	С
1	insert	*	а
0	(copy)	t	t

 Dr. R. Usbeck
 Extensives Beispiel
 16.07.2020
 47/48

		(С		a		t		С	,	a		t
	 0	1	1	2	2	3	3	4	4	5	5	6	6
С	1	0	2	2	3	3	4	3	5	5	6	6	7
	1	2	0	1	1	2	2	3	3	4	4	5	5
	2	2	1	0	2	2	3	3	4	3	5	5	6
а	2	3	1	2	0	1	1	2	2	3	3	4	4
_	3	3	2	2	1	0	2	2	3	3	4	3	5
۱ ۱	3	4	2	3	1	2	0	1	1	2	2	3	3

cost	operation	input	output
0	(copy)	С	С
0	(copy)	а	а
0	(copy)	t	t
1	insert	*	С
1	insert	*	а
1	insert	*	t

48 / 48

Dr. R. Usbeck Extensives Beispiel 16.07.2020