

Information Retrieval

Tutorial 2: Dictionary/Tolerant-retrieval

Professor: Michel Schellekens

TA: Ang Gao

University College Cork

2012-11-02

Terminology and concepts

- A **token** is a sequence of characters in a document
- Input: “*Friends, Romans, Countrymen*”
- Output: Tokens
 - *Friends*
 - *Romans*
 - *Countrymen*
- further processing: a **term** is a (normalized) word type, which is an entry in our IR system dictionary
- **We need to “normalize” words in indexed text as well as query words into the same form**

Terminology and concepts

- Lemmatization: Reduce inflectional/variant forms to base form
- E.g.,
 - *am, are, is* → *be*
 - *car, cars, car's, cars'* → *car*
- *the boy's cars are different colors* → *the boy car be different color*
- Lemmatization implies doing “proper” reduction to **dictionary headword form**

Terminology and concepts

- Stemming: Reduce terms to their “roots” before indexing
- “Stemming” suggest crude affix chopping
 - language dependent
 - e.g., *automate(s)*, *automatic*, *automation* all reduced to *automat*.

for example compressed and compression are both accepted as equivalent to compress.



for exampl compress and compress ar both accept as equal to compress

Exercise: True or False

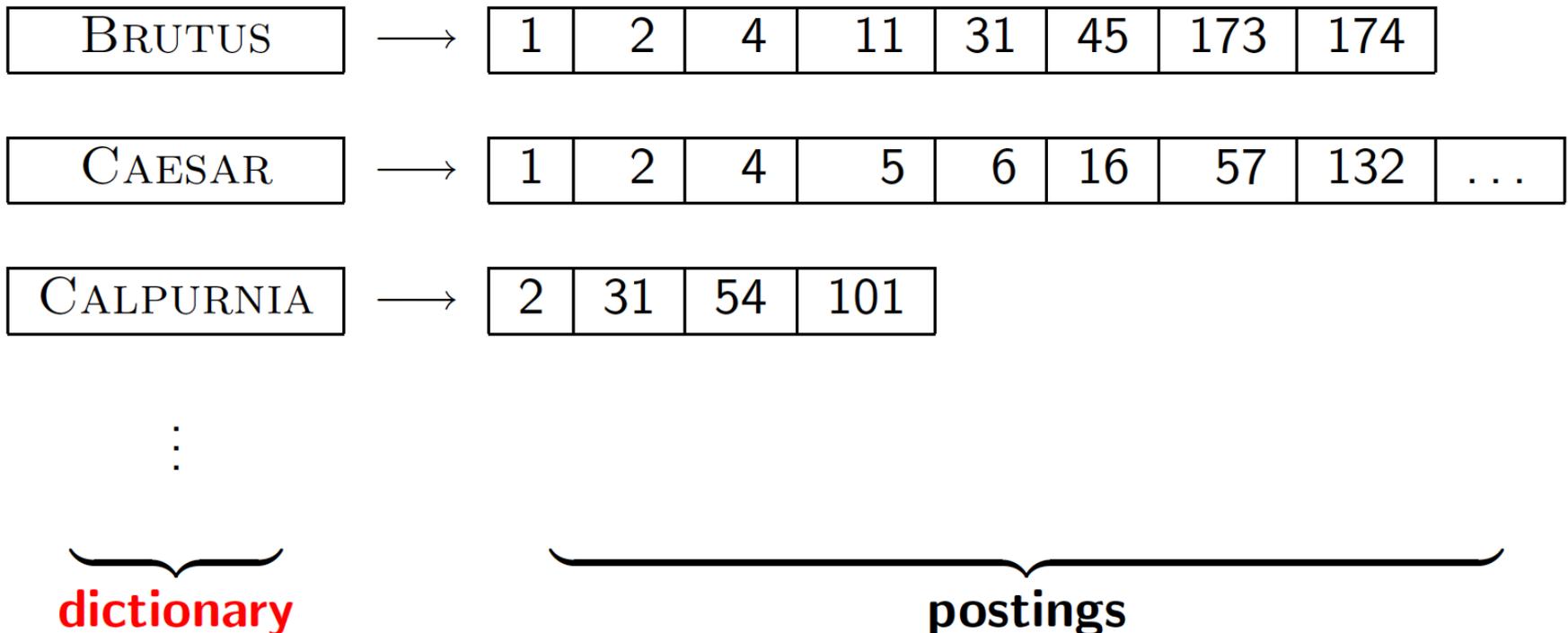
- In a Boolean retrieval system, stemming never lowers precision?
 - **False. Stemming can increase the retrieved set without increasing the number of relevant documents**
- In a Boolean retrieval system, stemming never lowers recall?
 - **True. Stemming can only increase the retrieved set, which means increased or unchanged recall.**
- Stemming increase the size of the vocabulary?
 - **False. Stemming decreases the size of the vocabulary.**
- Stemming should be invoked at indexing time but not while processing a query?
 - **False. The same processing should be applied to documents and queries to ensure matching terms.**

Solutions for wild card queries

- B-tree with extra reverse B-tree
- Permuterm index
- **Bigram index**

Dictionary data structures for inverted indexes

- Standard inverted indexes



Bigram (k -gram) indexes

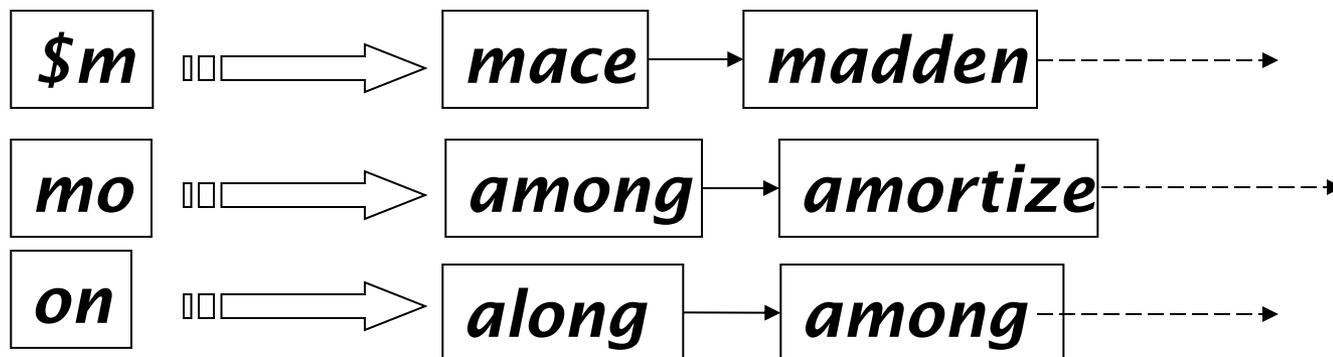
- Enumerate all k -grams (sequence of k chars) occurring in any term
- *e.g.*, from text “**April** **is** **the** **cruelest** **month**” we get the 2-grams (*bigrams*)

\$a,ap,pr,ri,il,l\$, \$i,is,s\$, \$t,th,he,e\$, \$c,cr,ru, ue,el,le,es,st,t\$, \$m,mo,on,nt,h\$

- \$ is a special word boundary symbol
- Maintain a *second* inverted index *from bigrams to dictionary terms* that match each bigram.

Bigram index example

- The k -gram index finds *terms* based on a query consisting of k -grams (here $k=2$).



All bigrams generated from
standard inverted index terms

All terms (standard inverted index) containing
that bigram, sorted in lexicographic order

Processing wild-cards

- Query *mon** can now be run as
 - *\$m AND mo AND on* ←
- *All terms start with m and containing mo and containing on*
- *Terms satisfy mon* will satisfy \$m AND mo AND on*
- *but reverse is not true.(False positive) eg: mormon*
- **Must post-filter these terms against query.**
- Surviving enumerated terms are then looked up in the term-document inverted index.
- Fast, space efficient (compared to permuterm).

Exercise

Give an example of a sentence that falsely matches the wildcard query `mor*ng` if the search were to simply use a conjunction of bigrams.

Solution

Query: \$m AND mo AND or AND ng AND g\$

Eg: motoring

`m,mo,ot,to,or,ri,in,ng,g`

Edit distance

- Given two strings S_1 and S_2 , the minimum number of operations to convert one to the other
- Operations are typically character-level
 - Insert, Delete, Replace, (Transposition)
- E.g., the edit distance from **dof** to **dog** is 1
 - From **cat** to **act** is 2 (Just 1 with transpose.)
 - from **cat** to **dog** is 3.
- Generally found by dynamic programming.

Edit distance

- The edit distance between string s_1 and string s_2 is the *minimum* number of basic operations that convert s_1 to s_2 .
- Levenshtein distance: The admissible basic operations are insert, delete, and replace
- Levenshtein distance *dog-do*: 1 (Deletion)
- Levenshtein distance *cat-cart*: 1 (Insertion)
- Levenshtein distance *cat-cut*: 1 (Replacement)
- Levenshtein distance *cat-act*: 2 (Replacements)

Levenshtein distance: Computation

	0	1	2	3	4	
			f	a	s	t
0		0	1	2	3	4
1	c	1	1	2	3	4
2	a	2	2	1	2	3
3	t	3	3	2	2	2
4	s	4	4	3	2	3

$m[|X|,|Y|]$ = edit distance between X and Y

$m[i,j]$ = edit distance between $X[1..i]$ and $Y[1..j]$

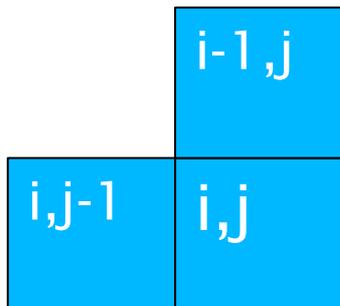
$$m(0,0) = 0$$

$$m(1,0)=1$$

$$m(k,0)=k \text{ (delete)}$$

$$m(0,k) = k \text{ (insert)}$$

	0	1	2	3	4	
0		0	1	2	3	4
1	c	1	1	2	3	4
2	a	2	2	1	2	3
3	t	3	3	2	2	2
4	s	4	4	3	2	3



Suppose we want to calculate $m[i, j]$
 Which is changing from $X[1..i]$ to $Y[1..j]$
 What about:

$m[i-1, j]$: the edit distance of changing from $X[1..i-1]$ to $Y[1..j]$

suppose this is optimal, how many steps you need to change from $x[1..i]$ to $Y[1..j]$ with the help of $m[i-1, j]$?

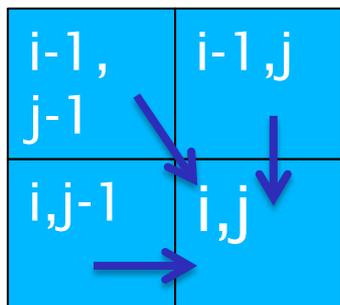
$$m[i, j] = m[i-1, j] + 1 \text{ (delete)}$$

$m[i, j-1]$: the edit distance of changing from $X[1..i]$ to $Y[1..j-1]$

suppose this is optimal, how many steps you need to change from $x[1..i]$ to $Y[1..j]$ with the help of $m[i, j-1]$?

$$m[i, j] = m[i, j-1] + 1 \text{ (insert)}$$

	0	1	2	3	4
		f	a	s	t
0	0	1	2	3	4
1	c	1	1	2	3
2	a	2	2	1	2
3	t	3	3	2	2
4	s	4	4	3	3



$$m[i,j]=m[i-1,j]+1 \text{ (delete)}$$

$$m[i,j]=m[i,j-1]+1 \text{ (insert)}$$

$m[i-1, j-1]$: the edit distance of changing from $X[1..i-1]$ to $Y[1..j-1]$

suppose this is optimal, how many steps you need to change from $x[1..i]$ to $Y[1..j]$ with the help of $m[i-1, j-1]$?

$$m[i,j]=m[i-1, j-1]+1 \text{ (replace)}$$

$$m[i,j]=m[i-1, j-1] \text{ (if } X[i]=Y[j] \text{)}$$

To find optimal solution:

$m[i,j]$ = minimum of three possible 'movements'

Each cell of Levenshtein matrix

cost of getting here from my upper left neighbor (copy or replace)	cost of getting here from my upper neighbor (delete)
cost of getting here from my left neighbor (insert)	the minimum of the three possible “movements”; the cheapest way of getting here

Levenshtein distance: Algorithm

LEVENSHTEINDISTANCE(s_1, s_2)

```
1  for  $i \leftarrow 0$  to  $|s_1|$ 
2  do  $m[i, 0] = i$ 
3  for  $j \leftarrow 0$  to  $|s_2|$ 
4  do  $m[0, j] = j$ 
5  for  $i \leftarrow 1$  to  $|s_1|$ 
6  do for  $j \leftarrow 1$  to  $|s_2|$ 
7     do if  $s_1[i] = s_2[j]$ 
8         then  $m[i, j] = \min\{m[i-1, j]+1, m[i, j-1]+1, m[i-1, j-1]\}$ 
9         else  $m[i, j] = \min\{m[i-1, j]+1, m[i, j-1]+1, m[i-1, j-1]+1\}$ 
10 return  $m[|s_1|, |s_2|]$ 
```

Operations: insert (cost 1), delete (cost 1), replace (cost 1), copy (cost 0)

Levenshtein distance: Algorithm

LEVENSHTEINDISTANCE(s_1, s_2)

```
1  for  $i \leftarrow 0$  to  $|s_1|$ 
2  do  $m[i, 0] = i$ 
3  for  $j \leftarrow 0$  to  $|s_2|$ 
4  do  $m[0, j] = j$ 
5  for  $i \leftarrow 1$  to  $|s_1|$ 
6  do for  $j \leftarrow 1$  to  $|s_2|$ 
7     do if  $s_1[i] = s_2[j]$ 
8         then  $m[i, j] = \min\{m[i-1, j]+1, m[i, j-1]+1, m[i-1, j-1]\}$ 
9         else  $m[i, j] = \min\{m[i-1, j]+1, m[i, j-1]+1, m[i-1, j-1]+1\}$ 
10 return  $m[|s_1|, |s_2|]$ 
```

Operations: insert (cost 1), delete (cost 1), replace (cost 1), copy (cost 0)

Levenshtein distance: Algorithm

LEVENSHTEINDISTANCE(s_1, s_2)

```
1  for  $i \leftarrow 0$  to  $|s_1|$ 
2  do  $m[i, 0] = i$ 
3  for  $j \leftarrow 0$  to  $|s_2|$ 
4  do  $m[0, j] = j$ 
5  for  $i \leftarrow 1$  to  $|s_1|$ 
6  do for  $j \leftarrow 1$  to  $|s_2|$ 
7     do if  $s_1[i] = s_2[j]$ 
8         then  $m[i, j] = \min\{m[i-1, j]+1, m[i, j-1]+1, m[i-1, j-1]\}$ 
9         else  $m[i, j] = \min\{m[i-1, j]+1, m[i, j-1]+1, m[i-1, j-1]+1\}$ 
10 return  $m[|s_1|, |s_2|]$ 
```

Operations: insert (cost 1), delete (cost 1), replace (cost 1), copy (cost 0)

Levenshtein distance: Algorithm

LEVENSHTEINDISTANCE(s_1, s_2)

```

1  for  $i \leftarrow 0$  to  $|s_1|$ 
2  do  $m[i, 0] = i$ 
3  for  $j \leftarrow 0$  to  $|s_2|$ 
4  do  $m[0, j] = j$ 
5  for  $i \leftarrow 1$  to  $|s_1|$ 
6  do for  $j \leftarrow 1$  to  $|s_2|$ 
7      do if  $s_1[i] = s_2[j]$ 
8          then  $m[i, j] = \min\{m[i-1, j]+1, m[i, j-1]+1, m[i-1, j-1]\}$ 
9          else  $m[i, j] = \min\{m[i-1, j]+1, m[i, j-1]+1, m[i-1, j-1]+1\}$ 
10 return  $m[|s_1|, |s_2|]$ 

```

Operations: insert (cost 1), delete (cost 1), **replace (cost 1)**, copy (cost 0)

Levenshtein distance: Algorithm

LEVENSHTEINDISTANCE(s_1, s_2)

```

1  for  $i \leftarrow 0$  to  $|s_1|$ 
2  do  $m[i, 0] = i$ 
3  for  $j \leftarrow 0$  to  $|s_2|$ 
4  do  $m[0, j] = j$ 
5  for  $i \leftarrow 1$  to  $|s_1|$ 
6  do for  $j \leftarrow 1$  to  $|s_2|$ 
7     do if  $s_1[i] = s_2[j]$ 
8         then  $m[i, j] = \min\{m[i-1, j]+1, m[i, j-1]+1, m[i-1, j-1]\}$ 
9         else  $m[i, j] = \min\{m[i-1, j]+1, m[i, j-1]+1, m[i-1, j-1]+1\}$ 
10 return  $m[|s_1|, |s_2|]$ 

```

Operations: insert (cost 1), delete (cost 1), replace (cost 1), **copy**
(cost 0)

Levenshtein distance: Example

		f	a	s	t
	<u>0</u>	<u>1</u> <u>1</u>	<u>2</u> <u>2</u>	<u>3</u> <u>3</u>	<u>4</u> <u>4</u>
c	<u>1</u> <u>1</u>	<u>1</u> <u>2</u> <u>2</u> <u>1</u>	<u>2</u> <u>3</u> <u>2</u> <u>2</u>	<u>3</u> <u>4</u> <u>3</u> <u>3</u>	<u>4</u> <u>5</u> <u>4</u> <u>4</u>
a	<u>2</u> <u>2</u>	<u>2</u> <u>2</u> <u>3</u> <u>2</u>	<u>1</u> <u>3</u> <u>3</u> <u>1</u>	<u>3</u> <u>4</u> <u>2</u> <u>2</u>	<u>4</u> <u>5</u> <u>3</u> <u>3</u>
t	<u>3</u> <u>3</u>	<u>3</u> <u>3</u> <u>4</u> <u>3</u>	<u>3</u> <u>2</u> <u>4</u> <u>2</u>	<u>2</u> <u>3</u> <u>3</u> <u>2</u>	<u>2</u> <u>4</u> <u>3</u> <u>2</u>
s	<u>4</u> <u>4</u>	<u>4</u> <u>4</u> <u>5</u> <u>4</u>	<u>4</u> <u>3</u> <u>5</u> <u>3</u>	<u>2</u> <u>3</u> <u>4</u> <u>2</u>	<u>3</u> <u>3</u> <u>3</u> <u>3</u>

Sample

- 1 Compute Levenshtein distance matrix for OSLO – SNOW

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$				
s	$\frac{2}{2}$				
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{\quad} 0$	$\frac{1}{1} 1$	$\frac{2}{2} 2$	$\frac{3}{3} 3$	$\frac{4}{4} 4$
o	$\frac{1}{1} 1$	$\frac{1}{2} 2$ $?$			
s	$\frac{2}{2} 2$				
l	$\frac{3}{3} 3$				
o	$\frac{4}{4} 4$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$			
s	$\frac{2}{2}$				
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{2}$ $\frac{3}{?}$		
s	$\frac{2}{2}$				
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{2}$ $\frac{3}{2}$		
s	$\frac{2}{2}$				
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{2}$ $\frac{3}{2}$	$\frac{2}{3}$ $\frac{4}{?}$	
s	$\frac{2}{2}$				
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	<u> </u> 0	<u> 1 </u> 1	<u> 2 </u> 2	<u> 3 </u> 3	<u> 4 </u> 4
o	<u> 1 </u> 1	<u> 1 2 </u> 2 1	<u> 2 3 </u> 2 2	<u> 2 4 </u> 3 2	
s	<u> 2 </u> 2				
l	<u> 3 </u> 3				
o	<u> 4 </u> 4				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{2}{4}$ $\frac{3}{2}$	$\frac{4}{5}$ $\frac{3}{?}$
s	$\frac{2}{2}$				
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{2}$ $\frac{3}{2}$	$\frac{2}{3}$ $\frac{4}{2}$	$\frac{4}{3}$ $\frac{5}{3}$
s	$\frac{2}{2}$				
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{2}{4}$ $\frac{3}{2}$	$\frac{4}{5}$ $\frac{3}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{?}$			
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{2}{4}$ $\frac{3}{2}$	$\frac{4}{5}$ $\frac{3}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$			
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{2}{4}$ $\frac{3}{2}$	$\frac{4}{5}$ $\frac{3}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{2}$ $\frac{3}{?}$		
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	<u>0</u>	<u>1 1</u>	<u>2 2</u>	<u>3 3</u>	<u>4 4</u>
o	<u>1</u> 1	<u>1 2</u> 2 1	<u>2 3</u> 2 2	<u>2 4</u> 3 2	<u>4 5</u> 3 3
s	<u>2</u> 2	<u>1 2</u> 3 1	<u>2 3</u> 2 2		
l	<u>3</u> 3				
o	<u>4</u> 4				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{2}{4}$ $\frac{3}{2}$	$\frac{4}{5}$ $\frac{3}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{3}{3}$ $\frac{3}{?}$	
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{2}$ $\frac{3}{2}$	$\frac{2}{3}$ $\frac{4}{2}$	$\frac{4}{3}$ $\frac{5}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{2}$ $\frac{3}{2}$	$\frac{3}{3}$ $\frac{3}{3}$	
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{2}$ $\frac{3}{2}$	$\frac{2}{3}$ $\frac{4}{2}$	$\frac{4}{3}$ $\frac{5}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{2}$ $\frac{3}{2}$	$\frac{3}{3}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{4}{?}$
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{2}$ $\frac{3}{2}$	$\frac{2}{3}$ $\frac{4}{2}$	$\frac{4}{3}$ $\frac{5}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{2}$ $\frac{3}{2}$	$\frac{3}{3}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{4}{3}$
l	$\frac{3}{3}$				
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{2}{4}$ $\frac{3}{2}$	$\frac{4}{5}$ $\frac{3}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{3}{3}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{4}{3}$
l	$\frac{3}{3}$	$\frac{3}{4}$ $\frac{2}{?}$			
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{2}{4}$ $\frac{3}{2}$	$\frac{4}{5}$ $\frac{3}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{3}{3}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{4}{3}$
l	$\frac{3}{3}$	$\frac{3}{4}$ $\frac{2}{2}$			
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{2}{4}$ $\frac{3}{2}$	$\frac{4}{5}$ $\frac{3}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{3}{3}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{4}{3}$
l	$\frac{3}{3}$	$\frac{3}{4}$ $\frac{2}{2}$	$\frac{2}{3}$ $\frac{3}{?}$		
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{\quad}$ 0	$\frac{1}{1}$ 1	$\frac{2}{2}$ 2	$\frac{3}{3}$ 3	$\frac{4}{4}$ 4
o	$\frac{1}{1}$ 1	$\frac{1}{2}$ 1 2 $\frac{2}{1}$ 1	$\frac{2}{2}$ 2 3 $\frac{2}{2}$ 2	$\frac{2}{3}$ 2 4 $\frac{4}{2}$ 2	$\frac{4}{3}$ 4 5 $\frac{5}{3}$ 3
s	$\frac{2}{2}$ 2	$\frac{1}{3}$ 1 2 $\frac{2}{1}$ 1	$\frac{2}{2}$ 2 3 $\frac{3}{2}$ 2	$\frac{3}{3}$ 3 3 $\frac{4}{3}$ 3	$\frac{3}{4}$ 3 4 $\frac{4}{3}$ 3
l	$\frac{3}{3}$ 3	$\frac{3}{4}$ 3 2 $\frac{2}{2}$ 2	$\frac{2}{3}$ 2 3 $\frac{3}{2}$ 2		
o	$\frac{4}{4}$ 4				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{2}{4}$ $\frac{3}{2}$	$\frac{4}{5}$ $\frac{3}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{3}{3}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{4}{3}$
l	$\frac{3}{3}$	$\frac{3}{4}$ $\frac{2}{2}$	$\frac{2}{3}$ $\frac{3}{2}$	$\frac{3}{4}$ $\frac{3}{?}$	
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2} \frac{2}{1}$	$\frac{2}{2} \frac{3}{2}$	$\frac{2}{3} \frac{4}{2}$	$\frac{4}{3} \frac{5}{3}$
s	$\frac{2}{2}$	$\frac{1}{3} \frac{2}{1}$	$\frac{2}{2} \frac{3}{2}$	$\frac{3}{3} \frac{3}{3}$	$\frac{3}{4} \frac{4}{3}$
l	$\frac{3}{3}$	$\frac{3}{4} \frac{2}{2}$	$\frac{2}{3} \frac{3}{2}$	$\frac{3}{3} \frac{4}{3}$	
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{\quad} 0$	$\frac{1}{1} 1$	$\frac{2}{2} 2$	$\frac{3}{3} 3$	$\frac{4}{4} 4$
o	$\frac{1}{1}$	$\frac{1}{2} \frac{2}{1}$	$\frac{2}{2} \frac{3}{2}$	$\frac{2}{3} \frac{4}{2}$	$\frac{4}{3} \frac{5}{3}$
s	$\frac{2}{2}$	$\frac{1}{3} \frac{2}{1}$	$\frac{2}{2} \frac{3}{2}$	$\frac{3}{3} \frac{3}{3}$	$\frac{3}{4} \frac{4}{3}$
l	$\frac{3}{3}$	$\frac{3}{4} \frac{2}{2}$	$\frac{2}{3} \frac{3}{2}$	$\frac{3}{3} \frac{4}{3}$	$\frac{4}{4} \frac{4}{?}$
o	$\frac{4}{4}$				

		s	n	o	w
	$\frac{\quad}{\quad}$ 0	$\frac{\quad}{\quad}$ 1 1	$\frac{\quad}{\quad}$ 2 2	$\frac{\quad}{\quad}$ 3 3	$\frac{\quad}{\quad}$ 4 4
o	$\frac{\quad}{\quad}$ 1 1	$\frac{\quad}{\quad}$ 1 2 $\frac{\quad}{\quad}$ 2 1	$\frac{\quad}{\quad}$ 2 3 $\frac{\quad}{\quad}$ 2 2	$\frac{\quad}{\quad}$ 2 4 $\frac{\quad}{\quad}$ 3 2	$\frac{\quad}{\quad}$ 4 5 $\frac{\quad}{\quad}$ 3 3
s	$\frac{\quad}{\quad}$ 2 2	$\frac{\quad}{\quad}$ 1 2 $\frac{\quad}{\quad}$ 3 1	$\frac{\quad}{\quad}$ 2 3 $\frac{\quad}{\quad}$ 2 2	$\frac{\quad}{\quad}$ 3 3 $\frac{\quad}{\quad}$ 3 3	$\frac{\quad}{\quad}$ 3 4 $\frac{\quad}{\quad}$ 4 3
l	$\frac{\quad}{\quad}$ 3 3	$\frac{\quad}{\quad}$ 3 2 $\frac{\quad}{\quad}$ 4 2	$\frac{\quad}{\quad}$ 2 3 $\frac{\quad}{\quad}$ 3 2	$\frac{\quad}{\quad}$ 3 4 $\frac{\quad}{\quad}$ 3 3	$\frac{\quad}{\quad}$ 4 4 $\frac{\quad}{\quad}$ 4 4
o	$\frac{\quad}{\quad}$ 4 4				

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{2}{4}$ $\frac{3}{2}$	$\frac{4}{5}$ $\frac{3}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{3}{3}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{4}{3}$
l	$\frac{3}{3}$	$\frac{3}{4}$ $\frac{2}{2}$	$\frac{2}{3}$ $\frac{3}{2}$	$\frac{3}{4}$ $\frac{3}{3}$	$\frac{4}{4}$ $\frac{4}{4}$
o	$\frac{4}{4}$	$\frac{4}{5}$ $\frac{3}{?}$			

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{2}{4}$ $\frac{3}{2}$	$\frac{4}{5}$ $\frac{3}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{3}{3}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{4}{3}$
l	$\frac{3}{3}$	$\frac{3}{4}$ $\frac{2}{2}$	$\frac{2}{3}$ $\frac{3}{2}$	$\frac{3}{4}$ $\frac{3}{3}$	$\frac{4}{4}$ $\frac{4}{4}$
o	$\frac{4}{4}$	$\frac{4}{5}$ $\frac{3}{3}$			

		s	n	o	w
	$\frac{\quad}{\quad}$ 0	$\frac{\quad}{\quad}$ 1 1	$\frac{\quad}{\quad}$ 2 2	$\frac{\quad}{\quad}$ 3 3	$\frac{\quad}{\quad}$ 4 4
o	$\frac{\quad}{\quad}$ 1 $\frac{\quad}{\quad}$ 1	$\frac{\quad}{\quad}$ 1 2 $\frac{\quad}{\quad}$ 2 1	$\frac{\quad}{\quad}$ 2 3 $\frac{\quad}{\quad}$ 2 2	$\frac{\quad}{\quad}$ 2 4 $\frac{\quad}{\quad}$ 3 2	$\frac{\quad}{\quad}$ 4 5 $\frac{\quad}{\quad}$ 3 3
s	$\frac{\quad}{\quad}$ 2 $\frac{\quad}{\quad}$ 2	$\frac{\quad}{\quad}$ 1 2 $\frac{\quad}{\quad}$ 3 1	$\frac{\quad}{\quad}$ 2 3 $\frac{\quad}{\quad}$ 2 2	$\frac{\quad}{\quad}$ 3 3 $\frac{\quad}{\quad}$ 3 3	$\frac{\quad}{\quad}$ 3 4 $\frac{\quad}{\quad}$ 4 3
l	$\frac{\quad}{\quad}$ 3 $\frac{\quad}{\quad}$ 3	$\frac{\quad}{\quad}$ 3 2 $\frac{\quad}{\quad}$ 4 2	$\frac{\quad}{\quad}$ 2 3 $\frac{\quad}{\quad}$ 3 2	$\frac{\quad}{\quad}$ 3 4 $\frac{\quad}{\quad}$ 3 3	$\frac{\quad}{\quad}$ 4 4 $\frac{\quad}{\quad}$ 4 4
o	$\frac{\quad}{\quad}$ 4 $\frac{\quad}{\quad}$ 4	$\frac{\quad}{\quad}$ 4 3 $\frac{\quad}{\quad}$ 5 3	$\frac{\quad}{\quad}$ 3 3 $\frac{\quad}{\quad}$ 4 ?		

		s	n	o	w
	<u>0</u>	<u>1 1</u>	<u>2 2</u>	<u>3 3</u>	<u>4 4</u>
o	<u>1</u> <u>1</u>	<u>1 2</u> <u>2 1</u>	<u>2 3</u> <u>2 2</u>	<u>2 4</u> <u>3 2</u>	<u>4 5</u> <u>3 3</u>
s	<u>2</u> <u>2</u>	<u>1 2</u> <u>3 1</u>	<u>2 3</u> <u>2 2</u>	<u>3 3</u> <u>3 3</u>	<u>3 4</u> <u>4 3</u>
l	<u>3</u> <u>3</u>	<u>3 2</u> <u>4 2</u>	<u>2 3</u> <u>3 2</u>	<u>3 4</u> <u>3 3</u>	<u>4 4</u> <u>4 4</u>
o	<u>4</u> <u>4</u>	<u>4 3</u> <u>5 3</u>	<u>3 3</u> <u>4 3</u>		

		s	n	o	w
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
o	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{2}{4}$ $\frac{3}{2}$	$\frac{4}{5}$ $\frac{3}{3}$
s	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{3}{3}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{4}{3}$
l	$\frac{3}{3}$	$\frac{3}{4}$ $\frac{2}{2}$	$\frac{2}{3}$ $\frac{3}{2}$	$\frac{3}{4}$ $\frac{3}{3}$	$\frac{4}{4}$ $\frac{4}{4}$
o	$\frac{4}{4}$	$\frac{4}{5}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{3}{3}$	$\frac{2}{4}$ $\frac{4}{?}$	

		s	n	o	w
	<u>0</u>	<u>1 1</u>	<u>2 2</u>	<u>3 3</u>	<u>4 4</u>
o	<u>1</u> <u>1</u>	<u>1 2</u> <u>2 1</u>	<u>2 3</u> <u>2 2</u>	<u>2 4</u> <u>3 2</u>	<u>4 5</u> <u>3 3</u>
s	<u>2</u> <u>2</u>	<u>1 2</u> <u>3 1</u>	<u>2 3</u> <u>2 2</u>	<u>3 3</u> <u>3 3</u>	<u>3 4</u> <u>4 3</u>
l	<u>3</u> <u>3</u>	<u>3 2</u> <u>4 2</u>	<u>2 3</u> <u>3 2</u>	<u>3 4</u> <u>3 3</u>	<u>4 4</u> <u>4 4</u>
o	<u>4</u> <u>4</u>	<u>4 3</u> <u>5 3</u>	<u>3 3</u> <u>4 3</u>	<u>2 4</u> <u>4 2</u>	

		s	n	o	w
	$\frac{\quad}{\quad}$ 0	$\frac{\quad}{\quad}$ 1 1	$\frac{\quad}{\quad}$ 2 2	$\frac{\quad}{\quad}$ 3 3	$\frac{\quad}{\quad}$ 4 4
o	$\frac{\quad}{\quad}$ 1 $\frac{\quad}{\quad}$ 1	$\frac{\quad}{\quad}$ 1 2 $\frac{\quad}{\quad}$ 2 1	$\frac{\quad}{\quad}$ 2 3 $\frac{\quad}{\quad}$ 2 2	$\frac{\quad}{\quad}$ 2 4 $\frac{\quad}{\quad}$ 3 2	$\frac{\quad}{\quad}$ 4 5 $\frac{\quad}{\quad}$ 3 3
s	$\frac{\quad}{\quad}$ 2 $\frac{\quad}{\quad}$ 2	$\frac{\quad}{\quad}$ 1 2 $\frac{\quad}{\quad}$ 3 1	$\frac{\quad}{\quad}$ 2 3 $\frac{\quad}{\quad}$ 2 2	$\frac{\quad}{\quad}$ 3 3 $\frac{\quad}{\quad}$ 3 3	$\frac{\quad}{\quad}$ 3 4 $\frac{\quad}{\quad}$ 4 3
l	$\frac{\quad}{\quad}$ 3 $\frac{\quad}{\quad}$ 3	$\frac{\quad}{\quad}$ 3 2 $\frac{\quad}{\quad}$ 4 2	$\frac{\quad}{\quad}$ 2 3 $\frac{\quad}{\quad}$ 3 2	$\frac{\quad}{\quad}$ 3 4 $\frac{\quad}{\quad}$ 3 3	$\frac{\quad}{\quad}$ 4 4 $\frac{\quad}{\quad}$ 4 4
o	$\frac{\quad}{\quad}$ 4 $\frac{\quad}{\quad}$ 4	$\frac{\quad}{\quad}$ 4 3 $\frac{\quad}{\quad}$ 5 3	$\frac{\quad}{\quad}$ 3 3 $\frac{\quad}{\quad}$ 4 3	$\frac{\quad}{\quad}$ 2 4 $\frac{\quad}{\quad}$ 4 2	$\frac{\quad}{\quad}$ 4 5 $\frac{\quad}{\quad}$ 3 ?

		s	n	o	w
	<u>0</u>	<u>1 1</u>	<u>2 2</u>	<u>3 3</u>	<u>4 4</u>
o	<u>1</u> <u>1</u>	<u>1 2</u> <u>2 1</u>	<u>2 3</u> <u>2 2</u>	<u>2 4</u> <u>3 2</u>	<u>4 5</u> <u>3 3</u>
s	<u>2</u> <u>2</u>	<u>1 2</u> <u>3 1</u>	<u>2 3</u> <u>2 2</u>	<u>3 3</u> <u>3 3</u>	<u>3 4</u> <u>4 3</u>
l	<u>3</u> <u>3</u>	<u>3 2</u> <u>4 2</u>	<u>2 3</u> <u>3 2</u>	<u>3 4</u> <u>3 3</u>	<u>4 4</u> <u>4 4</u>
o	<u>4</u> <u>4</u>	<u>4 3</u> <u>5 3</u>	<u>3 3</u> <u>4 3</u>	<u>2 4</u> <u>4 2</u>	<u>4 5</u> <u>3 3</u>

		s	n	o	w
	<u> </u> 0	<u> 1 </u> 1	<u> 2 </u> 2	<u> 3 </u> 3	<u> 4 </u> 4
o	<u> 1 </u> 1	<u> 1 2 </u> 2 1	<u> 2 3 </u> 2 2	<u> 2 4 </u> 3 2	<u> 4 5 </u> 3 3
s	<u> 2 </u> 2	<u> 1 2 </u> 3 1	<u> 2 3 </u> 2 2	<u> 3 3 </u> 3 3	<u> 3 4 </u> 4 3
l	<u> 3 </u> 3	<u> 3 2 </u> 4 2	<u> 2 3 </u> 3 2	<u> 3 4 </u> 3 3	<u> 4 4 </u> 4 4
o	<u> 4 </u> 4	<u> 4 3 </u> 5 3	<u> 3 3 </u> 4 3	<u> 2 4 </u> 4 2	<u> 4 5 </u> 3 3

Exercise

- 1 Compute Levenshtein distance matrix for do – redo

		r	e	d	o
	$\begin{array}{c c} & \\ \hline & 0 \end{array}$	$\begin{array}{c c} & \\ \hline 1 & 1 \end{array}$	$\begin{array}{c c} & \\ \hline 2 & 2 \end{array}$	$\begin{array}{c c} & \\ \hline 3 & 3 \end{array}$	$\begin{array}{c c} & \\ \hline 4 & 4 \end{array}$
d	$\begin{array}{c c} & 1 \\ \hline & 1 \end{array}$	$\begin{array}{c c} 1 & 2 \\ \hline 2 & 1 \end{array}$	$\begin{array}{c c} 2 & 3 \\ \hline 2 & 2 \end{array}$	$\begin{array}{c c} 2 & 4 \\ \hline 3 & 2 \end{array}$	$\begin{array}{c c} 4 & 5 \\ \hline 3 & 3 \end{array}$
o	$\begin{array}{c c} & 2 \\ \hline & 2 \end{array}$	$\begin{array}{c c} 2 & 2 \\ \hline 3 & 2 \end{array}$	$\begin{array}{c c} 2 & 3 \\ \hline 3 & 2 \end{array}$	$\begin{array}{c c} 3 & 3 \\ \hline 3 & 3 \end{array}$	$\begin{array}{c c} 2 & 4 \\ \hline 4 & 2 \end{array}$