

بازیابی اطلاعات

دکتر امین گلزاری اسکویی

a.golzari@azaruniv.ac.ir

a.golzari@tabrizu.ac.ir

<https://github.com/Amin-Golzari-Oskoue>



دانشگاه صنعتی ارومیه

پاییز ۱۴۰۲



فصل ۳

لغت نامه ها و بازیابی مقاوم

مطالب این فصل

بازیابی مقاوم

پرس و جوی جایگزین (Wildcard Query)

تصحیح املائی (Spelling Correction)



شاخص معکوس

For each term t , we store a list of all documents that contain t .

BRUTUS →

1	2	4	11	31	45	173	174
---	---	---	----	----	----	-----	-----

CAESAR →

1	2	4	5	6	16	57	132	...
---	---	---	---	---	----	----	-----	-----

CALPURNIA →

2	31	54	101
---	----	----	-----

⋮


dictionary


postings



لغت نامه‌ها

❖ لغت‌نامه، ساختمان داده‌ای برای ذخیره سازی واژگان اصطلاحی است.

❖ واژگان اصطلاحی: داده‌ها

لغت نامه به عنوان آرایه‌ای با طول ورودی‌های ثابت

❖ برای هر اصطلاح، ما باید یک زوج از موارد را ذخیره کنیم.

. فراوانی سند

. اشاره‌گر به پست لیست‌ها

...

❖ فعلا فرض کنید که ما می‌توانیم این اطلاعات را با یک ورودی طول ثابت ذخیره کنیم.

❖ فرض کنید که ما این ورودی‌ها را در آرایه ذخیره می‌کنیم.

term	document frequency	pointer to postings list
a	656,265	→
aachen	65	→
...
zulu	221	→

space needed: 20 bytes 4 bytes 4 bytes

در زمان پرس و جو، چگونه عبارت پرس و جویی q_i را در این آرایه جستجو می‌کنیم؟ به عبارت دیگر، چه ساختمان داده‌ای برای پیدا کردن ورودی (ردیف) در آرایه‌ای که q_i در آن ذخیره شده است، استفاده می‌کنیم؟

ساختمان داده‌هایی برای جستجو یک اصطلاح

❖ دو کلاس اصلی از ساختمان‌های داده: هش و درخت‌ها

❖ بعضی از سیستم‌های جستجوی اطلاعات از هش استفاده می‌کنند، بعضی دیگر از درخت‌ها

❖ معیارها برای زمانی که باید از هش یا درخت‌ها استفاده کرد:

. تعداد ثابتی از اصطلاح‌ها وجود دارد یا آیا این تعداد به مرور زمان افزایش خواهد یافت؟

. نسبت تکراری که انواع کلیدها به آن‌ها دسترسی دارند چگونه است؟

. احتمالاً چند اصطلاح خواهیم داشت؟

❖ هر اصطلاح واژگان به یک عدد صمیم هش می‌شود.

❖ سعی کنید از کالیزن‌ها (تصادم) بپرهیزید.

❖ در زمان پرس‌وجو، به ترتیب روبه‌رو عمل کنید: هش کردن عبارت پرس‌وجو، تصمیح کالیزن‌ها (تصادم)، یافتن ورودی‌هایی با طول ثابت در آرایه.

❖ مزایا: جستجوی یک هش سریع‌تر از جستجو در یک درخت است.
. زمان جستجو ثابت است.

❖ معایب:

. امکان نیافتن اختلافات جزئی مثل (*resume* vs. *résumé*)
. جستجو پیشوند ندارد (همه اصطلاح‌هایی که با “*automat*” شروع می‌شوند)
. اگر واژگان به مرور زمان افزایش می‌یابد، باید بطور دوره‌ای همه چیز را مجدداً هش کنیم.

❖ هر عبارت مجموعه واژگان، به یک عدد صمیم روی فضایی که به اندازه کافی بزرگ است و تصادم در آن بعید است، درهم سازی می‌شود.

❖ **نقاط قوت:**

❖ جستجو سریع تر از درخت انجام می‌گیرد $O(1)$

❖ **نقاط ضعف:**

❖ هیچ راه حل آسانی برای پیدا کردن گونه‌های یک کلمه وجود ندارد.

❖ مثلا judgment/judgement

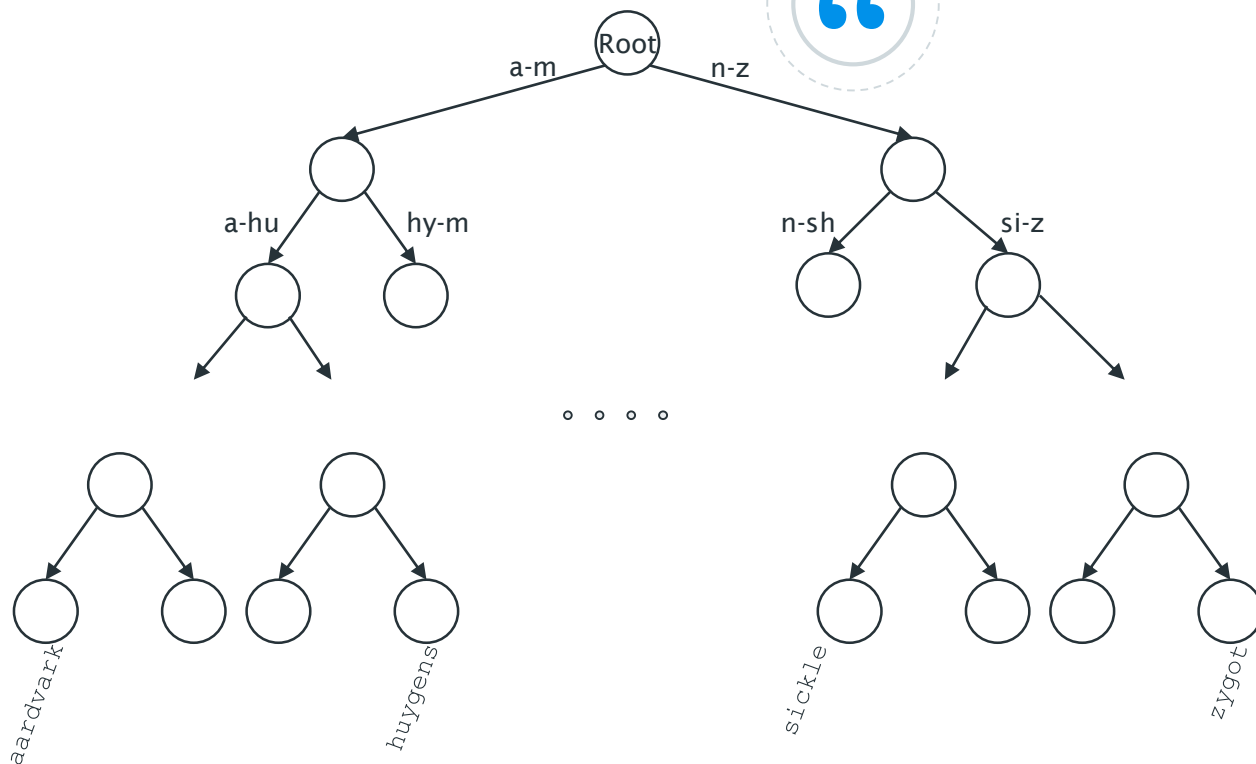
❖ جستجوی prefix وجود ندارد.

❖ اگر دامنه واژگان در حال افزایش باشد، نیاز است تا همه کلمات دوباره درهم سازی شود (عملیات پر هزینه)

درخت‌ها

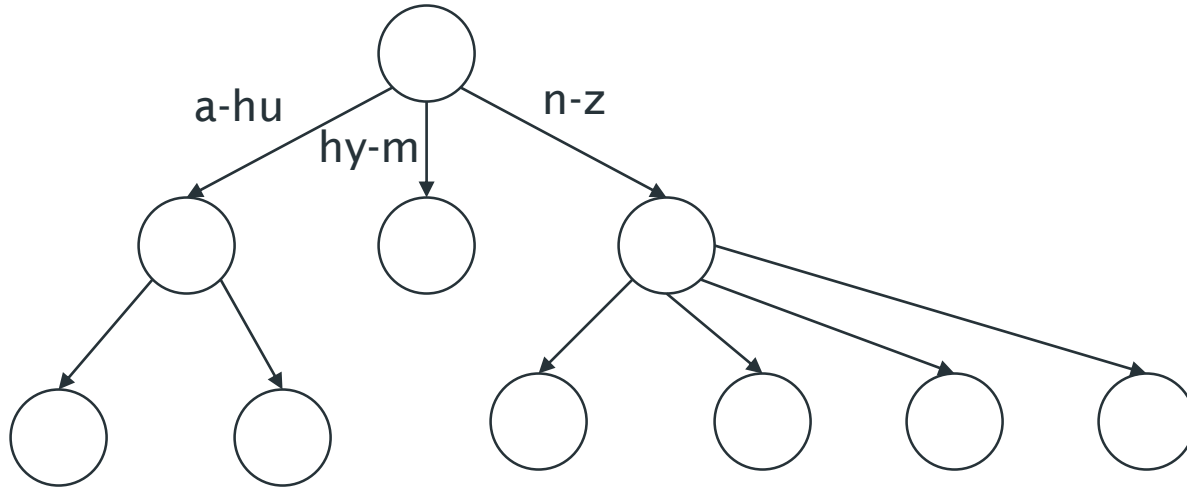
- ❖ درخت‌ها مشکل پیشوند را حل می‌کنند (همه اصطلاحات را که با automat شروع می‌شود پیدا کنید).
- ❖ ساده‌ترین درخت: درخت دودویی (Binary Tree).
- ❖ جستجو کمی‌کندتر از هاش است؛ $O(\log^n)$ که در آن N اندازه واژگان است.
- ❖ $O(\log^n)$ فقط برای درخت‌های متعادل است.
- ❖ متعادل کردن درخت‌های باینری پرهزینه است.
- ❖ B-trees مشکل تعادل را کاهش می‌دهد.
- ❖ تعریف B-tree : هر گره داخلی تعدادی فرزند در بازه $[a,b]$ دارد، که a,b اعداد صحیح مثبت هستند، به عنوان مثال $[2,4]$.

Binary tree



B-tree

“



پرس و جوهای جایگزین

❖ Mon^* : یافتن همه‌ی سندهایی که شامل هر اصطلاحی که با mon شروع می‌شوند.

❖ با B -tree میتوان همه اصطلاحات t که در بازه $mon \leq t < moo$ هستند به راحتی بازیابی کرد.

❖ $*Mon$: یافتن همه سندهایی که شامل هر اصطلاحی که به mon ختم می‌شوند.

- نگه داری یک درخت اضافی برای اصطلاحات معکوس.
 - سپس بازیابی همه اصطلاحات t که در بازه $nom \leq t < non$ هستند.
 - نیاز به داشتن یک B -tree وارونه است.
 - نتیجه: یک مجموعه‌ای از اصطلاحات که با پرس و جو جایگزین وصل می‌شوند.
- ❖ سپس بازیابی سندهایی که شامل هر یک از این اصطلاحات است.

چگونه را * در وسط یک اصطلاح به کار ببریم؟

- ❖ چگونه می‌توان co*tion در درخت جست و جو کنیم؟
- ❖ ما می‌توانیم co*tion و co* را در B-tree جستجو کرده و مجموعه دو اصطلاح را اشتراک بگیریم.

❖ هزینه‌بر

❖ جایگزین: شافص جایگردانی (Permuterm index)

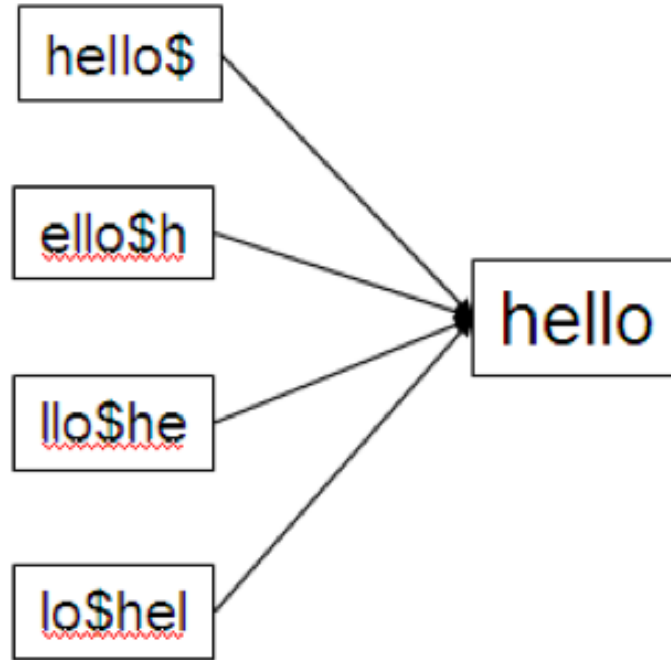
❖ ایده اولیه: هر پرس‌وجوی جایگزینی را پرفانده، تا * در انتها قرار گیرد.

شاخص جایگردانی

❖ نخست نماد ویژه \$ را برای علامت گذاری پایان هر عبارت به مجموعه‌ی کاراکتر اضافه می‌کنیم.

❖ برای کلمه hello شاخص جایگردانی را می‌سازیم که در آن دوران های مختلفی از هر عبارت همگی به عبارت مجموعه واژگان اصلی پیوند داده می شوند.

❖ For term HELLO: add *hello\$*, *ello\$h*, *llo\$he*, *lo\$hel*, and *o\$hell* to the B-tree
where \$ is a special symbol



⋮

❖ For HELLO, we've stored: *hello\$, ello\$h, llo\$he, lo\$hel, and o\$hell*

❖ Queries

For X, look up X\$

For X*, look up X*\$

For *X, look up X\$*

For *X*, look up X*

For X*Y, look up Y\$X*

Example: For hel*o, look up o\$hel*

❖ بهتر است که شاخص جایگردانی، درخت جایگردانی نامیده شود.

پردازش یک جستجو در شاخص جایگردانی

- ❖ پرفاندن پرس و جو جایگزین به راست.
- ❖ مثل قبل استفاده از جستجوی B-tree.
- ❖ مشکل: شاخص جایگردانی، اندازه لغت نامه را بیش از چهار برابر نسبت به یک B-tree معمولی افزایش می دهد.

شاخص‌های K گرمی

❖ فضای موثر تر از شاخص جایگردانی.

❖ شمارش تمامی کراکترهای k-gram (توالی از k کراکتر) موجود در یک واژه.

❖ 2-gram که به آن‌ها bigram می‌گویند.

❖ Example: from *April is the cruelest month* we get the 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\$

❖ مثل قبل \$ نماد مرزی واژه‌های ویژه است.

❖ نگهداری یک شاخص معکوس از bigrams برای اصطلاحاتی که شامل bigram هستند.

لیست‌پست‌ها در یک شاخص معکوس ۳-gram



شاخص‌های k-gram

- ❖ توجه کنید که اکنون دو نوع متفاوت از شاخص‌های معکوس داریم.
- ❖ شاخص معکوس اصطلاح-سند برای یافتن اسناد بر اساس یک پرس‌وجو متشکل از اصطلاحات.
- ❖ شاخص k-gram برای یافتن اصطلاحات بر اساس یک پرس‌وجو متشکل از k-gram‌ها.

پردازش اصطلاحات حاوی کاراکتر ویژه (*) در یک نمایه دوگرمی

- ❖ اکنون می‌توان پرس و جو mon^* را به عنوان: $m \text{ AND } mo \text{ AND } on$ اجرا کرد.
- ❖ تمام اصطلاحاتی که پیشوند mon را دارند را برای ما به دست می‌آورد...
- ❖ اما همچنین ممکن است تعداد زیادی نتایج غلط مانند MOON نیز داشته باشد.
- ❖ ما باید این اصطلاحات را نسبت به پرس و جو پس فیلتر کنیم.
- ❖ اصطلاحات باقی‌مانده سپس در شاخص معکوس اصطلاح-سند جستجو می‌شوند.
- ❖ شاخص k -gram در مقایسه با شاخص جایگردانی k -gram به ازای فضای بیشتری نیاز دارد.
- ❖ شاخص جایگردانی به postfiltering نیاز ندارد.

❖ گوگل حمایت بسیار محدودی از پرس و جو ویژه (*) دارد.

❖ به عنوان مثال، این پرس و جو در گوگل به خوبی عمل نمی کند: (gen* universit*)
. هدف: شما به دنبال دانشگاه Geneva هستید، اما نمی دانید که برای کلمات فرانسوی مربوط به دانشگاه و Geneva کدام لهجه ها را باید استفاده کنید.

❖ بر اساس مبانی جستجوی گوگل، 29-04-2010: توجه داشته باشید که اپراتور * فقط بر روی کلمات کامل کار می کند و نه برای بخش هایی از کلمات.

❖ اما این کاملاً درست نیست. (pythag*) و (m*nchen) را امتحان کنید.

❖ تمرین: چرا گوگل پرس و جو جایگزین را به طور کامل پشتیبانی نمی کند؟

پردازش اصطلاحات حاوی کاراکتر ویژه (*) در یک شاخص اصطلاح-سند

- ❖ مشکل 1: باید به احتمال زیاد تعداد زیادی پرس وجو بولی را اجرا کنیم.
- ❖ Most straightforward semantics: Conjunction of disjunctions
برای (gen* universit*) : geneva university یا geneva université یا genève university یا genève université یا general universities ...
- ❖ بسیار پرهزینه
- ❖ مشکل 2: کاربران از تایپ کردن متنفر هستند.
- ❖ اگر پرس وجوهای مختصر مانند (pyth* theo*) برای (pythagoras' theorem) مجاز باشند، کاربران از آنها به میزان زیادی استفاده می کنند.
- ❖ این باعث افزایش قابل توجه هزینه پاسخ به پرس وجوها خواهد شد.
- ❖ تا مدی توسط google suggest کاهش یافته است.

❖ دو کاربرد اصلی

. اصلاح اسنادی که در شافص گزاری هستند.

. اصلاح پرس و جوهای کاربر

❖ دو روش مختلف برای اصلاح املائی

❖ اصلاح املائی کلمات منفصل

. هر کلمه را به تنهایی برای اشتباه املائی بررسی کنید.

. اشتباههای املائی که کلمات به درستی نوشته شده منتج می شوند را گرفتار نمی کند، مثلاً،

یک سیارک که از آسمان افتاده.

❖ اصلاح املائی مساس به متن

. به کلمات مجاور نگاه کنید.

. می تواند اشتباه form/from بالا را اصلاح کند.

❖ ما در این کلاس به اصلاح املای تعاملی اسناد مثل (MS Word) علاقه‌ای نداریم.

❖ در بازیابی اطلاعات، اصلاح اسناد را به صورت اصولی برای اسناد OCR انجام می‌دهیم.
(optical character recognition = OCR)

❖ فلسفه‌ی عمومی‌در بازیابی اطلاعات این است: اسناد را تخییر ندهید.

اصلاح پرس وجوها

❖ اول : اصلاح املاي کلمات منفصل

❖ فرض 1: يك فهرست از کلمات صحيح وجود دارد که از آنها املاي صحيح به دست می آید.

❖ فرض 2: ما يك روش برای محاسبه فاصله بين يك کلمه با اشتباه املاي و يك کلمه صحيح داریم.

❖ الگوريتم ساده‌ي اصلاح املاي: کلمه 'صحيح' را که کمترین فاصله از کلمه با اشتباه املاي دارد را برگردان.

❖ مثال: “informaton → information”

❖ برای فهرست کلمات صحيح، می توانیم از واژگان تمامی کلماتی که در مجموعه مان ظاهر می شوند، استفاده کنیم.

❖ چرا اين مشکل ساز است؟

جایگزین‌هایی برای استفاده از واژگان اصطلاحی

- ❖ یک لغت‌نامه استاندارد (Webster's, OED ETC.).
- ❖ یک لغت‌نامه خاص برای صنعت (برای سیستم‌های خاص بازیابی اطلاعات).
- ❖ یک مجموعه از لغت‌نامه اصطلاحات، با وزن مناسب.

فاصله میان کلمه با اشتباه املائی و کلمه "صحیح"

❖ ما پندین جایگزین را مطالعه خواهیم کرد.

Edit distance and Levenshtein distance ❖

Weighted edit distance ❖

k -gram overlap ❖

Edit distance

❖ Edit distance بین رشته s_1 و رشته s_2 ، **کمترین** تعداد عملیات پایه‌ای است که s_1 را به s_2 تبدیل می‌کند.

❖ Levenshtein distance : عملیات‌های پایه‌ای مجازی که شامل **insert** , **delete** , **replace** است.

❖ Levenshtein distance *dog-do*: 1

❖ Levenshtein distance *cat-cart*: 1

❖ Levenshtein distance *cat-cut*: 1

❖ Levenshtein distance *cat-act*: 2

❖ Damerau-Levenshtein distance *cat-act*: 1

❖ Damerau-Levenshtein شامل عملیات تبدیل به عنوان یکی از چهار عملیات ممکن است.

Levenshtein distance: Computation

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

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

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

		f	a	s	t
	$\frac{\quad}{\quad}$ 0	$\frac{1}{1}$ 1	$\frac{2}{2}$ 2	$\frac{3}{3}$ 3	$\frac{4}{4}$ 4
c	$\frac{1}{1}$ 1	$\frac{1}{2}$ 2	$\frac{2}{2}$ 3	$\frac{3}{3}$ 4	$\frac{4}{4}$ 5
a	$\frac{2}{2}$ 2	$\frac{2}{3}$ 2	$\frac{1}{3}$ 3	$\frac{3}{2}$ 4	$\frac{4}{3}$ 5
t	$\frac{3}{3}$ 3	$\frac{3}{4}$ 3	$\frac{3}{4}$ 2	$\frac{2}{3}$ 3	$\frac{2}{3}$ 4
s	$\frac{4}{4}$ 4	$\frac{4}{5}$ 4	$\frac{4}{5}$ 3	$\frac{2}{4}$ 3	$\frac{3}{3}$ 3

- ❖ زیرساخت بهینه: راه حل بهینه برای مسئله شامل راه‌حل‌های فرعی بوده و به اصطلاح، راه‌حل‌های بهینه برای زیرمسائل می‌شود.
- ❖ هم‌پوشانی راه‌حل‌های فرعی: راه‌حل‌های فرعی هم‌پوشانی دارند. این راه‌حل‌های فرعی وقتی راه‌حل اصلی (global) بهینه داخل الگوریتم brute-force محاسبه می‌شود، به صورت مکرر محاسبه می‌شوند.
- ❖ یک زیرمسئله در مورد edit distance : edit distance دو پیشوند چه قدر است؟
- ❖ هم‌پوشانی راه‌حل‌های فرعی: ما بیشترین فاصله پیشوندها را سه بار نیاز داریم، این مربوط به حرکت کردن به راست، مورب و پایین است.

Weighted edit distance

- ❖ مشابه موارد بالا، اما وزن یک عملیات به کاراکترهای درگیر وابسته است.
- ❖ طراحی شده است به منظور گرفتن فضاهاى کیبورد، به عنوان مثال احتمالاً کلمه m به اشتباه به عنوان n نوشته شود تا به عنوان q .
- ❖ بنابراین، جایگزینی m با n در فاصله ویرایشی کوچک‌تری نسبت به جایگزینی با q است.
- ❖ اکنون نیازمند یک ماتریس وزن به عنوان ورودی هستیم.
- ❖ برنامه‌نویسی پویا را تخییر دهید تا بتواند با وزن‌ها برخورد کند.

استفاده از edit distance برای اصلاح املائی

- ❖ با توجه به پرس و جو، ابتدا تمام دنباله‌های مروفی داخل یک edit distance تعیین شده (ممکن است وزن دار) را فهرست می‌کنیم.
- ❖ این مجموعه را با فهرست کلمات "صحيح" تقاطع می‌دهیم.
- ❖ سپس کلمات موجود در چند تقاطع را به کاربر پیشنهاد می‌دهیم.

تمرین

❖ ماتریس Levenshtein distance را برای OSLO-SNOW محاسبه کنید.

❖ عملیات ویرایشی Levenshtein که کلمه cat را به catcat تبدیل می‌کند چیست؟

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

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

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

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

		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	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{3}{3}$ $\frac{3}{3}$	
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	$\frac{2}{2}$	$\frac{1}{3}$ $\frac{2}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{3}{3}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{4}{?}$
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	$\frac{1}{1}$	$\frac{1}{2}$ $\frac{2}{1}$	$\frac{2}{2}$ $\frac{3}{2}$	$\frac{2}{3}$ $\frac{4}{2}$	$\frac{4}{3}$ $\frac{5}{3}$
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	$\frac{3}{3}$	$\frac{3}{4}$ $\frac{2}{2}$			
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
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$
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	$\frac{3}{3}$	$\frac{3}{4}$ $\frac{2}{2}$	$\frac{2}{3}$ $\frac{3}{2}$		
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	$\frac{3}{3}$	$\frac{3}{4}$ $\frac{2}{2}$	$\frac{2}{3}$ $\frac{3}{2}$	$\frac{3}{3}$ $\frac{4}{?}$	
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	$\frac{3}{3}$	$\frac{3}{4}$ $\frac{2}{2}$	$\frac{2}{3}$ $\frac{3}{2}$	$\frac{3}{4}$ $\frac{3}{3}$	
o	$\frac{4}{4}$				

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

		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	$\frac{4}{4}$				

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

		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	$\frac{2}{2}$	$\frac{1}{2}$ $\frac{3}{1}$	$\frac{2}{3}$ $\frac{2}{2}$	$\frac{3}{3}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{4}{3}$
l	$\frac{3}{3}$	$\frac{3}{2}$ $\frac{4}{2}$	$\frac{2}{3}$ $\frac{3}{2}$	$\frac{3}{4}$ $\frac{3}{3}$	$\frac{4}{4}$ $\frac{4}{4}$
o	$\frac{4}{4}$	$\frac{4}{3}$ $\frac{5}{3}$	$\frac{3}{3}$ $\frac{4}{?}$		

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

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

		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	$\frac{4}{4}$	$\frac{4}{5}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{3}{3}$	$\frac{2}{4}$ $\frac{4}{2}$	$\frac{4}{5}$ $\frac{3}{?}$

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

		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	$\frac{4}{4}$	$\frac{4}{5}$ $\frac{3}{3}$	$\frac{3}{4}$ $\frac{3}{3}$	$\frac{2}{4}$ $\frac{4}{2}$	$\frac{4}{3}$ 5 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>

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

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

cost	operation	input	output
1	insert	*	w

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

cost	operation	input	output
0	(copy)	o	o
1	insert	*	w

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

cost	operation	input	output
1	replace	l	n
0	(copy)	o	o
1	insert	*	w

		s	n	o	w
	0	1 1	2 2	3 3	4 4
o	1 1	1 2 2 1	2 3 2 2	2 4 3 2	4 5 3 3
s	2 2	1 2 3 1	2 3 2 2	3 3 3 3	3 4 4 3
l	3 3	3 2 4 2	2 3 3 2	3 4 3 3	4 4 4 4
o	4 4	4 3 5 3	3 3 4 3	2 4 4 2	4 5 3 3

cost	operation	input	output
0	(copy)	s	s
1	replace	l	n
0	(copy)	o	o
1	insert	*	w

		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 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>

cost	operation	input	output
1	delete	o	*
0	(copy)	s	s
1	replace	l	n
0	(copy)	o	o
1	insert	*	w

		c	a	t	c	a	t
	0	1 1	2 2	3 3	4 4	5 5	6 6
c	1 1	0 2 2 0	2 3 1 1	3 4 2 2	3 5 3 3	5 6 4 4	6 7 5 5
a	2 2	2 1 3 1	0 2 2 0	2 3 1 1	3 4 2 2	3 5 3 3	5 6 4 4
t	3 3	3 2 4 2	2 1 3 1	0 2 2 0	2 3 1 1	3 4 2 2	3 5 3 3

		c	a	t	c	a	t
	$\begin{array}{ c } \hline 0 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 1 & 1 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 2 & 2 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 3 & 3 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 4 & 4 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 5 & 5 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 6 & 6 \\ \hline \end{array}$
c	$\begin{array}{ c } \hline 1 \\ \hline 1 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 0 & 2 \\ \hline 2 & 0 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 2 & 3 \\ \hline 1 & 1 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 3 & 4 \\ \hline 2 & 2 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 3 & 5 \\ \hline 3 & 3 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 5 & 6 \\ \hline 4 & 4 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 6 & 7 \\ \hline 5 & 5 \\ \hline \end{array}$
a	$\begin{array}{ c } \hline 2 \\ \hline 2 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 2 & 1 \\ \hline 3 & 1 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 0 & 2 \\ \hline 2 & 0 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 2 & 3 \\ \hline 1 & 1 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 3 & 4 \\ \hline 2 & 2 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 3 & 5 \\ \hline 3 & 3 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 5 & 6 \\ \hline 4 & 4 \\ \hline \end{array}$
t	$\begin{array}{ c } \hline 3 \\ \hline 3 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 3 & 2 \\ \hline 4 & 2 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 2 & 1 \\ \hline 3 & 1 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 0 & 2 \\ \hline 2 & 0 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 2 & 3 \\ \hline 1 & 1 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 3 & 4 \\ \hline 2 & 2 \\ \hline \end{array}$	$\begin{array}{ c c } \hline 3 & 5 \\ \hline 3 & 3 \\ \hline \end{array}$

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

		c		a		t		c		a		t	
		<div><div></div><div>0</div></div>	<div><div>1</div><div>1</div></div>	<div><div>2</div><div>2</div></div>	<div><div>3</div><div>3</div></div>	<div><div>4</div><div>4</div></div>	<div><div>5</div><div>5</div></div>	<div><div>6</div><div>6</div></div>					
c		<div><div>1</div><div>1</div></div>	<div><div>0</div><div>2</div><div>2</div><div>0</div></div>	<div><div>2</div><div>3</div><div>1</div><div>1</div></div>	<div><div>3</div><div>4</div><div>2</div><div>2</div></div>	<div><div>3</div><div>5</div><div>3</div><div>3</div></div>	<div><div>5</div><div>6</div><div>4</div><div>4</div></div>	<div><div>6</div><div>7</div><div>5</div><div>5</div></div>					
a		<div><div>2</div><div>2</div></div>	<div><div>2</div><div>1</div><div>3</div><div>1</div></div>	<div><div>0</div><div>2</div><div>2</div><div>0</div></div>	<div><div>2</div><div>3</div><div>1</div><div>1</div></div>	<div><div>3</div><div>4</div><div>2</div><div>2</div></div>	<div><div>3</div><div>5</div><div>3</div><div>3</div></div>	<div><div>5</div><div>6</div><div>4</div><div>4</div></div>					
t		<div><div>3</div><div>3</div></div>	<div><div>3</div><div>2</div><div>4</div><div>2</div></div>	<div><div>2</div><div>1</div><div>3</div><div>1</div></div>	<div><div>0</div><div>2</div><div>2</div><div>0</div></div>	<div><div>2</div><div>3</div><div>1</div><div>1</div></div>	<div><div>3</div><div>4</div><div>2</div><div>2</div></div>	<div><div>3</div><div>5</div><div>3</div><div>3</div></div>					

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

		c	a	t	c	a	t
	$\frac{\quad}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{4}{4}$	$\frac{5}{5}$	$\frac{6}{6}$
c	$\frac{1}{1}$	$\frac{0}{2}$	$\frac{2}{1}$	$\frac{3}{2}$	$\frac{4}{3}$	$\frac{5}{4}$	$\frac{6}{5}$
a	$\frac{2}{2}$	$\frac{2}{3}$	$\frac{0}{2}$	$\frac{2}{1}$	$\frac{3}{2}$	$\frac{3}{3}$	$\frac{5}{4}$
t	$\frac{3}{3}$	$\frac{3}{4}$	$\frac{2}{3}$	$\frac{0}{2}$	$\frac{2}{1}$	$\frac{3}{2}$	$\frac{3}{3}$

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

		c		a		t		c		a		t		
		0	1	1	2	2	3	3	4	4	5	5	6	6
c		1 1	0 2	2 0	2 1	3 1	3 2	4 2	3 3	5 3	5 4	6 4	6 5	7 5
a		2 2	2 3	1 1	0 2	2 1	3 1	3 1	3 2	4 2	3 3	5 3	5 4	6 4
t		3 3	3 4	2 2	2 3	1 1	0 2	2 0	2 1	3 1	3 2	4 2	3 3	5 3

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

برای اصلاح املائی شاخص‌های "k-gram"

❖ اکنون که ما می‌توانیم edit distance را محاسبه کنیم: چگونه از آن برای اصلاح املائی کلمات به

صورت جداگانه استفاده کنیم - این آخرین اسلاید در این بخش است.

❖ شاخص‌های k-gram برای اصلاح املائی کلمات به صورت جداگانه.

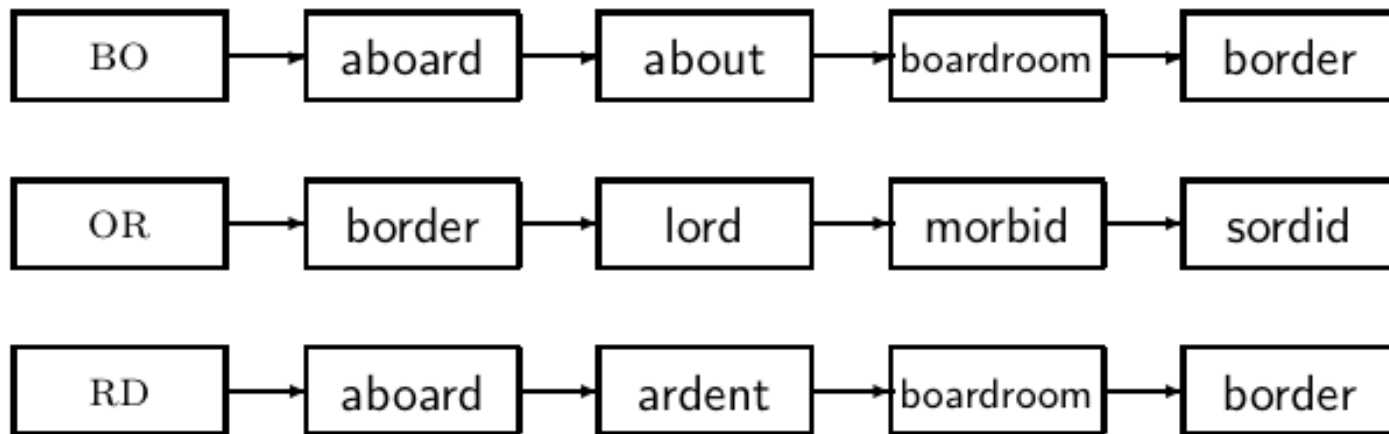
❖ اصلاح املائی محاسب به متن

❖ مسائل عمومی

برای اصلاح املائی شاخص‌های "k-gram"

- ❖ تمامی k-gram ها را در عبارت پرس‌وجو بشمارید.
- ❖ مثال: فهرست bigram، کلمه اشتباه نوشته شده boardroom
- ❖ bigram: bo, or, rd, dr, ro, oo, om
- ❖ از شاخص‌های k-gram برای بازیابی کلمات 'صمیم' که با k-gram های عبارت پرس‌وجو مطابقت دارند، استفاده کنید.
- ❖ حداقل تعداد k-gram های مطابقت‌دهنده را مشخص کنید.
- ❖ به عنوان مثال، فقط اصطلاحات واژگانی که حداقل در سه k-gram از یکدیگر متفاوت هستند.

k-gram indexes for spelling correction: boardroom



اصلاح املائی حساس به متن

❖ مثال ما: یک سیارکی که از آسمان افتاد.

❖ چگونه می‌توانیم فرم را در اینجا تصحیح کنیم؟

❖ یک ایده: *hit-base* spelling correction

❖ عبارات «درست» نزدیک به هر عبارت پرس‌وجو را بازیابی کنید.

❖ برای *flea for flew, from for form, munch for :flew form munich*

❖ *munich*

❖ اکنون تمام عبارات ممکن را به عنوان پرس‌وجو با یک کلمه "تثبیت" در یک زمان امتحان کنید.

❖ پرس‌وجوی «*flea form munich*» را امتحان کنید.

❖ پرس‌وجوی «*flew from munich*» را امتحان کنید.

❖ عبارت *flew form munch* را امتحان کنید.

❖ پرس‌وجو صحیح *Flew from Munich* بیشترین بازدید را دارد.

❖ فرض کنید 7 جایگزین برای پرواز، 20 گزینه برای فرم و 3 گزینه برای مونیف داریم، چند عبارت تصحیح شده را بر می‌شمریم؟



❖ الگوریتم hit-based که ما به آن اشاره کردیم خیلی کارآمد نیست.

❖ جایگزین کارآمدتر: به "مجموعه" پرس و جوها نگاه کنید، نه اسناد.

مسائل کلی در تصحیح املا

❖ (رابطا کاربري)

- تصميح خودکار در مقابل اصلاحي پيشنهادي .
- آيا منظور شما فقط براي يک پيشنهاد کار مي‌کند؟
- در مورد اصلاحيات متعدد ممکن چطور؟
- Tradeoff : رابفا کاربري ساده در مقابل قدرتمند

❖ هزينه

- تصميح املا به طور بالقوه گران است.
- از اجرائي هر پرس و جو اجتناب کنيد(؟).
- شايد فقط در مورد پرس و جوهايي که با اسناد کمي مطابقت دارند.
- مدس بزنيدي: تصميح املاي موتورهاي جستجوي اصلي به اندازه کافي کارآمد است که در هر پرس و جو اجرا شود.

Exercise: Understand Peter Norvig's spelling corrector

```
import re, collections
def words(text): return re.findall('[a-z]+', text.lower())
def train(features):
    model = collections.defaultdict(lambda: 1)
    for f in features:
        model[f] += 1
    return model
NWORDS = train(words(file('big.txt').read()))
alphabet = 'abcdefghijklmnopqrstuvwxyz'
def edits1(word):
    splits      = [(word[:i], word[i:]) for i in range(len(word) + 1)]
    deletes     = [a + b[1:] for a, b in splits if b]
    transposes  = [a + b[1] + b[0] + b[2:] for a, b in splits if len(b)>1]
    replaces    = [a + c + b[1:] for a, b in splits for c in alphabet if b]
    inserts     = [a + c + b      for a, b in splits for c in alphabet]
    return set(deletes + transposes + replaces + inserts)
def known_edits2(word):
    return set(e2 for e1 in edits1(word) for e2 in
               edits1(e1) if e2 in NWORDS)
def known(words): return set(w for w in words if w in NWORDS)
def correct(word):
    candidates = known([word]) or known(edits1(word)) or
                  known_edits2(word) or [word]
    return max(candidates, key=NWORDS.get)
```


Soundex

❖ Soundex مبنایی برای یافتن جایگزین‌های آوایی (در مقابل املای) است.

❖ مثال: chebyshev / tchebyscheff

❖ الگوریتم:

- . هر نشانه را برای ایندکس شدن به یک فرم کاهش یافته 4 کاراکتری تبدیل کنید.
- . همین کار را با اصطلاحات پرس و جو انجام دهید.
- . بر روی فرم‌های کاهش‌یافته یک فهرست بسازید و جستجو کنید.

Soundex algorithm

- ❖ حرف اول عبارت را حفظ کنید.
- ❖ همه موارد تکرار حروف زیر را به 0 (صفر) تغییر دهید: A, E, I, O, U, H, W, Y
- ❖ حروف را به صورت زیر به رقم تبدیل کنید:
 - ❖ 1 V, P, F, B
 - ❖ 2 Z, X, S, Q, K, J, G, C
 - ❖ 3 D, T
 - ❖ 4 L
 - ❖ 5 M, N
 - ❖ 6 R
- ❖ از هر جفت ارقام متوالی یکسان یکی را به طور مکرر حذف کنید.
- ❖ تمام صفرها را از رشته به دست آمده حذف کنید. رشته به دست آمده را با صفرهای دنباله‌دار بپوشانید و چهار جایگاه اول را برگردانید، که شامل یک حرف و به دنبال آن سه عدد خواهد شد.

Example: Soundex of HERMAN

- ❖ Retain H
- ❖ $ERMAN \rightarrow ORM0N$
- ❖ $ORM0N \rightarrow 06505$
- ❖ $06505 \rightarrow 06505$
- ❖ $06505 \rightarrow 655$
- ❖ Return $H655$
- ❖ Note: *HERMANN* will generate the same code

Soundex چقدر مفید است؟

❖ برای بازیابی اطلاعات خیلی مفید نیست.

❖ برای کارهای «high recall» در سایر برنامه‌ها (مانند Interpol) خوب است.

❖ Zobel و Dart (1996) جایگزین‌های بهتری برای تطبیق آوایی در IR پیشنهاد می‌کنند.

❖ Compute Soundex code of your last name.



تشکر

سوال؟

a.golzari@azaruniv.ac.ir

a.golzari@tabrizu.ac.ir

<https://github.com/Amin-Golzari-Oskouei>