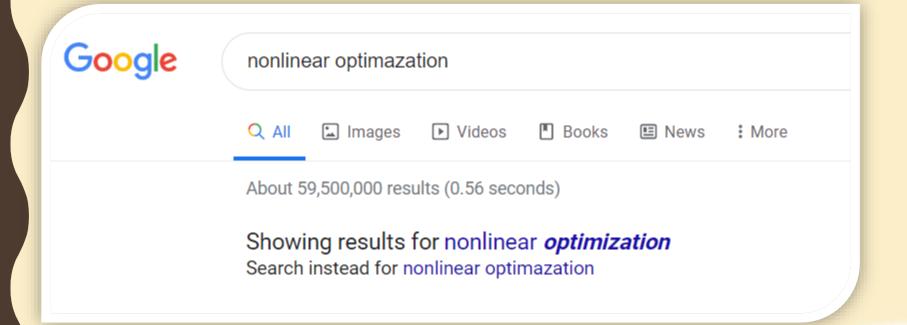
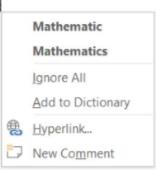
MINIMUM EDIT DISTANCE

FARZAD IMANPOUR



Mathemtic



چگونه دو متن شبیه هم هستند؟ ارتباط آن ها چه اهمیتی دارد؟

- . در علم زیست شناسی
- _ دو دنباله نوکلئوتیدها
- AGGCTATCACCTGACCTCCAGGCCGATGCCC -
- TAGCTATCACGACCGCGGTCGATTTGCCCGA C
 - _ پاسخ :
- -AGGCTATCACCTGACCTCCAGGCCGA--TGCCC--TAG-CTATCAC--GACCGC--GGTCGATTTGCCCGAC

اصلاح كلمات

_ وقتی کاربر تایپ میکند "graffe"

_ چه کلماتی به آن نزدیک هستند؟

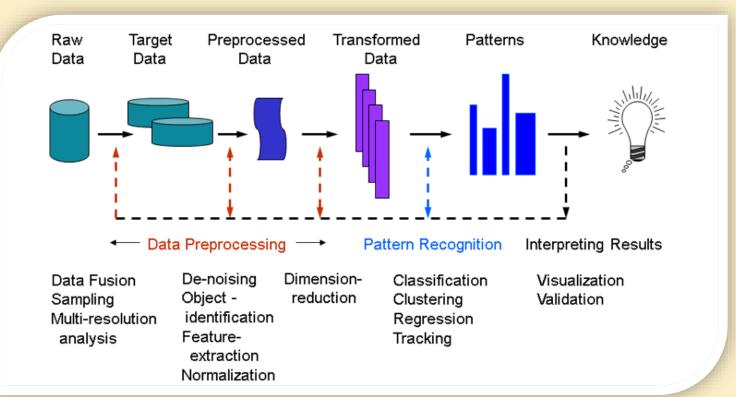
Graf.

Graft.

Grail.

Girrafe.

- موارد استفاده:
- تطبیق تقریبی رشته
- Google finding similar word –
- مقایسه توالی در DNA (استفاده بسیاز زیاد در علم ژنتیک)
 - تشخیص الگو (Pattern Recognition)





• مسئله:

بین دو رشته Edit distance بین دو رشته

- راه های حل ها و پیچیدگی:
- Brute Force $O(K^N) -$
- N = sI + s2, K = N (sI + s2(repetetive))
 - Greedy No Optimal Algorithms yet -
- Divide & Conquer None discovered yet
 - Dynamic Programming $-O(N^2)$

• چه تعداد تغیرات نیاز داریم تا از رشته داده شده به رشته هدف برسیم؟

odtifmiz optimize

- رشته داده شده:
 - هدف:
 - 3 تغير:
- (modify)p به d تغیر
 - رdelete) f حذف
- اضافه کردن e (add) e

TCGACGTCA

رشته داده شده: TGACGTGC هدف:

> if (strl[i-l] == str2[j-l]) d[i][j] = d[i-l][j-l] Or

else d[i][j] = I + min(d[i][j-I], d[i-I][j],d[i-1][j-1]);

		Т	С	G	Α	С	G	Т	С	Α
	0	I	2	3	4	5	6	7	8	9
Т	I	0		2	3	4	5	6	7	8
G	2		2		2	3	4	5	6	7
Α	3	2	3	2		2	3	4	5	6
С	4	3	2	3	2		2	3	4	5
G	5	4	3	2	3	2	ı	2	3	4
Т	6	5	4	3	4	3	2	ı	2	3
G	7	6	5	4	5	4	3	2	3	4
С	8	7	6	5	6	5	5	3	2	3

Replace	Add
Delete	You are here

- هیچ راهی بهینه ای به غیر این که تمام مسیر هارا چک کنیم وجود ندارد.
 - برای همین هم تمام مسیر ها تک به تک چک میشوند
 - تا به این ثانیه روش های اندکی با پیچیدگی کمتر ارائه شده اند که همگی پس از بررسی چند مثال رد شده اند.
 - پس بهترین روش برای حل این مسئله (O(nm) می باشد
 - اگر شما بتوانید روش بهتری ارائه دهید یک انسان شناخته شده بزرگ خواهید شد.