



Bölüm 5: Dizgi Algoritmaları

Algoritmalar



Dizgi Düzenleme Mesafesi Algoritmaları

- Levenshtein Mesafesi
 - İki dizgi arasındaki benzerliği ölçen bir metrik,
 - Bir dizgiden diğerine dönüştürmek için gereken minimum tek karakterli düzenleme sayısı olarak tanımlanır.
- En Uzun Ortak Alt Dizi (Longest Common Subsequence - LCS)
 - İki dizginin ortak olan en uzun alt dizisi,
 - Karakterlerin sıralı olmasını gerektirmez, ancak sıra korunmalıdır.
 - Dizgiler arasındaki benzerlik veya farkı belirlemek için kullanılır.



Levenshtein Mesafesi

- İki dizgi arasındaki farkı nicel olarak ölçen etkili bir metriktir.
- Otomatik düzeltme ve tahmin sistemlerinde kullanılır.
- İki dizgi arasındaki minimum işlem (ekleme, çıkarma veya değiştirme) sayısını belirtir.



Temel İşlemler

- Ekleme (*Insertion*):
 - Bir karakterin eklenmesi.
- Çıkarma (*Deletion*):
 - Bir karakterin çıkarılması.
- Değiştirme (*Substitution*):
 - Bir karakterin başka bir karakterle değiştirilmesi.



Levenshtein Mesafesi Hesaplama

- İki dizgi arasındaki minimum düzenleme işlemi sayısı olarak hesaplanır.
- Dinamik programlama yöntemiyle hesaplanır.
- İki dizgi arasındaki karakterlerin karşılaştırılması ve işlem maliyetlerinin belirlenmesi ile yapılır.

Levenshtein



H	O		N	D	A	
H	Y	U	N	D	A	I

H	Y	U	N	D	A	I
H		O	N	D	A	

Levenshtein



	""	P	A	I	R	S
""	0	1	2	3	4	5
C	1	1	2	3	4	5
A	2	2	1	2	3	4
R	3	3	2	2	2	3
S	4	4	3	3	3	2



Levenshtein

rain
sain
shin
shine

(a)

shine
rhine
raine
rain

(b)

shine
tshine
trhine
train

(c)



Substitution



Insertion



Deletion





En Uzun Ortak Alt Dizgi (LCS)

- İki dizgi içinde sıralı olarak bulunan ve mümkün olan en uzun dizgidir.
- Bu alt dizgi, dizgilerin karakter sırasını bozmaz ancak ardışık olmak zorunda değildir.
- İki dizgi arasındaki benzerlik seviyesini ölçer.
- Dinamik programlama yöntemiyle hesaplanır.
- İki dizgi arasındaki karakterlerin sıralı şekilde eşleştirilmesiyle elde edilir.

Örnek



- Dizi 1: AGGTAB
- Dizi 2: GXTXAYB
- LCS: GTAB



Dinamik Programlama Yaklaşımı

- İki boyutlu bir tablo oluşturulur.
- Her hücre, alt dizinin o noktaya kadar olan LCS uzunluğunu temsil eder.
- Tablonun sonunda LCS uzunluğu bulunur.



Algoritma

- İki dizinin uzunlukları m ve n olsun.
- $(m+1) \times (n+1)$ boyutunda bir tablo oluşturulur.
- İlk satır ve sütun sıfır ile doldurulur.
- $A[i] == B[j]$ ise, hücre değeri üst-sol köşedeki değerin 1 fazlasıdır.
- $A[i] != B[j]$ ise, hücre değeri üst veya sol hücrenin maksimum değeri olur.
- Tablonun son hücresi, LCS uzunluğunu verir.



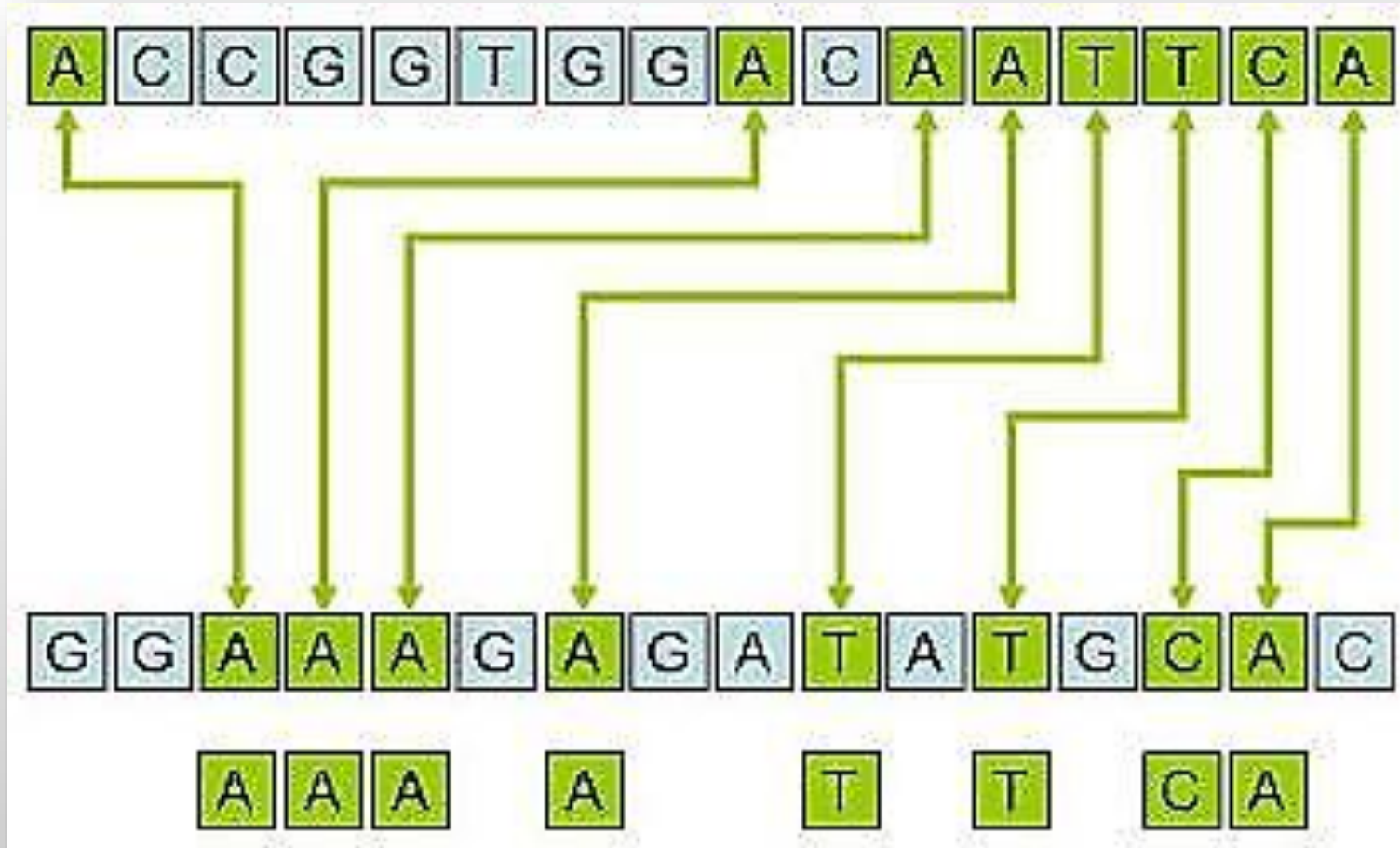
En Uzun Ortak Alt Dizi (LCS)

string 1	a	c	b	a	e	d
string 2	a	b	c	a	d	f

LCS: "acad" with length 4



En Uzun Ortak Alt Dizi (LCS)





SON