



Bölüm 5: Dizgi Algoritmaları

Algoritmalar



Dizgi Dönüşüm Algoritmaları

- Sonek Dizisi (*Suffix Array*)
 - Bir dizginin tüm son eklerinin bir dizisi.
 - Dizgi içindeki alt dizgilerin bir temsili olarak kullanılır.
- Burrows-Wheeler Dönüşümü (*BWT*)
 - Bir dizginin tersine dönüştürülmesiyle elde edilen yeni bir form,
 - Bzip2 gibi sıkıştırma algoritmaları için ön işlem adımı olarak kullanılır.



Sonek Dizisi (Suffix Array)

- Bir dizginin tüm soneklerinin alfabetik olarak sıralanmış halidir.
- Her bir sonek, dizginin belirli bir konumundan başlayan bir alt dizisidir.
- Dizgi arama, sıralama, genetik dizilim analizi benzeri işlemlerde kullanılır.



Oluşturma Yöntemleri

- Brute Force:
 - Tüm sonekleri oluşturup ardından sıralar.
 - $O(n^2 \log n)$ zaman karmaşıklığına sahiptir.
- Manber-Myers Algoritması:
 - Lineer zaman karmaşıklığına sahiptir. ($O(n \log n)$)
 - Soneklerin sıralanması sırasında tekrar tekrar harf karşılaştırması yapılmaz.
- Larsson-Sadakane Algoritması:
 - Sıralama işleminde art arda soneklerin karşılaştırılması esas alınır.





Burrows-Wheeler Dönüşümü (BWT)

- Girdi dizisini yeniden düzenler.
- Aynı karakterlerin bir araya toplanmasını sağlar.
- Sıkıştırma algoritmalarının performansını artırır.
- Tersine çevrilebilir: Orijinal veri, dönüşümden geri elde edilebilir.



Uygulama Adımları

- Girdi dizgisinin tüm döndürülmüş hallerini oluştur.
- Döndürülmüş dizgileri alfabetik olarak sırala.
- Sıralanmış dizgilerin son karakterlerinden yeni bir dizgi oluştur.



Örnek

- Girdi:
 - BANANA
- Döndürmeler:
 - BANANA, ANANAB, NANABA, ANABAN, NABANA, ABANAN
- Sıralanmış Döndürmeler:
 - ABANAN, ANABAN, ANANAB, BANANA, NABANA, NANABA
- Son Karakterler:
 - NNBAAA
- BWT Sonucu:
 - NNBAAA



Burrows-Wheeler

S = abraca

a	b	r	a	c	a
b	r	a	c	a	a
r	a	c	a	a	b
a	c	a	a	b	r
c	a	a	b	r	a
a	a	b	r	a	c

Initial Matrix

S' = caraab

a	a	b	r	a	c
a	b	r	a	c	a
a	c	a	a	b	r
b	r	a	c	a	a
c	a	a	b	r	a
r	a	c	a	a	b

Sorted Matrix

L





SON