

**Yasanın teorik anlatımı:**

Zipf (1935) yasası sayısal dilbilimci tarafından geliştirilerek enformetride yaygın kullanım alanı bulan bir tekniktir. Doğal dil kullanılan metinlerdeki kelime sıklığı ile kelimelerin sıra sayıları arasındaki ilişki üzerinde kurulan bu yasa bilgi erişimde yararlanılan bir teknik olarak yerini alır. Bir metinde geçen kelimelerin içerikleri bakımından temsil ettikleri konuyla ilişki düzeylerini belirlemeyi amaçlar. Kuramsal açılımında, anlamlı kelimelerin metin içinde kullanıldıkları "sıra" ile o metindeki kullanım sıklıkları çarpımı sabit bir değer üretir. Kaynak olarak her metnin ürettiği kelime ya da kelime grupları azalan verimlilik esasına göre sıralanır. Belirli bir konudaki dokümanların her biri için bu işlem sürdürülürse o konudaki dokümanların da aynı konuyu temsil etme verimliliklerine göre sıralanması olanaklı olacaktır.

**Aşağıda bir metinden seçilmiş anlamlı terimler için yapılan bir çalışmanın özetini bulacaksınız.**

$$R \cdot f(r) = C$$

- R: sıra
- F(r): kullanım sıklığı
- C: sabit değer (çarpan)

Kelime	Sıra ( R)	Kullanım Sıklığı f(r)	$r \cdot f = c$ Sabit d.
Kütüphane	7	9	63
Ödünç	8	8	64
Sağlama	9	7	63

$$V = 1 / K(r) \times C$$

V: Kelimelerin metindeki kullanım sıklığı

K: Metinde geçen toplam kelime sayısı

Yukarıdaki tabloda 1500 kelimelik bir metinden seçilmiş terimler örnek olarak verilmiştir. Sabit değer 63 alındığında, buna göre ;

$$V = 1 / 1500 \times 63$$

$$V = 0.042$$

**Uygulama çalışması:** İkinci metni de siz seçerek Zipf yasasının ampirik ifadesi ile bir karşılaştırma yapınız.