

Datasetler

Türkçe Vikipedi

Sütunlar = [id, title, content]

Satır Sayısı = 754265

	Sayfa Uzunlukları	
	Karakter Bazlı	Kelime Bazlı
min	3	1
mean	956.53	125.04
median	170	22
max	416681	55339

En çok geçen 10 kelime ve frekansları:

```
[('ㄣ喃', 2270156),  
( 'İmparatorluğu', 2149962),  
( 'doğum', 1400506),  
( 'adıyla', 1216859),  
( 'Temuçin', 1062081),  
( 'Moğol', 741796),  
( 'Han', 465491),  
( 'Ağustos', 132650),  
( '18', 11610),  
( '1227', 7360)]
```

Forum Donanım Haber

Sütunlar = [id, url, post, response]

Satır Sayısı = 1775

	Post Uzunlukları		Response Uzunlukları		Post + Response Uzunlukları	
	Karakter Bazlı	Kelime Bazlı	Karakter Bazlı	Kelime Bazlı	Karakter Bazlı	Kelime Bazlı
min	15	1	15	1	31	4
mean	2022.5	250.14	424.17	58.04	2447.68	308.18
median	1029	127	139	20	1293	164
max	181698	23935	13047	1709	185558	24458

Postlarda en çok geçen 10 kelime

[('한국정보디스플레이학회 youtubeIMID', 88092),
('çizim', 84573),
('Çok', 83042),
('yapıyorum', 79641),
('ve', 77735),
('tarzı', 74187),
('kullanacağım', 59014),
('SAI', 23024),
('Photoshop', 21762),
('PaintTool', 21440)]

Response'larda en çok geçen 10 kelime:

[('ターゲットに衝突すると小さな爆発を起こす火の玉を発射する', 31929),
('Öncelikle', 29896),
('yapalım', 28178),
('walmart', 27813),
('uc', 26927),
('marka', 20688),
('logic', 20436),
('gibi', 14767),
('ayrımını', 8058),
('Wacom', 5842)]

Post + Response'larda en çok geçen 10 kelime:

[('한국정보디스플레이학회 youtubeIMID', 102630),
('çizim', 98464),
('Çok', 96668),
('yapıyorum', 92612),
('ve', 90264),
('tarzı', 86045),
('kullanacağım', 68017),
('SAI', 25269),
('Photoshop', 23884),
('PaintTool', 23516)]