

Erişilebilir Bağlantılar ve Arama Motoru Sıralamaları: Ampirik Bir İnceleme

Göker CEBECİ

gokercebeci@std.yildiz.edu.tr;
Yıldız Teknik Üniversitesi,
Elektrik-Elektronik Fakültesi,
Bilgisayar Mühendisliği Bölümü,
İstanbul / Türkiye.

Banu DİRİ

diri@yildiz.edu.tr;
Yıldız Teknik Üniversitesi,
Elektrik-Elektronik Fakültesi,
Bilgisayar Mühendisliği Bölümü,
İstanbul / Türkiye.

Veri Seti

<https://goker.dev/datasets/hyperlinks>

Kaynak Kodu

<https://goker.dev/codes/hyperlinks>

Sunum Akışı

- Giriş ve Motivasyon
- Problem Tanımı ve Araştırma Sorusu
- Materyal ve Yöntem
 - Veri Toplama ve Etiketleme
 - Özellik Çıkarımı
 - Modelleme
- Bulgular ve Deneysel Sonuçlar
- Tartışma
- Sonuç ve Gelecek Çalışmalar
- Soru & Cevap

Motivasyon

- **SEO'nun Önemi:** SERP'in ilk sırası, onuncu sıradan 10 kat fazla tıklanma alıyor.
- **Erişilebilirliğin Rolü:** W3C standartları ve arama motorlarının erişilebilirliğe verdiği önem.
- **Boşluk:** SEO ve erişilebilirlik ilişkisi bilinse de, spesifik olarak iç bağlantıların erişilebilirlik özelliklerinin SERP'e etkisi üzerine ampirik bir çalışma eksikliği.

Problemin Tanımı ve Araştırma Sorusu

- **Hipotez:** Web sitelerindeki iç bağlantıların erişilebilirlik özellikleri, arama motoru sıralamasını etkiler.
- **Araştırma Sorusu:** Hangi erişilebilirlik özellikleri sıralama için daha önemlidir ve bu etki ne düzeydedir?
- **Yaklaşım:** Dış faktörleri (*anahtar kelime rekabeti, backlink vb.*) en aza indirmek için "**site:[adres]**" özel arama operatörü kullanılarak site içi hiyerarşiye odaklanılmıştır.

Materyal ve Yöntem

Veri Toplama ve Etiketleme

- **Hipotez:** Web sitelerindeki iç bağlantıların erişilebilirlik özellikleri, arama motoru sıralamasını etkiler.
- **Araştırma Sorusu:** Hangi erişilebilirlik özellikleri sıralama için daha önemlidir ve bu etki ne düzeydedir?
- **Yaklaşım:** Dış faktörleri (*anahtar kelime rekabeti, backlink vb.*) en aza indirmek için "**site:[adres]**" özel arama operatörü kullanılarak site içi hiyerarşiye odaklanılmıştır.

Materyal ve Yöntem

Veri Toplama ve Etiketleme

1. **Veri Kaynağı:** Google Trends'den popüler alışveriş anahtar kelimeleri ile 3419 web sitesi belirlendi.
2. **Veri Toplama:** Puppeteer tabanlı bir bot ile sitelerin iç bağlantıları ve HTML özellikleri toplandı.
3. **Sıralama Verisi:** Her site için Google ve Bing'de "site:" operatörü ile arama yapıp ilk 5 SERP sonucu kaydedildi.
4. **Etiketleme:** SERP'te listelenen bağlantılara pozisyonları (1-50), listelenmeyenlere ise sembolik olarak 100 değeri atandı.

Materyal ve Yöntem Özellik Çıkarımı

Toplam 11 erişilebilirlik özelliği incelendi.

- **kontrast_orani:** Metin ve arka plan rengi arasındaki kontrast (WCAG 2.0).
- **benzersiz_mi:** Bağlantının href değerinin site içinde tekil olup olmaması.
- **metin_netligi:** "read more", "click here" gibi genel ifadelerden arındırılmış, bağlantı metninin açıklığını ölçen özel bir metrik (0-1 arası).
- **Diğerleri:** gorunur_mu, gorunur_alanda_mi, gorsel_mi, metin_iceriyor, html_iceriyor, kalin_mi, alti_cizili_mi, klavye_ile_erisilebilir_mi

Bulgular Model Performansı

Problem: İki sınıflı ("listede var" / "listede yok") sınıflandırma.

En Başarılı Model: Rastgele Orman (RF).

Sonuç: Model, erişilebilirlik özelliklerine bakarak bir bağlantının sıralamada yer alıp almayacağını yüksek bir başarıyla tahmin edebiliyor.

Google için F1 Skoru: 0.82

Bing için F1 Skoru: 0.81

Bulgular Model Performansı

Gerçek

Bing	Tahmin	
	Listede Yok	Listede Var
Listede Yok	12.880	3.264
Listede Var	2.724	13.468

Gerçek

Google	Tahmin	
	Listede Yok	Listede Var
Listede Yok	11.125	2.521
Listede Var	2.336	11.239

Bulgular

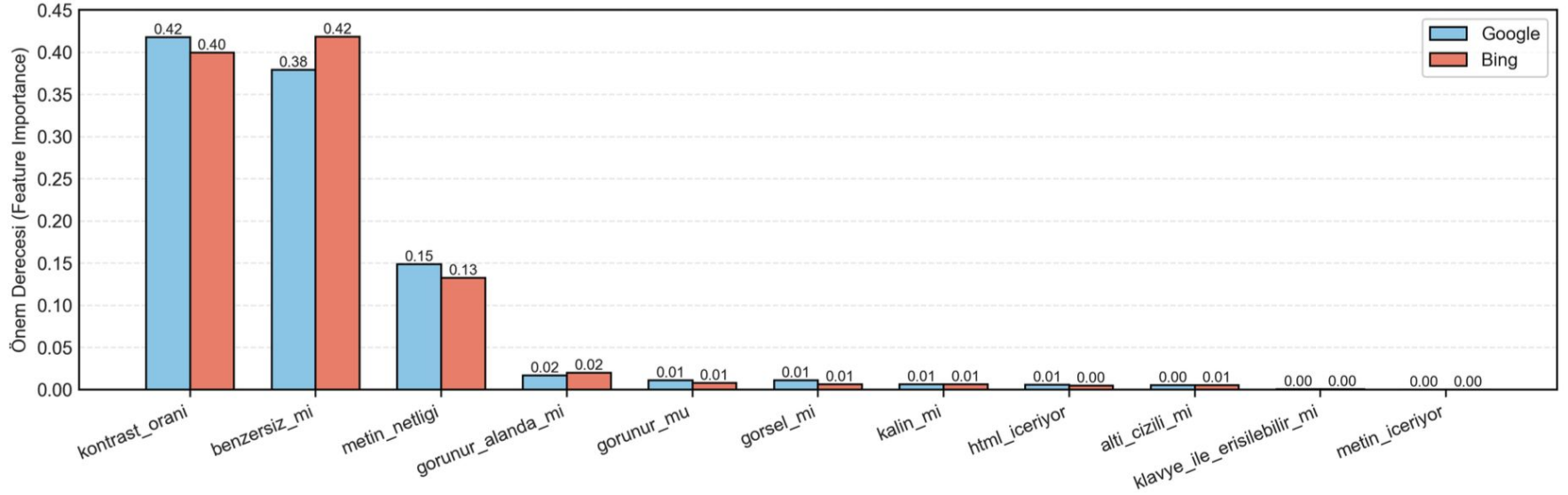
En Önemli

Özellikler

Her iki arama motoru için de en belirleyici 3 özellik:

- `kontrast_orani`
- `benzersiz_mi`
- `metin_netligi`

Bu üç özellik, diğer 8 özelliğin toplamından çok daha fazla etkiye sahip.



Tartışma

Beklenen ve Beklenmedik Sonuçlar

Beklenen Sonuç: metin_netligi özelliği beklendiği gibi sıralama ile doğru orantılı. Net ve açıklayıcı bağlantı metinleri daha iyi sıralama alıyor.

Tartışma

Beklenen ve Beklenmedik Sonuçlar

Beklenmedik Sonuç 1 (kontrast_orani):
Sıralamada yer alan bağlantıların ortalama kontrast oranı, almayanlardan daha düşük.

Yorum: Tüm bağlantıların ortalaması zaten WCAG minimum standardının (4.5, 7 - max: 21) çok üzerinde (~12). Standart karşılandıktan sonra aşırı yüksek kontrastın ek bir fayda sağlamadığı veya diğer faktörler yanında etkisiz kaldığı düşünülebilir.

Tartışma

Beklenen ve Beklenmedik Sonuçlar

Beklenmedik Sonuç 2 (benzersiz_mi):
Sıralamada yer alan bağlantıların büyük çoğunluğu "**benzersiz değil**".

Yorum: Bir alt sayfaya sitedeki birden çok yerden (*örneğin menü, alt bilgi, içerik*) link verilmesi, o sayfanın önemini belirten bir sinyal olabilir.

Tartışma

Sınıflandırma

Başarısı

Üzerine

İki Sınıflı Veri Seti: Erişilebilirlik özelliklerinin etkisi net bir şekilde görülüyor ($F1 \sim 0.82$).

Sayfa Bazlı / Pozisyon Bazlı Veri Setleri: Sınıf sayısı arttıkça (yani doğrudan pozisyonu tahmin etmeye çalışınca) modellerin başarısı dramatik şekilde düşüyor ($F1 \sim 0.09$).

Yorum: Bu beklenen bir sonuçtur. Erişilebilirlik, sıralamayı etkileyen onlarca faktörden sadece biridir. Tek başına kesin sıralamayı belirleyemez, ancak bir sayfanın "**dikkate alınıp alınmaması**" seviyesinde önemli bir rol oynar.

Sonuç

Ana Bulgular: İç bağlantı erişilebilirliği, Google ve Bing sıralamalarını etkileyen anlamlı bir faktördür.

Önemli Metrikler: Özellikle kontrast_orani, benzersiz_mi ve metin_netligi öne çıkmaktadır.

İlişkinin Doğası: Bu ilişki her zaman doğrusal değildir; bazen dolaylı ve karmaşık olabilir.

Katkılar: Alanında bu konuyu ampirik olarak inceleyen ilk çalışmalardan biri olması ve gelecekteki araştırmalar için açık kaynaklı bir veri seti ve yazılım sunması.

Gelecek Çalışmalar

Veri Setini Genişletme: Yandex gibi farklı arama motorlarını dahil etme.

Özellik Setini Zenginleştirme:
Hedef sayfanın performansı (*hız, vb.*).
Bağlantı metni ile sayfa içeriği arasındaki anlamsal uygunluk.

Yeni Teknikler: Bağlantının sayfadaki görsel konumu gibi özellikleri analiz etmek için görüntü işleme ve derin öğrenme modelleri kullanma

Teşekkürler!

Göker CEBECİ

gokercebeci@std.yildiz.edu.tr;
Yıldız Teknik Üniversitesi,
Elektrik-Elektronik Fakültesi,
Bilgisayar Mühendisliği Bölümü,
İstanbul / Türkiye.

Banu DİRİ

diri@yildiz.edu.tr;
Yıldız Teknik Üniversitesi,
Elektrik-Elektronik Fakültesi,
Bilgisayar Mühendisliği Bölümü,
İstanbul / Türkiye.

Veri Seti

<https://goker.dev/datasets/hyperlinks>

Kaynak Kodu

<https://goker.dev/codes/hyperlinks>