

DİJİTALLEŞTİRMEDEN DERLEME (KORPUS) GEÇİŞİ

İŞLENEBİLİR METİNLERDEN YAPAY ZEKÂ DESTEKLİ ARAŞTIRMALARA



Türk Dil Kurumu

Kütüphaneci

ONUR SELÇİN



ÖZET

Dijitalleştirme, belgeleri korur ve erişilebilir kılar; ancak tek başına araştırılabilir veri üretmez. Dijital görüntülerin **OCR/HTR ile metne dönüştürülmesi**, **doğrulanması** ve **metadata ile zenginleştirilmesi**, bu verilerin bir **korpus** hâline gelmesini sağlar.

Korpuslar; **yapay zekâ, anlamsal arama, veri analizi ve yeni araştırmalar** için işlenebilir bilgi altyapısı oluşturur.

Bu çalışma, **dijitalleştirmeden korpusa geçişi** ve bu dönüşüme örnek olarak **TDK Derlem Tabanlı Türkçe Sözlük Projesi**'ni ele almaktadır.



GİRİŞ

Dijitalleştirme anlayışı, belgelerin yalnızca sayısal kopyalarını üretmekten, **araştırılabilir ve makine tarafından işlenebilir veri üretmeye** doğru değişmektedir.

Bir belgenin taranması sürecin sonu değil, yeni bir bilgi üretim sürecinin başlangıcıdır.

Dijital görüntü → İşlenebilir metin → Yapılandırılmış veri → Korpus → Yeni bilgi

NEDEN KORPUS?

DİJİTALLEŞTİRME	DERLEM (KORPUS)
Dijital görüntü üretir	» İşlenebilir metin üretir
Dijital görüntü üretir	» Bilgi üretimini destekler
Görüntülenebilir	» Sorgulanabilir
Statik koleksiyon	» Dinamik araştırma altyapısı
İnsan odaklı kullanım	» İnsan + yapay zekâ kullanımı
Arşiv niteliğindedir	» Sürekli gelişen veri ekosistemidir

DİJİTALLEŞTİRMEDEN KORPUSA GİDEN YOL



İYİ BİR KORPUS NASIL ÇALIŞIR?



ÖRNEK ÇALIŞMA

TDK DERLEM TABANLI TÜRKÇE SÖZLÜK PROJESİ



“Proje; sözlük hazırlamanın yanı sıra Türkçe doğal dil işleme çalışmalarına ve Türkçe odaklı yapay zekâ uygulamalarına veri altyapısı sağlayabilecek olması bakımından önem taşımaktadır.”



TARTIŞMA

Dijitalleştirme, kültürel mirasın korunması için vazgeçilmezdir; ancak dijital görüntülerin tek başına araştırma verisi hâline gelmesi mümkün değildir. Metinlerin okunabilir, doğrulanabilir, yapılandırılabilir ve ilişkilendirilebilir olması gerekir.

"Bu nedenle geleceğin dijital koleksiyonlarında hedef yalnızca "daha fazla belgeyi dijitalleştirmek" değil, "daha nitelikli ve işlenebilir veri üretmek" olmalıdır."



SONUÇ

Dijitalleştirme kültürel mirası korurken; Korpus ise bu mirası araştırılabilir bilgiye dönüştürür.

İyi tasarlanmış bir korpus; metinleri depolamakla kalmaz, onları yapay zekâ, anlamsal arama, veri analizi ve dijital beşeri bilimler için kullanılabilir hâle getirir.

DİJİTALLEŞTİRMENİN GERÇEK DEĞERİ: GÖRÜNTÜDEN VERİYE → VERİDEN KORPUSA → KORPUSTAN BİLGİYE



KAYNAKLAR

Kalyoncuoğlu, Y. (2025). "TDK, 'derlem tabanlı' Türkçe sözlüğünü yapay zeka desteğiyle hazırlıyor." *Anadolu Ajansı*, 15 Ağustos 2025.

Philips, J., & Tabrizi, N. (2020). Historical document processing: A survey of techniques, tools, and trends. *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2020)*, 1, 341-349.



TEŞEKKÜR

Türk Dil Kurumu Kütüphanesi'nin çalışanlarına göstermiş oldukları destek ve çabadan dolayı teşekkür ederim.