

Teknik kılavuzlar için RAG parçalama stratejisi

Bir kılavuzu rastgele uzunlukta parçalara bölmek, doğru yanıtın yanlış bir bağlamda aranmasına neden olur. Başlık, tablo ve sürüm bilgisini koruyun.

Kısaca

- Sabit karakter sayısı ile bölmek tabloyu ortadan keser, bu yüzden kılavuz başlık sınırlarında bölünmeli ve her parçaya başlık yolu eklenmelidir.
- Hata kodu tablosundan başlık satırı olmadan alınan bir satır yanıltıcıdır, bu yüzden sütun başlıklarını her satıra ekleyin ya da tabloyu bir bütün olarak tutun.
- Her parçaya ürün ailesi, model, belge sürümü ve sayfa bilgisi eklenmelidir, çünkü benzer hata kodları üst veri olmadan ayrılamaz.

01 / Yapıyı koruyun

Kılavuzlar başlıklar, alt başlıklar, tablolar ve uyarı kutularından oluşur. Sabit karakter sayısı ile bölmek, bir tabloyu ortadan keser ya da bir uyarıyı ilgili adımdan ayırır. Başlık sınırlarına göre bölün ve her parçaya bağlı olduğu başlık yolunu ekleyin.

02 / Tabloları ve kodları ayrı ele alın

Hata kodu tabloları, satır satır anlam taşır. Bir satırı başlık satırı olmadan saklamak yanıltıcıdır. Her tablo satırına sütun başlıklarını ekleyin ya da tabloyu bir bütün olarak, parça sınırı içinde tutun. Çok büyük tablolarda satır grupları oluşturup her gruba başlığı tekrarlayın.

03 / Üst veri ekleyin

Her parçaya ürün ailesi, model, belge sürümü, dil ve sayfa bilgisini ekleyin. Bu alanlar hem aramada süzme hem de yanıtta kaynak gösterme için gereklidir. Farklı bir ürün ailesinden gelen benzer bir hata kodu, üst veri olmadan güvenilir biçimde ayrılamaz.

04 / Parçalamayı ölçün

Parça boyutunu ve örtüşmeyi sezgiyle değil, sabit bir soru kümesiyle karşılaştırın: beklenen belgenin ilk beş sonuçta çıkma oranını her ayarda ölçün. Bu yaklaşım önerilen bir mühendislik yöntemidir; en iyi ayar belgenin yapısına ve soru türüne bağlıdır.

Uygulayın

Bir kılavuz sayfasını iki farklı parça boyutuyla bölün ve aynı üç soru için hangi parçanın ilk sırada geldiğini karşılaştırın.

Kaynaklar ve çevrimiçi rehber

Güncel çevrimiçi rehberi açın

Introduction to Information Retrieval: Okapi BM25

Önerilen pilot adımları saha doğruluk iddiası içermez. Sentetik örnekler gerçek tesis ölçümü değildir.