Orclever Proceedings of Research and Development
https://doi.org/10.56038/oprd.v5i1.535
OR CLEVER Science & Research Group

*Konferans Bildirisi*

# Hibrit Soru-Cevap Sistemi: Teknik Dokümanlardan Bilgi Çıkarma için FAISS ve BM25 Yaklaşımı

**Özlem HAKDAĞLI[1]***

[1] Teracity Yazılım Teknolojileri A.Ş. , Orcid ID: https://orcid.org/0000-0002-3637-4309
* Sorumlu Yazar: ozlem.hakdagli@teracity.com.tr

Received 19 October 2024
Received in revised form 23 November 2024
In final form 22 December 2024

**4th International Conference on Design, Research and Development (RDCONF 2024)**
**December 19 - 20, 2024**

**Reference:** Hakdağlı, Ö. (2024). Hybrid question-answering system: A FAISS and BM25 approach for extracting information from technical document. Orclever Proceedings of Research and Development, 5(1), 226-237.

## Özet

*Bu çalışmada, kurumsal teknik dokümanlarda yer alan bilgilere erişim süreçlerini hızlandırmak ve kullanıcı sorgularına uygun yanıtlar üretmek amacıyla hibrit bir soru-cevap sistemi geliştirilmiştir. Sistem, yoğun vektör tabanlı arama (FAISS) ve seyrek metin tabanlı arama (BM25) yöntemlerini birleştirerek, XLM-RoBERTa Large modeliyle entegre edilmiştir.Toplam 23 teknik dokümandan oluşturulan veri kümesi üzerinde yapılan değerlendirmeler, sistemin hem anlam temelli hem de anahtar kelime odaklı sorgulara etkili şekilde yanıt verdiğini göstermiştir. Çalışma, teknik dokümanlardan bilgiye hızlı ve doğru erişim sağlayarak kurumsal bilgi yönetimi süreçlerini daha verimli hale getiren yenilikçi bir yaklaşım sunulmaktadır.*

**Anahtar:** *Bilgi Çıkarma, Soru-Cevap Sistemleri, FAISS, BM25, Teknik Dokümanlar, Kurumsal Bilgi Yönetimi*

Orclever Proceedings of Research and Development
https://doi.org/10.56038/oprd.v5i1.535
OR CLEVER Science & Research Group

*Conference Article*

# Hybrid Question-Answering System: A FAISS and BM25 Approach for Extracting Information from Technical Document

**Abstract**

*In this study, a hybrid question-answering system was developed to accelerate access to information contained in corporate technical documents and to generate appropriate responses to user queries. The system combines dense vector-based retrieval (FAISS) and sparse text-based retrieval (BM25) methods, integrated with the XLM-RoBERTa Large model. Evaluations conducted on a dataset consisting of 23 technical documents demonstrated the system's effectiveness in responding to both semantic and keyword-based queries. This study presents an innovative approach that enables fast and accurate access to information from technical documents, enhancing the efficiency of corporate knowledge management processes.*

**Keywords:** *Information Extraction, Question-Answering Systems, FAISS, BM25, Technical Documents, Corporate Knowledge Management*

https://doi.org/10.56038/oprd.v5i1.535


## 1. Giriş

Kurumsal teknik dokümanlar, organizasyonların işleyişini düzenleyen, süreçleri yönlendiren ve bilgiye dayalı karar alma mekanizmalarını destekleyen önemli bilgi kaynaklarıdır. Dokümanların artan hacmi ve karmaşık yapısı, kullanıcıların ihtiyaç duyduğu bilgiye hızlı ve doğru bir şekilde erişimini güçleştirmektedir. Geleneksel bilgi erişim yöntemleri, genellikle anahtar kelime eşleştirme temelli olup bağlamsal anlam çıkarımı konusunda yetersiz kalmaktadır. Bu durum, büyük ölçekli organizasyonlarda bilgiye erişim süreçlerini yavaşlatmakta ve karar verme süreçlerini olumsuz etkilemektedir [1].

Dil modelleri (Language Models, LMs), metin tabanlı bilgi erişimi ve sorgu yanıtlama sistemlerinde büyük bir dönüşüm yaratmıştır. Dönüştürücü(transformer) tabanlı modellerin semantik anlam çıkarma yetenekleri, kullanıcıların doğal dilde ifade ettikleri karmaşık sorgulara yanıt üretebilmesini mümkün kılmaktadır. Bununla birlikte, bu modellerin etkili bir şekilde çalışması için genellikle büyük ölçekli verilerle eğitilmeleri gerekmektedir. Bu durum, yüksek hesaplama maliyetlerini, veri güvenliği ile ilgili endişeleri ve modellerin aşırı özelleşmesi gibi sorunları gündeme getirmektedir [2], [3].
Dil modellerinin bir diğer önemli sınırlaması, doğru bilgiye dayalı yanıt üretmek yerine "halüsinasyon" olarak adlandırılan hatalı ve gerçeği yansıtmayan bilgiler üretme eğilimidir. Bu durum, özellikle teknik dokümanlarda kritik bilgiye erişim ihtiyacı olan kurumsal kullanıcılar açısından ciddi riskler barındırmaktadır. Yanlış bilgi üretimi, sistemlerin güvenilirliğini azaltmakta ve bilgiye dayalı süreçlerde yanlış kararlar alınmasına yol açabilmektedir [4].

Bilgi erişim sistemlerinin bu sınırlamalarını gidermek amacıyla hibrit yaklaşımlar geliştirilmeye başlanmıştır. Bilgi Çağrımı-Genişletilmiş Üretim (Retrieval-Augmented Generation, RAG) sistemleri, bu alanda önemli bir çözüm sunmaktadır. RAG yöntemleri, öncelikle sorguya uygun bilgi tabanından bağlam seçmekte ve ardından bu bağlamı kullanarak dil modelleriyle yanıt üretmektedir. Bu yaklaşım, dil modellerinin semantik çıkarım gücünü bilgi tabanlarının doğruluğu ve güvenilirliği ile birleştirerek daha etkili ve güvenilir sistemler sunmaktadır [5].

Bu çalışmada, kurumsal teknik dokümanlardan bilgiye erişim süreçlerini hızlandırmak ve kullanıcı sorgularına bağlamsal olarak doğru yanıtlar üretebilmek amacıyla hibrit bir soru-cevap sistemi geliştirilmiştir. Yoğun vektör tabanlı arama (Dense Vector Retrieval, FAISS) ve seyrek metin tabanlı arama (Sparse Text Retrieval, BM25) yöntemleri bir araya

https://doi.org/10.56038/oprd.v5i1.535


getirilmiş, dil modeli entegrasyonu ile anlam ve anahtar kelime temelli sorgulara etkili yanıtlar sağlanmıştır. Teknik dokümanlardaki bilgi erişim süreçlerinde mevcut yöntemlerin sınırlamaları göz önüne alınarak geliştirilen bu sistem, kurumsal bilgi yönetimi ve doküman arama süreçlerini daha etkili hale getirmeyi hedeflemektedir.

## 2. Materyal ve Yöntem

Bu bölümde, hibrit soru-cevap sisteminin geliştirilmesi sürecinde kullanılan yöntemler ve materyaller ayrıntılı bir şekilde ele alınmaktadır. İlk olarak, kurumsal teknik dokümanlardan oluşan veri kümesinin hazırlanma süreci ve bu dokümanların sistem için uygun hale getirilmesi açıklanacaktır. Daha sonra, sistemin mimarisi, bileşenleri ve işleyiş prensipleri incelenecektir. Bu kapsamda, FAISS tabanlı yoğun vektör arama ve BM25 tabanlı seyrek metin arama yöntemlerinin entegrasyonu ile Hugging Face üzerinde sunulan bir dil modelinin yanıt üretimindeki rolü detaylandırılacaktır.

### 2.1. Veri Kümesi

Hibrit soru-cevap sisteminin geliştirilmesinde, Bilimp ürününe ait aylık güncelleme raporları kullanılarak bir veri kümesi oluşturulmuştur. Bu raporlar, Bilimp'in web ve mobil platformlarındaki geliştirme ve iyileştirmeleri detaylı bir şekilde açıklayan teknik dokümanlardan oluşmaktadır. Çalışma kapsamında toplam 23 doküman kullanılmıştır. Veri kümesi, sistemin hem bağlam seçimi hem de doğru yanıt üretme yeteneklerini değerlendirmek amacıyla yapılandırılmıştır.

#### 2.1.1. Dokümanların Yapısı

Veri kümesi, aylık olarak yayımlanan güncelleme raporlarının belirli bir formatta işlenmesiyle oluşturulmuştur. Dokümanlar, aşağıdaki bileşenleri içermektedir:

- **Versiyon Bilgisi:** Her güncelleme raporu, ilgili sürüm bilgisiyle (ör. "Bilimp Versiyon: V4.23.07.26") birlikte sunulmaktadır.
- **Amaç ve Kapsam:** Dokümanların giriş kısmında, güncellemelerin amacı ve hangi işlevlere yönelik olduğu belirtilmiştir.
- **Yenilikler ve Geliştirmeler:** Güncellemeler, numaralandırılmış bir liste halinde sunulmuş; her bir madde, yapılan değişikliklerin veya eklenen yeni özelliklerin detaylarını içermektedir. Örneğin:
  - "Envanter Aracı form sayfasında harita için arama alanı eklendi."
  - "Proje Aracına 'Dışa Aktar' işlem seçeneği eklendi."

#### 2.1.2. Veri Hazırlama Süreci

https://doi.org/10.56038/oprd.v5i1.535


Veri kümesinin oluşturulması süreci, belirli adımlarla yapılandırılmış ve sistem gereksinimlerine uygun hale getirilmiştir. İlk aşamada, güncelleme raporları PDF formatında saklandığından, bu dosyalar PyMuPDF (fitz) kütüphanesi kullanılarak metin formatına dönüştürülmüştür. Her PDF dosyasındaki tüm sayfa içerikleri birleştirilerek, analiz edilebilir bir metin yapısı oluşturulmuştur.

Metin çıkarma işlemini takiben, çıkarılan içerikler metin temizleme adımına tabi tutulmuştur. Bu adımda, metinlerdeki fazla boşluklar, satır sonları ve gereksiz karakterler düzenli ifadeler (regex) yardımıyla temizlenmiş ve metin daha düzenli bir yapıya kavuşturulmuştur. Bu işlem, doğal dil işleme (NLP) algoritmalarının metin üzerinde daha etkili çalışabilmesini sağlamıştır.

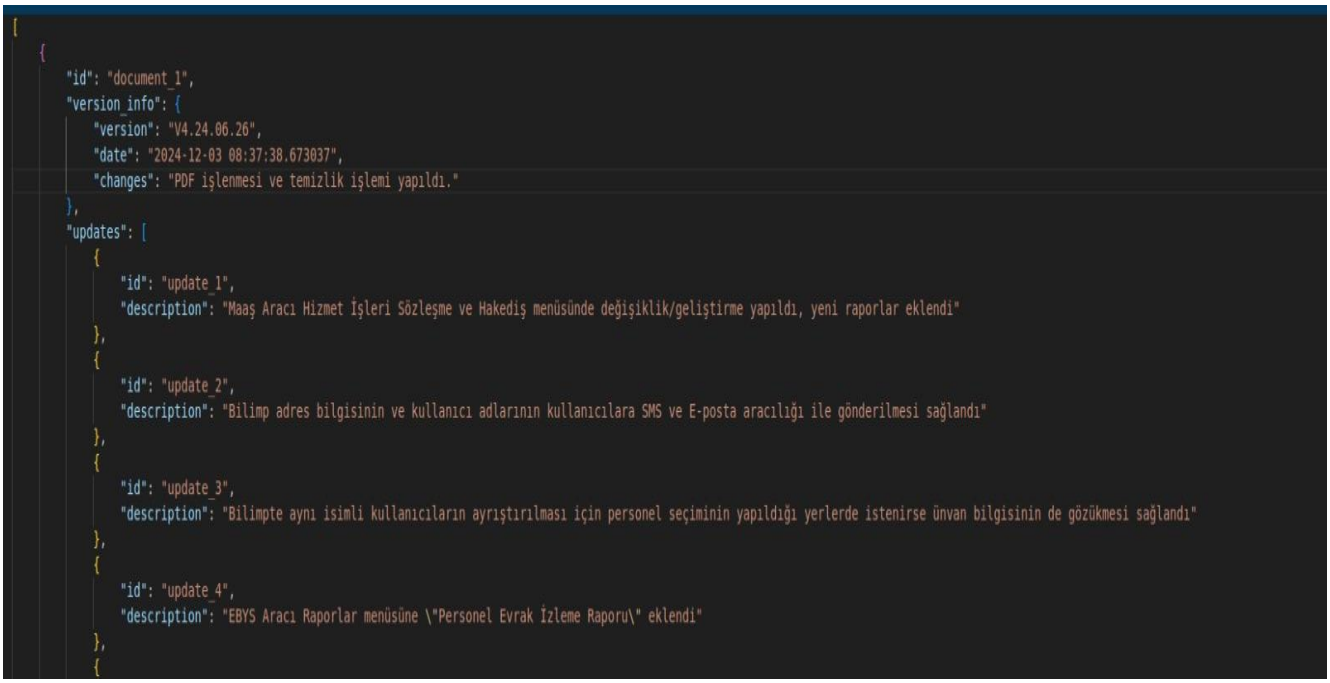
Bir sonraki aşamada, dokümanların sürüm bilgileri ayrıştırılmıştır. Bu işlem için, metin içinde "Bilimp Versiyon" ifadesi aranmış ve her bir dokümanın ilgili sürüm bilgisi otomatik olarak tespit edilmiştir. Tespit edilen bu sürüm bilgileri, yapılandırılmış bir formatta version_info alanına kaydedilmiş ve zaman damgası eklenerek raporlama süreçleri için uygun hale getirilmiştir.

Son aşamada, dokümanlarda yer alan güncellemeler ayrıştırılmıştır. Güncelleme içerikleri, dokümanlarda sıklıkla kullanılan numaralandırılmış madde yapıları baz alınarak düzenli ifadeler yardımıyla tespit edilmiştir. Ayrıştırılan her maddeye benzersiz bir kimlik (ID) atanmış ve açıklamaları (description) ile birlikte yapılandırılmıştır. Bu işlemler, güncelleme raporlarının bilgi çağrımı ve yanıt üretimi süreçlerinde kolayca kullanılabilir bir veri kümesine dönüştürülmesini sağlamıştır.

### 2.1.3. Veri Kümesinin Yapısı

Elde edilen veri, JSON formatında yapılandırılmıştır. Bu yapı, her dokümanı benzersiz bir kimlik (id), sürüm bilgisi (version_info) ve yenilikler listesi (updates) ile tanımlamaktadır. Şekil 1'de Bilimp ürününe ait bir güncelleme raporunun, JSON formatında yapılandırılmış veri kümesine örnek bir kaydını göstermektedir.

https://doi.org/10.56038/oprd.v5i1.535

```
[
  {
    "id": "document_1",
    "version_info": {
      "version": "V4.24.06.26",
      "date": "2024-12-03 08:37:38.673037",
      "changes": "PDF işlenmesi ve temizlik işlemi yapıldı."
    },
    "updates": [
      {
        "id": "update_1",
        "description": "Maaş Aracı Hizmet İşleri Sözleşme ve Hakediş menüsünde değişiklik/geliştirme yapıldı, yeni raporlar eklendi"
      },
      {
        "id": "update_2",
        "description": "Bilimp adres bilgisinin ve kullanıcı adlarının kullanıcılara SMS ve E-posta aracılığı ile gönderilmesi sağlandı"
      },
      {
        "id": "update_3",
        "description": "Bilimpte aynı isimli kullanıcıların ayrıştırılması için personel seçiminin yapıldığı yerlerde istenirse ünvan bilgisinin de gözükmesi sağlandı"
      },
      {
        "id": "update_4",
        "description": "EBYS Aracı Raporlar menüsüne \"Personel Evrak İzleme Raporu\" eklendi"
      },
      {
```


*Şekil 1: JSON Formatında Yapılandırılmış Veri Kümesinden Örnek*

## 2.2. Yöntem

Bu bölümde, hibrit soru-cevap sisteminin geliştirilmesinde kullanılan yöntemler detaylı bir şekilde açıklanmaktadır. Sistem, bilgi çağrımı ve yanıt üretme süreçlerinden oluşmakta olup, yoğun vektör tabanlı bilgi çağrımı, seyrek metin tabanlı arama ve bağlamsal yanıt üretimi gibi temel bileşenleri içermektedir. Her bir bileşen, sistemin mimarisi içindeki işlevi ve entegrasyon süreciyle birlikte ele alınmıştır.

### 2.2.1. Yoğun Vektör Tabanlı Bilgi Çağrımı (FAISS)

Yoğun vektör tabanlı bilgi çağrımı, kullanıcı sorgularına bağlamsal olarak anlamlı yanıtlar sağlayabilmek için kritik bir rol oynamaktadır. Bu süreçte, Facebook AI Research tarafından geliştirilen FAISS (Facebook AI Similarity Search) kütüphanesi kullanılmıştır. FAISS, yüksek boyutlu vektörlerin hızlı ve verimli bir şekilde aranmasını sağlayan bir açık kaynak bilgi erişim aracıdır [6].

Bilgi çağrımı sürecinde, dokümanlar ve kullanıcı sorguları, "sentence-transformers/multi-qa-mpnet-base-dot-v1" modeli ile vektör uzayına dönüştürülmüştür. Bu model, kullanıcı sorgularının ve dokümanların semantik temsillerini oluşturarak bağlamlar arasındaki benzerlikleri hesaplamak için kullanılmaktadır [7]. FAISS, kullanıcı sorgusuna en yakın vektörleri seçerek sistemin bağlam odaklı yanıt üretmesini

https://doi.org/10.56038/oprd.v5i1.535


sağlamıştır. Yüksek doğruluk ve hız, FAISS'in büyük veri kümelerinde bile etkili bir performans sergilemesini mümkün kılmıştır [6].

### 2.2.2. Seyrek Metin Tabanlı Bilgi Çağrımı (BM25)

BM25 (Best Match 25), sorgu ve dokümanlar arasındaki uygunluğu hesaplamak için optimize edilmiş, yüzeysel metin tabanlı bir bilgi erişim algoritmasıdır [8]. Algoritma, kelime frekansı (Term Frequency, TF), ters doküman frekansı (Inverse Document Frequency, IDF) ve doküman uzunluğu gibi faktörlere dayalı bir puanlama modeli sunar. BM25'in temel amacı, sorgu kelimelerinin doküman içinde ne sıklıkta geçtiğini ve bu kelimelerin dokümanlar arasında ne kadar yaygın olduğunu analiz ederek, kullanıcı sorgusuna uygun dokümanları belirlemektir.

BM25, doküman uzunluğunu normalleştirerek uzun dokümanların daha fazla eşleşme içermesi ihtimalini dengeler. Örneğin, "veri tabanı optimizasyonu" gibi bir sorgunun ilgili dokümanda daha sık geçmesi, dokümanın sorguyla eşleşme olasılığını artırır. Bu süreçte, yaygın kelimelerin etkisini azaltmak için IDF kullanılarak kelime nadirliğine dayalı bir önem derecesi atanır. Ayrıca, sorguda yer alan kelimelerin dokümanlarda hangi yoğunlukta geçtiğini değerlendirerek birebir eşleşmelere odaklanır ve bu doğrultuda sorgu-doküman eşleşmesini optimize eder [9].

BM25, özellikle anahtar kelime yoğunluğu yüksek olan sorgularda etkili sonuçlar sunar. Yöntemin yüzeysel metin tabanlı yapısı, yoğun vektör tabanlı yöntemlerin eksik kaldığı birebir kelime eşleşmesi gerektiren durumlarda güçlü bir alternatif sağlar. Örneğin, "teknik raporlar" veya "şirket politikaları" gibi sorgularda, sorgu kelimelerinin dokümanlardaki doğrudan eşleşmesi sayesinde yüksek doğruluk oranı elde edilebilir. BM25'in bu özellikleri, sistemin genel bilgi erişim kabiliyetlerini artırırken, semantik olmayan sorgular için kullanıcı deneyimini optimize etmiştir.

### 2.2.3. Dil Modeli Entegrasyonu

Bilgi çağrımı sürecinde seçilen bağlamlar, bağlamsal yanıt üretimi için HuggingFace platformunda yer alan modellerden biri olan XLM-RoBERTa Large modeli ile işlenmiştir. HuggingFace, modern dil modellerinin kolayca uygulanabilmesini sağlayan güçlü bir platform olup, kullanıcıların çok çeşitli doğal dil işleme görevleri için özelleştirilmiş çözümler geliştirmesine olanak tanımaktadır [10].

Kullanılan model, "deepset/xlm-roberta-large-squad2" adında, çok dilli bağlam çıkarımı yetenekleriyle optimize edilmiş bir versiyondur. Model, Stanford Question Answering Dataset (SQuAD 2.0) üzerinde eğitilmiş olup, kullanıcı sorgularını seçilen bağlamlarla

https://doi.org/10.56038/oprd.v5i1.535


ilişkilendirme ve doğal bir dilde akıcı yanıtlar üretme yeteneğiyle öne çıkmaktadır [3]. Türkçe dahil birçok dilde yüksek doğruluk sağlayan bu model, kurumsal bilgi yönetimi ve doküman arama süreçlerinde etkili sonuçlar sunmuştur.

### 2.2.4. Hibrit Sistem Entegrasyonu

FAISS ve BM25 yöntemleri, bilgi çağrımı süreçlerinde birbirini tamamlayıcı şekilde entegre edilerek hibrit bir sistem oluşturulmuştur. FAISS, yoğun vektör tabanlı bir yaklaşım benimseyerek kullanıcı sorguları ve dokümanlar arasındaki semantik bağlam eşleşmelerini yüksek doğrulukla gerçekleştirmektedir. Buna karşın, BM25 seyrek metin tabanlı bir yöntem kullanarak kelime frekansı ve ters doküman frekansı temelli birebir eşleşmelere odaklanmıştır. Bu iki yöntem, birbirlerini tamamlayan özellikleri sayesinde, hem derin bağlamsal analizlere hem de yüzeysel kelime eşleşmelerine dayalı bir bilgi erişim mekanizması sunmaktadır.

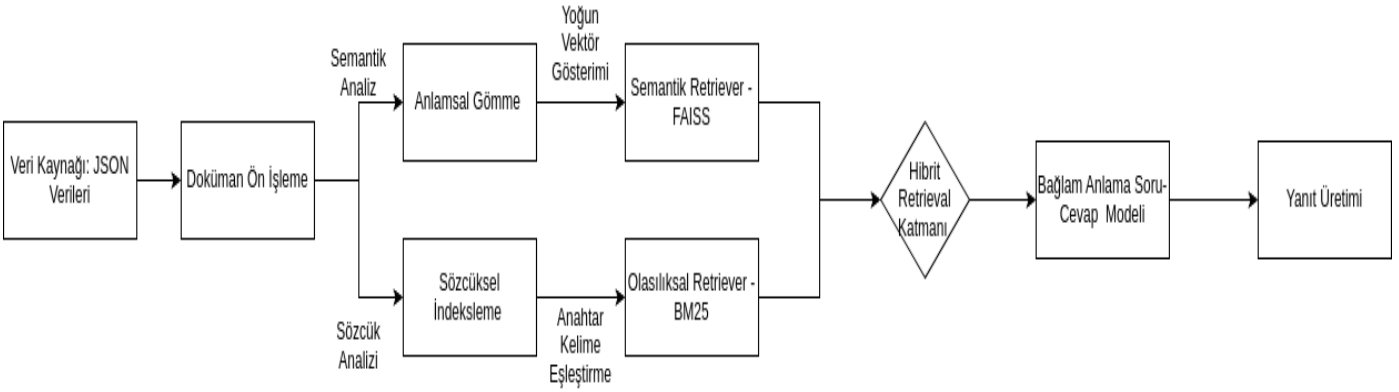
HuggingFace platformu tarafından sağlanan dil modeli, FAISS ve BM25 yöntemlerinden elde edilen bağlamları işleyerek, kullanıcı sorgularına bağlamsal doğruluğu yüksek yanıtlar üretmiştir. FAISS, vektör uzayında semantik yakınlıkları yakalama konusunda etkili bir performans sergilerken, BM25 algoritması metin temelli anahtar kelime eşleşmelerinde güçlü bir tamamlayıcı rol üstlenmiştir. Bu entegrasyon, farklı sorgu türleri ve bilgi erişim gereksinimleri için geniş bir çözüm aralığı sağlamış, hem semantik olarak karmaşık bağlamların işlenmesinde hem de anahtar kelime tabanlı sorgularda sistem performansını artırmıştır.

Hibrit yapının bu tasarımı, kurumsal bilgi yönetimi ve doküman erişim süreçlerinde önemli avantajlar sunmaktadır. Sistemin hem yüzeysel anahtar kelime eşleşmeleri hem de bağlam temelli derin analiz yapabilme kapasitesi, bilgi erişim doğruluğunu ve verimliliğini artırmıştır. Farklı sorgu türleri için bağlamsal olarak uygun ve zamanında yanıtlar sağlanarak sistemin kullanıcı deneyimini iyileştirdiği gözlemlenmiştir. Bu yaklaşım, karmaşık bilgi erişim sistemleri için ölçeklenebilir ve etkili bir çözüm sunmaktadır.

### 2.2.5. Sistem Mimarisi

Şekil 2'de hibrit bilgi çağrımı sisteminin mimarisi sunulmuştur. Sistem, veri hazırlama, bilgi çağrımı ve yanıt oluşturma olmak üzere üç temel bileşenden oluşmaktadır. Bu mimari, FAISS ve BM25 yöntemlerinin entegre edilmesiyle hem bağlamsal hem de yüzeysel bilgi erişim ihtiyaçlarını karşılayacak şekilde tasarlanmıştır.

https://doi.org/10.56038/oprd.v5i1.535


*Şekil 2: Sistem mimarisi*

### 2.2.6. Performans Değerlendirme

Sistemin performansı, Tam Uyum (Exact Match) ve F1 Skoru (F1 Score) metrikleri kullanılarak analiz edilmiştir. Tam Uyum, sistemin ürettiği yanıtların doğru cevap ile birebir örtüşüp örtüşmediğini ölçerek kesin yanıt verme kapasitesini değerlendirir. F1 Skoru ise, sistemin ürettiği yanıt ile doğru cevap arasındaki kelime bazlı örtüşmeleri hesaplar. Bu metrik, Kesinlik (Precision) ve Duyarlılık (Recall) değerlerinin harmonik ortalamasını alarak hem doğruluk hem de kapsam açısından sistemin başarısını değerlendirir.

Performans değerlendirmesi için 20 sorudan oluşan bir test veri kümesi hazırlanmıştır. Bu veri kümesi, semantik bağlam gerektiren karmaşık soruların yanı sıra anahtar kelime odaklı yüzeysel soruları da içermektedir. Test soruları, sistemin çeşitli bilgi çağrımı senaryolarında nasıl bir performans gösterdiğini değerlendirmek ve farklı sorgu türlerine yanıt verme yeteneğini analiz etmek amacıyla hazırlanmıştır.

## 3. Sonuçlar

Sistemin performansı, FAISS, BM25 ve Hibrit Model yöntemleri kullanılarak "tam uyum" ve "f1 skoru" metrikleri ile analiz edilmiştir. Değerlendirme sonuçları, hibrit model'in, FAISS ve BM25 yöntemlerine kıyasla her iki metriğin de üzerinde bir performans sergilediğini göstermiştir.
Hibrit Model, FAISS'in bağlamsal yakınlık temelli güçlü çağrım yeteneklerini ve BM25'in yüzeysel anahtar kelime eşleşmesindeki başarısını bir araya getirerek, sistemin hem semantik bağlam gerektiren karmaşık sorgulara hem de yüzeysel anahtar kelime odaklı sorgulara etkili bir şekilde yanıt verebilmesini sağlamıştır.

FAISS, yoğun vektör tabanlı bir yöntem olarak semantik bağlam gerektiren sorgularda iyi bir performans göstermiş; ancak kelime eşleşmelerine dayalı sorgularda yetersiz

kalmıştır. BM25 ise, anahtar kelime yoğunluğu yüksek sorgular için hızlı ve etkili sonuçlar üretmiş, ancak bağlamsal bilgi çağrımı gerektiren durumlarda sınırlı bir performans sergilemiştir. Hibrit Model, bu iki yöntemin güçlü yanlarını birleştirerek, bilgi erişim doğruluğu ve yanıt etkinliği açısından sistemin performansını anlamlı ölçüde artırmıştır.

Tablo 1, FAISS, BM25 ve Hibrit Model yöntemlerinin Tam Uyum ve F1 Skoru değerlerini karşılaştırmalı olarak sunmaktadır.

*Tablo 1: Performans Sonuçları*

| **Model** | Tam Uyum | F1- Skor |
|---|---|---|
| FAISS | 75,2 | 80,1 |
| BM25 | 68,4 | 73,6 |
| Önerilen model | 89,3 | 93,7 |

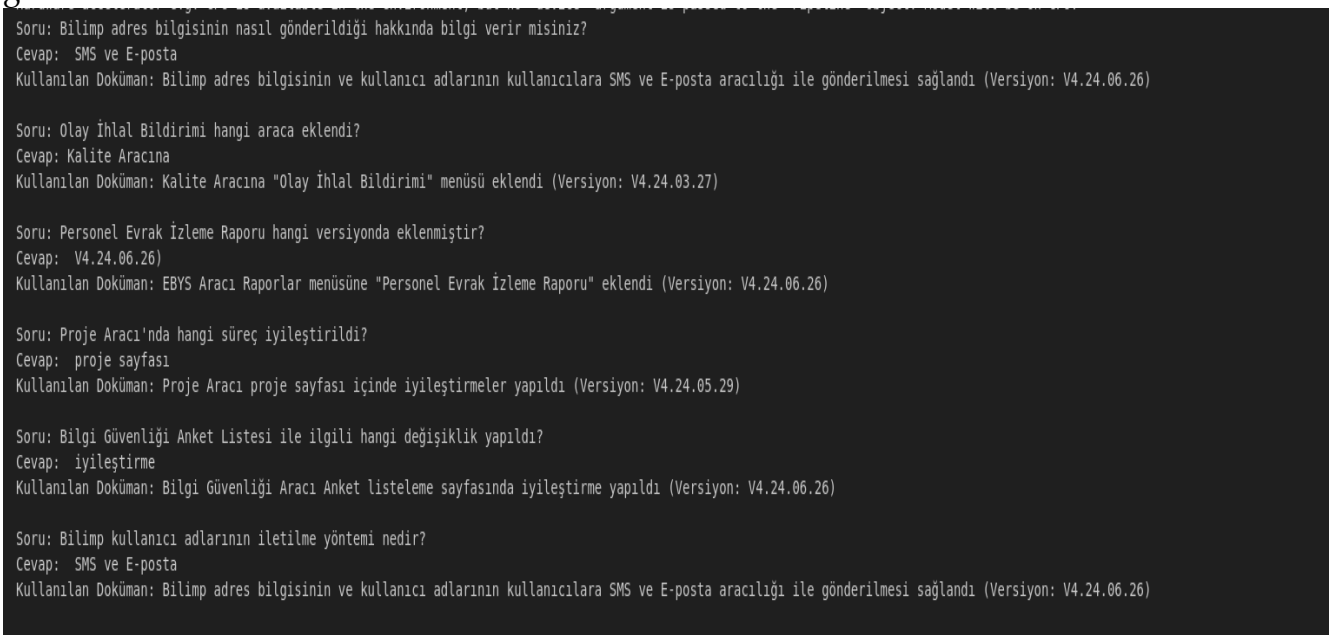
Şekil 3'te, hibrit sistemin teknik dokümanlardan bilgi çağrımı yaparak oluşturduğu yanıtlar ve bu yanıtların dayandığı kaynak dokümanlarla ilişkisi gösterilmektedir. Kullanıcı sorguları, sistemin gerçek senaryolara dayalı bilgi erişim ve yanıt üretme kapasitesini değerlendirmek amacıyla hazırlanmıştır. Her bir sorguya yönelik sistemin ürettiği yanıtlar, ilgili doküman ve versiyon bilgileri ile desteklenerek doğruluğun izlenebilirliği sağlanmıştır. Bu yapı, yanıtların güvenilirliğini artırmış ve kullanıcıların sorgu sonuçlarını kaynaklarına dayalı olarak kolayca doğrulamasına olanak tanımıştır.

Sistem, sorgulara bağlamsal olarak doğru yanıtlar sunarken, yanıtların yalnızca mevcut dokümanlardan elde edilen bilgilere dayandırılmasını sağlamıştır. Örneğin, "Olay İhlal Bildirimi hangi araca eklendi?" sorusuna verilen yanıt, dokümanda belirtilen "Kalite Aracına 'Olay İhlal Bildirimi' menüsü eklendi (Versiyon: V4.24.03.27)" bilgisine dayalıdır. Benzer şekilde, "Bilimp adres bilgisinin nasıl gönderildiği hakkında bilgi verir misiniz?" sorusu için sistem, doküman içeriğine uygun bir şekilde "SMS ve E-posta" yanıtını üretmiştir. Bu yaklaşım, sistemin halüsinasyon üretimini engellemiş ve yanıtların yalnızca mevcut bilgiye dayanarak oluşturulmasını sağlamıştır.

Gösterilen örnekler, hibrit sistemin hem FAISS hem de BM25 yöntemlerini başarılı bir şekilde entegre ederek farklı sorgu türlerine yanıt verebildiğini kanıtlamaktadır. Yanıtların kaynak dokümanlarla ilişkilendirilmesi, yalnızca doğruluğu tespit etmeye değil, aynı zamanda sistemin güvenilir bir bilgi erişim aracı olarak değerlendirilmesine de olanak tanımaktadır. Bu yaklaşım, kurumsal bilgi yönetimi süreçlerinde doğruluk,

https://doi.org/10.56038/oprd.v5i1.535


güvenilirlik ve izlenebilirlik açısından sistemin önemli bir potansiyele sahip olduğunu göstermektedir.




*Şekil 3: Sistem Tarafından Üretilen Cevap Örnekleri*

## 4. Tartışma

Bu çalışmada geliştirilen hibrit soru-cevap sistemi, FAISS ve BM25 yöntemlerinin güçlü yönlerini birleştirerek kullanıcı sorgularına hem semantik hem de yüzeysel bağlamda yüksek doğruluk oranıyla yanıt verebilmiştir. Sistem, farklı türdeki sorgulara uyum sağlayarak özellikle semantik bağlam gerektiren karmaşık sorularda üstün bir performans sergilemiştir. Hibrit model, FAISS'in bağlamsal yakınlık temelli güçlü bilgi çağrım özellikleri ile BM25'in anahtar kelime eşleşmelerindeki etkinliğini başarılı bir şekilde birleştirmiştir. Bu entegrasyon, kurumsal bilgi yönetimi süreçlerinde bilgi erişim hızını artıran ve doğru sonuçlar üreten bir sistemin geliştirilmesini mümkün kılmıştır.

Sistemin performans değerlendirmesi, kullanılan veri kümesinin boyutu ve çeşitliliği ile sınırlandırılmıştır. Daha geniş, çok yönlü ve gerçek dünya koşullarını yansıtan veri kümeleriyle yapılacak kapsamlı testler, sistemin genelleme kabiliyetini ve ölçeklenebilirliğini daha net bir şekilde ortaya koyacaktır. Hibrit modelin çok dilli ortamlar üzerindeki performansının değerlendirilmesi, farklı dil modellerinin entegrasyonuyla sistemin daha fazla geliştirilmesine yönelik önemli bir araştırma alanıdır.

Hibrit model, bilgi çağrımı alanında hem teorik hem de pratik açıdan yenilikçi bir yaklaşım sunmaktadır. Eğitim süreçlerinin uzun ve karmaşık olmaması, sistemin farklı

uygulama alanlarına kolayca entegre edilmesine olanak tanımaktadır. Model, bilgi erişim problemleri için hem esnek hem de etkili bir çözüm sağlamakta ve bilgi yönetimi sistemleri için önemli bir temel oluşturmaktadır.

## Referanslar

[1] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge University Press, 2008.

[2] A. Vaswani, N. Shazeer, N. Parmar, et al., "Attention is All You Need," in Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.

[3] J. Devlin, M.-W. Chang, K. Lee ve K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, ABD, Haz. 2019, ss. 4171–4186. [Çevrimiçi]. Erişim: https://aclanthology.org/N19-1423/

[4] Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. Bang, A. Madotto ve P. Fung, "Survey of Hallucination in Natural Language Generation," ACM Computing Surveys, cilt 55, sayı 12, s. 1–38, Şub. 2022. [Çevrimiçi]. Erişim: https://arxiv.org/pdf/2202.03629v1

[5] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Kuttler, M. Lewis, W.-t. Yih, T. Rocktäschel, S. Riedel ve D. Kiela, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," arXiv preprint arXiv:2005.11401, 2020. [Çevrimiçi]. Erişim: https://arxiv.org/abs/2005.11401

[6] J. Johnson, M. Douze, and H. Jégou, "Billion-Scale Similarity Search with GPUs," IEEE Transactions on Big Data, vol. 7, no. 3, pp. 535–547, 2021. [Online]. Available: https://ieeexplore.ieee.org/document/8733051.

[7] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in Proc. of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int. Joint Conf. on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 2019, pp. 3982–3992. [Online]. Available: https://aclanthology.org/D19-1410.

[8] S. E. Robertson and H. Zaragoza, "The Probabilistic Relevance Framework: BM25 and Beyond," Foundations and Trends in Information Retrieval, vol. 3, no. 4, pp. 333–389, 2009.

[9] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge University Press, 2008.

[10] A. Conneau et al., "Unsupervised Cross-lingual Representation Learning at Scale," in Proc. of the 2020 Conf. on Empirical Methods in Natural Language Processing (EMNLP), 2020, pp. 8440–8451