



FAMARA

(Fair Market Research Assistant)

Yapay Zeka Destekli Fuar & Pazar Araştırma Asistanı

softITO 4. DÖNEM 1. GRUP YAPAY ZEKA YAZILIMCILIĞI

BİTİRME PROJESİ RAPORU

Hazırlayanlar

Erdoğan Ersin Keskin

Çağan Erdem Üstündağ

Begüm Hilal Telli

Gamze Duran

Sezin Öztekin

Mehmet Alper Dinçeler

Temmuz 2026, İstanbul

İÇİNDEKİLER

1. GİRİŞ

- 1.1. Projenin Amacı
- 1.2. Proje Kısıtları

2. MATERYAL VE YÖNTEM

- 2.1. RAG (Retrieval-Augmented Generation) Kavramı
- 2.2. Python ve FastAPI
- 2.3. PostgreSQL ve pgvector
- 2.4. Gömme (Embedding) Modeli
- 2.5. Hibrit Arama ve RRF (Reciprocal Rank Fusion)
- 2.6. Büyük Dil Modelleri ve Fallback (Yedekleme) Zinciri
- 2.7. Web Kazıma (Web Scraping)
- 2.8. Tavily Web Arama API'si
- 2.9. React ve Vite ile Kullanıcı Arayüzü
- 2.10. LLM Güvenliği: Prompt Injection ve Hız Sınırlama

3. TASARIM VE GERÇEKLEŞTİRME

- 3.1. Sistem Mimarisi
- 3.2. Veritabanı Tasarımı
- 3.3. Veri Toplama Pipeline'ı
- 3.4. Hibrit Arama Modülü
- 3.5. RAG Süreci ve LLM Fallback Zinciri
- 3.6. Güvenlik Katmanı
- 3.7. API Uç Noktaları
- 3.8. Kullanıcı Arayüzü
- 3.9. Testler
- 3.10. Kurulum ve Çalıştırma

4. SONUÇ

5. KAYNAKÇA

1. GİRİŞ

Bu bölümde projenin amacı ve projede karşılaşılan kısıtlara yer verilmiştir.

1.1. Projenin Amacı

Bu projenin amacı, İstanbul Ticaret Odası (İTO) tarafından desteklenen ve yurt içi/yurt dışında düzenlenen ticari fuarlara ilişkin bilgilere erişimi kolaylaştıran, yapay zeka destekli bir fuar ve pazar araştırma asistanı (FAMARA) geliştirmektir. Fuar bilgileri günümüzde farklı web sitelerine dağılmış, yapılandırılmamış ve güncelliği takip edilmesi zor bir haldedir. Özellikle ihracat hedefleyen KOBİ'ler için doğru fuarı bulmak ciddi bir araştırma yükü oluşturmaktadır.

FAMARA, fuar verilerini otomatik olarak toplayan (web scraping), bu verileri anlamsal arama yapılabilir biçimde saklayan (PostgreSQL + pgvector) ve kullanıcı sorularını RAG (Retrieval-Augmented Generation) yaklaşımıyla, güncel ve doğrulanabilir bağlama dayanarak yanıtlayan uçtan uca bir web uygulamasıdır. Sistem; anahtar kelime araması, vektör tabanlı anlamsal arama ve web araması sonuçlarını hibrit olarak birleştirir; yanıt üretiminde ise birden fazla büyük dil modelini (LLM) yedekli bir zincir halinde kullanarak servis sürekliliği sağlar.

Bu çalışma ile hem güncel yapay zeka mühendisliği pratiklerinin (RAG, hibrit arama, çok modelli yedekleme, prompt injection savunması) gerçek bir iş problemine uygulanması hem de İTO'nun 800 binden fazla üyesiyle birlikte Türkiye'deki fuara katılım sağlamak isteyen bütün işletmeler için değer üretebilecek, kullanıma hazır bir prototip ortaya konması hedeflenmiştir.

1.2. Proje Kısıtları

Kısıt 1: Veri kaynağı olarak ito.org.tr fuar takvimi ile Ticaret Bakanlığı ve TOBB'un yayınladığı fuar listeleri kullanılmıştır. Bu kaynakların sayfa yapısı değiştiğinde kazıma (scraper) modülünün güncellenmesi gerekebilir, kapsam, bu kaynaklarda yayımlanan fuar bilgileri ile sınırlıdır.

Kısıt 2: Yapay zeka yanıtları harici LLM servislerine (Gemini, LLMapi, Claude, ChatGPT) bağımlıdır. API anahtarlarının bulunmadığı veya tüm servislerin erişilemez olduğu durumlarda sistem, veritabanı bağlamından ekstraktif (özetleyici) bir yedek yanıt üretir, bu yanıtın dil kalitesi LLM yanıtlarına göre daha sınırlıdır.

Kısıt 3: Anlamsal arama, 384 boyutlu çok dilli bir gömme (embedding) modeli (paraphrase-multilingual-MiniLM-L12-v2) ile gerçekleştirilmiştir. Model boyutu ile veritabanı şemasındaki vektör boyutu birbirine bağlıdır, farklı bir model kullanılmak istendiğinde şema da güncellenmelidir.

Kısıt 4: Proje test/geliştirme aşamasında olduğundan bazı uç noktalarda (örn. POST /api/scraper) kimlik doğrulama bulunmamaktadır. Uygulama gerçek kullanıcılara açılmadan önce bu uç noktalara koruma eklenmesi planlanmaktadır.

2. MATERYAL VE YÖNTEM

Bu bölümde, projede kullanılan temel teknolojiler ve yöntemler kavramsal düzeyde açıklanmıştır.

2.1. RAG (Retrieval-Augmented Generation) Kavramı

RAG, büyük dil modellerinin yanıt üretirken yalnızca eğitim verisine değil, sorguya göre dinamik olarak getirilen (retrieval) güncel ve alana özgü belgelere dayanmasını sağlayan bir mimari yaklaşımdır. Kullanıcının sorusu önce bir arama katmanından geçirilir, bulunan ilgili içerikler modele "bağlam" olarak verilir ve modelden yalnızca bu bağlama dayanarak yanıt üretmesi istenir. Bu yaklaşım, modelin halüsinasyon (uydurma bilgi) üretme riskini azaltır ve yanıtların kaynağa dayandırılabilmesini sağlar. FAMARA'da fuar soruları, veritabanından ve web aramasından getirilen fuar kayıtları bağlam olarak kullanılarak yanıtlanır.

2.2. Python ve FastAPI

Projenin sunucu tarafı Python ile geliştirilmiştir. FastAPI, tip güvenliği (Pydantic şemaları), otomatik API dokümantasyonu (OpenAPI/Swagger) ve asenkron istek işleme desteği sunan modern bir web çatısıdır. Uygulama, uvicorn ASGI sunucusu üzerinde çalışır. İstek/yanıt gövdeleri Pydantic modelleri ile doğrulanır, böylece hatalı veriler iş mantığına ulaşmadan reddedilir.

2.3. PostgreSQL ve pgvector

Veri katmanında PostgreSQL kullanılmıştır. pgvector eklentisi, veritabanına vektör veri tipi ve benzerlik arama yeteneği kazandırır, böylece hem klasik ilişkisel sorgular hem de anlamsal (vektör) arama tek bir veritabanında yapılabilir. Vektör araması için HNSW (Hierarchical Navigable Small World) indeksi kosinüs benzerliği ile yapılandırılmıştır, bu indeks, büyük veri kümelerinde yaklaşık en-yakın-komşu aramasını milisaniyeler düzeyine indirir.

2.4. Gömme (Embedding) Modeli

Metinlerin anlamsal karşılaştırılabilmesi için sentence-transformers kütüphanesindeki paraphrase-multilingual-MiniLM-L12-v2 modeli kullanılmıştır. Model, bir metni 384 boyutlu bir vektöre dönüştürür, anlamca yakın metinler vektör uzayında birbirine yakın konumlanır. Çok dilli olması sayesinde Türkçe fuar adları ve açıklamaları ile Türkçe kullanıcı sorguları aynı uzayda karşılaştırılabilir. Model, uygulama açılışında arka planda önceden yüklenir (warm-up); böylece ilk aramada model yükleme gecikmesi yaşanmaz.

2.5. Hibrit Arama ve RRF (Reciprocal Rank Fusion)

Tek başına anahtar kelime araması eş anlamlıları kaçırır; tek başına anlamsal arama ise kesin eşleşmelerde (örn. fuar adı) zayıf kalabilir. Bu nedenle FAMARA, üç arama kanalını birlikte kullanır: (1) veritabanında anahtar kelime araması, (2) pgvector ile anlamsal arama, (3) Tavily API üzerinden canlı web araması. Farklı kanallardan gelen sıralı sonuç listeleri, RRF (Reciprocal Rank Fusion) yöntemiyle birleştirilir. RRF'te her sonuç, listelerdeki sırasına göre $1/(k+rank)$ puanı alır ($k=60$) ve toplam puana göre yeniden sıralanır. Veritabanı yeterli sayıda

benzersiz sonuç döndürdüğünde (eşik: 5), gecikmeyi ve API maliyetini azaltmak için web araması atlanır.

2.6. Büyük Dil Modelleri ve Fallback (Yedekleme) Zinciri

Tek bir LLM sağlayıcısına bağımlılık, kota aşımı, servis kesintisi veya model hatası durumunda uygulamanın tamamen yanıtız kalmasına yol açar. FAMARA'da yanıt üretimi dört asistanlı bir yedekleme zinciri ile gerçekleştirilir: **Gemini** → **LLMApi** → **Claude** → **ChatGPT**. Zincirdeki her asistan sırayla denir; biri başarısız olur veya boş yanıt dönerse bir sonrakine geçilir. Tüm asistanlar başarısız olursa, sistem kullanıcıya bilgilendirici bir hata mesajı ve bağlamdan üretilen ekstraktif bir özet sunar. Tüm asistanlar ortak bir arayüzü (soyut temel sınıf) uygular, böylece zincire yeni model eklemek yalnızca yeni bir sınıf yazmayı gerektirir.

2.7. Web Kazıma (Web Scraping)

Fuar verileri, requests + BeautifulSoup (statik HTML ayrıştırma) ve gerektiğinde Playwright (JavaScript ile yüklenen içerikler için başsız tarayıcı) kullanılarak toplanır. Kazıma süreci, fuar listelerinin çekilmesi, detay sayfalarının işlenmesi, fuar afişlerinin indirilmesi ve kayıtların gömme vektörlerinin üretilmesi adımlarından oluşan bir pipeline (boru hattı) olarak tasarlanmıştır.

2.8. Tavily Web Arama API'si

Tavily, LLM uygulamaları için tasarlanmış bir web arama API'sidir. Veritabanında yeterli sonuç bulunamadığında, kullanıcının sorgusu Tavily'ye iletilir ve güncel web sonuçları hibrit arama havuzuna dahil edilir. Bu sayede sistem, veritabanına henüz girmemiş güncel fuar duyurularına da erişebilir.

2.9. React ve Vite ile Kullanıcı Arayüzü

Ön yüz, React, TypeScript ve Vite ile geliştirilmiş tek sayfa uygulamasıdır (SPA). Tailwind CSS ile şekillendirilen arayüz, arama çubuğu, fuar kartları, karşılaştırma tablosu, istatistik kartları ve çok turlu sohbet bileşenlerinden oluşur. Ayrıca projeyi tanıtan ayrı bir tanıtım (landing) sayfası bulunmaktadır. API adresi, derleme zamanı değişkeni (VITE_API_BASE_URL) ile yapılandırılır, üretim dağıtımında frontend ve backend aynı alan adı arkasında sunulabilir.

2.10. LLM Güvenliği: Prompt Injection ve Hız Sınırlama

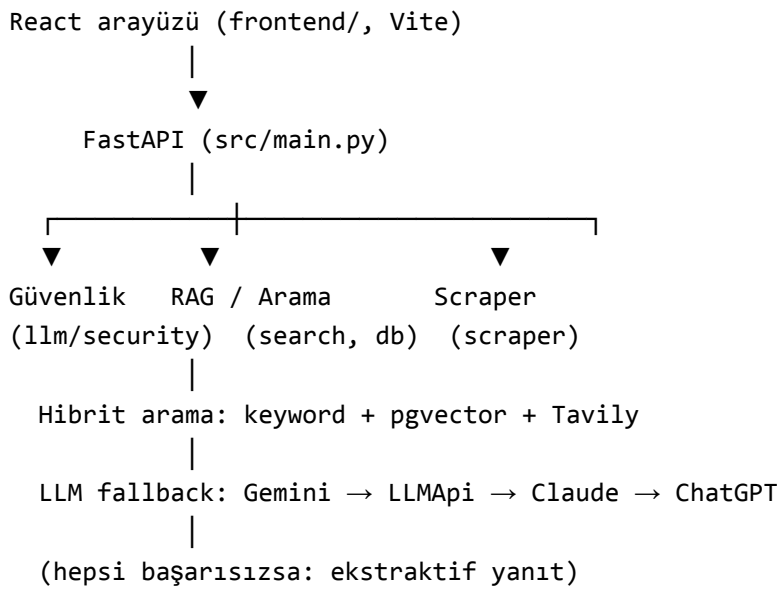
LLM tabanlı uygulamalarda kullanıcı girdisi, modele verilen talimatları manipüle etmeye çalışabilir (prompt injection). Ayrıca kötü niyetli istemciler, servisleri aşırı istekle meşgul edebilir. Bu risklere karşı, kullanıcı mesajlarını süzen bir güvenlik katmanı (kalıp tabanlı engelleme, HTML/karakter temizleme, uzunluk sınırı) ve oturum başına hız sınırlayıcı (rate limiter) kavramları uygulanmıştır. Genel yapı Bölüm 3.6'da anlatılmıştır.

3. TASARIM VE GERÇEKLEŞTİRME

Bu bölümde sistemin mimarisi, modüllerin gerçekleştirimi ve kurulum/çalıştırma adımları anlatılmıştır.

3.1. Sistem Mimarisi

FAMARA, katmanlı bir mimariyle tasarlanmıştır. React arayüzünden gelen istekler FastAPI uygulamasına ulaşır, burada önce güvenlik katmanından geçer, ardından hibrit arama katmanı ile ilgili fuar kayıtları toplanır ve RAG bağlamı oluşturularak LLM yedekleme zincirine iletilir. Veri toplama tarafında ise scraper modülü, kaynak sitelerden fuar verilerini düzenli aralıklarla toplayarak veritabanını güncel tutar.



Şekil 3.1.1. FAMARA sistem mimarisi

3.2. Veritabanı Tasarımı

Veritabanı şeması tek bir SQL göç (migration) dosyasında tutulur ve uygulama açılışında otomatik olarak uygulanır. Başlıca tablolar şunlardır:

- **fairs:** Fuar kayıtları (ad, tarih aralığı, sektör dizisi, konum, ülke, İTO desteği, afiş yolu vb.) ile 384 boyutlu embedding sütunu. Ad, ülke, yıl, tarih ve sektör alanları için B-tree/GIN indeksleri, embedding için kosinüs benzerlikli HNSW indeksi tanımlıdır.
- **search_logs:** Yapılan aramaların sorgu metni, sonuç sayısı, kaynak ve çalışma süresi (ms) ile günlüklenmesi.
- **scrape_logs:** Scraper çalışmalarının kaynak, kayıt sayıları (eklenen/güncellenen/hatalı) ve durum bilgisiyle izlenmesi.
- **chat_messages:** Çok turlu sohbet için oturum (session_id) bazlı kullanıcı/asistan mesaj geçmişi.

3.3. Veri Toplama Pipeline'ı

Veri toplama, src/scrapper paketinde uçtan uca bir pipeline olarak gerçekleştirilmiştir. Pipeline,ito.org.tr fuar takviminin kazınması, Ticaret Bakanlığı ve TOBB fuar listelerinin işlenmesi, fuar afişlerinin indirilmesi ve yeni/güncellenen kayıtlar için embedding üretimi adımlarından oluşur. Pipeline komut satırından (python -m src.scrapper.run_pipeline) veya API üzerinden (POST /api/scrape) tetiklenebilir, ayrıca uygulama içindeki zamanlayıcı her gece 02:00'de pipeline'ı otomatik çalıştırır. Aynı anda yalnızca tek bir kazıma işinin çalışmasına izin verilir ve iş durumu GET /api/scrape/{id} ile takip edilebilir.

3.4. Hibrit Arama Modülü

src/search paketi, anahtar kelime araması (SQL ILIKE + OR/sıralama mantığı), pgvector anlamsal araması ve Tavily web aramasını yürütür. Kanallardan gelen sonuçlar RRF (k=60) ile birleştirilir ve tarih önceliğine göre sıralanır. Veritabanından 5'ten fazla benzersiz sonuç dönerse web araması atlanarak gecikme ve API maliyeti azaltılır. Arama katmanı ayrıca basit yazım düzeltmesi yapar ve her aramayı search_logs tablosuna kayıt eder.

3.5. RAG Süreci ve LLM Fallback Zinciri

Sohbet isteği (POST /api/chat) geldiğinde, hibrit aramadan dönen fuar kayıtları, karakter limiti gözetilerek bir RAG bağlamına dönüştürülür ve oturum geçmişiyle (chat_messages) birlikte sistem prompt içerğine eklenir. Yanıt üretimi dört asistanlı zincirle yapılır: Gemini (varsayılan model gemini-3.5-flash) → LLMapi (OpenAI uyumlu erişim) → Claude (claude-sonnet-4-6) → ChatGPT (gpt-4o). Her asistan ortak soyut arayüzü uygular, başarısızlık veya boş yanıt durumunda zincir bir sonraki asistana geçer. Tüm zincir başarısız olursa kullanıcıya İTO web sitesine yönlendiren bilgilendirici bir mesaj ve bağlamdan üretilen ekstraktif özet gösterilir. Bu tasarım sayesinde tek bir sağlayıcı kesintisi hizmeti durdurmaz.

3.6. Güvenlik Katmanı

LLM katmanına ulaşan her mesaj src/llm/security.py modülünden geçer. Katman şu korumaları uygular:

- **Hız sınırlama:** Oturum başına 60 saniyelik pencerede en fazla 20 istek kabul edilir.
- **Uzunluk sınırı:** Tek mesaj en fazla 1000 karakterdir.
- **Kalıp tabanlı engelleme:** Prompt injection ("ignore previous instructions" vb.), sistem promptu sızdırma, XSS, SQL injection ve kod enjeksiyonu kalıpları düzenli ifadelerle engellenir. Türkçe karakter normalleştirilmesi ile kalıpların farklı yazımları da yakalanır.
- **Sanitizasyon:** HTML etiketleri temizlenir, fazla boşluklar sadeleştirilir; istemciden gelip LLM bağlamına eklenen tüm metinler (örn. fuar açıklamaları) da aynı süzgeçten geçirilir.
- **Konu dışı tespiti:** Fuar dışı konular (hava durumu, borsa vb.) tespit edilir ve kullanıcı kibarca kapsam hakkında bilgilendirilir.

Blok-liste yaklaşımının tek başına yeterli olmadığı bilinmekte, asıl koruma, bağlamın yapısal yalıtımı ve sistem promptu sertleştirmesi ile birlikte sağlanmaktadır.

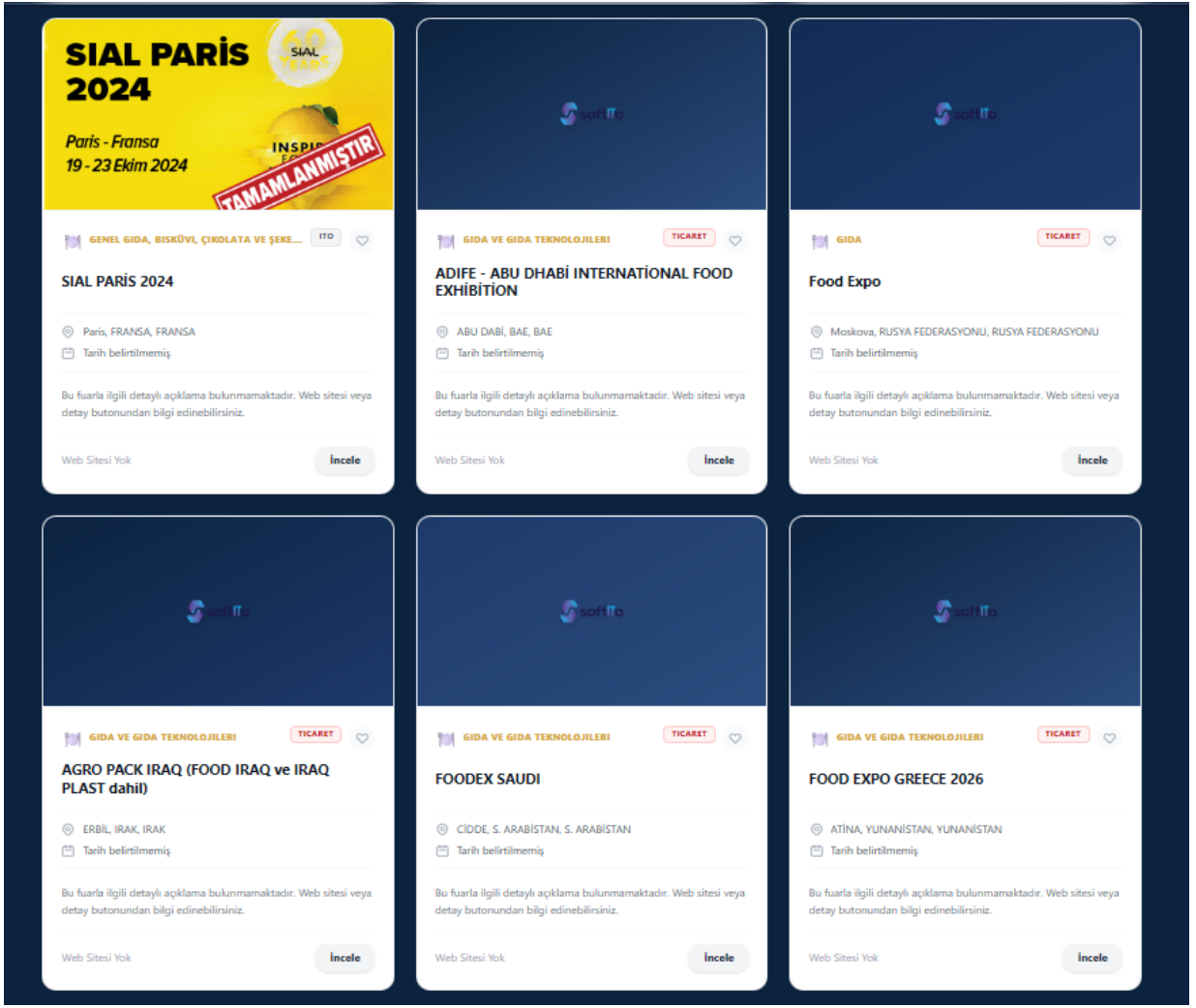
3.7. API Uç Noktaları

FastAPI uygulamasının sunduğu başlıca uç noktalar aşağıdaki tabloda özetlenmiştir:

Metot	Yol	Açıklama
GET	/api/health	Sistem ve veritabanı durumu
POST	/api/search	Hibrit arama (keyword + semantik + web); LLM'siz, anlık sonuç
POST	/api/chat	RAG destekli asistan yanıtı (çok turlu, session_id ile)
POST	/api/fairs	Veritabanı anahtar kelime araması (ülke / İTO desteği filtreli)
POST	/api/scrape	Scraper pipeline'ını tetikler (aynı anda tek iş)
GET	/api/scrape/{id}	Pipeline iş durumu sorgulama
POST	/api/posters/sync	Eksik fuar afişlerini indirir ve yollarını günceller

3.8. Kullanıcı Arayüzü

React arayüzü iki ana bölümden oluşur. Uygulama ekranında arama çubuğu, fuar kartları (afiş, tarih, konum, sektör ve İTO desteği rozetiyle), fuar detay penceresi, karşılaştırma tablosu, istatistik kartları ve yapay zeka sohbet alanı yer alır. Sohbet bileşeni, yazıyor efekti, sohbet geçmişi listesi ve devam sorusu önerileri ile çok turlu diyalogu destekler. Ayrıca projeyi İTO paydaşlarına ve fuara katılım sağlamak isteyen herkese tanıtmak amacıyla hazırlanan tanıtım sayfası (/landing) özellik bölümleri, mimari özeti ve canlı demo bağlantısı içerir.



Şekil 3.8.1. Uygulama arayüzü, fuar kartları



İTO Asistan AI

Düşünce Süreci Tamamlandı (6 adım incelendi)

Göster ▼

Küresel Gıda ve Gıda Teknolojileri Sektörü Fuar Analizi

Gıda ve gıda teknolojileri sektörü, dünya genelinde en dinamik ve geniş katılım oranlarına sahip fuar gruplarının başında gelmektedir. Arama sonuçları incelendiğinde, küresel gıda pazarının farklı alt kırılımlara (gıda bileşenleri, gıda işleme teknolojileri, şekerleme, su ürünleri, helal gıda ve bitkisel bazlı gıdalar) ayrıldığı ve çok geniş bir coğrafyaya yayıldığı görülmektedir.

Önemli Noktalar ve Sektörel Eğilimler:

Coğrafi Yoğunlaşma ve Bölgesel Fırsatlar:

Ortadoğu ve Körfez Bölgesi: Suudi Arabistan, Birleşik Arap Emirlikleri (Abu Dabi) ve Irak gıda üretimi, tüketimi ve özellikle "Private Label" (özel markalı ürünler) ile şekerleme/atıştırmalık sektörlerinde çok güçlü birer ticaret merkezi haline gelmiştir.

Avrupa: Fransa, Almanya, İspanya ve İtalya; hem genel gıda (SIAL gibi dev organizasyonlar) hem gıda teknolojileri (Foodtech/Cibustec) hem de alternatif gıda akımları (Veggie World) için ana pazar konumundadır.

Asya ve Gelişmekte Olan Pazarlar: Çin, Vietnam, Endonezya ve Özbekistan; özellikle gıda bileşenleri (ingredients), tarım-hayvancılık teknolojileri ve su ürünleri alanında yüksek büyüme potansiyeli sunmaktadır.

Niş Alanların Yükselişi: Sektör artık sadece nihai gıda ürünleriyle sınırlı kalmayıp; gıda bileşenleri (Fi), evcil hayvan gıdaları ve teknolojileri (Interzoo) ile bitki bazlı/vegan beslenme gibi spesifik alanlarda kendi ihtisas fuarlarını yaratmıştır.

İTO Yönlendirmesi ve Devlet Teşvikleri:

Sistemde listelenen gıda fuarlarında bu dönem için doğrudan İTO milli katılım organizasyonu bulunmamaktadır. Ancak bu durum devlet desteği alamayacağınız anlamına gelmez:

Bireysel Katılım Desteği: Bu fuarların çok büyük bir bölümü **T.C. Ticaret Bakanlığı Bireysel Katılımı Desteklenen Sektörel Fuarlar** listesinde yer almaktadır.

Teşvik Oranları: İTO üyesi firmalarımız, bireysel katılım sağladıkları takdirde yer kirası, stant kurulumu ve nakliye harcamaları için bakanlığın belirlediği limitler dahilinde **%50 ile %70 arasında** devlet desteği geri ödemesi alabilmektedir.

Sonraki Adımlarınız:

Katılmak istediğiniz fuarı belirledikten sonra, fuar tarihinden en az 1 ay önce İTO Fuar Hizmetleri Şubesi veya üyesi olduğunuz ihracatçı birliği ile iletişime geçerek bireysel katılım desteği başvuru prosedürlerini başlatmanız, teşviklerden sorunsuz yararlanabilmeniz açısından kritik önem taşımaktadır.

Bulunan fuarların detaylı konum, tarih ve içerik bilgilerini aşağıdaki fuar kartlarından ve karşılaştırma tablosundan inceleyebilirsiniz.



38 fuar bulundu - 0% İTO destekli

 PDF Olarak İndir

Şekil 3.8.2. Yapay zeka rapor ekranı



FAMARA
Yapay Zeka Destekli Fuar & Pazar Araştırma Asistanı

Hakkında İletişim

Yapay Zeka Destekli Pazar Araştırma Asistanı

Yapay Zekâ ile Küresel Pazarları Keşfedin

İTO öncülüğünde; İTO, TOBB ve Ticaret Bakanlığı veritabanları ile anlık web araması gücüyle, hedef fuarlarınızı, katılım maliyetlerinizi ve B2B fırsatlarını saniyeler içinde analiz edin.

İli + Araştır

HIZLI ARAMA ÖNERİLERİ

Avrupa'daki Tekstil Fuarları

Ortadoğu Gıda Fuarları

İTO Destekli Otomotiv Fuarları



Fuar & Pazar Araştırma Asistanı
YAPAY ZEKÂ DİSTİLLİ

 wefito Öğrenciler Grup Projesi
İletişim & Destek: 0212 231 10 10

© 2024 İstanbul Ticaret Odası - Yapay Zeka Destekli Ön Gözetim

Hakkımızda Destek

Şekil 3.8.3. Tanıtım (landing) sayfası

Devam Eden Araştırma
Bu fuarlarla ilgili merak ettiğiniz diğer konuları keşfedin.

Gıda fuarlarında bireysel katılım devlet desteği başvuru süreci ve gerekli belgeler nelerdir?

Ortadoğu gıda pazarı ve "Private Label" (Özel Markalı Ürünler) sektörü için güncel pazar araştırma raporları mevcut mu?

Gıda işleme, paketlenme ve gıda bileşenleri alanında İTO'nun önümüzdeki dönemde düzenleyeceği bir milli katılım organizasyonu var mı?

Fuar Başvuru Rehberi

Farklı bir sorunuz varsa buraya yazın...

Gönder

Şekil 3.8.4. Follow-up Soruları ve Sohbet ekranı

3.9. Testler

tests/ dizinindeki pytest paketi tamamen mock'ludur, veritabanı bağlantıları ve LLM SDK istemcileri sahte nesnelerle değiştirildiğinden testler ağ erişimi veya gerçek API anahtarı gerektirmez. Kapsanan alanlar, anahtar kelime arama ve sıralama mantığı (repository), güvenlik katmanı (hız sınırı, zararlı içerik engeli, sanitizasyon), fallback zinciri (asistan sırası, hata durumunda geçiş, ekstraktif yedek), RAG bağlam hazırlama (karakter limiti) ve her LLM asistanının başarı/hata/boş-yanıt senaryoları. Gerçek API anahtarlarıyla elle çalıştırılan tanı araçları ise pytest kapsamı dışında tutulmuştur.

```
pytest tests/ -v
```

3.10. Kurulum ve Çalıştırma

1. Sanal ortam oluşturulur ve bağımlılıklar yüklenir: `python -m venv .venv → pip install -r requirements.txt`
2. Proje kökündeki .env dosyasında DATABASE_URL (pgvector destekli PostgreSQL) ve en az bir LLM API anahtarı tanımlanır. EMBED_MODEL/VECTOR_DIM değerlerinin şema ile uyumu kontrol edilir.
3. Backend başlatılır: `uvicorn src.main:app --reload` (şema açılışta otomatik uygulanır, embedding modeli arka planda yüklenir).
4. Frontend başlatılır: `cd frontend → npm install → npm run dev`; üretim için `npm run build`.
5. Veri toplamak için pipeline çalıştırılır: `python -m src.scrapers.run_pipeline` (veya `POST /api/scrape`).
6. Sağlık kontrolü `GET /api/health` ile doğrulanır.

4. SONUÇ

Bu çalışmada İTO, Ticaret Bakanlığı ve TOBB fuarlarına ilişkin dağınık ve yapılandırılmamış bilgiyi tek bir noktadan, doğal dille sorgulanabilir hale getiren **FAMARA** adlı yapay zeka destekli fuar ve pazar araştırma asistanı geliştirilmiştir. Sistem, web kazıma ile veri toplama, PostgreSQL + pgvector üzerinde anlamsal saklama, RRF ile birleştirilen hibrit arama ve çok modellenli LLM yedekleme zinciri üzerine kurulu uçtan uca bir RAG mimarisi olarak gerçekleştirilmiştir.

Proje kapsamında, tek sağlayıcıya bağımlılığın yedekleme zinciriyle aşılabildiği, hibrit aramanın tek kanallı aramaya göre hem kesin eşleşmelerde hem de anlamsal sorgularda daha isabetli sonuç verdiği ve prompt injection gibi LLM'e özgü saldırı yüzeylerine karşı katmanlı savunmanın uygulanabilir olduğu gösterilmiştir. Mock tabanlı test paketi sayesinde sistemin kritik bileşenleri, harici servislere ihtiyaç duymadan sürekli olarak doğrulanabilmektedir.

Sonuç olarak **FAMARA**, güncel yapay zeka mühendisliği pratiklerini gerçek bir iş problemine uygulayan, çalışır durumda ve genişletilebilir bir prototip ortaya koymuştur. Gelecek çalışmalar olarak; scrape uç noktasına kimlik doğrulama eklenmesi, veri kaynaklarının çeşitlendirilmesi, kullanıcı hesapları ile kişiselleştirilmiş fuar önerileri ve üretim ortamına dağıtım planlanmaktadır.

5. KAYNAKÇA

1. FastAPI Dokümantasyonu - <https://fastapi.tiangolo.com/>
2. PostgreSQL Dokümantasyonu - <https://www.postgresql.org/docs/>
3. pgvector: Open-source vector similarity search for Postgres - <https://github.com/pgvector/pgvector>
4. Sentence-Transformers Dokümantasyonu - <https://www.sbert.net/>
5. Cormack, G. V., Clarke, C. L., & Buettcher, S. (2009). Reciprocal Rank Fusion outperforms Condorcet and individual rank learning methods. SIGIR '09.
6. Lewis, P. vd. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS 2020.
7. Tavily Search API - <https://tavily.com/>
8. Google Gemini API - <https://ai.google.dev/>
9. Anthropic Claude API - <https://docs.anthropic.com/>
10. OpenAI API - <https://platform.openai.com/docs/>
11. React Dokümantasyonu - <https://react.dev/>
12. Vite Dokümantasyonu - <https://vitejs.dev/>
13. İstanbul Ticaret Odası Fuar Takvimi - <https://www.ito.org.tr/>
14. OWASP Top 10 for LLM Applications - <https://owasp.org/www-project-top-10-for-large-language-model-applications/>