

Yazılım Gereksinim Analizi

Giriş

Amaç

Bu projenin amacı, dijital belge türlerini tanıyan ve bunları otomatik olarak sınıflandıran bir sistem geliştirmektir. Evrak Tanıma ve Kategorilendirme Sistemi, PDF, Word gibi dijital belgeleri tarayarak, içerik analizine dayalı olarak belge türlerini tanıyacak ve uygun kategorilere ayıracaktır. Bu sistem, belgelerin manuel olarak sınıflandırılması sürecindeki insan hatalarını azaltarak işletmelere hızlı ve doğru bir çözüm sunmayı amaçlamaktadır.

Döküman Standartları

Model eğitimi sırasında kullanılan tüm parametreler, değerleri ve elde edilen performans metrikleri detaylı olarak kaydedilecektir.

Farklı sınıflandırma modelleri (Naive Bayes, SVM, LSTM gibi) ile yapılan kıyaslamalara yer verilecektir.

Kod yazımında açıkça anlaşılır fonksiyon ve değişken isimlendirmeleri kullanılacak, snake_case tercih edilecektir. Sınıf isimlendirmelerinde ise PascalCase kullanılacaktır.

Her fonksiyonun aldığı parametreler ve döndürdüğü sonuçlar ayrıntılı olarak açıklanacak, gerekli durumlarda tablolarla gösterilecektir.

Her versiyon güncellemesinde eklenen, değiştirilen veya silinen özellikler detaylı bir dökümantasyonla kaydedilecektir.

Hedef Kitle

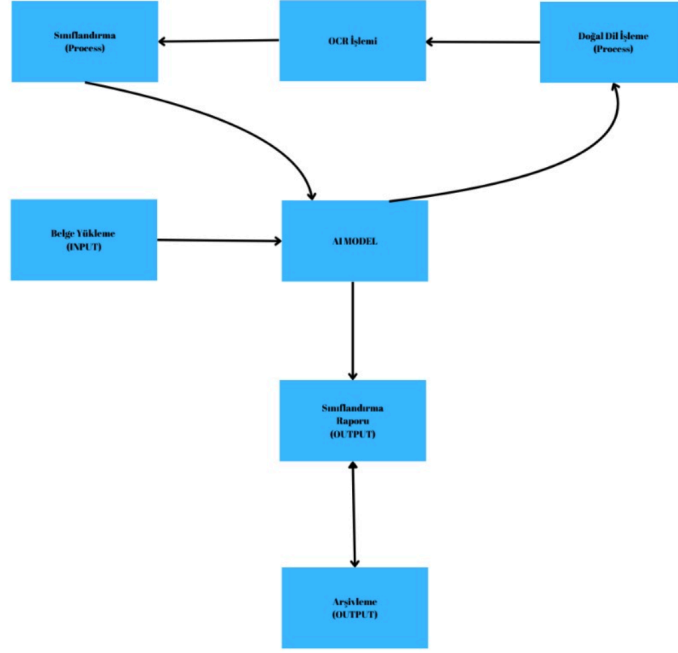
- Küçük ve Orta Ölçekli İşletmeler
- Doküman Yönetimi Departmanları
- Serbest Çalışanlar ve Kişisel Kullanıcılar

Ürün Kapsamı

Evrak Tanıma ve Kategorilendirme Sistemi, dijital belgeleri otomatik olarak tanıyıp sınıflandırarak belge yönetim sürecini kolaylaştıran bir yazılımdır. Bu sistem, fatura, sözleşme, teklif, rapor gibi farklı belge türlerini hızlı bir şekilde kategorilere ayırarak kullanıcıların belge arşivleme, erişim ve raporlama süreçlerini hızlandırmayı hedefler. Geleneksel manuel sınıflandırma süreçlerinin ötesine geçerek, kullanıcılara doğruluk oranı yüksek bir sınıflandırma çözümü sunar.

Genel Tanımlama

Yazılım Bakış Açısı – IPO (Input-Process-OutPut) Diyagramı

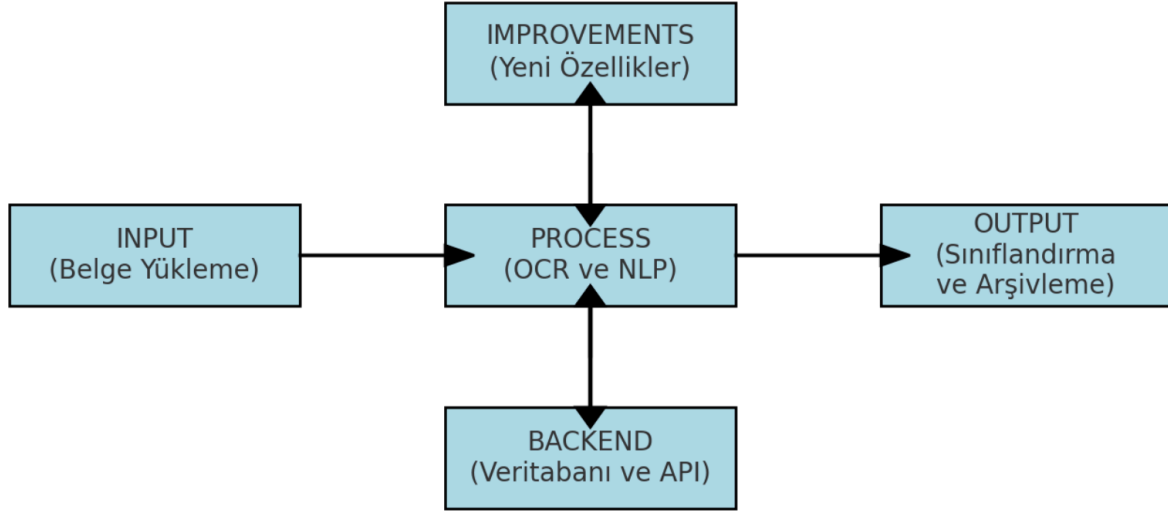


Bu proje kapsamında geliştirilen yazılım, dijital belgelerin otomatik olarak tanınması ve sınıflandırılması üzerine odaklanan, özgün bir sistem olarak tasarlanmıştır. Sistem, mevcut bir belge yönetim çözümünün yerini almamakta; daha ziyade bağımsız bir ürün olarak işletmelerin belge düzenleme ve arşivleme ihtiyaçlarına özel olarak geliştirilmiştir.

Sistem, kullanıcıların belge yüklemelerini ve türlerini tanımlamasını kolaylaştırarak belge yönetim sürecini hızlandırmayı hedeflemektedir. Ürünün temel işleyişi, belgeleri otomatik olarak tanıyarak fatura, sözleşme, rapor gibi farklı kategorilere ayırmaktır. Böylece, işletmelerde belge yönetimi süreçleri daha verimli hale gelir.

Bu bağlamda sistem, OCR teknolojisi kullanarak belgelerdeki metinleri çıkarır ve doğal dil işleme teknikleriyle analiz eder. Belgelerin kategorilerine göre sınıflandırılmasının ardından, kullanıcıya rapor olarak sunulmak üzere sistemde depolanır. Aşağıdaki diyagram, sistemin genel yapısını ve bileşenler arasındaki bağlantıları görselleştirmektedir.

Ürün Fonksiyonları, Veri Akış Diyagramı (DFD)



Girdi Tarafında Fonksiyonellik

Belge Yükleme: Kullanıcı, PDF, Word gibi belgeleri sisteme yükler.

İşlem Tarafında Fonksiyonellik

OCR İşlemi: Yüklenen belgelerden metin bilgisi çıkarılır.

Doğal Dil İşleme (NLP): Metin verisi analiz edilerek belge türüne göre anahtar kelimeler ve yapısal özellikler belirlenir.

Sınıflandırma: Belge, içerik analizi sonucuna göre fatura, sözleşme, teklif gibi kategorilere ayrılır.

Çıktı Tarafında Fonksiyonellik

Sınıflandırma Raporu: Belgenin türü ve ilgili kategori bilgisi kullanıcıya rapor olarak sunulur.

Arşivleme: Belgeler, ilgili kategorilerde arşivlenir ve daha sonra erişilebilir şekilde saklanır.

Yetkili Kullanıcı Erişimi: Yetkili kullanıcılar sisteme giriş yaparak geçmiş belgelere erişebilir ve kategorilendirme sonuçlarını inceleyebilir.

Veri Yönetimi Tarafında Fonksiyonellik

Veri Alışverişi: Sisteme yüklenen belgeler, OCR ve NLP analizlerinden geçtikten sonra veritabanına kaydedilir.

Kayıt ve Erişim: Kullanıcıların önceki belge analizlerine hızlı erişimi sağlanır.

Geliştirme Tarafında Fonksiyonellik

Sonuç Filtreleme ve Karşılaştırma: Belirli kriterlere göre sınıflandırma sonuçlarını filtreleme ve karşılaştırma.

Versiyon Değişiklikleri: Güncellemeler sonrası kullanıcılara versiyon değişiklikleri bildirilir.

Kullanıcı Sınıfları ve Davranışları Use Case Diyagramı

Kullanıcılar şu şekilde sınıflandırılabilir:

- İşletme Yöneticisi
- Doküman Yönetim Uzmanı
- Çalışan (Genel Kullanıcı)

Davranış	İşletme Yöneticisi	Doküman Yönetim Uzmanı	Çalışan (Genel Kullanıcı)
Detaylı Klasör Görüntüleme	Evet	Evet	Hayır
Nihai Klasör Görüntüleme	Evet	Evet	Evet
Klasör Düzenleme ve Ekleme	Evet	Hayır	Hayır

Ortam, Teknoloji ve Donanım

Durum	Tercih Edilen Spesifikasyon	Durum	Tercih Edilen Spesifikasyon
Yapay Zeka Modeli	TensorFlow (OCR ve NLP)	Yapay Zeka Modeli	TensorFlow (OCR ve NLP)
Girdi Cihazı	USB, Ağ Sürücüsü	Girdi Cihazı	USB, Ağ Sürücüsü
Kullanıcı Arayüzü	Web Tabanlı	Kullanıcı Arayüzü	Web Tabanlı
Önceliklendirilen Kullanıcı Ortamı	Windows, Mac OS	Önceliklendirilen Kullanıcı Ortamı	Windows, Mac OS
Proje Programlama Ortamı	Python ekosistemi	Proje Programlama Ortamı	Python ekosistemi
SQL Veritabanı	PostgreSQL	SQL Veritabanı	PostgreSQL
NoSQL Veritabanı	Redis	NoSQL Veritabanı	Redis

Harici Arayüz Gereksinimleri

Kullanıcı Arayüzü

Kullanıcı arayüzleri, kullanıcı dostu ve erişilebilir bir deneyim sunmak için tasarlanacaktır. Arayüz, kullanıcıların belgelerini kolayca yükleyebilmesi ve analiz sonuçlarına hızlı erişim sağlaması amacıyla sade ve işlevsel bir yapıda olacaktır.

Kullanıcılar, belgeye dair parametreleri belirleyebilecekleri etkileşimli formlar ve ayar panelleri ile karşılaşacaklardır. Sınıflandırma sonuçlarının görselleştirilmesi için grafikler ve renkli işaretçiler kullanılacaktır. Ayrıca, analiz sonuçlarını raporlama ve kaydetme araçları ile kullanıcı dostu bir şekilde sunulacaktır.

Donanım Arayüzü

Donanım arayüzü, yazılımın işlevselliğini destekleyecek şekilde optimize edilmiş bir yapı ile tasarlanacaktır. Bu arayüz, kullanıcıların sistemle sorunsuz bir şekilde etkileşim kurmasını sağlamak için gerekli bağlantı noktalarını ve veri iletişim standartlarını içerecektir.

Donanım arayüzü, kullanıcıların belgeleri çeşitli cihazlardan (USB, ağ sürücüler veya bulut depolama alanları) sisteme yükleyebilmeleri için gerekli fiziksel ve sanal bağlantıları sağlayacaktır. Ayrıca, belgelerin sistemde işlenmesini kontrol edebilmeleri için kullanıcılara işlevsel bir kontrol paneli sunulacaktır. Bu panel üzerinden belge yükleme, işleme başlatma ve sonuçları görüntüleme gibi işlemler kolayca gerçekleştirilebilecektir.

Ek olarak, donanım arayüzü veri güvenliğini sağlamak amacıyla gerekli güvenlik önlemleri ile donatılacak; yetkilendirme mekanizmaları ve veri şifreleme yöntemleri ile kullanıcıların belgeleri ve kişisel bilgileri koruma altına alınacaktır. Bu bütünlüklü yaklaşım sayesinde, yazılım ve donanım arasındaki etkileşim optimize edilerek, kullanıcıların daha verimli bir deneyim yaşaması mümkün kılınacaktır.

Yazılım Arayüzü

Proje kapsamında geliştirilecek yazılım arayüzü, modern ve kullanıcı dostu bir deneyim sunmak amacıyla web tabanlı olarak tasarlanacaktır. Bu arayüz, kullanıcıların belgelerini hızlı bir şekilde yüklemesine, analiz başlatmasına ve sınıflandırma sonuçlarını görüntülemesine olanak tanıyacaktır.

Arayüz, Python ile geliştirilen arka planda çalışan algoritmalar ile entegre edilecektir ve veri alışverişini sağlamak için Restful API kullanılacaktır. Belgeler analiz edilip sınıflandırıldıktan sonra, sonuçlar kullanıcıya görsel olarak sunulacak ve grafiklerle desteklenecektir. Kullanıcılar, sınıflandırma sürecini gerçek zamanlı olarak izleyebilecek ve sonuçları detaylı bir şekilde inceleyebilecektir.

Veri yönetimi için PostgreSQL veritabanı, önbellekleme işlemleri için Redis kullanılacaktır. Bu sayede kullanıcılar geçmiş analiz sonuçlarına hızlıca erişebilirken, sistemin performansı optimize edilecektir.

İletişim Arayüzü

Proje kapsamında geliştirilecek iletişim arayüzü, sistem bileşenleri arasındaki veri akışını sağlamak ve kullanıcılarla sistem arasındaki etkileşimi optimize etmek için tasarlanacaktır. Bu arayüz, belgelerden elde edilen verilerin etkili bir şekilde işlenmesi ve analiz sonuçlarının kullanıcıya sunulması için gerekli tüm iletişim protokollerini içerecektir.

Web tabanlı kullanıcı arayüzü ile Python tabanlı arka uç arasında, Restful API ve WebSocket gibi iletişim yöntemleri kullanılacak, böylece kullanıcılar analiz sonuçlarını gerçek zamanlı olarak görüntüleyebilecektir.

PostgreSQL ve Redis ile veritabanı yönetimi ve önbellekleme mekanizmaları da etkin bir şekilde yönetilecek, böylece kullanıcıların geçmiş analiz sonuçlarına hızlı erişimi sağlanacaktır. Bu entegre iletişim yapısı, kullanıcı deneyimini geliştirecek ve sistem performansını artıracaktır.

Sistem Özellikleri

Girdi Tarafı

Bu aşamada, kullanıcılar PDF veya Word gibi dijital belgelerini sisteme yükler. Belgeler, OCR (Optik Karakter Tanıma) teknolojisi kullanılarak işlenebilir metin formatına dönüştürülür. Elde edilen metin verisi, sınıflandırma ve analiz işlemleri için hedef lokasyona aktarılır ve veri tabanında saklanır. Bu sayede, sisteme yüklenen belgeler, sınıflandırma için hazır hale getirilir.

İşlem Tarafı

OCR ve NLP (Doğal Dil İşleme) teknikleri ile işlenen metin verisi, sınıflandırma modeli tarafından analiz edilir. Yapay zeka modeli, belgeleri içerdikleri bilgilere göre otomatik olarak sınıflandırır ve belirli anahtar kelimeler veya yapı özelliklerine göre kategorilere ayırır. Sınıflandırma sonuçları, analiz sürecinin sonunda veritabanına kaydedilir ve kullanıcıların erişimine sunulmak üzere arayüze gönderilir.

Çıktı Tarafı

Analiz ve sınıflandırma sonuçları, tasarlanan kullanıcı arayüzünde çeşitli grafikler ve raporlar halinde sunulur. Eğer belgeyi inceleyecek kullanıcı, belgeyi yükleyen kişi değilse (örneğin, belgeyi görüntülemek isteyen bir iş ortağı), teknik terimler ve ayrıntılar daha sade bir formatta sunulacaktır. Böylece, kullanıcıların belge bilgilerini kolayca anlaması sağlanır. Raporlama sistemi, her kullanıcının ihtiyacına göre özelleştirilebilir bir yapıda olacaktır.

Veri Yönetimi Tarafı

Tasarlanan Restful API ile veri akışı, belge yükleme, OCR/NLP işlemi, sınıflandırma modeli ve kullanıcı arayüzü arasında kesintisiz bir biçimde sağlanır. Bu API ayrıca, tüm işlemlerin veritabanlarına güvenli ve hızlı bir şekilde kaydedilmesini sağlar. Sistemin veri yönetimi tarafında PostgreSQL ve Redis kullanılarak güçlü bir altyapı oluşturulacaktır.

Belgeler ve sınıflandırma sonuçları gibi büyük boyutlu dosyalar veritabanında doğrudan saklanmak yerine dosya sisteminde tutulacak; veritabanında yalnızca referans bilgileri yer alacaktır. Bu yapıda, verilerin güvenli ve verimli bir şekilde depolanması için FILESYSTEM felsefesi benimsenmiştir.

Diğer Fonksiyonel Olmayan Gereksinimler

Performans Gereksinimleri

Belgelerin yüklenmesi ve OCR ile analiz edilmesi sırasında işlem süresinin 1 saniyeden kısa bir sürede tamamlanması hedeflenmektedir. Kullanıcıların veritabanına erişim sürelerinin ise 50 milisaniye içinde gerçekleştirilmesi amaçlanmaktadır. Redis önbellekleme mekanizması sayesinde sık erişilen belgelerin ve analiz sonuçlarının hızlı bir şekilde sunulması sağlanarak, sistemin genel performansı artırılacak ve kullanıcı deneyimi iyileştirilecektir.

Güvenlik Gereksinimleri

Kullanıcı kimlik doğrulama sürecinde çok faktörlü kimlik doğrulama (MFA) yönteminin kullanılması planlanmaktadır. Bu sayede yetkisiz erişimlerin önüne geçilecek ve kullanıcıların güvenliği sağlanacaktır. Ayrıca, tüm veri iletimi SSL/TLS protokolleri kullanılarak şifrelenerek güvenli hale getirilecektir.

Veritabanında saklanan tüm hassas bilgiler, endüstri standartlarına uygun güçlü şifreleme algoritmalarıyla korunacaktır. Yalnızca yetkili kullanıcıların erişimine izin verilecek ve sistem düzenli güvenlik güncellemeleri ile olası tehditlere karşı sürekli olarak izlenecektir. Bu güvenlik gereksinimleri, kullanıcı gizliliğini korumak ve yasal düzenlemelere uyumu sağlamak açısından hayati önem taşımaktadır.

Yazılım Kalite Özellikleri

Uygulamada hiçbir bilgi veya içerik hardcoded olarak yer almayacak, tüm veriler dinamik olarak sunulacaktır. Kullanıcı arayüzü, farklı dil seçeneklerini destekleyecek şekilde çoklu dil desteği sunacaktır. Çeviriler profesyonel çevirmenler tarafından yapılacak ve kullanıcı deneyimini zenginleştirmek amacıyla sade bir dil kullanılacaktır.

Yapay zeka modeli, farklı veri türleri ve sınıflandırma ihtiyaçlarına göre sürekli güncellenebilir ve özelleştirilebilir bir yapıda olacaktır. Parametreler esnek şekilde tasarlanacak, böylece yeni sınıflandırma ihtiyaçlarına ve değişen analiz gereksinimlerine kolayca adapte olabilecektir.

Uygulama, ilerleyen zamanlarda ortaya çıkacak yeni gereksinimleri karşılayabilecek şekilde modüler ve sürdürülebilir bir mimariyle inşa edilecektir. Yazılımın güvenilirliği, çeşitli test senaryolarıyla sürekli olarak denetlenecek ve sistem performansı ile doğruluğu en üst seviyede tutulacaktır.

Gelecekte Yapılması Planlananlar

Kısa Vadede

- Şirket içi ağ veya bulut tabanlı bir sistem üzerinden kullanıcılara doğrudan erişim sağlanacak bir hizmet modeli geliştirilmesi.
- Sınıflandırma sonuçlarının kullanıcı geri bildirimleri ile daha doğru hale getirilmesi; kullanıcıların belge kategorilerini onaylayarak veya düzenleyerek modelin zamanla kendini iyileştirmesine olanak tanınması.

Orta Vadede

- Farklı belge türleri ve alanlara yönelik özelleştirilmiş sınıflandırma modellerinin geliştirilmesi ve bu hizmetlerin ayrı platformlar veya modüller olarak sunulması.
- Yerel sunuculara bağımlılığın azaltılması; tüm veri işleme ve depolama işlemlerinin merkezi bir sunucu veya bulut altyapısı üzerinden sunulması.

Uzun Vadede

- Tüm belge yönetimi ve sınıflandırma hizmetlerinin tek bir kapsamlı platform altında birleştirilmesi; böylece kullanıcıların tüm ihtiyaçlarını tek bir çatı uygulama üzerinden karşılaması.
- Modüler, iyi dökümente edilmiş ve esnek bir Restful API tasarımı; böylece uygulamanın farklı sistemlerle entegrasyonunun kolaylaştırılması ve kullanım alanlarının genişletilmesi.

Uzatma Planı

Projenin tüm geliştirme süreçlerinin bir yıllık bir periyotta tamamlanması hedeflenmektedir. Ancak, olası gecikmeler göz önünde bulundurularak bu sürenin altı ay daha uzatılma ihtimali öngörülmektedir.

Proje Yürütücüleri

- Zeynep KİCİKOĞLU
- Mehmet Eren EKİZ
- Abdullah KARATAŞ