1. **Veri İşleme ve İndeksleme (data\_management.py):**
   * **Dosya İçeriklerinin Okunması:** Belirtilen klasördeki tüm dosyalar taranır ve uygun fonksiyonlarla içerikleri okunur.
   * **Metin Ön İşleme:** Okunan içerikler, preprocess\_text fonksiyonu ile temizlenir (noktalama işaretlerinin kaldırılması, küçük harfe dönüştürme, stop kelimelerin çıkarılması vb.).
   * **Vektörleştirme:**
     + **TF-IDF Vektörleri:** TfidfVectorizer kullanılarak kelime frekanslarına dayalı vektörler oluşturulur.
     + **SBERT Vektörleri:** SentenceTransformer modeli ile semantik anlamı yakalayan vektörler elde edilir.
   * **FAISS İndeksi Oluşturma:** Yüksek boyutlu vektörlerin hızlı bir şekilde aranabilmesi için FAISS indeksi kullanılır.
   * **Verilerin Kaydedilmesi:** İndeksler ve gerekli veriler disk üzerinde saklanır ve ihtiyaç duyulduğunda yüklenir.
2. **Arama Fonksiyonları:**
   * **TF-IDF Araması (tfidf\_search.py):** Sorgu ile belgeler arasındaki kelime frekanslarına dayalı benzerliği hesaplar.
   * **SBERT Araması (sbert\_search.py):** Sorgu ve belgeler arasındaki anlamsal benzerliği hesaplar.
   * **Kombine Arama (combined\_search.py):** TF-IDF ve SBERT skorlarını belirli ağırlıklarla birleştirerek daha hassas sonuçlar sunar.
3. **Model Seçimi ve Ağırlıklandırma (model\_selection.py):**
   * **Model Seçimi:** Sorgunun özelliklerine göre en uygun arama modeli belirlenir (TF-IDF, SBERT veya Kombine).
   * **Ağırlık Belirleme:** Kombine arama için TF-IDF ve SBERT ağırlıkları hesaplanır.
4. **Grafik Kullanıcı Arayüzü (GUI) (gui.py):**
   * **Kullanıcı Etkileşimi:** Kullanıcıdan arama sorgusunu alır ve sonuçları görsel olarak sunar.
   * **Arka Plan İşlemleri:** search\_query fonksiyonunu çağırarak arama işlemini gerçekleştirir ve sonuçları işler.

**Çalışma Prensibi**

1. **Veri Yükleme ve İşleme:**
   * Uygulama başlatıldığında, daha önce oluşturulmuş indeksler varsa bunlar yüklenir; yoksa dosyalar taranır ve içerikleri işlenir.
   * **Dosya Okuma:** Her dosya türü için uygun okuma fonksiyonları kullanılır (örneğin, PDF için pdfplumber, Word için python-docx).
   * **Ön İşleme:** Metin verisi temizlenir ve analiz için uygun hale getirilir.
   * **Vektörleştirme ve İndeksleme:** İşlenmiş metinler TF-IDF ve SBERT modelleri ile vektörleştirilir ve FAISS indeksi oluşturulur.
2. **Arama İşlemi:**
   * Kullanıcı bir sorgu girdiğinde, sorgu ön işlenir ve vektörleştirilir.
   * **Model Seçimi:** select\_model fonksiyonu ile sorgunun yapısına göre en uygun arama modeli seçilir.
   * **Ağırlıklandırma (Kombine Model için):** get\_similarity\_weights fonksiyonu ile TF-IDF ve SBERT ağırlıkları belirlenir.
   * **Arama Fonksiyonunun Çalıştırılması:** Seçilen modele uygun arama fonksiyonu çağrılır ve en benzer belgeler bulunur.
3. **Sonuçların Gösterimi:**
   * Elde edilen sonuçlar skorları ile birlikte kullanıcıya sunulur.
   * **Eşik Değeri Uygulama:** Belirli bir skorun altındaki sonuçlar gösterilmeyebilir (örneğin, skor eşik değeri %70).

**Kodun Temel Bileşenleri ve Fonksiyonları**

* **get\_file\_contents(folder):** Belirtilen klasördeki tüm dosyaları okuyarak içeriklerini ve dosya isimlerini döndürür.
* **preprocess\_text(text):** Metin verisini temizleyerek analiz için uygun hale getirir (noktalama işaretlerini kaldırma, küçük harfe dönüştürme, stop kelimeleri çıkarma).
* **vectorize\_and\_index\_content(contents):** Metinleri TF-IDF ve SBERT modelleri ile vektörleştirir ve FAISS indeksine ekler.
* **search\_query(query, top\_k=5):** Kullanıcının sorgusunu işleyerek en benzer top\_k sayıda belgeyi bulur.
* **tfidf\_search(query\_tfidf, top\_k, tfidf\_matrix, filenames):** TF-IDF benzerlik skorlarına göre arama yapar.
* **sbert\_search(query\_embedding, top\_k, index, filenames):** SBERT benzerlik skorlarına göre arama yapar.
* **combined\_search(query\_tfidf, query\_embedding, tfidf\_weight, sbert\_weight, top\_k, tfidf\_matrix, index, filenames):** TF-IDF ve SBERT skorlarını ağırlıklandırarak kombine arama yapar.
* **select\_model(query):** Sorgunun uzunluğuna veya yapısına göre en uygun arama modelini seçer.
* **get\_similarity\_weights(query):** Kombine model için TF-IDF ve SBERT ağırlıklarını belirler.