

Metin Birleřtirme Uygulaması

Kocaeli Üniversitesi

Bilgisayar Mühendislięi Bölümü

Rana Dudu Kabak

Muhammed Taha Metin

ranadudukabak@gmail.com

tahametin341@gmail.com

I. ÖZET

Bu projede, kitap, dergi vb. gibi metinlerden elde edilen cümleler/kelimeler arasındaki benzerlik, frekans gibi özelliklerin bulunması, metinlerin birleştirilip birleştirilmeyeceğine karar verilmesi ve elde edilen verilerden karakter tabanlı cümlelerin/kelimelerin birleştirilmesi hedeflenmektedir. Projenin ana amacı, iki veya daha fazla metnin uygun bir şekilde birleştirilmesi problemi için yeni algoritmalar geliřtirmek ve bunları görsel bir arayüzle sunmaktır.

Kullanıcının sisteme girdięi metinleri kaydedip listeleyebilmesi için girilen metinler .JSON formatında döküman olarak NoSQL tabanlı saklanmalıdır.

Görsel arayüzdeki iki adet metin giriř etiketi ve bu etiketlere ait “ekle”, “birleřtir” ve “kaydet” butonları sayesinde kullanıcılar rahatlıkla metinleri girebilir, birleřtirebilir ve kaydedebilirler.

II. GİRİř

Proje kapsamında, en az iki adet metin giriř butonu, girilen metinleri birleřtirme butonu, girilen metinleri kaydetme (.json formatında) butonu ve veriden elde edilen en uzun dizi listelerinin tümü bir liste halinde kaydedilmesi gerekmektedir. Ayrıca, iki metnin birleřtirme iřlemi için geçen süre de hesaplanacaktır.

Java dili kullanılacak olan bu projede, veri tabanı olarak MongoDB (NoSQL) kullanılacak ve web arayüzü için istenilen bir programlama dili kullanılabilecektir.

III. YÖNTEM

İlgili proje veri tabanı, önyüz (frontend) ve arka yüz (backend) kısımlarından oluşmaktadır.

Öncelikle bir web projesi oluşturulmak istendiğinden geliřtirme ortamı kullanılarak bir Java Spring projesi oluşturulmalıdır. Bu şekilde web uygulaması ekstra bir web server ayaęa kaldırılmadan localhost:8080 adresinde çalışabilmektedir.

Oluřturulan projede MongoDB veri tabanı bağlantısı için öncelikle MongoDB Desktop uygulaması local’e indirilir ve uygulamadan gerekli connection string’ler

alınır. Kod tarafında bir DatabaseConnection class'ı oluşturulur ve ilgili bağlantı kodları burada yazılır.

Bu aşamadan sonra projenin önyüz tarafına geçilir. Bu kısım için bir html dokümanı içerisinde gerekli işlemlerin yapılması tercih edilmiştir. Html dokümanında arayüzde görünecek component'lerin görünüm ayarları, arka plan görüntüleri, formdaki butonlarla etkileşime girildiğinde yapılacak işlemler ayarlanır.

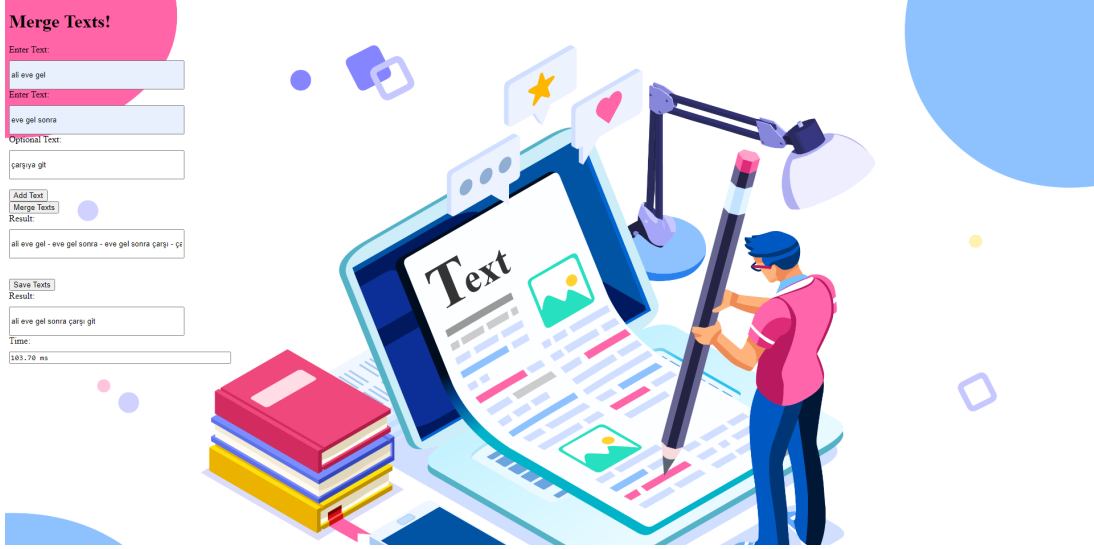
Son durumda ise arka yüz geliştirme işlemleri yapılır. Bu aşamada proje isterlerinde bulunan şekilde arayüzden girilen iki metin ve ek olarak girilmişse opsiyonel metin alınır ve AddText butonu yardımıyla birleştirilir. Daha sonrasında Merge Text butonu kullanıldığında, **VI. Akış Diyagramı** bölümünde akış diyagramı paylaşılan algoritmalar yardımıyla girilen metinler birleştirilir. Metin birleştirme algoritmasının çalışma süresi hesaplanır. MergeSimilarText sınıfında, benzerliği onaylanmış metinlerin ortak kelimeleri silinerek birleştirilir; SimilarityData sınıfında, metinler arasındaki benzerlik skoru ve bu benzerliklerin hangi metinler arasında olduğu tutulur; TextMergeManager sınıfında, diğer sınıflardaki fonksiyonlar kullanılarak süreç yönetilir ve TextSimilarity sınıfında CosineSimilarity, JaccardSimilarity ve CosineSimilarity'nin ilk 3 harfi baz alınacak şekilde modifiye edilmiş hali kullanılarak metinler arasındaki benzerlik sayısal bir skora çevrilir ve skora göre metinlerin benzer olup olmadıklarına karar verilir.

IV. SONUÇ

Yapılan tüm işlemler sonucunda, MongoDB'ye bağlantı sağlandı. Arayüzden alınan metinler ekrandaki bir textbox'ta gösterildi. Bu metinler kendi aralarındaki benzerlikleri hesaplanılarak uygun biçimde birleştirildi. Aynı zamanda birleştirilen metin MongoDB veri tabanında .json formatında kaydedildi. Bu birleştirme işleminin ne kadar sürdüğü hesaplandı ve arayüzde gösterildi.

İlgili proje Windows 11 işletim sisteminde, IntelliJ IDEA 2022.2.1 IDE, arka yüzde Java programlama dili ve önyüzde HTML kullanılarak başarıyla test edilmiş ve yapılan testlerde herhangi bir runtime hatasıyla karşılaşılmamıştır.

V. DENEYSEL SONUÇLAR



Görsel 1: Web sitesi arayüzü.

Merge Texts!

Enter Text:

Enter Text:

Optional Text:

Add Text

Merge Texts

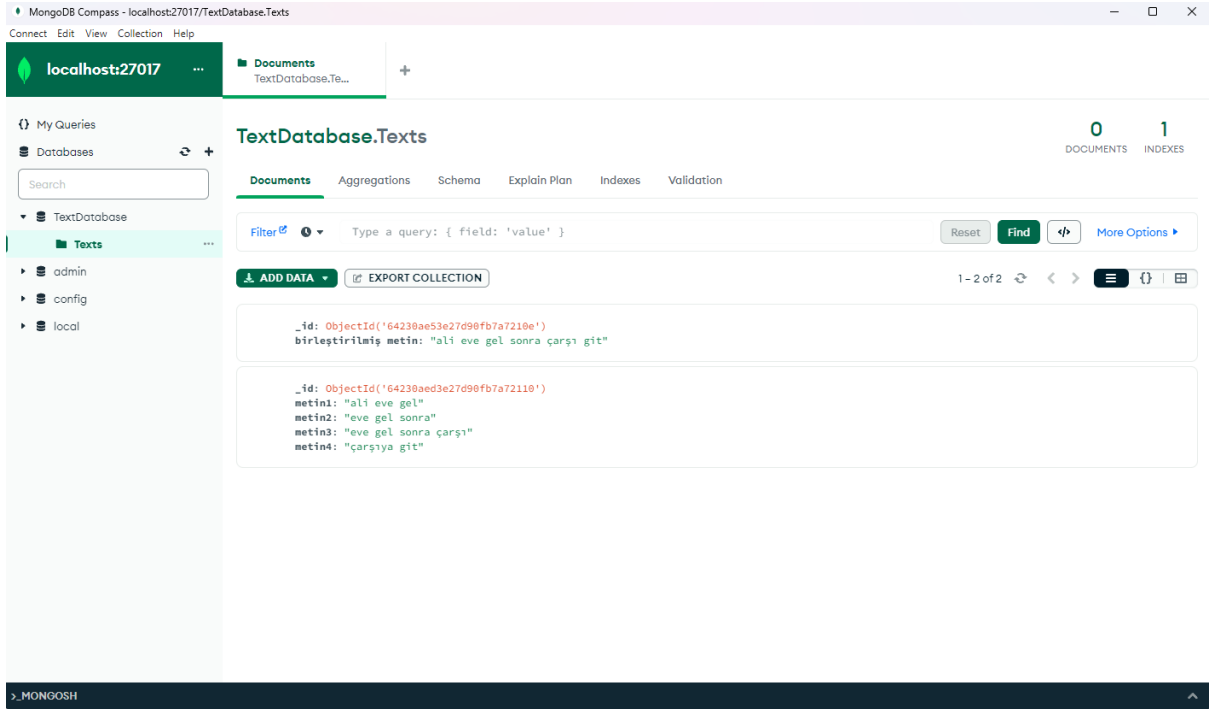
Result:

Save Texts

Result:

Time:

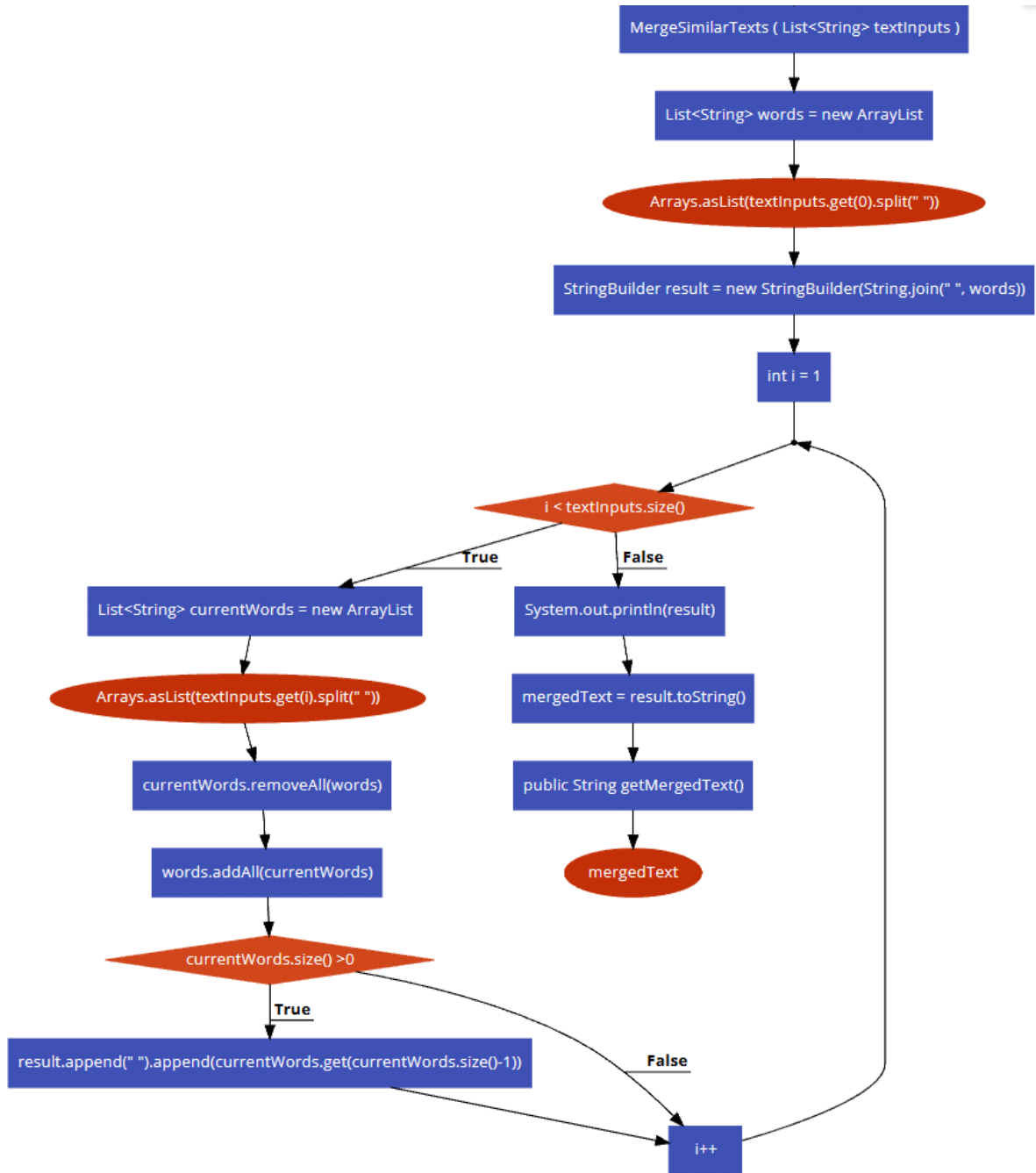
Görsel 2: Çalıştırılmış uygulamadan örnek bir ekran görüntüsü.



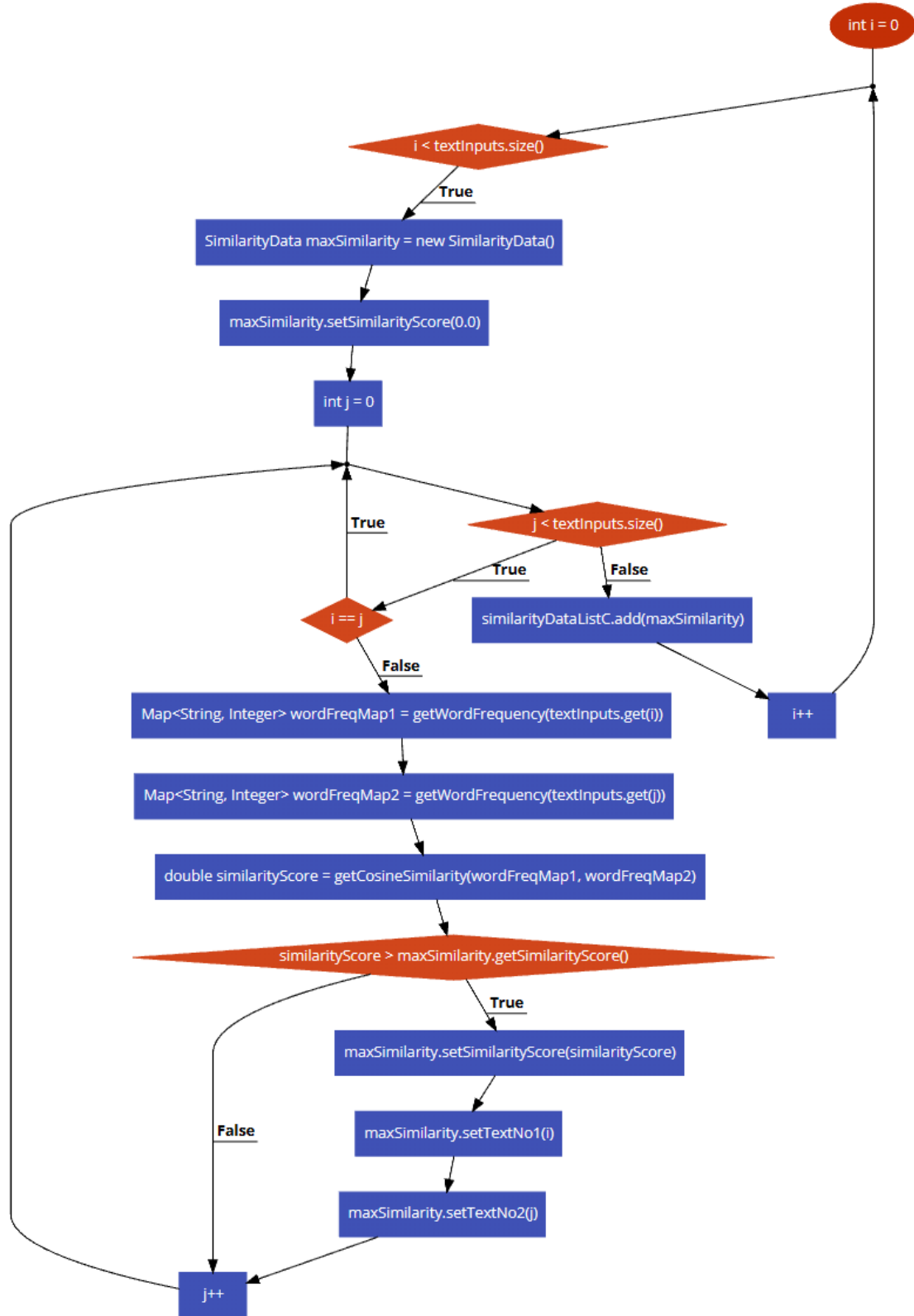
Görsel 3: MongoDB Compass'te görüntülenen kaydedilmiş veriler.

VI. AKIŞ DİYAGRAMI

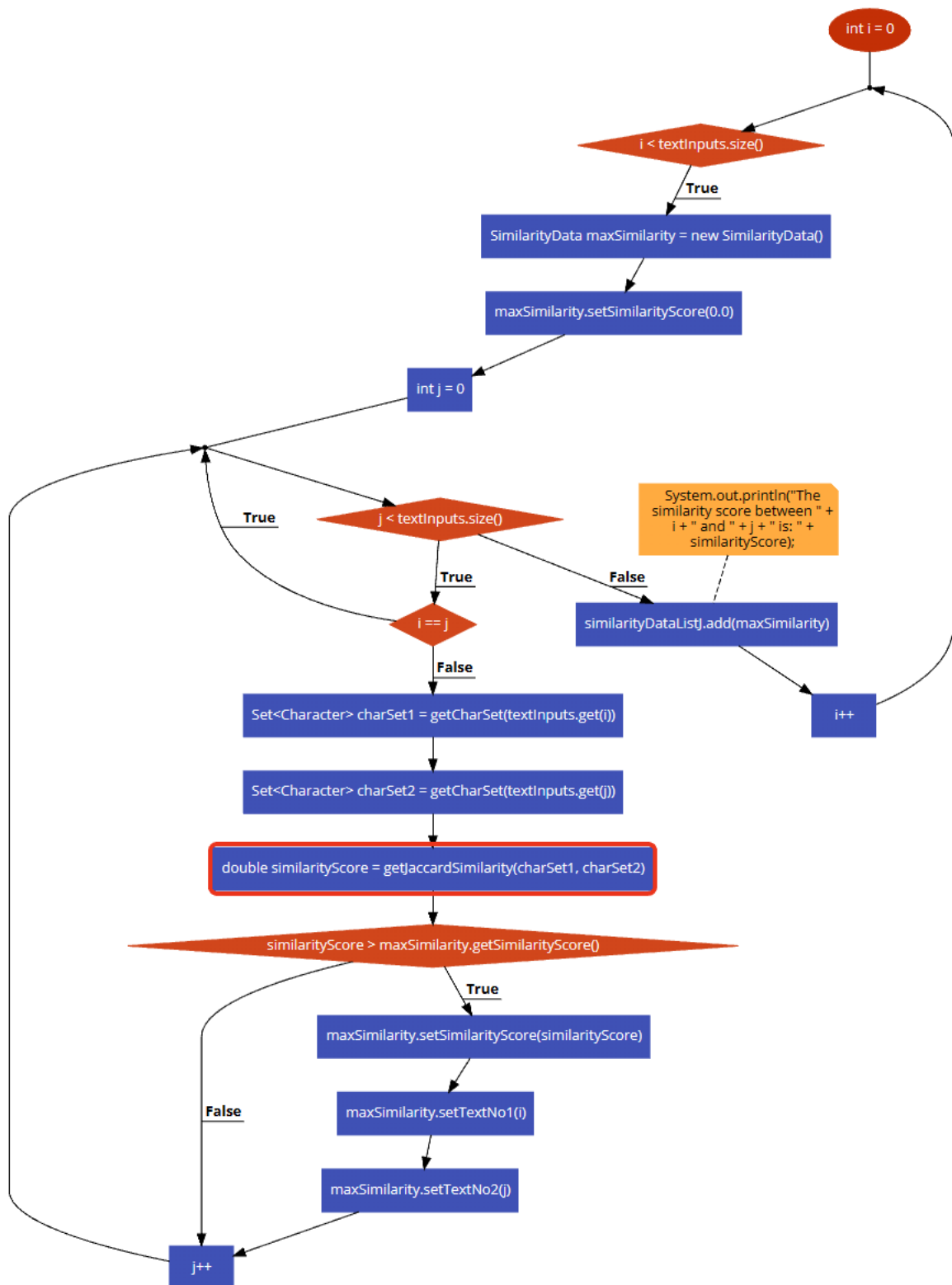
VI.I. Metin Birleştirme Akış Diyagramı



VI.II. Cosine Similarity Akış Diyagramı



VI.III. Jaccard Similarity Akış Diyagramı



VII. KAYNAKÇA

1. [HTTPS://WWW.JETBRAINS.COM/GO/GUIDE/TIPS/MONGODB-CONNECT-FROM-IDE/](https://www.jetbrains.com/go/guide/tips/mongodb-connect-from-ide/)
2. [HTTPS://EN.WIKIPEDIA.ORG/WIKI/COSINE_SIMILARITY](https://en.wikipedia.org/wiki/Cosine_similarity)
3. [HTTPS://EN.WIKIPEDIA.ORG/WIKI/JACCARD_INDEX](https://en.wikipedia.org/wiki/Jaccard_index)
4. [HTTPS://MONGODB.GITHUB.IO/MONGO-JAVA-DRIVER/3.4/DRIVER/TUTORIALS/CONNECT-TO-MONGODB/](https://mongodb.github.io/mongo-java-driver/3.4/driver/tutorials/connect-to-mongodb/)
5. [HTTPS://WWW.MONGODB.COM/DOCS/DRIVERS/JAVA/SYNC/V4.3/FUNDAMENTALS/CONNECTION/CONNECT/](https://www.mongodb.com/docs/drivers/java/sync/v4.3/fundamentals/connection/connect/)
6. [HTTPS://SPRING.IO/GUIDES/GS/SPRING-BOOT/](https://spring.io/guides/gs/spring-boot/)
7. [HTTPS://WWW.YOUTUBE.COM/WATCH?V=OFME2o29NgU](https://www.youtube.com/watch?v=OFME2o29NgU)
8. [HTTPS://WWW.MONGODB.COM/COMPATIBILITY/SPRING-BOOT](https://www.mongodb.com/compatibility/spring-boot)
9. [HTTPS://WWW.YOUTUBE.COM/WATCH?V=SSJ0CGxv60k](https://www.youtube.com/watch?v=SSJ0CGxv60k)
10. [HTTPS://PLUGINS.JETBRAINS.COM/PLUGIN/7179-MAVEN-HELPER](https://plugins.jetbrains.com/plugin/7179-maven-helper)