

YAZILIM LABORATUVARI 1

PROJE 2

TARIK GÖREN
Bilgisayar Mühendisliği
200202022@kocaeli.edu.tr

ÖMER ARAN
Bilgisayar Mühendisliği
190202012@kocaeli.edu.tr

ÖZET

Bu proje ile amacımız thread ve multithread kavramlarını öğrenmek, uygulamak ve büyük verilerle çalışırken dikkat etmemiz gereken noktaları keşfetmek aynı zamanda tecrübe edinmektir. Yaklaşık bir milyon verinin bulunduğu bir şikayet listesini önce csv formatında localimize alarak ardından tercih ettiğimiz java dilinin bir kütüphanesi olan common-csv kütüphanesiyle ayıklamak gereksiz sütunları yok edip yeni csv dosyasına yazdırıp ardından bu şikayet listesindeki değerleri java arayüzü olan swing ile masaüstü uygulama haline getirerek istelere göre listeleme ve filtreleme yapmaktır. Multithreadler ise burada devreye giriyor. Milyonlarca veriyi kendi belirlediğimiz thread sayısı ile işlemek inanılmaz bir hız katmalı ve zamandan kazandırıp kaliteli bir hizmet sunmaya yardımcı olmalıdır. Temel amacımız multithread, ayıklama ve masaüstü uygulama geliştirmektir.

I. GİRİŞ

Bu projede backend kısmında java dilini, arayüz tasarımı Java Swing'i verileri ayıklamak için ve gerektiğinde tekrar yazdırmak için common-csv kütüphanesini ve işlemleri hızlandırmak için threadleri kullandık.

Bu proje kapsamında bizden aşağıdaki amaçların gerçekleştirilmesi istenmiştir:

- Elde edilen tabloda 6 farklı sütun bulunmalıdır: Product (Ürün), Issue (Konu), Company (Şirket), State, Complaint ID, Zip Code.
- Null değer içeren kayıtlar bulunmamalıdır.
- Kayıtlardaki noktalama işaretleri kaldırılmalıdır
- Kayıtlardaki stop word'ler kaldırılmalıdır (nltk kütüphanesi kullanılabilir)

II. İSTERLER

- Verilen veri seti istenen şekilde yeniden düzenlenmelidir.
- Düzenlenmiş veri setindeki kayıtlar arasında benzerlik kontrolü yapılmalıdır. Kontrol sırasında mutlaka multithreading kullanılmalıdır. Multithreading için kullanılacak thread sayısı uygulama arayüzünden girilmelidir
- Her thread'in çalışma zamanı ve tüm thread'ler için toplam çalışma zamanı bilgileri uygulama arayüzünde gösterilmelidir.
- İstenilen sütun ya da sütunlar arasındaki girilen benzerlik oranı (threshold) ve üzerinde benzerliğe sahip kayıtlar masaüstü uygulamasında gösterilmelidir..
- Uygulamanızı sunmak üzere basit bir arayüz geliştirmeniz istenmektedir. Bu arayüz aşağıdaki isteleri içermelidir:
 - Benzerlik oranının (Threshold değeri) seçilebileceği / girilebileceği bir araç
 - Benzerliklerinin araştırılması istenen sütun veya sütunların seçilebileceği bir araç.
 - Kaç tane thread kullanılacağını seçilebileceği / girilebileceği bir araç,
 - Her bir thread'in çalışma zamanını ve toplam çalışma zamanını gösteren araçlar
 - Sonuçların açıkça ekranda gösterilebileceği bir araç.
- Uygulamada aşağıdaki ve benzer senaryolardan elde edilen sonuçlar ekranda gösterilmelidir:
 - Senaryo 1: Ürün (Product) sütununda ekranda gösteriniz.
 - Senaryo 2: Aynı ürünler (Product) için (Issue) içeren Şirket (Company) isimlerini ekranda gösteriniz.
 - Senaryo 3: 'Complaint Id' = 3198084 olan şikayet kaydı için benzerlikteki konuları (is-

sue) içeren kayıtları ekranda gösteriniz.

- Senaryo 4: 5 Thread ile Konular(Issue) sü-tununda kayıtları ekranda gösteriniz

III. YÖNTEM

A. Başlangıç

Öncelikle bizden istenen programları ve prog-ramların içerisinde kullanacağımız kütüphaneleri entegre ettik. İsterleri gerçekleştirmeye başladık. Bizden istenen ilk isterden başladık. Common-CSV dokümanları ve ilgili eğitim videoları ile ayıklama yapacağımız verileri oluşturduğumuz maven projesinin test kısmında denemeler yaparak complaint(kaggle'den alınan veri kümesinin complaintlerden oluşması üzerine class ismi olarak bunu uygun gördük.) verilerini csv dosyamızdan çekmeyi denedik.

B. Geliştirme Ortamı

Projenin backend tarafı Java'nın Maven mimari proje tipi ile geliştirilirken arayüz tarafı Swing-GUI ile geliştirilmiştir. Ayıklama aracı olarak Common-csv ve Java csv okuma-yazma komutları kullanılmıştır.

Macos-Windows sistemlerde geliştirilip IDE olarak IntelliJ Idea Ultimate Version ve Netbeans 15 kullanılmıştır. Projenin versiyon kontrolü tamamen github üzerinden yapılmıştır. IDE'lerimizde yazdığımız kodları ayıklama ve senaryo gerçeklemelerini sağladıktan sonra çalıştırıp önlem açısından debug ve testler yapılmıştır.

IV. PSUEDO KOD

Bir ayıklama ve multithread projesinin yalancı kodu senaryoları gerçekleşme sırasına uygun bir durumda açıklayarak gerçekleşir.

Uygulama Çalıştırılır:

Arayüz(swing) açılır: seçebileceğimiz tablalar ve senaryolar

senaryo 1 : seçtik

gerekli verileri girip getir butonuna bas

button similarity classının similarity1 fonksiyonunu çağırıp

ayıklanmış veriyi load et

load olan veriyi istenen kolon ve benzerlik oranına göre çalıştır

şartları sağlayanları bir listeye at

listeyi döndür

arayüzde table'da sayı ve kolonları listele

senaryo1 işi biter:

senaryo2 çalıştır:

gir thread sayısı

gir benzerlik oranı

gir talep edilen kolon

gir aynı olanlar

butona bas : getir 2

arka planda similarity2 fonksiyonu çalışır

ayıklanan verileri bir listeye ata

listedeki verileri giriş koşullarına uygun olanları al yeni listeye at

liste tamamlanınca threadleri kapat

listeyi döndür

arayüzde thread ve veri listeleme bölümleri var

bu bölümlerde elindeki verileri sergile

senaryo 2 : bitti.

senaryo 3 :

gir thread sayısı

gir benzerlik oranı

gir talep edilen kolon

gir aynı olanlar

butona bas : getir 3

arka planda similarity3 fonksiyonu çalışır

ayıklanan verileri bir listeye ata

arayüze döndür

senaryo 3 : bitti

senaryo 4 :

gir thread sayısı

gir benzerlik oranı

gir talep edilen kolon

gir aynı olanlar

butona bas : getir 4

arka planda similarity4 fonksiyonu çalışır

ayıklanan verileri bir listeye ata

senaryo 4 : bitti

V. SONUÇLAR

Proje amacı yerine getirilerek multithreading ile verileri daha hızlı işleme, ayıklama, atama ve şartları sağlayanları arayüze döndürerek baştan sona end to end bir masaüstü uygulaması geliştirdik. Multithread kavramının temel amacı ve gerekliliği hakkında bilgi edindik. Geçmiş yıllarda da kullandığımız üzere javanın masaüstü app geliştirmek maksadıyla sunmuş olduğu swing gui ile projemizin önuç kısmını senaryoları gerçekleştirmek için ve aynı zamanda kullanıcıya daha iyi bir deneyim sunmak amacıyla combobox'lar, textfieldlar ve sade tasarım ile geliştirdik. Verileri işleme kısmı bir kez olacağı için projenin main yerine test kısmında bir class içerisinde parse adındaki class ile ayıklayıp yeni bir csv dosyasına attık bu şekilde her seferinde yeniden parse işlemi yapmak zorunda kalmadık. Bir yazılım hizmeti sunmanın çeşitli aşamaları olduğunu bu aşamaların hepsinin projenin sunumunda ve kullanıcı deneyimi kısmında önemini aynı zamanda zamandan ve maliyetten kazanmanın multithreadler aracılığı ile yapılabileceği ve milyonlarca veriye sahip şirketlerin bu verileri işlerken ya da efektif bir veri kullanımı sağlamak adına girişimlerde bulunurken maliyeti en aza indirme uğraşlarının yazılım kısmında çok büyük avantaj sağlayabileceğini gördük.

KAYNAKLAR

[1] Multithreading dokümanları

<https://www.geeksforgeeks.org/multithreading-in-java/>

https://www.tutorialspoint.com/java/java_multithreading.htm

<https://medium.com/@sinanselimoglu/java-8-streams-paralel-i-CC%87%C5%9Flemler-3-a01233c8fd9f>

https://www.tutorialspoint.com/java/java_multithreading.htm

<https://medium.com/s%C4%B1f%C4%B1rdan-i-CC%87leri-d%C3%BCzeye-java-e%C4%9Fitim-serisi/multithreaded-programlama-1-k%C4%B1s%C4%B1m-40904a219a18>

[2] Veri Seti Düzenleme

<https://commons.apache.org/proper/commons-csv/>

<https://www.baeldung.com/apache-commons-csv>

<https://www.callicoder.com/java-read-write-csv-file-apache-commons-csv/>

<https://commons.apache.org/proper/commons-csv/user-guide.html>

<https://stackoverflow.com/questions/35319544/removing-stopwords-java>

[3] Swing Arayüz dokümanları

<https://www.javatpoint.com/java-swing>

<https://docs.oracle.com/javase/7/docs/api/javax/swing/package-summary.html>

<https://docs.oracle.com/javase/7/docs/api/javax/swing/ComboBoxModel.html>

<http://www.baskent.edu.tr/~tkaracay/etudio/ders/prg/java/ch27/JavaSwing.pdf>

[4] Spring Framework Doküman

<https://spring.io/projects/spring-boot>

<https://www.baeldung.com/learn-spring-course>

EKRAN GÖRÜNTÜLELRLİ

#	Kayıt-1	Product	Kayıt-2	Product	Benzerlik Oranı
983245	2948152	Credit reporting credit repair services...	664444	Credit reporting	37.5
983246	2948152	Credit reporting credit repair services...	664471	Credit reporting	37.5
983247	2948152	Credit reporting credit repair services...	662634	Credit reporting	37.5
983248	2948152	Credit reporting credit repair services...	662640	Credit reporting	37.5
983249	2948152	Credit reporting credit repair services...	664424	Credit reporting	37.5
983250	2948152	Credit reporting credit repair services...	662584	Credit reporting	37.5
983251	2948152	Credit reporting credit repair services...	664842	Credit reporting	37.5
983252	2948152	Credit reporting credit repair services...	806809	Credit reporting	37.5
983253	2948152	Credit reporting credit repair services...	2671290	Credit reporting credit repair services...	100.0
983254	2948152	Credit reporting credit repair services...	2672043	Credit reporting credit repair services...	100.0
983255	2948152	Credit reporting credit repair services...	2282589	Credit reporting	37.5
983256	2948152	Credit reporting credit repair services...	807704	Credit reporting	37.5
983257	2948152	Credit reporting credit repair services...	806703	Credit reporting	37.5
983258	2948152	Credit reporting credit repair services...	808168	Credit reporting	37.5
983259	2948152	Credit reporting credit repair services...	807595	Credit reporting	37.5
983260	2948152	Credit reporting credit repair services...	806882	Credit reporting	37.5
983261	2948152	Credit reporting credit repair services...	806814	Credit reporting	37.5
983262	2948152	Credit reporting credit repair services...	808224	Credit reporting	37.5

S1 S2 S3

Seçilen Sütun: Product

Veri Sayısı: 5

Benzerlik Oranı: 30

Thread Sayısı: 10

Getir

Toplam Geçen Süre: 5281 ms

#	Thread ID	Thread Adı	Çalışma Süresi
1	67	ThresholdThread0	5135 ms
2	68	ThresholdThread1	5232 ms
3	69	ThresholdThread2	5232 ms
4	70	ThresholdThread3	5135 ms
5	71	ThresholdThread4	5275 ms
6	72	ThresholdThread5	4837 ms
7	73	ThresholdThread6	4245 ms
8	74	ThresholdThread7	4064 ms
9	75	ThresholdThread8	4392 ms
10	76	ThresholdThread9	4341 ms

Şekil 1: Senaryo 1

#	Kayıt-1	Kayıt-2	Product-1	Product-2	Oran	Issue-1	Issue-2	Oran	Company-1	Company-2
1	2837577	2837577	Mortgage	Mortgage	100.0	Applying mortgage	Applying mortgage	100.0	Seterus	Seterus
2	2837577	2948424	Mortgage	Mortgage	100.0	Applying mortgage	Applying mortgage ...	60.0	Seterus	MILEND
3	2837577	2669770	Mortgage	Mortgage	100.0	Applying mortgage	Applying mortgage ...	60.0	Seterus	AMERICAN ADVIS...
4	2837577	2670028	Mortgage	Mortgage	100.0	Applying mortgage	Applying mortgage ...	60.0	Seterus	WYNDHAM CAPITA...
5	2837577	2836616	Mortgage	Mortgage	100.0	Applying mortgage	Applying mortgage ...	60.0	Seterus	QUICKEN LOANS
6	2837577	3115735	Mortgage	Mortgage	100.0	Applying mortgage	Closing mortgage	50.0	Seterus	Professional Mortg...
7	2837577	3115600	Mortgage	Mortgage	100.0	Applying mortgage	Closing mortgage	50.0	Seterus	HUNTINGTON NA...
8	2837577	3172440	Mortgage	Mortgage	100.0	Applying mortgage	Closing mortgage	50.0	Seterus	M T BANK CORPO...
9	2837577	3173004	Mortgage	Mortgage	100.0	Applying mortgage	Applying mortgage ...	60.0	Seterus	Real Time Resolut...
10	2837577	3171965	Mortgage	Mortgage	100.0	Applying mortgage	Closing mortgage	50.0	Seterus	WELLS FARGO C...
11	2837577	2947897	Mortgage	Mortgage	100.0	Applying mortgage	Applying mortgage	100.0	Seterus	NAVY FEDERAL C...
12	2837577	3115791	Mortgage	Mortgage	100.0	Applying mortgage	Closing mortgage	50.0	Seterus	Dtech Financial LLC
13	2837577	3116318	Mortgage	Mortgage	100.0	Applying mortgage	Applying mortgage ...	60.0	Seterus	J G Wentworth Co...
14	2837577	2837250	Mortgage	Mortgage	100.0	Applying mortgage	Applying mortgage	100.0	Seterus	BANK AMERICA N...
15	2837577	2947166	Mortgage	Mortgage	100.0	Applying mortgage	Applying mortgage ...	60.0	Seterus	Selene Finance LP
16	2837577	2959038	Mortgage	Mortgage	100.0	Applying mortgage	Closing mortgage	50.0	Seterus	CITIZENS FINANCI...
17	2837577	3116326	Mortgage	Mortgage	100.0	Applying mortgage	Applying mortgage ...	60.0	Seterus	PNC Bank N
18	2837577	3116309	Mortgage	Mortgage	100.0	Applying mortgage	Closing mortgage	50.0	Seterus	WELLS FARGO C...

S1 S2 S3

Sütun 1: Product

Benzerlik Oranı: 100

Sütun 2: Issue

Benzerlik Oranı: 50

Thread Sayısı: 30

İstenen Sütun: Company

Veri Sayısı: 10

Getir

Toplam Geçen Süre: 11982 ms

#	Thread ID	Thread Adı	Çalışma Süresi
1	84	ThresholdThread0	11598 ms
2	85	ThresholdThread1	11653 ms
3	86	ThresholdThread2	11559 ms
4	87	ThresholdThread3	11852 ms
5	88	ThresholdThread4	11290 ms
6	89	ThresholdThread5	11453 ms
7	90	ThresholdThread6	11966 ms
8	91	ThresholdThread7	11070 ms
9	92	ThresholdThread8	11931 ms
10	93	ThresholdThread9	11915 ms
11	94	ThresholdThread10	11017 ms
12	95	ThresholdThread11	11674 ms
13	96	ThresholdThread12	11131 ms
14	97	ThresholdThread13	10999 ms
15	98	ThresholdThread14	10593 ms

Şekil 2: Senaryo 2

S1S2S3

Seçilen Sütun:

Issue

Complaint ID:

2280669

Benzerlik Oranı:

10

Thread Sayısı:

20

Getir

Toplam Geçen Süre: 1365 ms

#	Thread ID	Thread Adı	Çalışma Süresi
1	196	ThresholdThread0	1046 ms
2	197	ThresholdThread1	1007 ms
3	198	ThresholdThread2	1258 ms
4	199	ThresholdThread3	1035 ms
5	200	ThresholdThread4	1157 ms
6	201	ThresholdThread5	1163 ms
7	202	ThresholdThread6	1063 ms
8	203	ThresholdThread7	990 ms
9	204	ThresholdThread8	1160 ms
10	205	ThresholdThread9	1160 ms
11	206	ThresholdThread10	936 ms
12	207	ThresholdThread11	1165 ms
13	208	ThresholdThread12	1169 ms
14	209	ThresholdThread13	1046 ms
15	210	ThresholdThread14	1170 ms

#	Kayıt-1	Issue	Kayıt-2	Issue	Benzerlik Oranı
1	2280669	Making receiving payments sending ...	400965	Loan servicing payments escrow acc...	20.0
2	2280669	Making receiving payments sending ...	403059	Loan servicing payments escrow acc...	20.0
3	2280669	Making receiving payments sending ...	3172629	Problem making payments	40.0
4	2280669	Making receiving payments sending ...	401142	Loan servicing payments escrow acc...	20.0
5	2280669	Making receiving payments sending ...	399552	Loan servicing payments escrow acc...	20.0
6	2280669	Making receiving payments sending ...	2281622	Making receiving payments sending ...	100.0
7	2280669	Making receiving payments sending ...	2280181	Money promised	20.0
8	2280669	Making receiving payments sending ...	2280544	Loan servicing payments escrow acc...	20.0
9	2280669	Making receiving payments sending ...	2279987	Loan servicing payments escrow acc...	20.0
10	2280669	Making receiving payments sending ...	2280669	Making receiving payments sending ...	100.0
11	2280669	Making receiving payments sending ...	2280767	Loan servicing payments escrow acc...	20.0
12	2280669	Making receiving payments sending ...	2281432	Making receiving payments sending ...	100.0
13	2280669	Making receiving payments sending ...	3172891	Problem making payments	40.0
14	2280669	Making receiving payments sending ...	2947483	Problem making payments	40.0
15	2280669	Making receiving payments sending ...	2947018	Problem making payments	40.0
16	2280669	Making receiving payments sending ...	401744	Loan servicing payments escrow acc...	20.0
17	2280669	Making receiving payments sending ...	401633	Loan servicing payments escrow acc...	20.0
18	2280669	Making receiving payments sending ...	401297	Loan servicing payments escrow acc...	20.0

Şekil 3: Senaryo 3