

Kocaeli Üniversitesi

Bilgisayar Mühendisliği Bölümü

Yazılım Laboratuvarı II

Akademik Makale Öneri Sistemi

Oğuzhan Çelik-190202105

Can Güneri-190202094

Projenin Özeti

Akademik Makale Öneri Sistemi projesi Visual Studio Code ve MongoDB kullanılarak hazırlanmıştır. Proje, öğrenci ve akademisyenlerin makale seçimlerini kişiselleştirmeyi ve kolaylaştırmayı amaçlayan bir Akademik Makale Öneri Sistemi'ni kapsıyor. Flask framework kullanılarak Python'da geliştirilen bu sistem, kullanıcıların ilgi alanlarına ve geçmiş okuma alışkanlıklarına göre özelleştirilmiş makale önerileri sunar. Veri tabanı olarak MongoDB kullanılıyor ve kullanıcı bilgileri ile okuma geçmişleri bu veri tabanında saklanıyor.

Öneri motoru, TfIdfVectorizer ile metin işleme ve kosinüs benzerliği hesaplaması kullanarak çalışır. Bu süreç, kullanıcının ilgi alanlarına, arama sorgularına ve okuma geçmişine dayanarak en alakalı makaleleri seçer. Sistem, kullanıcıya en uygun on makaleyi önermek için gelişmiş algoritmalarından yararlanır.

Sistem ayrıca kullanıcıların kaydolmasını, giriş yapmasını, profil bilgilerini güncellemesini ve önerilen makaleleri incelemesini sağlayan çeşitli web sayfaları içerir. Kullanıcı yönetimi, Flask ve Werkzeug ile güvenli bir şekilde yönetilirken, kullanıcı bilgileri MongoDB'de şifreli olarak saklanır.

Proje, makale önerilerinin doğruluğunu artırmak amacıyla kullanıcı etkileşimlerinden ve geri bildirimlerden yararlanarak sürekli olarak iyileştirilmeye açıktır. Bu dinamik yapı, akademik araştırma sürecini daha verimli ve etkili hale getirerek öğrenci ve araştırmacıların zamanlarını daha iyi kullanmalarını sağlar.

1. GİRİŞ

Günümüzde bilimsel araştırmalar, hızla artan bilgi birikimi ve yayınlanan makalelerin çeşitliliği ile sürekli genişlemektedir. Bu durum, öğrencilerin ve araştırmacıların ilgili literatürü takip etmelerini zorlaştırmakta ve zaman yönetimini önemli bir sorun haline getirmektedir. Etkili bir literatür taraması yapabilmek ve alandaki güncel gelişmelerden haberdar olabilmek için kişiselleştirilmiş öneri sistemlerine olan ihtiyaç gün geçtikçe artmaktadır.

Bu çalışma, akademik makale önerilerini kişiselleştiren bir sistem olan "Akademik Makale Öneri Sistemi"ni sunmaktadır. Sistem, kullanıcıların ilgi alanlarına ve okuma alışkanlıklarına göre uygun makaleleri önermek için gelişmiş metin işleme teknikleri ve makine öğrenimi algoritmalarını kullanmaktadır. Projenin temel amacı, kullanıcıların araştırma sürecini daha verimli hale getirerek, onların bilgiye hızlı ve etkili bir şekilde ulaşmalarını sağlamaktır.

Sistem, Python programlama dili kullanılarak Flask web framework üzerinde geliştirilmiştir ve MongoDB, veri saklama ve kullanıcı yönetimi için tercih edilmiştir. Metin ön işleme için NLTK kütüphanesinden yararlanılmış, TfIdfVectorizer ile metinler vektörel forma dönüştürülmüş ve makaleler arası benzerlik, kosinüs benzerliği metodu ile hesaplanmıştır.

2. TEMEL BİLGİLER

A. Veri İşleme

Projede Inspec ve Krapivin2009 veri setleri kullanılmıştır. Veriler, Python pandas ile işlenerek, NLTK kütüphanesi ile metin ön işleme uygulanmıştır. Bu süreç, metinlerin temizlenmesi, küçük harfe dönüştürülmesi ve kelime köklerinin çıkarılmasını içerir.

B. Metin Vektörleştirme ve Benzerlik Hesaplama

Metinler, TfidfVectorizer ile vektörel forma dönüştürülür ve makaleler arası benzerlikler, kosinüs benzerliği ile hesaplanır. Bu yöntem, kullanıcıya uygun makale önerileri yapmak için kullanılır.

C. Kullanıcı Veri Yönetimi

Kullanıcı bilgileri, MongoDB'de güvenli bir şekilde saklanır ve işlenir. Kullanıcıların kayıt, giriş ve profil güncellemeleri bu veri tabanı üzerinden yönetilir. Şifreler, güvenlik için hash'lenir.

D. Kosinüs Benzerliği ile Makale Tavsiyesi

Oluşturulan TF-IDF matrisi üzerinden, makaleler arası kosinüs benzerliği hesaplanarak benzerlik skorları elde edilir. Kullanıcının ilgi alanlarına ve daha önce okuduğu makalelere göre özelleştirilmiş öneriler sunulur. Bu süreçte kullanıcı profiline göre ağırlıklandırılmış sorgu vektörleri oluşturulur ve benzer makaleler dinamik olarak önerilir.

E. Kullanıcı Etkileşimi ve Geri Bildirim

Kullanıcılar, sistem üzerinde kayıt olabilir, giriş yapabilir ve ilgi alanlarını güncelleyebilir. Kullanıcıların etkileşimleri, sistemde saklanır ve öneri algoritmasının sürekli olarak iyileştirilmesi için kullanılır. Kullanıcıların okudukları ve tercih ettikleri makaleler, gelecekteki önerileri şekillendirmek için önemli bir veri kaynağıdır.

3. YÖNTEM

A. Veri Toplama

Proje, Inspec ve Krapivin2009 veri setlerinden toplanan akademik makaleleri kullanmaktadır. Bu veri setleri zip dosyaları halinde indirilip Python ile açılarak işlenmiştir. Veri setleri, makale başlıkları ve özetleri gibi temel metin verilerini içermektedir.

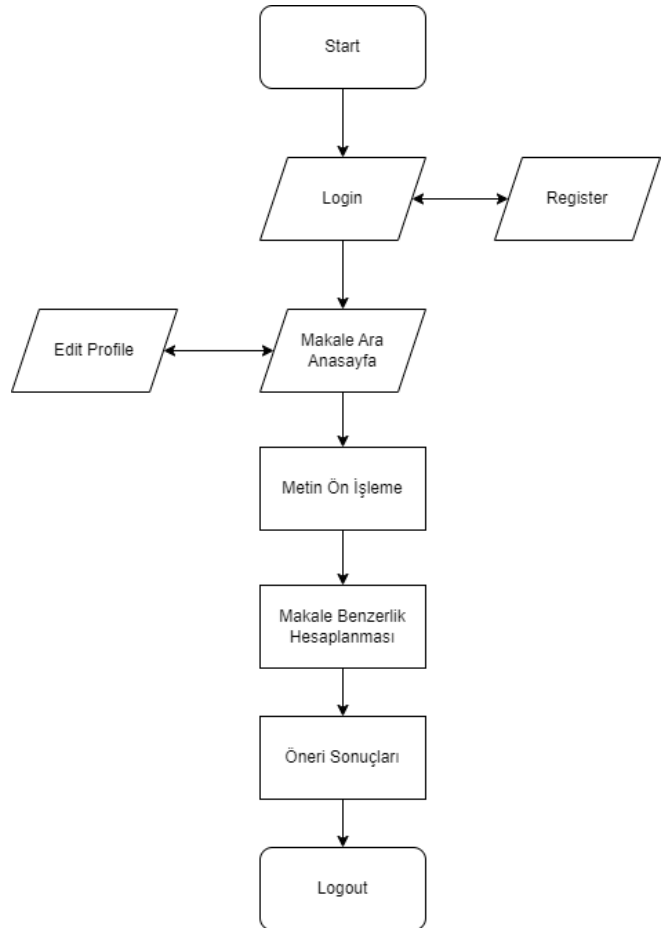
B. Veri Ön İşleme

Veri ön işleme aşamaları, metin verilerinin temizlenmesi ve standardizasyonu işlemlerini içerir. Metinler den noktalama işaretleri kaldırılır, tüm metinler küçük harfe çevrilir ve durak kelimeler çıkarılır. NLTK'nin PorterStemmer modülü ile kelime kökleri elde edilir, böylece metinlerin daha etkili bir şekilde işlenmesi sağlanır.

C. TF-IDF Vektörleştirme

Metin vektörleştirme işlemi, TfidfVectorizer kullanılarak gerçekleştirilir. Bu yöntem, belge içerisindeki her kelimenin önemini, belgenin bütünü içindeki sıklığına göre değerlendirerek vektörel bir formatta sunar. Bu vektörler, makaleler arası benzerlik hesaplamalarında kullanılmak üzere hazırlanır.

Veri Akış Diyagramı;



Veritabanı ER Diyagramı;

Users	
1	id
2	email
3	password
4	interests
5	reading_history
6	name
7	surname

4. SONUÇ

Bu çalışmada, akademik makale öneri sistemi geliştirilmiş ve başarıyla uygulanmıştır. Sistem, kullanıcıların ilgi alanlarına ve okuma alışkanlıklarına göre kişiselleştirilmiş makale önerileri sunarak, bilimsel araştırma süreçlerini kolaylaştırmaktadır. İmplementasyon sürecinde, Inspec ve Krapivin2009 veri setleri kullanılarak zengin bir veri kaynağı oluşturulmuş ve bu veriler üzerinden etkili bir şekilde metin işleme yapılmıştır.

TF-IDF vektörleştirme ve kosinüs benzerliği yöntemleri kullanılarak, makaleler arasındaki ilişkiler başarılı bir şekilde tespit edilmiş ve bu bilgiler kullanıcıya uygun önerilerin yapılmasında kullanılmıştır. Sistemin kullanıcı arayüzü Flask ile geliştirilmiş ve MongoDB ile entegre edilmiş olup, kullanıcılar sisteme kolayca kayıt olabilmekte ve makale önerilerini etkileşimli bir şekilde alabilmektedir.

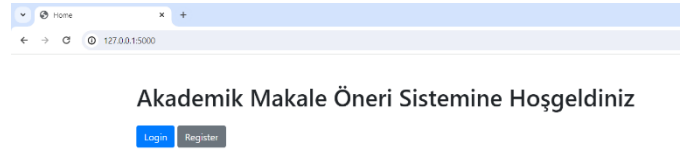
Kullanıcı geri bildirimleri, sistemin performansının ölçülmesi ve sürekli iyileştirilmesi açısından kritik bir rol oynamaktadır. İlk testler ve kullanıcı yorumları, sistemin doğruluğu ve kullanıcı memnuniyeti açısından umut verici sonuçlar ortaya koymuştur. Ancak, daha fazla veri seti ile sistem daha da geliştirilebilir ve öneri algoritmaları daha da rafine edilebilir.

Sonuç olarak, geliştirilen akademik makale öneri sistemi, bilimsel araştırmaların daha verimli ve odaklanmış bir şekilde yürütülmesine olanak sağlamaktadır. Bu sistem, özellikle bilgi aşırı yükünün olduğu akademik dünyada, araştırmacıların ve öğrencilerin zamanlarını daha etkin kullanmalarına yardımcı olmaktadır. Gelecekte, sistem üzerinde yapılacak geliştirmeler ile daha geniş bir kullanıcı kitlesine hitap etmesi ve daha geniş veri setleri ile daha doğru öneriler yapılabilmesi hedeflenmektedir.

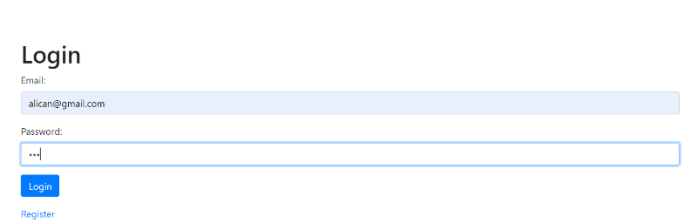
5. DENEYSEL SONUÇLAR

5.1. Uygulama Sonuçlarımız

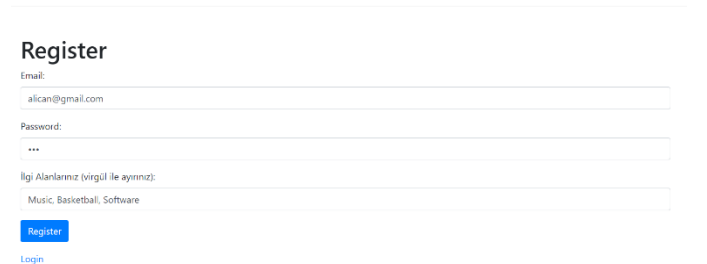
5.1.1. Başlangıç sayfamız;



5.1.2. Login sayfamız;



5.1.3. Register sayfamız;



5.1.4. Dashboard(Anasayfa) sayfamız; (Arama yapılan sayfa)

Dashboard

Hosgeldiniz!

İlgili Alanlarınız : [Music, Basketball, Software]

Makale Ara:

İNGİS

Ara

Edi Profile

Logout

5.1.5. Profil sayfamız(Kullanıcının bilgilerini güncelleyebildiği sayfa);

Profile

Ad:

Alican

Soyad:

Kaplan

Email:

alican@gmail.com

İlgili Alanlarınız:

Technology, Music, Art

Profil Güncelle

Anasayfaya Geri Dön

5.1.6. Arama sonuçları sonucu önerilen makaleler(recommendations.html);

Önerilen Makaleler

The acquisition of out-of-print music
Automatic Segmentation of Acoustic Musical Signals Using Hidden Markov Models
Software processes are software too
Software design
Software architecture
Software Engineering
Software process modeling
Software agents
Software prefetching
Software pipelining

Gerisi Dön

5.2. .CSV Sonuçlarımız

Veri Ön İşlemleri sonucunda Inspec ve Krapivin2009 datasetlerimizin temizlenmiş verilerimizin listelendiği dosyamız(cleaned_combined_articles.csv);

cleaned_combined_articles.csv - Trio Office Calc

Doğa | Dişleme | Görünüm | Ekle | Biçim | Bilgiler | Çalışma Sayfası | Yeri | Araçlar | Pencere | Yardım

Yeni • Aç • Kaydet • PDF • Yazdır | Yazdırma Özelleme | Kes | Kopyala | Yapıştır • Çıktı | Temizle | Geri Al • Tekrarla • Bul ve Değiştir | İmla Denetimi | Satır • Sütun • Sırala | Artan Sıralama

A16206

Function Wizard | Sum | Formula | The MPS Computer System Simulator

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

154

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

204

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

254

255

256

257

258

259

260

261

262

263

264

265

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

327

328

329

330

331

332

333

334

335

336

337

338

339

340

341

342

343

344

345

346

347

348

349

350

351

352

353

354

355

356

357

358

359

360

361

362

363

364

365

366

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405

406

407

408

409

410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

427

428

429

430

431

432

433

434

435

436

437

438

439

440

441

442

443

444

445

446

447

448

449

450

451

452

453

454

455

456

457

458

459

460

461

462

463

464

465

466

467

468

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

875

876

877

878

879

880

881

882

883

884

885

886

887

888

889

890

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

917

918

919

920

921

922

923

924

925

926

927

928

929

930

931

932

933

934

935

936

937

938

939

940

941

942

943

944

945

946

947

948

949

950

951

952

953

954

955

956

957

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

1000

5.3. MongoDB Database sonuçlarımız;

My Queries Users X +

localhost:27017 > AkademikMakaleOnerisiSistem > Users

Documents 4 Aggregations Schema Indexes 1 Validation

Type a query: (field: 'value') or [Generate query](#)

1 - 9 of 9

name: "Tugay"
surname: "Kaplan"
• email: Array (10)
• password: "pbkdf2:sha256:66066656d479f80f2e9a948f96e356e7c4c52d8f687eca7ed55b..."
• interests: Array (1)
• reading_history: Array (50)

• id: ObjectId("664c7f33a0d0eccc1d0b6c")
• email: "tugay@gmail.com"
• password: "pbkdf2:sha256:66066656d479f80f2e9a948f96e356e7c4c52d8f687eca7ed55b..."
• interests: Array (1)
• reading_history: Array (40)

• id: ObjectId("664c7f33a0d0eccc1d0b6c")
• email: "tugay@gmail.com"
• password: "pbkdf2:sha256:66066656d479f80f2e9a948f96e356e7c4c52d8f687eca7ed55b..."
• interests: Array (1)
• reading_history: Array (40)

6. KAYNAKÇA

<https://www.youtube.com/watch?v=wNeykqfSYZA>

<https://dergipark.org.tr/tr/download/article-file/1594872>

<https://medium.com/web-mining-is688-spring-2021/article-recommendation-system-using-python-8b0fec6e6de8>

<https://www.kaggle.com/datasets/jainilcoder/article-recommendation-system>

<https://github.com/boudinfl/krapivin-2009-pre>

<https://paperswithcode.com/dataset/krapivin>

<https://paperswithcode.com/dataset/inspec>

<https://medium.com/@ilkbaharnaz/data-preprocessing-veri-ön-işleme-85236484f913>

<https://medium.com/@cmukesh8688/tf-idf-vectorizer-scikit-learn-dbc0244a911a>