

RSS ile Kişiyeye Özel Haber Sistemini Büyük Veri Araçları ile Gerçekleştirme

Sümeyye Şahin¹, Ahmet Anıl Müngen², Gökhan Yılmaz³

1-Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim, Sivas

2-Fırat Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Elazığ

3-Bilecik Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Bilecik

sumeyyekayaokay@gmail.com, amungen@gmail.com, gokhanyilmazceng@gmail.com

Özet: Türkiye’de ve Dünya’da milyonlarca insan bilgisayarları ve mobil cihazlarından internete erişmektedir. İnterneti kullanan kullanıcıların yaklaşık %70’i haber siteleri ve araçlarından günlük haber okumaktadır[1]. Çalışmamız onlarca farklı kaynaktan toplanan haberlerin aynı olanlarını birleştirip, kullanıcının daha önce okuduğu haberlere göre ilgi alanı tespit ederek kullanıcıya özel haberler sunmaktadır. Her gün üretilen milyonlarca haberin sağlıklı şekilde depolanması ve kullanılması için Büyük Veri araçları kullanarak oluşturulan sistem, görme engelli kullanıcılara da ses algılama ve sesli haber okuma özellikleri ile hizmet vermektedir. Veri Madenciliği algoritmaları ile kullanıcıya özel haber servisi sunarak kullanıcı deneyimlerinin artırılması amaçlanmıştır.

Anahtar Kelimeler: RSS, Haber Okuma, Büyük Veri, Kişiyeye Özel Haber, Görme Engelliler için Haber

Personalized News System With Rss Via Big Data Tools

Abstract : Everyday millions of people access to internet via computer and mobile devices all over the world. Nearly 70 percentage of all internet users check and read daily news everyday. Our project present that gather news from dozens news websites via RSS, determine interest areas of users with user's readed news, and provide specialized news for user. The project uses big data tools to store millions of news and project that has special page for blind people with navigate webpage via speech. Data mining algorithms are used to increase user experience with provide news for interest area of users.

Keywords: RSS, News Reading, Big Data, Personalized News, News for the Visually Impaired

1. Giriş

TÜİK'in "Hanehalkı Bilişim Teknolojileri Kullanım Araştırması, 2015 verilerine göre Türkiye nüfusunun %97'si akıllı bir cep telefonu ile internete bağlanmakta bununla birlikte %65'i de evlerindeki bilgisayarla internete erişebilmektedir.[1] TÜİK'in aynı raporunda İnternete bağlanan insanların %70'inin çevirim-içi haber sitelerini ziyaret ettiği ve haber okuduğu tespit edilmiştir.[1] Avrupa genelinde internete erişim sağlayan hane oranı da %78.5'dir.[2][3] Hem Türkiye'de hem Dünya'da internet kullanımının yarısı mobil cihazlar ile gerçekleştirilmektedir.[4] Amazon firmasının Alexa servisi verilerine göre Türkiye'de en çok ziyaret edilen 50 web sitenin 20'si haber siteleri olmaktadır.[5] Haber siteleri haberleri, haber kaynaklarından veya kendi kaynaklarından alan, özetleyen resim grafik gibi görseller ekleyip kullanıcılara ücretli veya ücretsiz sunan, haberleri değişik kategorilere ayıran web siteleridir.[6]

RSS ise genellikle haber sağlayıcıları, bloglar ve podcastlar tarafından üretilen, yeni eklenen içeriğin kolaylıkla takip edilmesini sağlayan bir içerik besleme yöntemidir.[7] RSS ile haber sitelerinde yayınlanan en güncel haberler alınabilmektedir. Alınan genelde XML veya JSON tipinde olan RSS veriler çeşitli çevirim içi, masaüstü veya mobil okuyucu uygulamaları ile okunulabilmektedir.[8]

Birçok haber sitesi haberleri güncelliği ve sabit kategorilere göre sıralayıp kullanıcıya sunmaktadırlar. Bu sistemler farklı kaynaklardan aldığı aynı haberleri farklı haber olarak sunabilmekte, haberleri herhangi bir şekilde puanlandırma ve içeriği bakımından benzer olan haberleri aynı kategoriye alma gibi bir sınıflandırma yapmamaktadırlar. Bu bağlamda gündemi takip eden kullanıcılar için hem sağlıklı hem de ilgilendikleri haber çeşitlerine kısa sürede erişmelerine ihtiyaç duyulmaktadır.

Çalışma 5 bölümde ele alınmıştır. Genel Dizayn ve mimari: Gerçekleştirilen projenin çalışma mekanizması, Modüller: Kullanılan teknolojilerin sistemdeki yeri ve neler olduğu, Tartışma ve Sonuç bölümlerinden oluşmaktadır.

2. Genel Dizayn, Mimari ve Modüller

Çalışmanın tamamı özgür yazılım teknolojileri ve açık yazılım mimarisi ile geliştirilmiştir. Çalışmayı altı ana kısma

ayırmaktayız. Bunlar; JSF/Arayüz, veri saklama modülü, arama modülü, görme engelli modülü, kinect modülü ve web sitesi gösterimidir.

A. JSF / Arayüz

Web 1.0 teknolojisi ile hazırlanmış siteler kullanıcıya sadece bilgi vermek amacıyla kullanılmaktaydı. Web 2.0 teknolojilerine geçiş ile kullanıcılar sitelerden hem bilgi alabilir hem de bilgi verebilir şekilde interaktif bir etkileşim içinde bulunmuş oldular.[9] Bu interaktifin sağlanması için yapılan çalışmada dinamik web site oluşturmak için java tabanlı ve MVC yapısına uygun web teknolojisi olan JSF kullanılmıştır.[10] JSF, arayüz tasarımı kolaylaştıran bileşen yapısına sahiptir. Kullanıcı kolaylığı sağlaması, tema çeşitliliği, mobil ve ajax desteğine sahip olduğundan JSF in Primefaces kütüphanesinin zengin bileşenlerinden faydalanılmıştır.[11] Primefaces ajax sayesinde sayfa tekrar yüklenmesi yerine değiştirilecek kısmın güncellenmesi sayesinde hızlı bir sunum sağlanarak hazırlanmıştır.[12] Primefaces Mobile ile hibrit bir yapı oluşturarak mobil platform da popüler sayılan Android, Windows Phone ve IOS işletim sistemlerinde çalıştırılabilmektedir.[12]

B. Veri Saklama Modülü

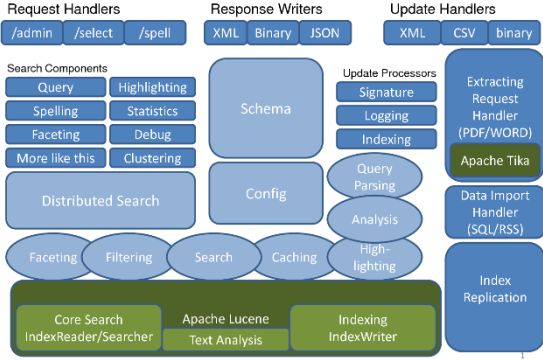
Veri tabanında ilişkisel veri tabanında ilk sıralarda yer alan MySql veri tabanı kullanılmıştır. Platformdan bağımsız, ücretsiz ve kullanımı kolaydır. Kullanıcı listesi ve haberlerin aynı ortamda tutulması ve ilişkisel yapıya sahip olmasıyla kolay erişim sağlamaktadır.

C. Arama Modülü

Haber sitelerinde alınmakta olan haber verileri gittikçe artmaktadır. Sürekli artış gösteren bu veriler üzerinde standart metotlarla arama yapmak zamansal olarak uzun sürmektedir. Bunun yerine Çalışmanın arama kısmında Apache Lucene altyapısına sahip, açık kaynak kodlu, ölçeklenebilir bir tam metin arama motoru olan Apache Solr kullanılmıştır. RestFull API desteği sayesinde birçok programlama dili ile entegre olabilmektedir.[13] Apache Solr içerisinde yer alan Apache Tika kütüphanesi sayesinde Apache Solr'a word, ppt, pdf gibi bir çok veri türü de indexlenebilmektedir[14] Apache solr moreLikeThis desteği mevcuttur. Bu sayede indexlenen dokümanlar arasında benzerlik

kurarak kullanıcıya içerik bakımından birbirine benzeyen dokümanları verebilmektedir.

Lucene/Solr Architecture



Şekil 2.C. Lucene Tabanlı Arama Motorları Mimarisi [15]

D. Görme Engelli Modülü

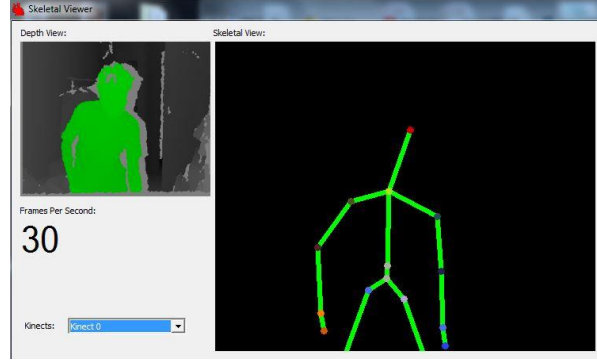
Bunun yanında sistem görme engelliler için menüler ve haberleri ses ile dinleyip bilgisayardaki tuş takımı ile haberler arası geçişi kolayca yapabilecekleri şekilde de hazırlanmıştır. MIT tarafından geliştirilen WAMI Toolkit, JS altyapısı ile sisteme kolayca eklenmekte ve kullanıcıların sesini dinlemek sureti ile kullanıcıların sesini yazıya dönüştürmektedir.[16] Çalışmamızda haber başlığı okunduktan sonra haber numarası da okunmaktadır. Kullanıcı detayını dinlemek istediği haberin numarasını söylemek sureti ile ilgili haberi açabilmektedir. Yine “Geri”, “Bir Sonraki Haber” gibi komutlar ile de web sayfaları arasında ses ile geçiş yapabilmektedir.

E. Kinect Modülü



Şekil 2.e.1. Microsoft Kinect Ön Yüzden Görüntüsü

Kullanıcıların fare, klavye veya diğer fiziksel araçlara dokunmadan bilgisayarı ve/veya siteyi kullanması kullanıcı deneyimini arttıracak, kullanıcının bilgisayardan uzak olduğu zamanlarda da sistemi yönetmesine izin verilecektir. Kullanıcının el hareketlerinin algılanması bir görüntü işleme çalışması olup Microsoft tarafından üretilen XBOX Kinect 360 cihazı el hareketlerini otomatik olarak tespit edip bir API aracılığı ile geliştiricilere hareketlerin matematiksel değerlerini, iskeletteki eklem noktalarını yani elin konumunu sunmaktadır[17].

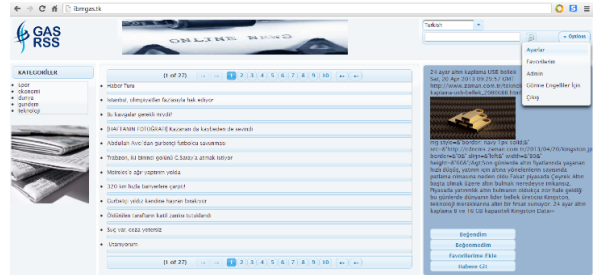


Şekil 2.e.2. Microsoft Kinect İskelet Algılama Görüntüsü

Çalışmamızda Kinect kullanıcıların el hareketlerini takip etmekte ve el hareketlerine göre bir önceki habere geçme, bir sonraki habere geçme ya da haber kategorisi değiştirme gibi işlemler yapılabilmektedir.

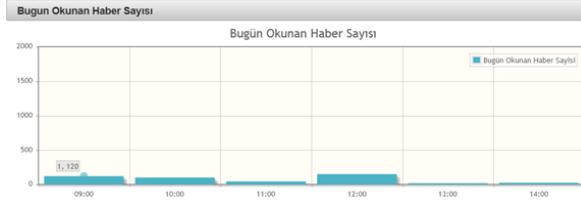
F. Web Sitesinde Gösterim

Web sitesin de RSS ile çekilen verileri kategori şeklinde ve bu kategorilerde de haberler listelenmektedir. Kullanıcının ilgisini çekebilecek haber türleri detaylı şekilde ekrana yansıtılmakta ve okunan haber üzerinde geri bildirimler gönderilmektedir. Haber içeriklerinde arama yapılabilmekte, haber sitesi görme engelliler için ses ve tuş takımlarıyla yönetilebilmekte ve tüm kullanıcılar için kinect desteğiyle haberler arası geçişleri el hareketi ile sağlanabilmektedir.



Şekil 2.f.1. Proje ana sayfası görüntüsü, haber özeti ve menüler.

Okunan haberlerin dönülen geri bildirimlerin istatistikleri tutulmakta ve grafikler ile kullanıcıya yansıtılmaktadır. Bu sayede hangi haberin ön plana çıktığı tespit edilebilmektedir.



Şekil 2.f.2. Proje ana istatistik görüntüsü, gün bazlı haber okuma sayıları.

4. Tartışma

Arayüzde ajax desteği sağlanarak müşteri bekleme süresi ve ağ trafiği azaltılırken mobil cihazlarda veya tarayıcılarda geri tuşunun işlevini kaybetmesine yol açmaktadır.[18] Apache Solr'ın morelikethis desteği ile birbirine benzeyen dosyalara erişim sağlanabilmektedir.[13] Bu sayede benzer dosyaları sunabilmek için ekstra bir algoritma yazmaya gerek duyulmaz. Veri miktarı büyüdükçe veri tabanından arama yapmak yerine Apache Solr üzerinde arama yapmak zaman kaybını azaltır. Ama Apache Solr üzerinde arama yapabilmek için verilerin Apache Solr'a indexlenmesi ekstra bir işleme sebep olmaktadır.[15]

5. Sonuç

Sonuç olarak hem normal kullanıcılara hem de görme engelli kullanıcılara haber kaynakları sesli ve görsel olarak sunulabilmektedir. Kullanıcı profillerinde okudukları haber türlerinden çıkararak önerilerde bulunabilir ve kitle yoğunluğundan yola çıkarak popüler olarak okunan haber içeriklerini sunmaktadır. Haber içeriklerinde arama yaparak seçilen habere benzeyen haberlere de erişim sağlanabilmektedir.

6. Kaynakça

- [1] <http://www.tuik.gov.tr/PreHaberBultenleri.do?id=18660>
- [2] <http://www.internetworldstats.com/stats9.htm>
- [3]Thompson S.H Teoa, , Vivien K.G Limb, Raye Y.C Laia;Intrinsic and extrinsic motivation in Internet usage;Omega Volume 27, Issue 1, February 1999, Pages 25–37
- [4]Ronald E Rice, , James E Katz;Comparing internet and mobile phone usage: digital divides of usage, adoption, and dropouts;Telecommunications Policy Volume 27, Issues 8–9, September–October 2003, Pages 597–623
- [5]<http://www.alexa.com/topsites/countries;1/TR>
- [6]<http://www.pcmag.com/encyclopedia/term/49528/portal>
- [7]<https://tr.wikipedia.org/wiki/RSS>

[8] Benzinger, Brian. "Google Reader Reviewed." *Retrieved from*, (Oct. 7, 2005).

[9] o'Reilly, Tim. *What is web 2.0*. " O'Reilly Media, Inc.", 2009.

[10]Distante, Pedone, Rossi, Canfora, "Model-Driven Development of Web Applications with UWA, MVC and JavaServer Faces",2007 7th International Conference, ICWE 2007 Como, Italy, July 16-20, 2007 Proceedings

[11] Kod'unun Dünyası Primefaces, <http://www.kodumundunyasi.net/java/jsf-ve-primefaces-giris.html>, 09.11.2015

[12] Varaksin, Oleg. *PrimeFaces Cookbook*. Packt Publishing Ltd, 2013.

[13] Smiley, David, et al. *Apache Solr Enterprise Search Server*. Packt Publishing Ltd, 2015.

[14]Mattmann, Chris, and Jukka Zitting. *Tika in action*. Manning Publications Co., 2011.

[15]Kodcu.com Blog, <http://kodcu.com/2013/03/apache-solr-kurulumu-ornek-sorgulama/> 09.11.2015

[16] WAMI | MIT CSAIL, <https://www.csail.mit.edu/research/playground/wami09.11.2015>

[17] Microsoft Kinect, <https://www.microsoft.com/en-us/kinectforwindows/develop/>, 09.11.2015

[18] Alper Çalışmalar, http://people.sabanciuniv.edu/alper/calismalar_ax.html, 09.11.2015