İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

ÇOKLU İÇERİK YÖNETİMİ SİSTEMLERİNDE ORTALAMA EĞİMLİ İŞBİRLİKÇİ SÜZGEÇLEME

YÜKSEK LİSANS TEZİ

Zübeyde Deniz ALAY

Anabilim Dalı : Bilgisayar Mühendisliği

Programı : Bilgisayar Mühendisliği

HAZİRAN 2010

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

ÇOKLU İÇERİK YÖNETİMİ SİSTEMLERİNDE ORTALAMA EĞİMLİ İŞBİRLİKÇİ SÜZGEÇLEME

YÜKSEK LİSANS TEZİ

Zübeyde Deniz ALAY
(504031550)

Tezin Enstitüye Verildiği Tarih : 07 Mayıs 2010
Tezin Savunulduğu Tarih : 14 Haziran 2010

Tez Danışmanı : Prof. Dr. Eşref ADALI (İTÜ)
Diğer Jüri Üyeleri : Doç. Dr. Sabih ATADAN (İTÜ)
Yrd. Doç. Dr. Banu DİRİ (YTÜ)

HAZİRAN 2010

ÖNSÖZ

Tez çalışmam öncesinde ve çalışmam sırasında beni yönlendirdiği, her konuda yol gösterdiği ve yapıcı önerileri ile çalışmamı tamamlamamı sağladığı için danışmanım Sayın Prof. Dr. Eşref ADALI'ya teşekkür ederim.

Her zaman bana destek olan annelerim Müzeyyen ALAY ve Hikmet ÇULCUOĞLU'na ve babalarım Nagip ALAY ve Tayfun ÇULCUOĞLU'na ve çalışmamın en zor zamanlarında desteğini esirgemeyen eşim Çağatay M. ÇULCUOĞLU'na teşekkür ederim.

Haziran 2010

Zübeyde Deniz ALAY

Bilgisayar Mühendisi

İÇİNDEKİLER

Sayfa

ÖNSÖZ ... **iii**
İÇİNDEKİLER ... **v**
KISALTMALAR ... **vii**
ÇİZELGE LİSTESİ ... **ix**
ŞEKİL LİSTESİ ... **xi**
SEMBOL LİSTESİ ... **xiii**
ÖZET ... **xv**
SUMMARY ... **xvii**
1. GİRİŞ ... **1**
2. ÇOKLU SİTE İÇERİK YÖNETİM SİSTEMİ ... **5**
2.1 İçerik Yönetim Sistemi ... 5
2.2 İçerik Yönetimi Sistemlerinin Karşılaştırılmalı Çözümlemesi ... 9
2.3 Çoklu Site İçerik Yönetim Sistemi ... 11
2.3.1 Bileşen nedir? ... 11
2.3.2 Parça nedir? ... 11
2.3.3 Joomla bot nedir? ... 12
2.4 Çoklu Site İçerik Yönetim Sistemi Uygulaması: MANTAR ... 12
2.4.1 MANTAR kurulumu ... 13
2.4.2 Kurulu MANTAR sisteminin ayarlanması ... 16
2.4.3 Sitelerin tanımlanması ... 16
2.4.3.1 Şablonlar 16
2.4.3.2 Parçalar 16
2.4.3.3 Menü öğeleri 16
2.4.3.4 Bölümler 17
2.4.3.5 Kategoriler 17
2.4.3.6 İçerik öğeleri 17
3. ÇEVRİMİÇİ UYGULAMALARDA ÖNERİ SİSTEMLERİ ... **19**
3.1 Öneri Sistemleri ... 19
3.2 İçerik Tabanlı Süzgeçleme Sistemleri ... 23
3.3 İçerik Tabanlı Süzgeçleme Sisteminde En Yakın Komşu Yöntemi ... 25
3.4 İçerik Tabanlı Süzgeçleme Sisteminde Olasılık Modelleme ... 26
3.5 Karma Süzgeçleme Sistemleri ... 27
4. İŞBİRLİKÇİ SÜZGEÇLEME SİSTEMLERİ (İFS) ... **29**
4.1 Eğimli İşbirlikçi Süzgeçleme (EİS) ... 29
4.2 Bellek Tabanlı Süzgeçlemeler ... 30
4.3 Bellek Tabanlı ve Model Tabanlı Tasarımlar ... 30
4.4 Süzgeçleme Tasarımlarında Kullanılan Tablo ... 32
4.5 Kullanıcı Başına Ortalama İşbirlikçi Süzgeçleme Tasarımı ... 33
4.6 Kişiselleştirilmemiş İşbirlikçi Süzgeçleme Tasarımı ... 34
4.7 Pearson Tasarımı ... 35
4.8 Tek Eğimli İşbirlikçi Süzgeçleme Algoritmaları (TEİS) ... 36

4.9 Ortalama Eğimli İşbirlikçi Süzgeçleme Algoritmaları (OEİS) 37
4.10 Süzgeçleme Tasarımlarına Ait Ortalama Hata Hesaplaması 38
5. ORTALAMA EĞİMLİ İŞBRİLİKÇİ SÜZGEÇLEME UYGULAMASI 41
5.1 OEİS Uygulaması İçin Kullanılan Veri Kümesi 41
5.1.1 Kullanıcılar 41
5.1.2 Kitaplar 42
5.1.3 Kitap oyları 43
5.2 OEİS Uygulaması 44
6. SONUÇ VE ÖNERİLER 53
KAYNAKLAR 57
EKLER 61

KISALTMALAR

CMS : Content Management System
EM : Expectation Maximization
GPL : General Public Licence
HCA : Hierarchical Cluster Analysis
IR : Information Retrieval
İYS : İçerik Yönetim Sistemi
KOBİ : Küçük Orta Büyük İşletmeler
MAE : Mean Average Error
MSCMS : Multi Site Content Management System
MSM : Multi Site Management
OEİS : Ortalama Eğimli İşbirlikçi Süzgeçleme
OGK : Olay Güçlendirme Kuvveti
ROI : Return on Investment
RSS : Real Simple Syndication
SEO : Arama Motorları Optimizasyonu
TEİS : Tek Eğimli İşbirlikçi Süzgeçleme

ÇİZELGE LİSTESİ

Sayfa

Çizelge 2.1 : İçerik yönetim sistemleri karşılaştırması 10
Çizelge 3.1 : Öneri sistemlerinin kullanıldığı ticari veya ücretsiz siteler 20
Çizelge 3.2 : İçerik süzgeçleme sistemlerinin kullanıldığı siteler 24
Çizelge 4.1 : Süzgeçleme tasarımlarına ait hata ölçümleri 39

ŞEKİL LİSTESİ

Sayfa

Şekil 2.1 : Birbirinden farklı içerik yönetimi sistemleri ile kurulmuş farklı siteler 7
Şekil 2.2 : Çoklu site içerik yönetim sistemi: MANTAR..... 13
Şekil 3.1 : Amazon.com da işbirlikçi süzgeçleme yönteminin uygulanışı..... 21
Şekil 3.2 : Google Web Master Tools da içerik süzgeçleme yönteminin uygulanışı..... 22
Şekil 3.3 : Facebook'ta içerik süzgeçleme algoritmasının uygulanışı..... 24
Şekil 3.4 : 5 adımdan oluşan içerik tabanlı öneri sistemi şeması 26
Şekil 3.5 : İçerik ve işbirlikçi süzgeçlerinin beraber uygulandığı çalışma 28
Şekil 4.1 : Süzgeçleme tasarımlarında kullanılan oylama tablosu 32
Şekil 4.2 : Kullanıcı başına ortalama işbirlikçi süzgeçleme kestirim tablosu 33
Şekil 4.3 : Kişiselleştirilmemiş işbirlikçi süzgeçleme tasarımı kestirim tablosu..... 34
Şekil 4.4 : Pearson sonucunda kestirim tablosu 36
Şekil 5.1 : BX-Crossing veri kümesindeki kullanıcılar tablosu 42
Şekil 5.2 : BX-Crossing veri kümesindeki kitaplar tablosu 43
Şekil 5.3 : BX-Crossing veri kümesindeki oylar tablosu 44
Şekil 5.4 : OEİS uygulamasındaki oylama tablosu 45
Şekil 5.5 : Kullanıcılar arasındaki eksik bilgi deseni 46
Şekil 5.6 : www.bizede.com/oyver.php sayfasındaki ürün oylama ara yüzü..... 47
Şekil 5.7 : OEİS uygulamasındaki güncelleme tablosu..... 48
Şekil 5.8 : Genel ürün tavsiyeler listesi 49
Şekil 5.9 : Ortalama eğimli süzgeçleme oy kestirimi..... 50
Şekil 5.10 : OEİS uygulaması akış şeması 51

SEMBOL LİSTESİ

d	: Kitap
z	: Demetlemeye karşılık olarak gizli bir değişken
s	: Alan, başlık veya özet
w	: Bir kelime
$n_{s,d}(w)$	: d kitabında s alanında ki w kelimesinin sayısı
$p(z)$	: Olasılık fonksiyonu
$p_s(w \mid z)$	: Koşullu Olasılık fonksiyonu
b	: Bir sabit
x	: Değişken oy değerleri
$f(x)$	: Bir öğenin oyundan diğer bir öğenin oyunu öngören işlev
U	: Değerlendirme dizisi, bir üye tarafından verilen oylar dizisi
i	: Öğe
u_i	: u dizininde i öğesinin oy değeri
$S(u)$	: u dizinin de oy verilen öğeler alt kümesi
χ	: Bütün oylamaları içeren çalışma kümesi
card(S)	: S kümesindeki öğelerin sayısı
$\overline{u}$	: u dizindeki oyların ortalama değeri
$S_i(X)$	: $S(u)$ alt kümesinde olan i öğelerinin sabit oy değer kümesi
$Corr(u,v)$	: İlişki (Korelasyon)
$P(u)$	: Pearson kestirim değeri
v_i	: Değerlendirme dizini
w_i	: Değerlendirme dizini
$dev_{j,i}$	: i öğesinin j öğesinden ortalama sapması
R_j	: Tüm öğelerin kümesidir
$P^{S1}(u)_j$	: Tek eğimli işbirlikçi tasarım için kestirim değeri
$P^{wS1}(u)_j$	: Ortalama eğilimli işbirlikçi süzgeçleme kestirim değeri

ÇOKLU İÇERİK YÖNETİMİ SİSTEMLERİNDE ORTALAMA EĞİMLİ İŞBİRLİKÇİ SÜZGEÇLEME

ÖZET

Verilerin özelliklerine göre öbeklere ayrılması, süzgeçlenmesi ve verilerin kullanılarak kestirim yapılması çevirim içi sistem, içerik yönetimi, veri madenciliği, görüntü işleme, finansal çözümleme gibi pek çok birbirinden değişik uygulama için önemli bir problemdir. Çevrimiçi uygulamalarda işbirlikçi süzgeçleme, kullanıcıların vermiş olduğu oylara dayalı olarak yeni oy veya ürün kestiriminde kullanılır. Önceden verilen ortak oyların benzerliği, oy öbekleri arasındaki değer farkları hesaplanması ve hata ölçümleri bulunarak uygun oy kestirimleri gerçekleştirilir.

Bu çalışmada, çoklu site içerik yönetim sistemlerinde kullanıcıların çeşitli ürünlere verebilecekleri oylara dayanarak kestirim yapabilme probleminin çözümü için Ortalama Eğimli İşbirlikçi Süzgeçleme (OEİS) kullanarak gerçeklenmiştir. Çoklu site içerik yönetim sistemi geliştirilmesi karşılaştırmalı veri çözümleme sonucu seçilen uygun bir içerik yönetimi sistemi üzerine oluşturulmuştur. Sistemde birden fazla alan adı üzerine tek bir veri tabanı üzerinden erişilebilecek çok kullanıcılı bir yapı gerçeklenmiştir.

Gerçekleştirilen Ortalama Eğimli İşbirlikçi Süzgeçleme algoritması yeni oluşturulan ve var olan yöntemler kullanılarak tasarlanmıştır. Yapılan sınamalarda çeşitli algoritmaların oy öğeleri arasındaki ilişkileri daha doğru tanımladığı ve büyük veritabanlarının niteliğine bağlı olarak daha iyi başarım sağladığı gözlemlenmiştir. Aynı zamanda kullanıcı başına Ortalama İşbirlikçi, Kişiselleştirilmemiş, Pearson ve Ortalama Eğimli İşbirlikçi Süzgeçleme yöntemlerinin öngörü kabiliyetleri karşılaştırılmıştır. Çoklu site içerik yönetim sistemi ve bu sistem üzerine uygulanabilecek olan Ortalama Eğimli İşbirlikçi Süzgeçleme yöntemi kurulum maliyetinin diğer içerik yönetim sistemlerine üzerine kurulabilecek süzgeçleme yöntemine göre çok düşük olması da algoritmanın yaralarından biridir.

PHP ve mySQL kullanılarak yapılan uygulamada süzgeçlenecek oylar ve kullanıcı belgileri düzenlenerek iki boyutlu kullanıcı oy süzgeçlemesi oluşturulmuştur. Bu işlemde amaç, kayıtlı kullanıcı çoklu sitelerden herhangi birisine bağlandığında kullanıcının hangi ürünü en çok isteyebileceğini kestirerek, kullanıcıya daha önceden oy vermediği ürünü göstermektir. Çoklu site içerik yönetim sistemi oluşturulması için Joomla içerik yönetim sistemi seçilmiştir. Bu sistemin gelişmiş içerik yönetim ve PHP dili ile birlikte kullanılabilmesini sağlayan kod yapısı sistemin çoklu hale getirilmesini sağlar.

WEIGHTED SLOPE ONE COLLABORATIVE FILTERING IN MULTI SITE CONTENT MANAGEMENT SYSTEM

SUMMARY

Grouping of data, filtering of data and prediction based on this data is a known common problem for several distinct applications such as online systems, content management, database mining, image processing, and financial analysis. In Online applications, collaborative filtering can be used for rating or product prediction based on the user's ratings. Through calculation and analysis of the similarity and distance of the past ratings, rating differences and mean and absolute errors predictions can be achieved.

In this study, the rating prediction problem based on the product votes or ratings on a multi-site is solved by weighted slope one filtering implementation. The design and implementation of multi site content managent system is decided by comparison of available systems and information analysis. In multi-site content management setup, the content management system works over one database managing multiple domain names.

The realized weighted slope one filtering algorithm is implemented by reviewing existing and newly designed algorithms. In the tests conducted, it is observed that this methodology gives better results in defining the relationship among rating pairs and performs better in bigger databases. In the analysis, prediction capabilities, error rates of Pearson, collaborative, non-personalized filtering are compared. The advantage of implementing weighted slope one filtering on an online site is that it is less-costly in terms of time and effort and provides acceptable performance vs. other solutions.

The implemenation of weighted slope one filtering is achieved by using PHP, MySQL, and two-dimentional data. The purpose is to have the ability to predict what a user of any of the sites would like as a product and accordingly show the result to the user to improve accuracy. Also, to design multi-site application Joomla CMS is used with PHP coding.

1. GİRİŞ

Günümüzde internet artan bir hızla ve büyüklükte hayatımızın parçası olmaya devam ediyor. Birçok kişi aktif olarak bilgi alma, alışveriş, bilgi paylaşımı ya da iş için internet kullanıyor. Her geçen gün internet kullanıcılarının, web sitesi sahiplerinin ve geliştiricilerinin sayısı katlanarak artıyor.

İnternet kullanıcılarını doğru içerik, ürün sitelerine ve web ana kapılarında uygun ürünlere yönlendirme, işletmeler için önemli bir sorundur ve başarı için çözülmesi gereken bir konudur. KOBİ'ler için bu tarz yapı ve düzeneklerin uygulanabildiği veya uygulanabileceği uygun siteler iyi yatırım getirisi için çok daha önemlidir.

KOBİ'ler ve büyük şirketler de bu gelişmenin farkında ve son sürat yeni öncellikleri ve bilgisayar teknolojilerindeki yeni fırsatları kullanarak web siteleri oluşturmakta veya var olan sitelerini geliştirmektedir. Bilgisayar teknolojilerine ve internet evrimine baktığımızda da insanların internet üzerinden web sitelerine bağlanarak şirketler, kişiler, ürünler hakkında bilgi aradıklarına, bilgi paylaşımında bulunduklarını gözlemliyoruz. Her web sitesi mevcut müşterisine hizmet sunmayabilir ancak potansiyel müşterileriler ve geniş müşteri kitlesi ile iletişim kurmak için siteler kullanılabilir ve genişletilebilir.

İşletmeciler rekabetçi olabilmek için, kullanıcılara doğru ve ilginç ürünleri daha hızlı bulmasında yardımcı olmalıdırlar. Aynı zamanda, kullanıcıları kendi seçeneklerine göre ve kişiselleştirilmiş bilgiler ışığında, önerilen içeriklerin sunulduğu web sitelerine yönlendirmeleri daha da önem kazanmaktadır. Bundan dolayı işletmeler otomatik öneri sistemlerini mümkün hale getirebilecek şekilde kendi kullanıcıları ile ilgili çok büyük verileri bir araya getirirler. Öneri sistemleri dinamik bağlantılar oluşturarak, bireysel kullanıcılar için web sitelerinin kullanılırlığını arttırır. Bu konuya daha geniş açıdan bakarsak modern web'in en değerli özelliğinden biri, çok büyük ölçüdeki veri içerisinde arama yapabilmesidir ve bu araştırmayı destekleyen herhangi bir süreç kullanıcılar için değerlidir.

Süzgeçlenmiş dinamik bağlantılarla çoklu içerik sitesi kullanıcısı herhangi bir sitede kendisine uygun ilginç öğelere daha kolay ulaşabilir ve böylece sistem ve kullanıcı arasındaki etkileşim geliştirilmiş olur. Çoklu site içerik yönetim sistemleri, pazarlama kampanyaları, ürün tanıtımları, daha çok kullanıcıya ulaşma çabaları için tam teşekküllü birden fazla web sitesini yönetme ihtiyacında olan kuruluşlar için oluşturulabilecek bir sistemdir.

İki farklı öneri sistemi bulunmaktadır: Kullanıcı verilerine bağlı olan öneri sistemleri işbirlikçi süzgeçleme sistemleri, önerilecek verilerin özelliklerine bağlı olan öneri sistemleri ise içerik tabanlı süzgeçleme sistemleridir. İşbirlikçi süzgeçlemedeki temel kural kullanıcıların bildikleri ürün veya özelliklere oy vererek kendileri ile ilgili bazı bilgileri paylaşmaları gerekmektedir ki böylece daha doğru öneri yapılabilsin. İçerik tabanlı öneri sistemleri ise nesnelerle daha iyi çalışma eğilimi gösterirler ki bu nesnelerin içerikleri metin gibi işlem görebilir [1]. Film, kitap gibi diğer nesnelerin içeriklerine çevrimiçi ulaşmak her zaman kolay olmayabilir. İçeriklere kolaylıkla ulaşabilsek bile içerik süzgeçleme yöntemi ile çok da doğru sonuç elde edilemeyebilir. Aynı zamanda, içerik süzgeçleme yöntemi bazen kullanıcı istediği ürün veya servis hakkında fazla bilgi sahibi değilse zor olabilir. Bazen kullanıcılar internette gezinirken özel olarak aradıkları bir ürün veya servis olmadan gezinirler ve de web sitesinden değişik önerilerde bulunmasını beklerler. Böyle durumlarda, kullanıcıdan açık veya kapalı olarak oy değerleri alınabilir, bu durumda da işbirlikçi süzgeçleme tercih edilir.

Ancak, bu durumda da karşı karşıya kaldığımız zorluklardan biri kullanıcıların az sayıda nesneye oy vermeleridir. Bu da seyrek ve az miktarda veri ile çalışmamıza neden olmaktadır [2]. Birçok veriyi geri getirme veya düzeltme sürecinde yazılım büyük orandaki doğru verilerden ve özel sorgulamalardan eşleşenleri getirmeye çalışır. Buna karşılık, işbirlikçi süzgeçleme az verinin üstesinden gelmek zorundadır ve bu veriler genelde kesin ve doğru olmayan verilerdir.

Bu çalışmada, farklı öneri süzgeçleme sistemleri karşılaştırılarak kullanıcı tercihleri ve öğeler arasındaki ilişkiye dayalı sorunlar incelenmekte ve çoklu içerik sitesi yönetim sistemleri üzerindeki bağlantıların dinamik bağlantı listelerinin en uygun şekilde oluşturulması sorunu çözülmektedir [3].

İlk olarak bu tezde öneri sistemlerinin karşılaştırılmış ve ardından çoklu site içerik yönetimi sistemine çevrilen uygun bir öneri sistemi oluşturulmuştur. Bu yeni sistem özellikleri açısından uygulayıcı şirketlere kritik çözümler sağlamakta ve bilgisayar mühendisliği sayesinde uygulanabilir çözümleri farklılaştırmaktadır. Öneri sistemleri ve çoklu site içerik sistemlerinin birlikte uygulanması web sitesi kullanıcılarının giderek az zaman harcayarak ve daha kolay aradıkları bilgilere ulaşmasını sağlar.

2. ÇOKLU SİTE İÇERİK YÖNETİM SİSTEMİ

2.1 İçerik Yönetim Sistemi

İçerik Yönetim Sistemi (İYS) işbirliği ortamı içinde iş akışını yönetmek için kullanılan süreçler toplamıdır. İYS herhangi bir biçimdeki bilgi kaynaklarını yaratma, toplama, bir araya getirme, sınıflandırma ve yapılandırma süreleri olarak tanımlanabilir. Böylece bilgi istenilen herhangi bir yöntem ile saklanmış, geri getirilmiş, yayınlanmış, güncellenmiş ve tekrardan tasarlanmış olur [4].

İYS de içerik, İYS'nin yaratmada, düzenlemede, kayıt etmede ve yayınlamada yardımcı olduğu yazı, görüntü, medya gibi farklı dosya tiplerinde olabilir. İYS kullanmanın yararlarından biri de, herhangi bir web sitesi yaratırken iyi bir kodlama bilgisine sahip olmak gerekmez. İYS deneyimli kodlayıcı ile hafta sonları kişisel sayfa yazan kişi arasında bir köprü oluşturur. Böylece her iki kişi farklı teknoloji bilgisine sahip oldukları halde kendi web sitelerinde kontrolleri tek başına sürdürebilirler. Web 2.0 teknolojileri geldikten sonra, İYS, kişisel sayfalar, Wikis ve RSS gibi kullanılması geleneksel araçlardan daha kolay ve esnek olan bir çok yeni araç oluşmuştur. Joomla, Drupal, Bitweaver, Mambo, Pligg, Plone, Post Nuke gibi birçok İYS vardır.

Bugünkü rekabetçi ortam KOBİ'lerin maliyetlerini kısmalarına, süreçlerini geliştirmelerine, sürekli verimli yönetmelere ve büyümeyi dağıtmalarına neden olmaktadır. Bu nedenle, büyüme, kurumsal iletişim çözümleri bulmayı, toplu ve hareketli gerçeklemeyi, kalite sistemini ve de iyi bir güvenlik yapısını gerektirir. Bu yüzden KOBİ'ler açık kaynak kodlu İYS'lerini tercih edeceklerdir [5].

2007 yılında ilk kez, internet reklam geliri, Türkiye'deki dergi gelirini geçti. KOBİ'ler geliştirilmiş ve reklamı yapılmış web sitesinin en ucuz ve en etkili iletişim, reklam ve pazarlama aracı olduğunu fark ettikten sonra kendi web sitelerine sahip olmaya başladılar.

Buna ek olarak, Avrupa Birliği kurallarına uymak ve KOBİ'leri desteklemek amacı ile Türkiye Ticaret ve Sanayi Bakanlığı 2003 de "KOBİ Stratejisi ve İş Planı" çıkardı [6]. Bu iş planı aşağıdaki stratejik iş alanlarını içermektedir:

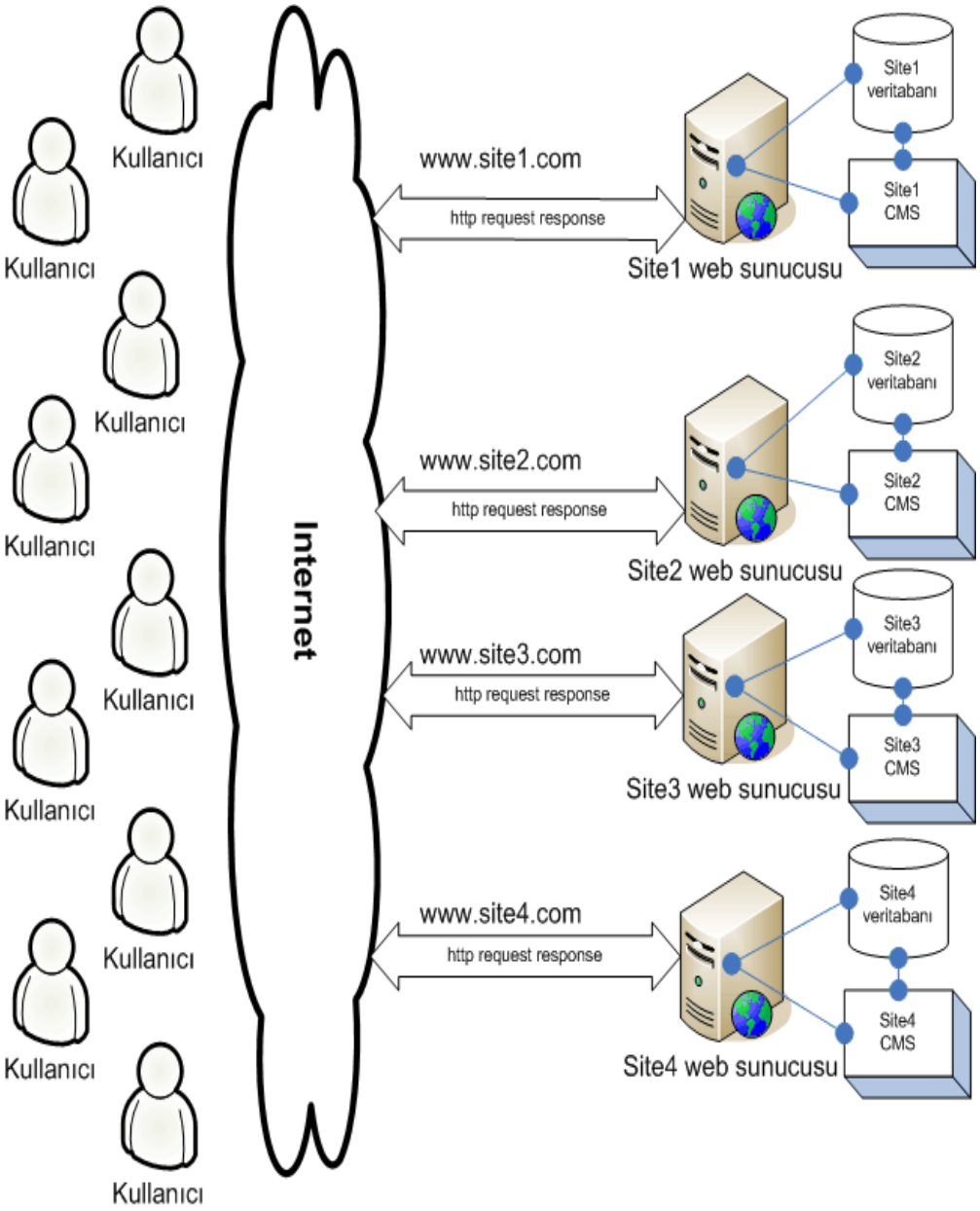
- İnternet erişimi ve internet ile alışma fırsatlarını geliştirmek
- İnternet pazarında yer almak
- KOBİ'lerin teknolojisini ve yenilikçi kapasitesini geliştirmek
- İnternetin en iyi uygulamaları hakkında bilgi sahibi olmak
- Üst düzey KOBİ'lere desteğin arttırılması
- Ulusal ve uluslararası internet ortamlarda KOBİ'lerin etkili olarak temsil edilmesi

Bu stratejik iş planını uyarlamak ve de uygulamak için ve aynı zamanda pazar payından büyük pay almak için, işletmelerin birden fazla web sitesi üzerinde çalışmaları gerekmektedir. Birden fazla web sitesi olan işletmeler aşağıdaki özelliklere sahip birçok müşteriye ulaşabilirler:

- Farklı kültür (bölgesel kültür)
- Farklı dil
- Farklı gelir düzeyleri
- Farklı türdeki ürünlere ilgisi olan

Eğer işletmeler sadece tek bir müşteri belgisine yönelik çalışırlarsa kendilerinin farklı web siteleri aracılığı ile ulaşabilecekleri farklı müşterilere sahip olmalarını kısıtlamış olurlar. Çoklu İYS ile işletmeler müşterilerinin beklentilerine karşılık verebilme olanağına sahip olabilirler. Aynı zamanda birden fazla web sitesine sahip olmanın diğer bir yararı ise, web sitelerinin birbiri ile bağlantılı olmasından dolayı kullanıcılar o işletmeye ait farklı iş ürünlerine yönlendirilebilir.

Son kullanıcıların kullanım alışkanlıklarına belirlemek ve kısa bir zaman içerisinde doğru ürün ile ilgili bilgiyi sunabilmek için işletmelerin birden fazla web sitesine ihtiyaçları vardır.



Şekil 2.1 : Birbirinden farklı içerik yönetimi sistemleri ile kurulmuş farklı siteler

İYS özellikleri sitenin yapısını yönetmek için sağlanır, yayınlanan sayfaların görünümünü ve kullanıcılara sağlanan dolaşım biçimi oluşturur. İYS kullanarak web sitelerine uygulanan özellikler aşağıdaki yararları oluştururlar:

- İçerik yazma işlemlerinin kolaylaştırılması
- Yeni sayfalar ve değişiklikler için daha hızlı gerçekleştirme süresi
- Daha fazla tutarlılık ve yüksek kalite

- Geliştirilmiş site içi dolaşım
- Site esneklik
- Ara yüz esnekliği
- Merkezi olmayan yazım özellikleri
- Küresel ve ücretsiz destek
- Yüksek düzeyde güvenlik
- Azaltılmış bilgi tekrarları
- Büyüme için daha fazla kapasite
- Azaltılmış site bakım maliyetleri
- Çok miktarda kullanıcı
- İstenildiğinde ücretli yazılım yaptırabilme olanağı
- Bilgisayar ve internet teknolojilerinin kısa zamanda uygulanabilmesi

Bunun ötesinde, İYS, müşterilerin iş hedefleri ve stratejileri destekleme konusunda büyük yarar sağlayabilir. Örneğin, İYS kullanıcı memnuniyeti artışını, satış geliştirilmesini veya kamu ile iletişime yardımcı olur.

Açık Kaynak İçerik Yönetim Sistemleri de yaygın olarak hemen her uzman programcı tarafından anlaşılabilir. Fortune 500 şirketleri, üniversiteler, devlet ve özel kuruluşlarca sıkça kullanılmaktadır. Açık kaynak kod İYS yazılımlarının yaygın kullanımının birkaç nedeni vardır:

- İçerik
- Ücretsiz program
- Daha geniş kitlelere ulaşım ve geliştiricilerin küresel ağı
- Düzenli güncellemeler
- Kapsamlı sınama
- Sosyal kullanıcı ağı
- Gelişmiş görsellik ve özellikler
- Başka sistemlere kısa sürede bedava uyum

- Az riskli yüksek ROI olanağı

2.2 İçerik Yönetimi Sistemlerinin Karşılaştırılmalı Çözümlemesi

İçerik yönetim sistemleri çok daha güçlü sistemler gerektiren internet uygulamalarının hızla ilerlemesi sonucu gelişmiştir. Üç ana içerik yönetim sistemi; Drupal, Joomla ve Plone'dur. Bu üç sistem 2000, 2005 ve 2007 yıllarında sırasıyla ortaya çıkmışlardır [9,10,11]. Drupal ve Joomla'da genel kontrol kullanıcıda olup, sistemin farklı yönlerini özelleştirmek için kullanıcıya izin verirler. Kullanıcıların genel olarak açık kaynak kodlu ortamlar üzerinde çalışmak ve bu ortamları kullanmak gibi eğilimleri oluğundan bu sistemler hızla gelişmişlerdir [7,10].

Açık kaynak sistemlerin genellikle daha genişletilebilir yapıları ve binlerce kullanıcıları vardır. Drupal 2007 yılında en iyi genel İYS ödülünü, Joomla ise 2008 yılında en iyi PHP İYS ödülünü kazanmışlardır. Bu iki sistem işlevsellik ve genişleme açısından, hem de modernlikleri bakımından yeni İYS için ölçüt oluştururlar [8].

Bu tezde, başlangıçta, değerlendirme amacı ile Joomla ve Drupal yerel bir sunucuya yüklenmiş ve karşılaştırmalı çözümleme yapılmıştır. Karşılaştırmalı çözümleme için seçilen çeşitli kıstaslar aşağıdadır:

- Toplam Maliyeti (Tüm ücretler ve insan kaynakları)
- Yükleme ve Yükseltme (nasıl kolay kurulum, yükseltme yer ve zaman çerçevesi)
- Ortam desteği
- Tarayıcılar desteği
- Ücretsiz Parçalar, Eklentiler
- Ücretli Parçalar, Eklentiler
- Belgeler (Bilgi tabanı, Bilgi kullanılabilirlik.)
- Destek (Forumlar)
- Kurulu sistem adedi
- Veri tabanı ve işletim sistemi uyumluluğu

- Dil ve yerelleştirme özellikleri
- Kullanıcı yönetimi (İYS yetenekleri, kullanıcı erişim seviyeleri, yetkilendirme ve kimlik doğrulama)
- Arama (İçerik arama özellikleri)
- SEO (Arama Motorları Optimizasyonu)

Çizelge 2.1 : İçerik yönetim sistemleri karşılaştırması

Özellik	**Drupal**	**Joomla**	**Plone**
Toplam Maliyet	0	0	0
Yükleme	4 saat	3 saat	8 saat
Yükseltme	2 saat	2 saat	3 saat
Ortam desteği	4 adet	4 adet	2 adet
Tarayıcılar desteği	5 tarayıcı	3 tarayıcı	4 tarayıcı
Ücretsiz Parçalar, Eklentiler	1800	3200	700
Ücretli Parçalar	350	2800	300
Belgeler	1800 forum	4500 forum	250 forum
Destek	Anlık	Anlık	Günlük
Kurulu sistem adedi	100000ler	100000ler	10000ler
Veri tabanı ve işletim sistemi uyumluluğu	3	4	2
Dil ve yerelleştirme özellikleri	Her dil	Her dil	Ana diller
Kullanıcı yönetimi	Gelişmiş	Gelişmiş	Orta
Arama	Gelişmiş	Çok gelişmiş	Gelişmekte
SEO	Gelişmiş	Çok gelişmiş	Gelişmiş
Sürüm	5.12	1.5.6	3

İçerik yönetim sistemlerinin yükleme süreleri bilgisayar ve veri tabanı bilgisi olup sistemi hiç kurmamış kullanıcılar göz önüne alınarak hesaplandı. Bu değerler internet arama motorlarından gelen sonuçlara ve İYS 'deki açıklamalar göz önüne alınarak derlendi.

Joomla sahip olduğu özellikler ve yukarıda yapılan karşılaştırmalı çözümleme sonucunda çoklu site içerik yönetim sistemi uygulaması için uygun sistem olarak seçildi.

Aynı zamanda KOBİ'lerin en kolay ve etkili şekilde uygulayabileceği ve bakımını sağlayabileceği bir sistem olarak belirlendi.

Bu çözüm sayesinde, KOBİ'ler rakiplerine karşı daha büyük bir yarar kazanımında olacaktır. Bunun yanında birden çok web sitesi aracılığıyla çoklu web sitelerine ait bir KOBİ en ucuz şekilde pazarlama yönetimi yapabilecektir. KOBİ'ler web sitelerinin sayısını arttırarak, gelirlerini koruma, geliştirme ve yeni alanlarda ilerleme olanağı bulacaklardır.

2.3 Çoklu Site İçerik Yönetim Sistemi

Joomla Bölüm 2.2 de yapılan çözümleme sonucunda, bu tezde çoklu site içerik yönetim sistemi için uygun bir içerik yönetim ortamıdır. Joomla, PHP ve MySQL veri tabanını önbelleğe alma gibi özellikleri içerir. Ayrıca Joomla'da; RSS, içerik sayfalarının yazdırılabilir sürümleri, flaş haber, bloglar, anketler, arama yayınları ve uluslararası dil desteği bulunmaktadır.

Joomla birden çok ve farklı içeriği tek bir yerden yönetmenize olanak sağlayan hazır bir web sitesi şablonudur. Joomlanın özelliği bileşen, parça ve mambot olarak 3 farklı bölümden oluşmasıdır.

2.3.1 Bileşen nedir?

Genel olarak, geniş kapsamlı ve üyeden üyeye değişecek içerikler kullanılmak istendiğinde, bileşen kullanılır. Mesela; bir forum, bir dosya yöneticisi veya ayrıntılı bir üye belgisi (profili) bileşen sınıfına girmektedir. Çoklu site içerik yönetimi sisteminde kullanılmak istenen özelliğe bağlı olarak istenen bileşen siteye eklenebilir veya çıkarabilir.

Bunu yapmak için de bileşenin hazırlanmış paketini yönetici arayüzündeki bileşen ekle/kaldır kısmında seçilir ve yükle düğmesi ile yüklenir. Yüklenen bileşenin ayarlarını yapmak için ise yine bileşenler bağlantısından eklenen bileşenin ismi bulunup bileşene ait özel ayarlar yapılır.

2.3.2 Parça nedir?

Parça, daha çok sabit içerik veya üyeden bağımsız içerikleri görüntülemek için kullanılan bir eklentidir.

Örneğin bir site sayacı, forumda son gönderilenler veya sitede çevrimiçi olan üyeler gibidir. İçerik sitesinde kullanılmak istenen özelliğe bağlı olarak istenen parçayı siteye eklemeye veya çıkarmaya yarar. Bunu yapmak için de parçanın hazırlanmış paketi yönetici arayüzündeki parça ekle/kaldır kısmından seçilir ve yükle düğmesi ile yüklenir. Burada dikkat edilecek önemli bir nokta yüklenen parça kaldırılmak istendiğinde parça yönetiminden değil gene parça yükle/kaldır kısmından olması gerektiğidir.

Parçalar da kendi aralarında 2 gruba ayrılırlar. Birincisi ve en çok kullanılan site parçalarıdır. İkincisi ise yönetici parçalarıdır. Parçaları bileşenlerden farklı olarak sitede istenen yerde gösterebilme şansı vardır, yeter ki kullanılan şablon buna olanak sağlasın. Bunu yapabilmek için de yönetici arayüzünden site parçaları bölümüne gidip ayarların yapılacağı parçanın ismini tıklamak gerekir. Parçanın ismi ile birlikte eğer varsa değiştirgeler gelecektir ve istendiği gibi parça yönetilebilecektir. Aynı yerden parçanın pozisyonu ve ismini de değiştirmek mümkündür.

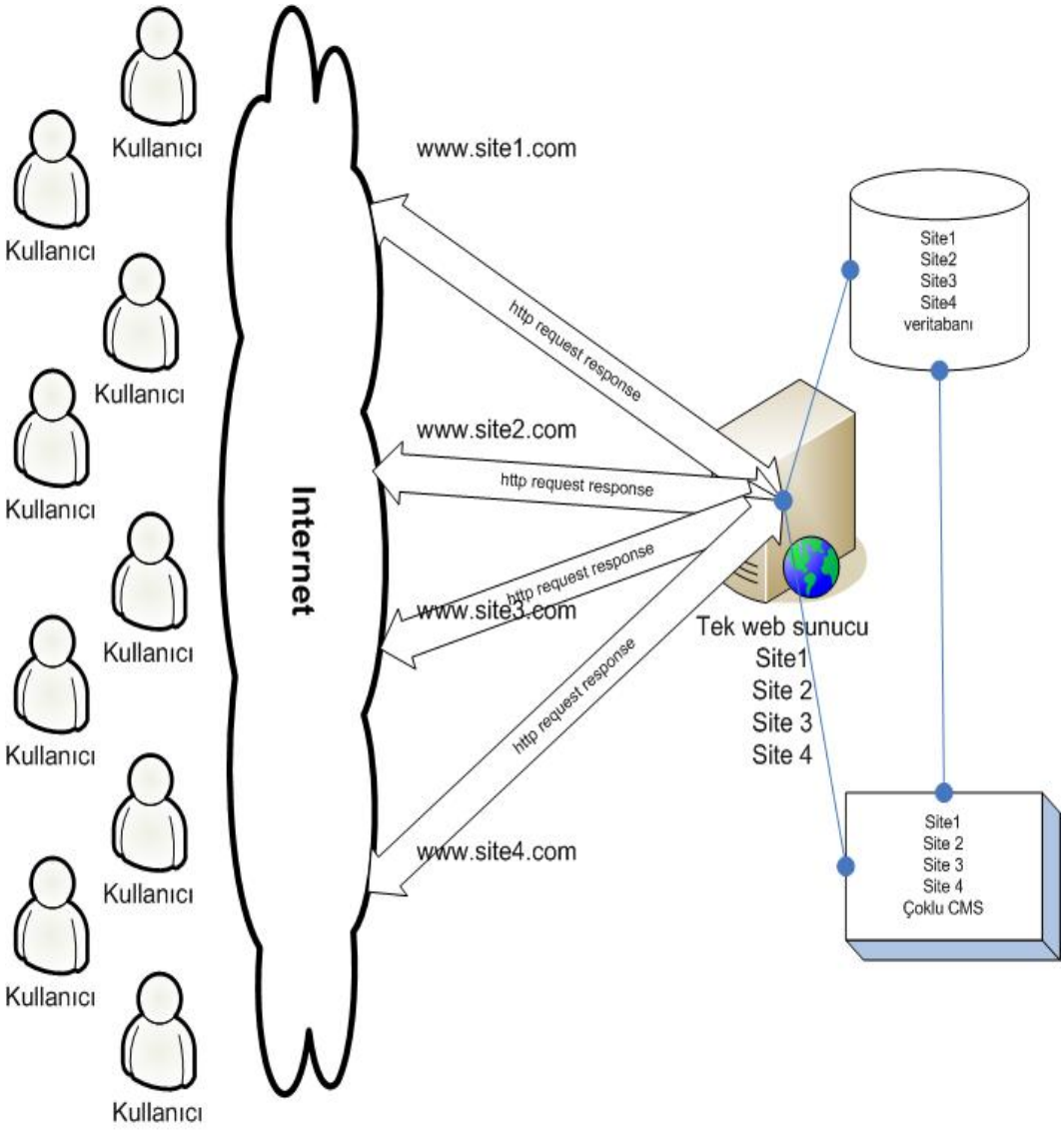
2.3.3 Joomla bot nedir?

Joomla bot'lara ufak programcık denebilir. Sitenin işleyişine katkı sağlamak amacıyla tasarlanmışlardır. Örneğin; bir html editörü, kategoriler içerisinde arama yapmayı sağlayan bir bot'tur. Bunlar da yönetici arayüzünde mambots yükle/kaldır kısmından yüklenebilir veya kaldırabilir.

Bileşen, parça ve bot bölümlerinin yanında ayrıca siteye sabit içerik eklenebilir. İçerikler sınıflandırabilir ve ana sayfada görünmesi veya görünmemesi sağlanabilir [9].

2.4 Çoklu Site İçerik Yönetim Sistemi Uygulaması: MANTAR

Bu tezin bir parçası olarak, bir php4.x tabanlı Linux sunucusuna, 3 web sitesi, tek bir İYS (yalnızca bir etki alanı) uygulanması ile (3 farklı etki) kurulmuştur. Bu şekilde üç veya istenilen adet siteyi tek bir İYS kurulumuyla oluşturma ve yönetme olanağı oluşmuştur. Bu aynı veya farklı tarz sitelerden toplanmış olduğu varsayılan oy bilgileri ile istenilen biçimde işbirlikçi süzgeçleme sistemi kurulabilir.



Şekil 2.2 : Çoklu site içerik yönetim sistemi: MANTAR

2.4.1 MANTAR kurulumu

Çoklu site içerik yönetim sistemi oluşturulurken Joomla 1.5.x veriyonları ve daha üst versiyonları uygulanabilecek şekilde düzenlenmiştir. Uygulama kodu Joomla'ya ait 15 adet parça, bileşen ve botlara uygulanmıştır Bu sistem çok kolay şekilde genişletilebileceğinden MANTAR adı verilmiştir.

Çoklu site içerik yönetim sistemi PHP ortamında geliştirilip, MySQL veri tabanı üzerinde çalışacak şekilde ayarlanmıştır.

Tek Joomla kurulumu üzerinden birden fazla siteyi yönetebilmek için aşağıdaki adımlar uygulanır.

1) Alan adı isimleri belirlenir. MANTAR kuramsal olarak n adet siteyi tek seferde yönetme olanağı sağlar. Bu tezde 4 adet site örnek alınarak uygulama yapılmıştır. Alan adları olarak www.site1.com, www.site2.com, www.site3.com ve www.site4.com seçilmiştir.

2) Seçilmiş olan alan adları sunucu servisi veren sistemde ayrı ayrı tanımlanır. Çoklu site içerik yönetimi sistemini www.site1.com alan adı adresinde kurulacağını varsayılarak www.site2.com, www.site3.com ve www.site4.com alan adları www.site1.com sitesine yönlendirilir.

3) www.site1.com adresine Joomla 1.5.x normal bir kurulmuş gibi kurulur. Bir önceki adımdan hareketle diğer alan adı isimleri, yani www.site2.com, www.site3.com ve www.site4.com tarayıcılara girildiğinde www.site1.com sitesi karşımıza çıkar. Ancak sistem MANTAR sistemi uygulanmadığından doğru çalışmaz.

4) www.site1.com üzerine kurulmuş olan Joomla içerik yönetim sisteminde ilk önce com_multisitemanager.zip dosyası ile gelen MSM bileşeni yüklenir. Bu bileşen otomatik olarak yönetici arayüzünden eklenir. MSM bileşeni ile www.site2.com, www.site3.com ve www.site4.com www.site1.com da kurulu Joomla sistemi üzerinde tanımlanır. Bu şekilde www.site2.com, www.site3.com ve www.site4.com siteleri tarayıcıda girildiğinde artık www.site1.com üzerine kurulu olan Joomla her site için farklı bir Joomla kurulumu varmış gibi sonuç gösterir.

5) www.site1.com üzerine kurulmuş olan Joomla içerik yönetim sisteminde plg_multisitemanager.zip dosyasıyla gelen MSM Şablon Yükleme uyumlu eki yüklenir ve yayınlanır. Böylece kodlarında değişiklik yapılacak parçalar, bileşenler ve şablonlar her bir site için düzgün çalışacak şekilde tanımlanmış olur.

6) MANTAR. php dosyası public_html/libraries klasörü altına kopyalanmalıdır. MANTAR. php dosyası çoklu site içerik yönetim sisteminde dinamik alan adı isimlerinin doğru yönlendirilip tek bir Joomla SQL veri tabanı üzerinden bütün sitelerin çalışmasını sağlar.

7) Joomla içerisinde kod değişikliği yapılmalıdır. Kod değişikliği yapılması gereken kısımlar aşağıdaki şekildedir:

- Site ismi
- Menü elemanı
- Ana sayfa elemanı
- Bölüm elemanı
- Kategori elemanı
- Çoklu sistemde sonucu alınamayan dosyalar ile ilgili çıkan mesaj
- Arama bileşeni - içerik için
- Arama bileşeni - bölüm için
- Arama bileşeni - kategori için
- Haberler parçası
- Son Haberler parçası
- Yorumlar parçası
- Son yorumlar parçası
- Yorumlar ve açıklamalar bileşenleri
- En çok okunan yazılar parçası
- En beğenilen yorumlar bileşeni
- Resim ve küçük resim gösterme bileşeni
- Etiket bulutu bileşeni
- Etiket bulutu parçası
- Joomla içerik oluşturma bileşeni
- Joomla içerik anahtar kelime bileşeni
- Joomla web haritası bileşeni

Her bir bileşen, parça veya dosya üzerinde yapılması gereken değişiklikler detaylı olarak CD de teslim edilen MANTAR-Joomla1.5.x-Duzenlemesi. doc dokümanında belirtilmiştir.

2.4.2 Kurulu MANTAR sisteminin ayarlanması

MANTAR bileşeni ve parçası kurulup MANTAR dosyası kopyalandıktan sonra normal Joomla içeriği temel alınarak her site için farklı görünümler tanımlanabilir. Burada önemli olan nokta şudur: MANTAR bileşimi ile yeni içerik eklenmez, yeni bölüm tanımlanmaz, yeni parça ya da yeni menü oluşturulmaz. MANTAR bileşeni sadece mevcut bilgilerin sitelerle ilişkilerini tanımlamayı sağlar. Joomla kodu üzerinde yapılan değişiklikler de bu tanımlamaları kullanarak sitelerin oluşturulmasını sağlar. Dolayısıyla Joomla içerisinde önceden yayımlanmış durumdaki bir parça, menü elemanı ya da içerik elemanı yayından kaldırılırsa, bu parça, menü elemanı yada içerik elemanının ilişkilendirildiği tüm sitelerde bu içerik elemanı yayından kaldırılmış olur.

2.4.3 Sitelerin tanımlanması

İlk önce Yönetici->Bileşenler->Multi Site Management->Sites seçilerek siteler tanımlanır. Site tanımlanırken var olan tanımları kopyala seçilirse Joomla içerisindeki yayınlanmış parçalar, menü öğeleri, bölüm, kategori ve içerik öğeleri site ile otomatik olarak ilişkilendirilir ve birebir normal sitenin kopyası oluşturulmuş olur.

2.4.3.1 Şablonlar

Yönetici->Bileşenler->Multi Site Management->Şablonlar adımları ile tanımlanan siteye hali hazırdaki şablonlardan bir tanesi atanarak sitenin farklı görünmesi sağlanır.

2.4.3.2 Parçalar

Yönetici->Bileşenler->Multi Site Management->Parçalar adımları ile Joomla içerisindeki yayınlanan parçalar siteye pozisyon bilgileri de verilerek atanabilir. Örneğin Ana-Menü parçası A sitesine sol konumda olacak şekilde atanarak bu sitede Ana-Menü-'nün solda gözükmesi sağlanır.

2.4.3.3 Menü öğeleri

Yönetici->Bileşenler->Multi Site Management->Menü öğeleri adımları ile published durumdaki menü öğeleri sitelerle ilişkilendirilebilir. Bu ilişkilendirme işlemi parça tanımlarını dikkate alarak çalışır; yani Ana-Menü parçası aktif iken bu parçanın oluşturduğu Ana-Menü içerisinde tanımlı Haber Kaynakları menü elemanı bu adımda tanımlanıp yayınlandıktan sonra Ana-Menü içerisinde görünür hale gelir.

2.4.3.4 Bölümler

Yönetici->Bileşenler->Multi Site Management->Sections adımları ile yayınlı durumdaki bir bölüm herhangi bir siteye atanabilir.

2.4.3.5 Kategoriler

Yönetici->Bileşenler->Multi Site Management->Categories adımları ile daha önce siteye atanan bir bölüme ait yayınlı durumdaki bir kategori siteye atanabilir.

2.4.3.6 İçerik öğeleri

Yönetici->Bileşenler->Multi Site Management->Content Items adımları ile daha önce siteye atanan bir bölüme ait kategoriye ait yayınlı durumdaki bir içerik elemanı siteye atanabilir ve içeriğinin gözükmesi sağlanabilir.

3. ÇEVRİMİÇİ UYGULAMALARDA ÖNERİ SİSTEMLERİ

3.1 Öneri Sistemleri

Öneri sistemleri (Recommendation System) kullanıcının ilgisini çeken bilgi öğelerini göstermeye yönelik özel bir bilgi süzgeçleme (Information Filtering (IF)) tekniğidir. Öneri sistemleri film, kitap, topluluklar, haberler, makale, vb. uygulama sistemleri ve siteleri için oluşturulmuştur. Tipik olarak bir öneri sistemi kullanıcı belgisini bazı ilgi özellikleri ile karşılaştırır ve kullanıcı tarafından bir öğeye verilmiş fakat henüz değerlendirilmemiş olan oylamayı öngörmeye çalışır. Bu özellikler bilgi elemanından veya kullanıcının sosyal çevresinden olabilir. [8].

Kullanıcı belgisi oluşturulurken açık ve kapalı veri yapıları arasında ayırım yapılmalıdır.

Açık veri örnekleri şunlar olabilir:

- Kullanıcıya değişebilen değerler arasından bir öğeye oy vermesi istenebilir.
- Kullanıcıya en çok beğendiği öğeden en az beğendiğine doğru sıralama yapması istenebilir.
- Kullanıcıya iki farklı öğe verilip en sevdiğini seçmesi istenebilir.
- Kullanıcıya en sevdiği öğelere ait bir liste yapması istenebilir.

Kapalı veri örnekleri şunlar olabilir:

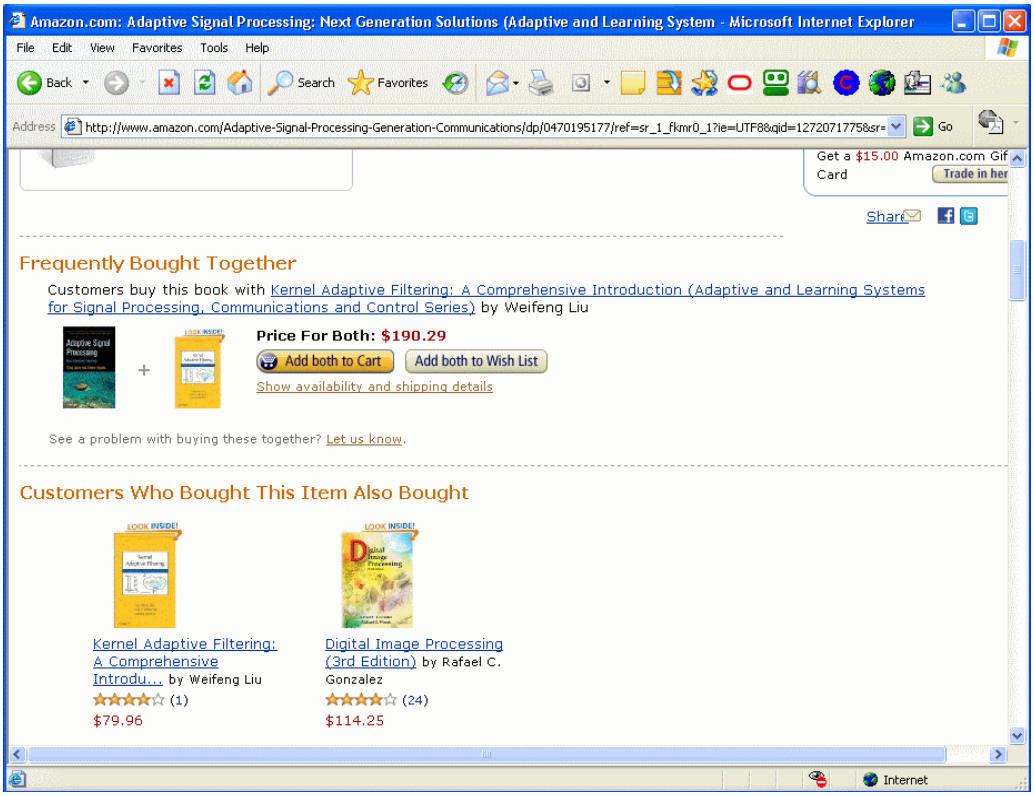
- Çevrim içi mağazalarda kullanıcının baktığı öğeler incelenebilir.
- Kullanıcının ve öğenin görüntülenme süresi incelenebilir.
- Kullanıcın çevrim içi sayfalarda satın aldığı öğelerin kayıtları tutulabilir.
- Kullanıcının sosyal ağları gözden geçirilerek sevdikleri veya sevmedikleri bulunabilir.

Öneri sistemi, toplanmış veri ile diğer yerlerden toplanmış benzer verileri karşılaştırır ve kullanıcı için önerilmiş öğeler listesini hesaplar. Çizelge 3.1 de öneri sisteminin kullanıldığı ticari ve ticari olmayan popüler örnek siteler verilmiştir [13]. Ticari sitelerin çoğu daha fazla ürün satışı sağlayabilmek için ve daha çok kullanıcı çekebilmek için işbirlikçi veya içerik süzgeçleme sistemleri kullanırlar.

Çizelge 3.1 : Öneri sistemlerinin kullanıldığı ticari veya ücretsiz siteler

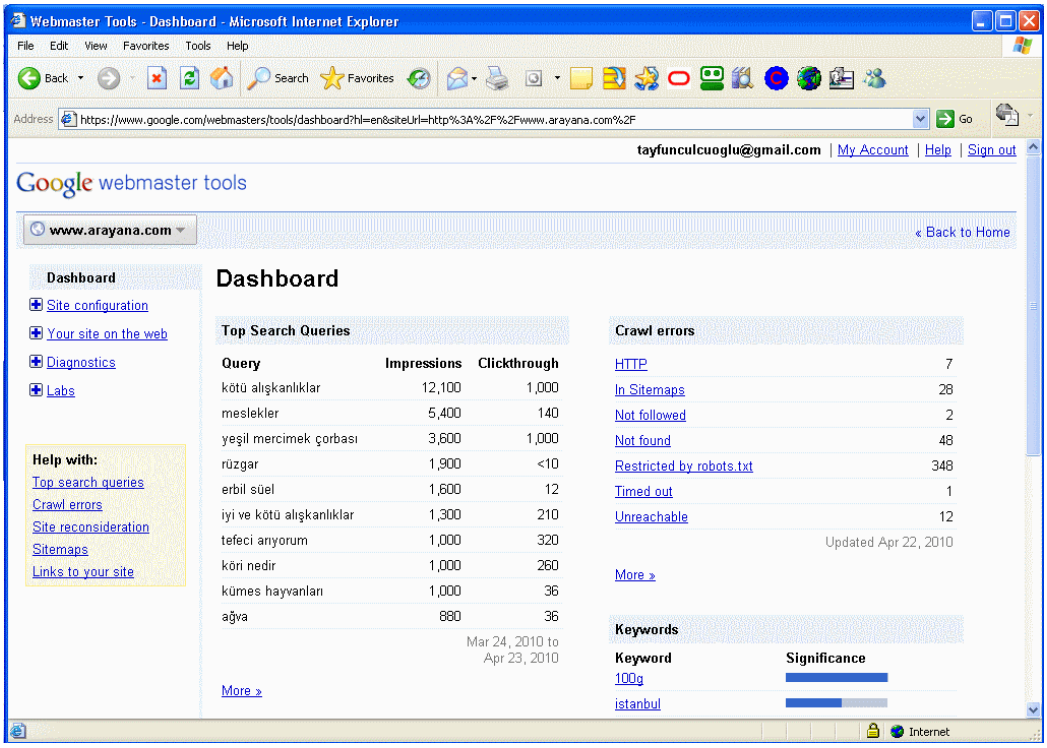
Site İsmi	**Türü**	**Ticari/Ücretsiz**
Amazon	E-Ticaret	Ticari
Amie Street	Müzik	Ticari
AmphetaRate	Etkileşim	Ücretsiz
Barnes and Noble	Kitap	Ticari
Baynote	Müzik	Ticari
ChoiceStream	E-Ticaret	Ticari
Collarity	E-Ticaret	Ticari
Digg.com	Etkileşim	Ticari
eBay	E-Ticaret	Ticari
Everyone's a Critic	Film	Ücretsiz
GiveALink.org	Etkileşim	Ücretsiz
Google News	Etkileşim	Ticari
half.ebay.com	Kitap	Ticari
Hollywood Video	Film	Ticari
Hulu	Film	Ticari
iLike - music	Müzik	Ticari
Internet Movie Database - movies	Film	Ticari
iTunes - music	Müzik	Ticari
Last.fm - music	Müzik	Ticari
LibraryThing - books	Kitap	Ticari
MovieLens	Film	Ücretsiz
Musicmatch	Müzik	Ticari
Neflix	Film	Ticari
Rate Your Music	Müzik	Ücretsiz
Simania	Kitap	Ticari
Strands	Kitap	Ticari
StumbleUpon	Etkileşim	Ticari
Threadless	E-Ticaret	Ticari
TiVo	Film	Ticari

Öneri sistemleri kullanıcıların kendi başlarına bulamayacağı öğelerin keşfinde yardımcı olan başka araştırma algoritmalarından yararlı olan bir tanesidir. Bir öneri sistemi oluşturmak için birçok yaklaşım vardır. Bunlardan işbirlikçi süzgeçleme (Collaborative filtering) ve içerik tabanlı süzgeçleme (Content Based Filtering) en çok kullanılan süzgeçleme sitemleridir. Ortak süzgeçleme, iki kullanıcı arasındaki benzerliği kullanıcı değerlendirme belgisine göre hesaplar ve benzer kullanıcılar tarafından yorumlanan öğeleri önerir. Öte yandan içerik tabanlı süzgeçleme herhangi bir veri tercihini kullanmaz ve doğrudan öğelerin benzerliklerine dayalı öneri sağlar ve benzerlik öğeye göre hesaplanır [14]. Her iki sistemin de yararları olmasına rağmen iki sistemin de kullanıldığı veya işbirlikçi süzgeçlemenin bellek tabanlı ve model tabalı türevlerinin beraber kullanıldığı karma sistemlerin yapılandırılması ile daha verimli sonuçlar alınabileceği düşünülmüştür [15].



Şekil 3.1 : Amazon.com da işbirlikçi süzgeçleme yönteminin uygulanışı

Kullanıcı ile etkileşimi gerektiren Web uygulamaları için modern öneri sistemleri ortak bir senaryodur. Tipik olarak, bir sistem öğelerin özet listesini kullanıcıya sunar ve kullanıcı bir öğe hakkında daha fazla bilgi almak için öğeler arasında bir seçim yapar veya bir şekilde öğe ile etkileşime geçer. Örneğin; e-ticaret siteleri genellikle bireysel ürünlerin listesi ile birlikte bir sayfa sunar ve sonra sayfada kullanıcının seçtiği veya satın almak istediği ürün ile ilgili daha ayrıntılı bilgi görebilmesini sağlar. Web sunucusunda genellikle öğelere ait bir veri tabanı yer almaktadır ve dinamik olarak web sayfalarını bu öğe listesi ile oluşturur. Çoğu zaman bir web sayfasındaki öğelerden çok daha fazla öğe veri tabanında mevcuttur, bu yüzden bunu kullanıcıya göstermek için öğelerin bir alt kümesini seçmek veya hangi öğelerin görüntüleneceğinin sırasını belirlemek gerekir.



Şekil 3.2 : Google Web Master Tools da içerik süzgeçleme yönteminin uygulanışı

Son beş senede yapılan bilimsel çalışmalarda, her iki sistemin yapı yöntemleri kullanılarak oluşturulacak karma öneri sistemi iki sistemin de açıklarını kapatacak yapıda olabileceği vurgulanmıştır. Ancak bu sistemlerin uygulanması kuramsal açıdan mümkün görünse de uygulanabilirlik açısından zor ve yüksek maliyetlidir [16]. Netflix şirketi film oylamalarını daha iyi hale getirebilmek için $1.000.000'lık bir yarışma yapmış ve bunu sonucunda işbirlikçi ve içerik süzgeçleme sistemlerinin beraber kullanıldığı karışık bir algoritmayı uygulamaya karar vermiştir [17].

3.2 İçerik Tabanlı Süzgeçleme Sistemleri

İçerik tabanlı süzgeçleme sistemi bir kullanıcıya, kullanıcının ilgisine ait belgiye veya öğenin tanımına dayalı bir öğe önerir. İçerik tabanlı süzgeçleme sistemleri web sayfaları, haber makaleleri, restoran, televizyon öneri programları ve çevrimiçi satış siteleri gibi çeşitli alanlarda kullanılabilir. Her ne kadar sistemler birbirinden farklı olsa da, içerik tabanlı sistemler önerilmiş öğenin tanımlanması, kullanıcı belgisinin oluşturulması ve de kullanıcı belgisindeki öğelerin karşılaştırılması için ortak bir araçtır. Kullanıcı belgisi genellikle kullanıcıya sunulmuş öğelere ait isteklerin geri cevaplarına göre otomatik olarak oluşturulur ve güncellenir. İçerik tabanlı öneri sistemi öğeleri kullanıcı tercihlerine ve öğelerin içerikleri arasındaki ilintiye göre seçerken, ortak süzgeçleme sistemi ise öğeleri benzer tercihleri yapan kişiler arasındaki ilintiye göre seçer [18].

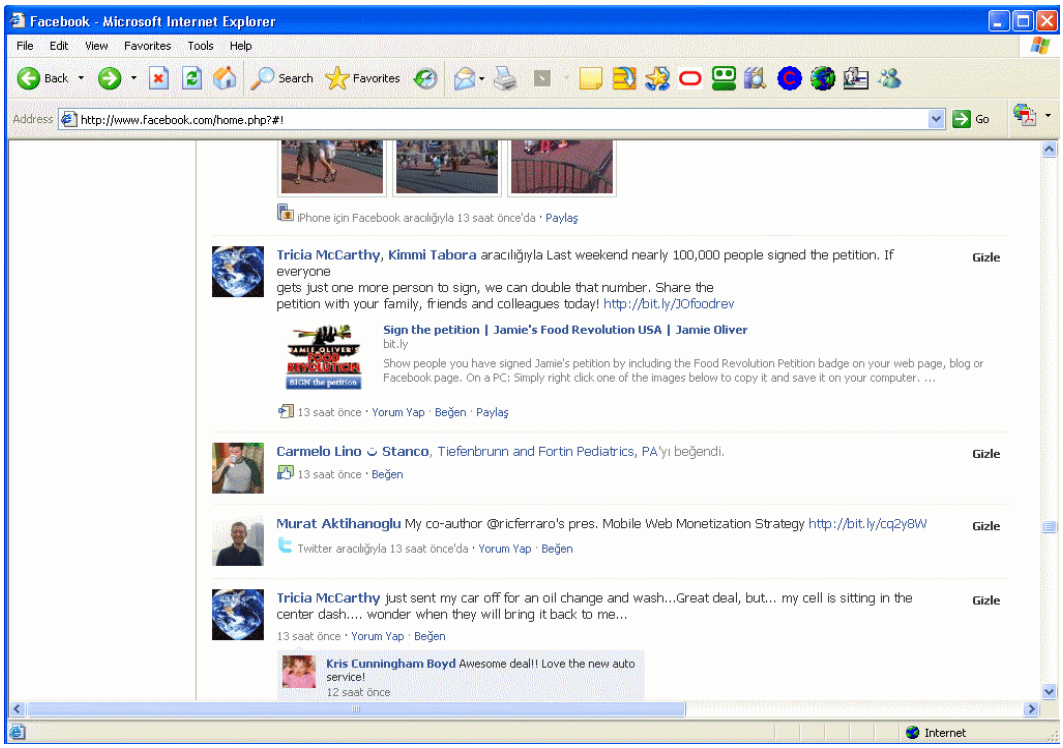
İçerik tabanlı öneri sistemlerinde sınıflandırma algoritmaları anahtar bileşenlerdir. Çünkü her kullanıcının ilgi alanlarını modelleyen işlevleri öğrenirler. Yeni bir öğe ve kullanıcı modeli verildiğinde işlev kullanıcın öğe ile ilgilenip ilgilenmeyeceğini öngörür. Çoğu sınıflandırma algoritmaları kullanıcının görmediği bir öğeyi sevip sevmeyeceğinin olasılığını kestirecek işlevi sağlar. Bu olasılık öneri listesinin sıralanmasında kullanılabilir. Ek olarak, algoritma ilginin derecesi gibi doğrudan bir değer kestiren işlev yaratabilir [19].

Aşağıdaki Çizelge 3.2 de çeşitli içerik süzgeçleme sistemlerinin kullanıldığı ticari ve ticari olmayan popüler örnek siteler verilmiştir. Ticari sitelerin kullanıcılara daha ilgili ürünleri gösterebilmek için içerik süzgeçleme sistemleri kullanırlar. Ticari veya ücretsiz uygulamalar ayrıca son kullanıcıların görmesini, ulaşmasını istenmeyen içerikleri süzgeçlemek için kullanılırlar [13].

Çizelge 3.2 : İçerik süzgeçleme sistemlerinin kullanıldığı siteler

Site İsmi	Türü	Ticari/Ücretsiz
K9 Web Protection	Sosyal	Ücretsiz
Dans Guardian	Sosyal	Ücretsiz
CyberySitter	Sosyal	Ticari
Facebook	Sosyal	Ücretsiz

Algoritmaların bazıları geleneksel makine öğrenme algoritmalarıdır. Bu algoritmalar yapılandırılmış veriler üzerinde çalışır. Yapılandırılmamış veriler ile çalışıldığında önce yapılandırılmamış veriler farklı özelliklerinden oluşturulmuş küçük alt kümelere bölünerek yapılandırılmış verilere dönüştürülürler. Tam tersi, diğer algoritmalar çok boyutlu uzayda çalışırlar ve önceden verilerin yapılandırılmasına gerek yoktur.



Şekil 3.3 : Facebook'ta içerik süzgeçleme algoritmasının uygulanışı

İşte, bu süzgeç diğer kullanıcıların bilgilerini kullanmadan yalnızca tek bir kullanıcı tarafından tamamlanabilir. Başarım, öğelerin içeriğinin kalitesine bağlıdır ve eğer yeterince öğe içeriği getirilebilinirse, içerik tabanlı süzgeçleme yeterince iyi çalışacağı beklenebilir.

3.3 İçerik Tabanlı Süzgeçleme Sisteminde En Yakın Komşu Yöntemi

En yakın komşu algoritması tüm açık veya kapalı etiketlenmiş olan eğitim verilerini bellekte tutar. Etiketlenmemiş yeni bir öğeyi sınıflamak yerine algoritma benzerlik işlevini kullanarak tüm kayıt edilmiş öğeler ile karşılaştırır ve "en yakın" ve ya "k en yakın " komşusunu belirler. Daha önceden görülmemiş öğe için verilen rakamsal değer veya sınıf etiketi en yakın komşu sınıfındaki sınıfın etiketinden türetilebilir.

Benzerlik işlevi, verinin tipine bağlı olarak en yakın komşu algoritmasında kullanılır. Yapılanmış veri için "Öklid Uzaklık Ölçütü" sıklıkla kullanılır. "Vektör uzay modeli" kullanıldığında kosinüs benzerlik ölçüsü kullanılır [20].

Öklid uzaklık işlevinde iki örnekte küçük değer taşıyan aynı özellik, her iki örnekte de büyük değer taşıyan aynı özellik olarak ele alınır. Tam tersi, kosinüs benzerlik işlevi eğer iki örneğin ilgili özellikleri küçük değerliyse büyük değer olarak ele almayacaktır.

Şekil 3.4 de içerik öneri sitemine ait örnek bir uygulamanın beş adımdan oluşan iş şeması verilmiştir.

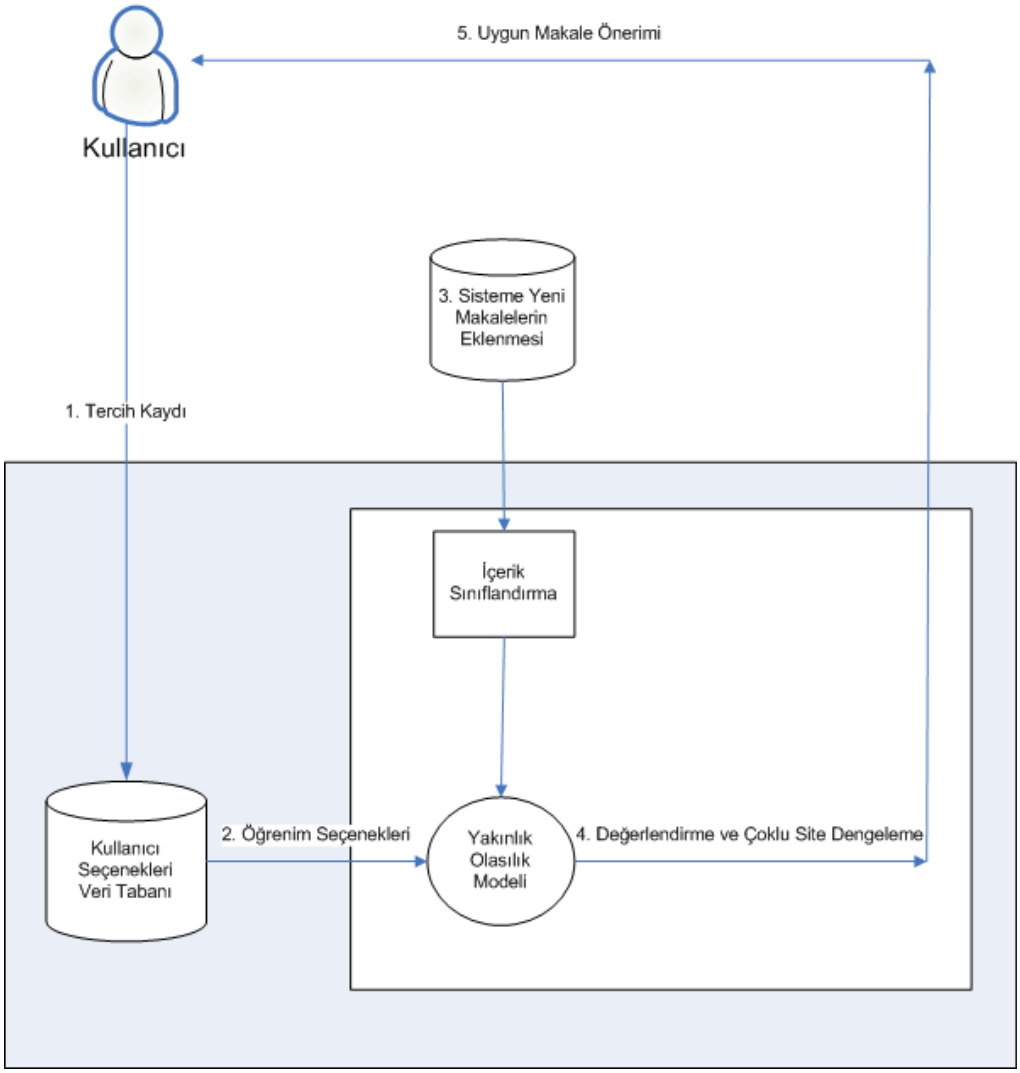
1) Kullanıcı tercih kaydı yapılır ve tercihler veri tabanına kayıt edilir.

2) Tercihler olasılık modeli ile eğitilir.

3) Sisteme kayıt edilen makaleler anahtar kelimelere yakınlıklarına göre sınıflandırılır.

4) Kestirim, eğitilmiş model kullanılarak yüklenmiş makalelerin değerlendirilmesi ile yapılır ve çoklu site dengelemesi yapılır.

5) Son olarak uygun makale önerilir.



Şekil 3.4 : 5 adımdan oluşan içerik tabanlı öneri sistemi şeması

3.4 İçerik Tabanlı Süzgeçleme Sisteminde Olasılık Modelleme

Bu bölümde içerik tabanlı Süzgeçleme sistemlerinde olasılık modellemesini anlatmak için kitap öneri sistemi örnek bir uygulama olarak ele alınmıştır. Bu kitap öneri sisteminde, seçilmiş anahtar kelimeler ve tercih edilmiş kitaplardaki değerler yeni kitapları değerlendirmek üzere olasılık modelinde kullanılır. Tercih edilen kitaplar sadece tercihlerin öğrenilmesinde kullanılır. Bu öğrenmede kullanıcı sadece tercihlerini vermek zorundadır. Bu tercih edilen kitapların elde edilmesinde yararlı bir yöntemdir [21]. Aşağıda öğrenme sürecindeki olasılık modellemesinin detayları verilmiştir.

İçerik Süzgeçleme gösteriminde d kitap, z demetlemeye karşılık olarak gizli bir değişken, s alan ve ya başlık veya özet, w bir kelime, $n_{s,d}(w)$ ise d kitabında s alanında ki w kelimesinin sayısı olarak gösterilmektedir. Kitabın olasılık yapısı aşağıdaki sonlu karışım modeli **(3.1)** ile modellenmiştir.

$$p(d) = \sum_z p(d,z) = \sum_z p(z) \prod_{s \in d} \prod_{w \in s} p_s(w \mid z), \qquad \textbf{(3.1)}$$

Daha sonra tercih edilmiş kitaplardan olasılığı olan $p(z)$ ve $p_s(w \mid z)$ eğitilir. Bu genellikle demetleme için kullanılan bir sözde karışım modelidir. Bu modeli normal bir karışım modelinden farklı kılan önemli bir özellik ise bu modelde alanlar ayrılmaktır yani her alandaki tercihler birbirinden bağımsız olarak yakalanabilir. Aslında, burada iki alan kullanılmaktadır: başlık ve özet.

Olasılık değerleri bir EM tarafından (Beklenti-Maksimizasyon) algoritması ile eğitilir ve bu algoritmada aşağıdaki E- ve M-adımları durum memnun edici olana kadar dönüşümlü olarak devam eder.

E-Adımı:

$$p(z \mid d) = \frac{p(d,z)}{\sum_z p(d,z)} \qquad \textbf{(3.2)}$$

M- Adımı:

$$\hat{p}(z) = \frac{\sum_d p(z \mid d)}{\sum_{z,d} p(z \mid d)} \qquad \textbf{(3.3)}$$

$$\hat{p}(w \mid z) = \frac{\sum_d n_{s,d}(w) p(z \mid d)}{\sum_w \sum_d n_{s,d}(w) p(z \mid d)} \qquad \textbf{(3.4)}$$

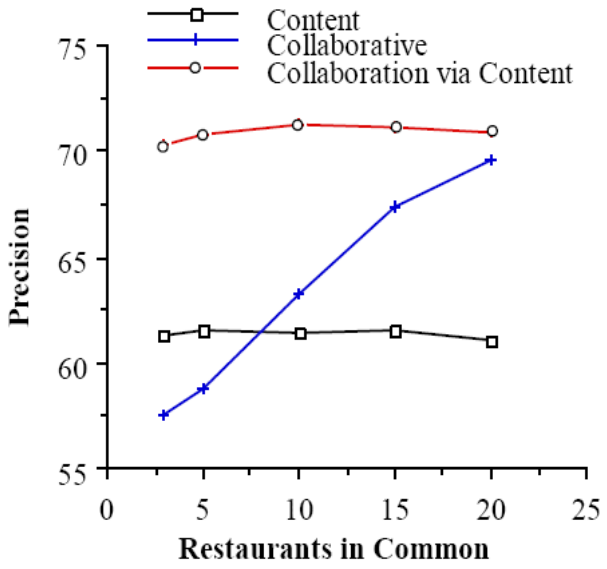
3.5 Karma Süzgeçleme Sistemleri

Öneri sisteminin kalitesini arttırmak için işbirlikçi ve içerik tabanlı yaklaşımı birleştirerek *Karma* yaklaşım gerçekleştirilebilir. Karma süzgeci kullanan öneri sistemleri 3 ayrı kümede toplanır.

İlk kümedeki sistem, içerik tabanlı ve işbirlikçi süzgeçler ile gerçekleştirilmiş öneri sonuçlarını basitçe birleştirmeye çalışır [22]. İkinci kümedeki sistem, iki yaklaşımı ürün bilgisinin ve işlem bilgisinin bir arada olduğu tek sunum katında birleştirir. Bu sistemlerin bazıları işbirlikçi süzgeçlemeye dayanır ve kullanıcı gösteriminde öğelerin içerik bilgilerini ekler veya öğelerin içeriklerine bağlı olarak oyları ekler [23]. Üçüncü kümedeki ise farklı bilgi kaynakları ile birleşmiş kapsamlı modellerden oluşmuş sistemleri içermektedir.

Kapsamlı karma yaklaşım algoritmasının önemli bir sınıfı da demetleme modellerinden oluşturulmuştur. Kullanıcı ve öğe özellikleri ve aynı zamanda buna uygun ek varlıklar ve bunların özellikleri daha yüksek kalitede öneri sağlayabilmek için demetleme modelleri ile birlikte çalışırlar. Bu alandaki Hoffman'ın çalışmaları bulunmaktadır [24]. Şekil 3.5 den de görüldüğü gibi işbirlikçi süzgeçleme içerik süzgeçlemesine göre daha kesin sonuçlar verebilmektedir. Karma sistemler ise veri ve bilginin az olduğu tasarımlarda başarı göstermekte iken örnek sayısı çoğaldığında işbirlikçi sistemler karma sistemler kadar başarılı olabilmektedirler [25].

Şekil 3.5'de görüldüğü üzere içerik süzgeçleme öğe sayısı bilgisi artsa da aynı şekilde başarım gösteriyor. İşbirlikçi süzgeçleme genelde az veri olduğu durumlarda düşük başarım gösterir. İki süzgeçleme yönteminin beraber kullanıldığı karma süzgeçleme yöntemi ile ise en yüksek başarım sağlanır [25].



Şekil 3.5 : İçerik ve işbirlikçi süzgeçlerinin beraber uygulandığı çalışma

4. İŞBİRLİKÇİ SÜZGEÇLEME SİSTEMLERİ (İFS)

4.1 Eğimli İşbirlikçi Süzgeçleme (EİS)

İşbirlikçi süzgeçleme sistemleri bir kullanıcın bir ürüne ne kadar oy verebileceğini o kullanıcının önceki oylarını göz önüne alarak kestirme sürecidir. Bu tezde, $f(x) = x + c$ şeklinde tek eğimli yapı kullanarak kullanıcıların oylarının kestirilebileceğini göstereceğiz. Bu işlem aynı kişi tarafından oy almış birden fazla öğenin birbirleri arasında ortalama farklara bakılarak gerçekleştirilir.

Tek Eğimli İşbirlikçi Süzgeçleme (TEİS) yapısı kullanılarak bellek tabanlı (Memory-Based) işbirlikçi süzgeçleme sonuçlarına benzer sonuçlar elde edilebilir. TEİS algoritmalarının denenmek istenmesinin sebebi kolayca uygulanabilir, verimli sorgulama yapılabilir, kabul edilebilir derecede doğru ve çevrimiçi sorgulama ve güncellemeleri destekleyebilmelerinden ötürüdür. Bu özellikler günümüzün çevrimiçi dinamik gerçekleme sistemlerine çok uygundur.

Kullanıcıların yüksek oy verdikleri öğeleri ve düşük oy verdikleri öğeleri göz önüne alarak uyguladığımız TEİS daha yavaş olan bellek tabanlı içerik süzgeçleme yöntemlerine oranla karşılaştırılabilir sonuçlar verdi.

Çevrimiçi oylama tabanlı TEİS sorgusu tek bir kullanıcı için öğe/oy çiftlerinden oluşan bir diziden oluşur. Bu sorguya verilen cevap ise kullanıcın daha önceden oy vermemiş olduğu öğeler için oluşturulan öğe/oy çiftlerinin oluşturduğu bir dizidir. Bu tezde TEİS yapısını oluşturmamızın sebebi aşağıda belirttiğimiz kıstasların işbirlikçi süzgeçleme sistemlerin TEİS yapısı tarafından karşılanabilecek olmasıdır:

- Dinamik olarak yenilenebilme: Bir kullanıcı tarafından eklenen yeni bir oy sistem tarafından anında gözden geçirilebilir ve süzgeçleme yenilenir.
- Verimli sorgulama: Sorgulamalar ve cevaplar hızlı ve sistemde çok yer kaplamayacak şekilde olmalıdır.
- Ziyaretçilerden beklenti azdır: Az oy vermiş ziyaretçiler bile geçerli ve mantıklı öneriler alabilir.

- Kolayca uygulanabilir ve yönetilebilir: Toplanan bütün veriler bir mühendis tarafından kolayca anlaşılabilmeli ve algoritmalar kolayca uygulanabilir ve sınanabilir olmalıdır.
- Kabul edilebilir derecede gerçekçi ve doğru sonuçlar sağlamalı: Sonuçlar kabul edilebilir derecede doğru olduktan sonra daha da iyi sonuçları elde etmek için basitlikten ve ölçeklenebilirlikten ödün vermeyecek yapıda olmalıdır.

TEİS ve oluşturduğumuz uygulamalar ile yukarıda belirtilmiş olan sorunları çözerek MANTAR biçimindeki çoklu sistemlerde veya tekli içerik yönetimi sistemlerinde uygulanabilirlik sorununu çözmüş oluruz.

4.2 Bellek Tabanlı Süzgeçlemeler

Bellek tabanlı işbirlikçi süzgeçleme sistemleri kullanıcı öbekleri arasındaki benzerliklere bakarak kestirimlerde bulunurlar. Bu genellikle ağırlıklı ortalama ile olur [26]. Seçilen yakınlılık ölçütleri kestirimlerin doğruluğunu etkiler ve bu konularla ilgili birçok çalışma yapılmıştır. Ancak bu yapılar verilerin çok olduğu çevrimiçi uygulamalarda doğruluk açısından yetersiz kalırlar. Bunun dışında bu tarz süzgeçleme yapıların düzgün sonuç sağlayabilmesi için minimum miktarda kullanıcı ve oy bilgisinin sistemde olması lazımdır. Bu tarz bilenen yöntemlerden birisi Pearson modelidir.

4.3 Bellek Tabanlı ve Model Tabanlı Tasarımlar

Bellek tabanlı işbirlikçi süzgeçleme tasarımları, kullanıcı çiftleri arasındaki benzerliği ölçerek kestirimde bulunmaya çalışırlar. Bu benzerliği oluştururken ortalama değerler kullanılır [27]. Seçilen benzerlik ölçümleri ve hesaplamalar kestirimin geçerliliğini etkiler. Bu yöntemde hızlı çevrimiçi uygulamalarda çok sayıda sorgulama yapılacağından uygun olmaz. Ayrıca, eğer kullanıcı sayısı veya oy miktarı az ise sonuçlar doğru çıkmayabilir.

Bölüm 3'te belirtildiği gibi birçok içerik süzgeçleme yöntemleri vardır. Bunlar düzlemsel matematik, Bayes yöntemlerinin kullanıldığı yapay zeka yöntemleri, sinir ağları, demetleme algoritmalarıdır. Bellek tabanlı yapılara oranla model tabanlı süzgeçleme yöntemleri daha hızlı sorgulama yapmaya olanak sağlar. Model tabanlı uygulamalar bellek tabanlı uygulamalara göre çevrimiçi kullanımlar için daha uygun olabilir.

Uygulamayı planladığımız EİS ve TEİS yöntemlerini bilimsel çalışmalarda ve araştırmalarda kestirim sistemleri ile matematiksel olarak karşılaştırabiliriz. Oluşturduğumuz kestirim sistemleri

$$f(x) = x + b \quad \textbf{(4.1)}$$

şeklindedir. Bu sebepten ötürü tek eğimli süzgeçleme adı verilir.

Bu eşikte *b* bir sabit ve *x* değişken oy değerlerini belirtmektedir. Her bir eleman çifti için bir öğenin oyundan diğer bir öğenin oyunu kestiren en uygun *f* işlev değerini bulmaya çalışırız. Bu işlev her bir çift için farklı bir değer oluşturabilir. İçerik süzgeçleme modeli her bir kestirim için ortalama değer alarak doğru kestirim değerini oluşturmaya çalışır.

Bilimsel çalışma ve araştırmalardaki bazı örneklerde kullanılan algoritmada kestirim sistemleri

$$f(x) = x \quad \textbf{(4.2)}$$

şeklinde kullanılmıştır [28]. Bu algoritmanın istatistiksel regresyon tarzı ise

$$f(x) = ax + b \quad \textbf{(4.3)}$$

şeklindedir. Bu çalışmaların devamı yapıldığında

$$f(x) = ax^2 + bx + c \quad \textbf{(4.4)}$$

şeklinde sonuçlar elde edilebilir.

Bu tez çalışmasında $f(x) = x + b$ biçiminde Naïve kestirimler kullanılmıştır. Ayrıca Naïve ortalama da kullanılmıştır. Yapılan karşılaştırmalardan görülmüştür ki bağlanım tabanlı $f(x) = ax + b$ bellek tabanlı uygulamaya karşı $f(x) = x + b$ model tabanlı süzgeçleme çok yakın sonuçlar sağlamıştır.

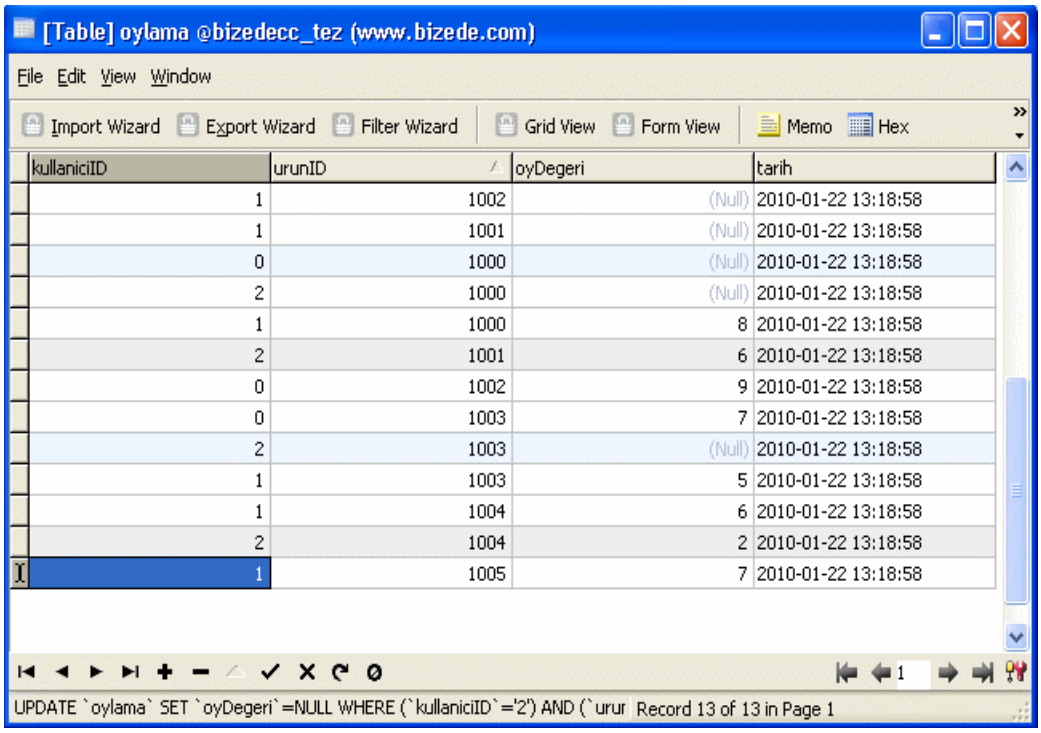
4.4 Süzgeçleme Tasarımlarında Kullanılan Tablo

Süzgeçleme tasarımlarında kullanmak üzere oylama tablosu oluşturulmuştur. Bu tablo;

- KullanıcıID:
- Ürün ID
- Oy Değeri
- Tarih

alanlarından oluşmaktadır.

Oylama tablosunda kullanıcıların ilgili ürünlere verdikleri oylamalar tarihleriyle birlikte saklanmaktadır. Şekil 4.1 de Oylama tablosunda 0 numaralı kullanıcı 1002 ve 1003 numaralı ürünlere sırasıyla 9 ve 7 oylarını verdiği, 1 numaralı kullanıcı 1000, 1003, 1004 ve 1005 numaralı ürünlere sırasıyla 8, 5, 6 ve 5 oylarını verdiği ve 2 numaralı kullanıcının 1001 ve 1004 numaralı ürünlere sırasıyla 6 ve 2 oylarını verdiği görülmektedir.



Şekil 4.1 : Süzgeçleme tasarımlarında kullanılan oylama tablosu

4.5 Kullanıcı Başına Ortalama İşbirlikçi Süzgeçleme Tasarımı

En basit kestirim algoritmalarından olan kullanıcı başına Ortalama İşbirlikçi Süzgeçleme tasarımı $P(u) = \bar{u}$ eşitliği ile gösterilir.

Bunun anlamı bir kullanıcı gelecekte vereceği oyu var olan oy ortalamasına bağlı olaraktan verecektir.

Şekil 4.1 deki tablo değerlerinden her kullanıcını ortalaması bulunduğunda:

0 numaralı kullanıcı (9+7)/2 = 8

1 numaralı kullanıcı (8+5+6+3)/4 = 6

2 numaralı kullanıcı (6+2)/2 = 4

olarak hesaplanır. Bu değerlere bağlı olaraktan kestirim tablosu aşağıdaki Şekil 4.2 deki gibi oluşturulmuştur.

[Table] tahmin-kullanici-basina-ortalama-dolu @bizedecc_tez (www.bizede.com)

File Edit View Window

Import Wizard | Export Wizard | Filter Wizard | Grid View | Form View | Memo | Hex | Image

kullaniciID	urunID	oyDegeri	tarih
2	1001	6	2010-01-22 13:18:58
1	1000	8	2010-01-22 13:18:58
1	1005	7	2010-01-22 13:18:58
0	1003	7	2010-01-22 13:18:58
1	1004	6	2010-01-22 13:18:58
2	1003	4	2010-04-22 14:13:14
0	1002	9	2010-01-22 13:18:58
1	1003	6	2010-04-22 14:13:35
2	1004	2	2010-01-22 13:18:58
2	1000	4	2010-04-22 14:13:13
0	1000	8	2010-04-22 14:13:47
1	1001	6	2010-04-22 14:13:33
1	1002	6	2010-04-22 14:13:36
0	1001	8	2010-04-22 14:13:45
0	1004	8	2010-04-22 14:13:44
0	1005	8	2010-04-22 14:13:38
2	1002	4	2010-04-22 14:13:11
2	1005	4	2010-04-22 14:13:10

SELECT * FROM `tahmin-kullanici-basina-ortalama-dolu` LIMIT 0, 1000 Record 16 of 18 in Page 1

Şekil 4.2 : Kullanıcı başına ortalama işbirlikçi süzgeçleme kestirim tablosu

4.6 Kişiselleştirilmemiş İşbirlikçi Süzgeçleme Tasarımı

Kullanıcı başına ortalama yöntemi uygulanabilecek en basit tasarımdır. Bu tasarımın bir adım geliştirilmiş hali diğer kullanıcıların da oy değerlerini de dikkate alarak yapılan ortalama eğilimli veya kişiselleştirilmemiş süzgeçleme sistemidir. Kişiselleştirilmemiş tasarım denklemi (4.5) de verilmiştir.

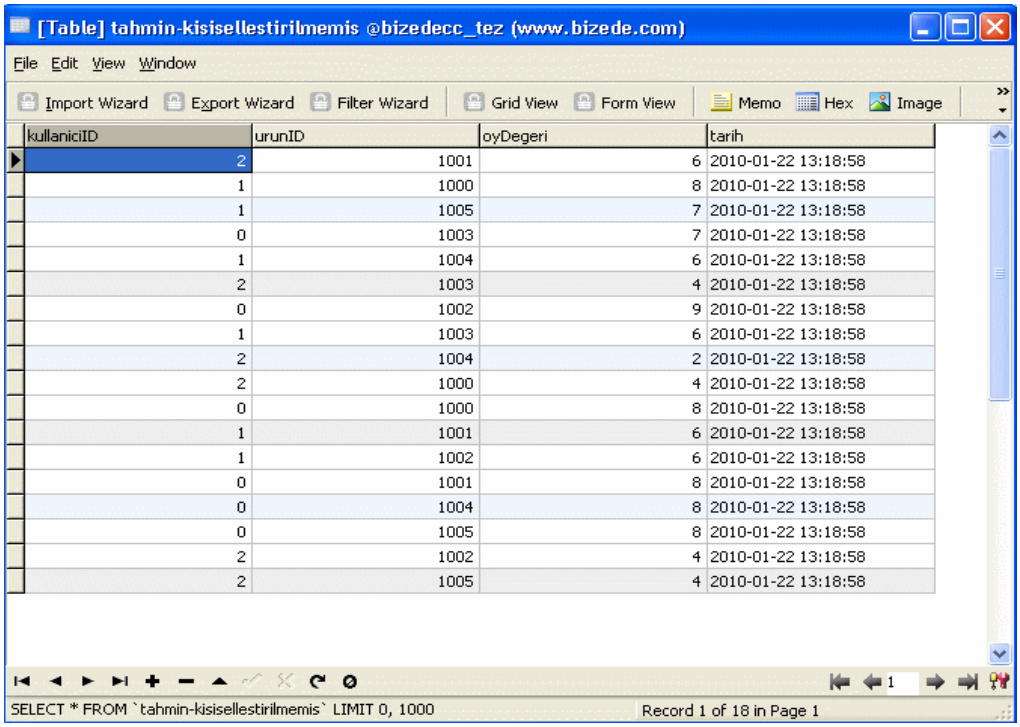
$$P(u)_i = \bar{u} + \frac{1}{\text{adet}(S_i(X))} \sum_{v \in S_i(X)} v_i - \bar{v} \qquad \textbf{(4.5)}$$

Bu kestirim tasarımı kullanıcının Bölüm 4.2 de belirtilen ortalama oy değerine ek olarak diğer kullanıcıların ilgili öğeye veya ürüne vermiş oldukları oyların ortalama sapma değerlerini de ekler.

Şekil 4.1 deki tabloda 0 numaralı kullanıcı 1004 numaralı ürüne oy vermediği halde 1 ve 2 numaralı kullanıcılar 6 ve 2 olarak oy verdiklerini görüyoruz. 0 numaralı kullanıcı için $\bar{u}$ 8 olarak önceden hesaplanmıştı. 4.5deki eşitlik uygulandığında

$P(u)_{1004} = 8 + ½ (2 – 4 + 6 – 4) = 8$ olarak hesaplanmıştır.

Şekil 4.1 deki tablo üzerine bu tasarım uygulandığında ortaya çıkan sonuç Şekil 4.3 deki gibi olur.



kullaniciID	urunID	oyDegeri	tarih
2	1001	6	2010-01-22 13:18:58
1	1000	8	2010-01-22 13:18:58
1	1005	7	2010-01-22 13:18:58
0	1003	7	2010-01-22 13:18:58
1	1004	6	2010-01-22 13:18:58
2	1003	4	2010-01-22 13:18:58
0	1002	9	2010-01-22 13:18:58
1	1003	6	2010-01-22 13:18:58
2	1004	2	2010-01-22 13:18:58
2	1000	4	2010-01-22 13:18:58
0	1000	8	2010-01-22 13:18:58
1	1001	6	2010-01-22 13:18:58
1	1002	6	2010-01-22 13:18:58
0	1001	8	2010-01-22 13:18:58
0	1004	8	2010-01-22 13:18:58
0	1005	8	2010-01-22 13:18:58
2	1002	4	2010-01-22 13:18:58
2	1005	4	2010-01-22 13:18:58

Şekil 4.3 : Kişiselleştirilmemiş işbirlikçi süzgeçleme tasarımı kestirim tablosu

4.7 Pearson Tasarımı

Bellek tabanlı olan Pearson süzgeçleme modeli bu tezde kullanılan EİS modeliyle karşılaştırmak amacı ile kullanılmıştır. Pearson modelinde kullanılan ilişki denklemi aşağıdaki gibidir:

$$Corr(u,v) = \frac{\langle u - \overline{u}, w - \overline{w} \rangle}{\sqrt{\sum_{i \in S(u) \cap s(w)} (u_i - \overline{u})^2 \sum_{i \in S(u) \cap s(w)} (w_i - \overline{w})^2}} \quad \textbf{(4.6)}$$

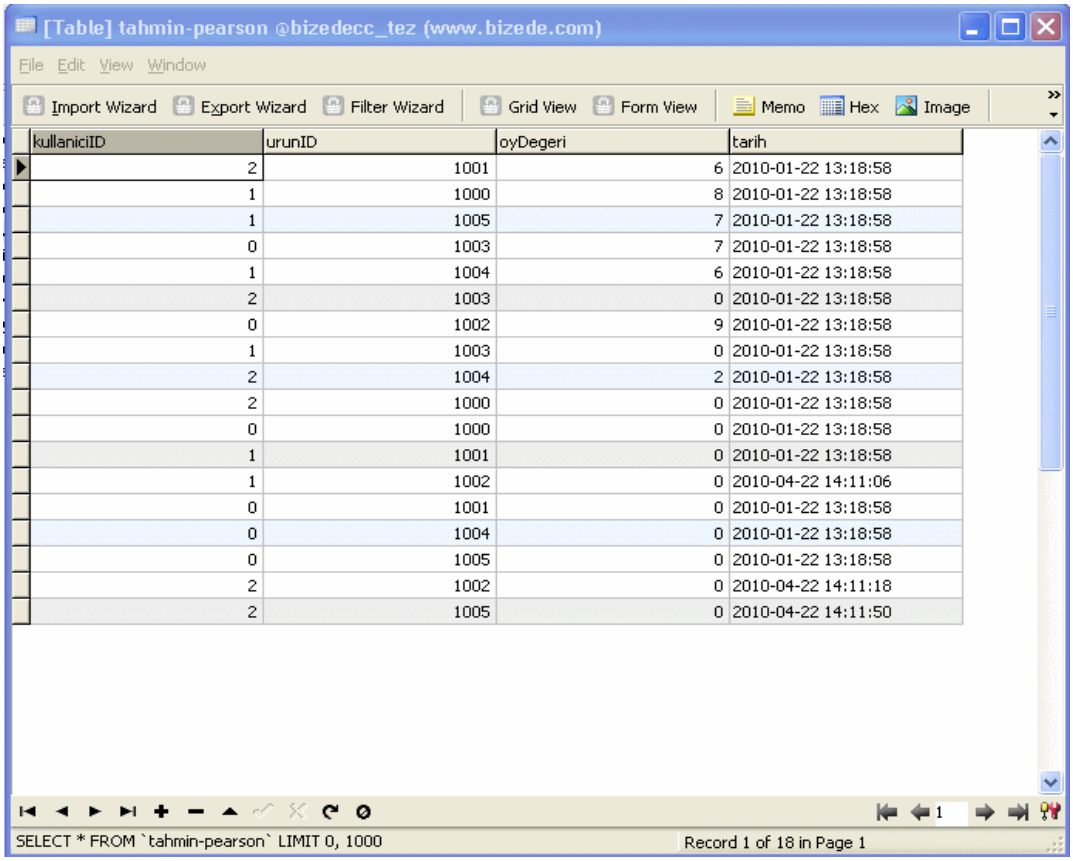
Pearson denklemi ise aşağıdaki şekildedir:

$$P(u)_i = \overline{u} + \frac{\sum_{v \in S_i(X)} \gamma(u,v)(v_i - \overline{v})}{\sum_{v \in S_i(X)} |\gamma(u,v)|} \quad \textbf{(4.7)}$$

Herlocker ve diğ. (1999) nin çalışmalarını göz önüne alarak ρ = 2.5 alınmıştır; ρ değeri olay güçlendirme kuvvetidir (OGK) (Case Amplificatin Power). OGK verideki gürültüyü azaltır. Eğer ilinti yüksek ise, örneğin Corr = 0.86, OGKdan sonra o zaman yüksek kalır ($0.86^{2.5}$ = 0.8), eğer Corr = 0.1 ise göz ardı edilebilir ($0.1^{2.5}$ = 0.003).

Pearson modeli OGK kullanılarak bu tarz uygulamalar için kabul edilebilir bir model olduğu ispatlanmıştır.

Pearson uygulanarak yapılan sınamalarda oylama tablosundaki sonuçlar Şekil 4.4 de verilmiştir.



kullaniciID	urunID	oyDegeri	tarih
2	1001	6	2010-01-22 13:18:58
1	1000	8	2010-01-22 13:18:58
1	1005	7	2010-01-22 13:18:58
0	1003	7	2010-01-22 13:18:58
1	1004	6	2010-01-22 13:18:58
2	1003	0	2010-01-22 13:18:58
0	1002	9	2010-01-22 13:18:58
1	1003	0	2010-01-22 13:18:58
2	1004	2	2010-01-22 13:18:58
2	1000	0	2010-01-22 13:18:58
0	1000	0	2010-01-22 13:18:58
1	1001	0	2010-01-22 13:18:58
1	1002	0	2010-04-22 14:11:06
0	1001	0	2010-01-22 13:18:58
0	1004	0	2010-01-22 13:18:58
0	1005	0	2010-01-22 13:18:58
2	1002	0	2010-04-22 14:11:18
2	1005	0	2010-04-22 14:11:50

Şekil 4.4 : Pearson sonucunda kestirim tablosu

4.8 Tek Eğimli İşbirlikçi Süzgeçleme Algoritmaları (TEİS)

Tek Eğimli İşbirlikçi Süzgeçleme (TEİS) modelinin yararı diğer kullanıcılar tarafından oylanmış aynı öğeyi ve de aynı kullanıcı tarafından oylanmış diğer öğeleri dikkate almasıdır. Bu modelin önemli bir özelliği diğer modellerde hesaba katılmayan verileri de dikkate almasıdır. Özellikle, sadece kestirim edilmesi istenen kullanıcı ile aynı öğelere oy verenlerin değerleri ve de kestirim edilen kullanıcı tarafından oy verilen öğelerin değerleri göz önüne alınır.

İki değerlendirme dizini olduğunu v_i ve w_i i=1….n olduğunu varsayarsak ve en iyi ve uygun kestirim modelinin tezimde belirttiğimiz $f(x) = x + b$ modeli olduğunu öngörebiliriz. $\sum_i (v_i + b - w_i)^2$ azaltarak da w ve v yi bulabiliriz. b ye bağlı olarak türev alındığında ve de türev 0'a eşitlendiğinde $b = \frac{\sum_i w_i - v_i}{n}$ olarak bulunur. Başka bir deyişle, b değeri iki dizini arasında ortalama fark olarak seçilir.

Çalışma kümesi χ, i ve j öğeleri, bazı kullanıcıların hesaplanmış u_j ve u_i oylari verilerek i öğesinin j öğesinden ortalama sapmasını aşağıdaki şekilde ifade ederiz:

$$dev_{j,i} = \sum_{u \in S_{j,i}(\chi)} \frac{u_j - u_i}{card\left(S_{j,i}(\chi)\right)} \quad \textbf{(4.8)}$$

Burada dikkat edilmesi gereken, herhangi bir kullanıcı değerlendirmesi olan u, u_j ve u_i içermiyorsa toplama katılmaz. $dev_{j,i}$ ile tanımlanan sistematik matris sisteme yeni bir veri eklendiğinde ve ya veri güncellendiğinde kolaylıkla hesap edilebilir.

$dev_{j,i} + u_i$ u_j ve u_i için bir kestirim olduğundan kabul edilebilir bir öngörü bütün bu öngörülerin ortalaması olabilir:

$$P(u)_j = \frac{1}{card\left(R_j\right)} \sum_{i \in R_j} \left(dev_{j,i} + u_i\right) \quad \textbf{(4.9)}$$

Burada $R_j = \left\{i \mid i \in S(u), i \neq j, card(S_{j,i}(\chi)) > 0\right\}$ ilgili tüm öğelerin kümesidir.

Tüm öğe çiftlerinin oylarının olduğu yoğun veri kümeleri için, yani $card(S_{j,i}(\chi)) > 0$, çoğu zaman $j \notin S(u)$ için $R_j = S(u)$ dur ve $j \in S(u)$ için $R_j = S(u) - \{j\}$ dir.

Birçok j için $\bar{u} = \sum_{i \in S(u)} \frac{u_i}{card(S(u))} \approx \sum_{i \in R_j} \frac{u_i}{card(R_j)}$ dir.

Tek eğimli işbirlikçi tasarım için öngörü eşitliği aşağıdaki şekilde basitleştirilir:

$$P^{S1}(u)_j = \bar{u} + \frac{1}{card(R_j)} \sum_{i \in R_j} dev_{j,i} \quad \textbf{(4.10)}$$

Tek eğimli işbirlikçi uygulamasında görüldüğü gibi modelimiz kullanıcının her bir öğeyi nasıl oyladığına bağlı değil, sadece kullanıcının oy ortalamasına ve kullanıcın hangi öğeleri oyladığına bağlıdır.

4.9 Ortalama Eğimli İşbirlikçi Süzgeçleme Algoritmaları (OEİS)

Bölüm 4.4 de açıklanan TEİS başarılı sonuçlar vermiş olsa da bütün yapılan oylamaları hesaba katmadığından ötürü tam istenen sonucu vermeyebilir.

Bir kullanıcının *j* öğesinin oyunu kestirebilmesi için ve kullanıcının *j* ve *k* öğeleri için vermiş olduğu oyların bilindiği bir durumdan, eğer 500 kullanıcı *j* ve *l* öğesine oy vermişse ve de 20 kullanıcı *k* ve *l* öğelerine oy vermiş ise o zaman kullanıcının *l* için *j*'ye bağlı kestirilen oyu *k*'ya bağlı kestirilen oyundan daha iyi olacaktır. Bu sebepten TEİS'i ortalama alarak uygulamaya koyduğumuzda ortalama eğilimli işbirlikçi süzgeçleme (OEİS) modeli ortaya çıkar.

$$P^{wS1}(u)_j = \frac{\sum_{i \in S(u)-\{j\}} \left(dev_{j,i} + u_i\right) c_{j,i}}{\sum_{i \in S(u)-\{j\}} c_{j,i}} \qquad \textbf{(4.11)}$$

$c_{j,i} = card(S_{j,i}(\chi))$ dir.

4.10 Süzgeçleme Tasarımlarına Ait Ortalama Hata Hesaplaması

Önceki bölümlerde açıkladığımız işbirlikçi süzgeçlerin ne kadar başarılı olduklarını ve etkilerini ölçebiliriz. İşbirlikçi süzgeçleme için bilimsel çalışmalarda ve araştırmalarda kullanılan birçok ölçüt bulunmaktadır [27].

Birçok bilim adamı ve araştırmacı MAE'yi kestirim hatalarını hesaplamak için kullanırlar [27,28]. Kullanıcı kümelerinde gerçek ve kestirilen değerler arasında en yaygın olarak kullanılan Ortalama Hata (All But One Mean Average Error (MAE)(OHÖ)) ölçütüdür. Biz bu tezde MAE'yi en yaygın kullanılan ölçüt olduğu, büyük ölçüde veri kümelerinde karşılaştırma yapmaya olanak sakladığı ve de diğer ölçütler ile ilişkili olduğu için kullandık.

MAE hesaplanırken, sınama kümesindeki bütün değerlendirmelerden her seferinde oylar saklanır. Bu adım yapılırken kullanıcı tarafından verilmemiş veya gizlenmiş oylar ön görülür.

$$MAE = \frac{1}{card(X')} \sum_{u \in X'} \frac{1}{card(S(u))} \sum_{i \in S(u)} | u^{(i)} - u_i | \qquad \textbf{(4.12)}$$

Bu denklemde $u^{(i)}$ u kullanıcısının i öğesine verdiği oylama değerlendirmesini göstermektedir.

$$MAE(u,\sigma) = \frac{\sum_{i \in S(u)-\sigma} \left| P(u_{|\sigma})_i - u_i \right|}{card(S(u)-\sigma)} \qquad \textbf{(4.13)}$$

$\sigma \subset S(u)$ bilinen kümenin öğesidir ve diğer gizli oylanmamış öğeler bu kümenin elemanı değildir.

Önemli bir MAE hata ölçütü, $S(u)$ kümesinden i öğesi çıkartılarak ve $\sigma = S(u) - \{i\}$ olarak eşitlendiğinde elde edilir. Yani $S(u)$ - $\sigma = \{i\}$ dir.

$$ALLBUT1(P,u,i) = \left|P(u_{|s(u)-\{i\}})\right|_i - u_i \quad \textbf{(4.14)}$$

AllBut1 hata ölçütü büyük veritabanları üzerinde yapılan çözümlemelerde en gerçekçi ölçümlerini vermektedir. Vektör öğeleri için, AllBut1 dışında başka hata ölçümleri de vardır [30].

AllBut1 MAE öngörülen P kümesindeki kestirim hatası olarak belirlendiğinde eşitlik aşağıdaki şekilde olur:

$$\frac{1}{card(X')}\sum_{u\in X'}\frac{1}{card(S(u))}\sum_{i\in S(u)} AllBut1(P,u,i) \quad \textbf{(4.15)}$$

Bölüm 4 de açıklanan kullanıcı başına ortalama, kişiselleştirilmiş işbirlikçi, Pearson, tek eğilimli işbirlikçi ve ortalama eğilimli işbirlikçi süzgeçleme tasarımları Bölüm 5.1 de verilen BX -Kitaplar tablosu üzerinde sınanmıştır. BX-Kitap-Oyları tablosundaki veriler kullanılarak 50,000 oy içeren χ çalışma kümesi ve 100,000 oydan oluşan χ' sınama kümeleri oluşturulmuştur.

Çizelge 4.1 aynı hata ölçümlerinin aynı veri kümeleri üzerine uygulanmış sonuçlarını gösterir. Beklendiği gibi Kullanıcı Başına Ortalama Süzgeçleme tasarımı en kötü sonucu vermiştir. Ortalama Eğimli İşbirlikçi Süzgeçleme tasarımı Pearson tasarımı dışında diğer bütün tasarımlardan daha başarılı sonuçlar vermiştir ve aynı zamanda Pearson tasarımına çok yakın bir değer ile sonuçlanmıştır.

Çizelge 4.1 : Süzgeçleme tasarımlarına ait hata ölçümleri

Süzgeçleme Tasarım	**AB1 MAE Kitap Verisi**
Kullanıcı Başına Ortalama	0.223
Kişiselleştirilmiş İşbirlikçi	0.211
Pearson	0.167
Tek Eğimli İşbirlikçi	0.179
Ortalama Eğimli İşbirlikçi	0.169

Çizelge 4.1 deki sonuçlardan da görülmektedir ki Ortalama Eğimli İşbirlikçi tasarımı gayet iyi sonuçlar vermektedir. Bu çözümlemelerin sonucunda, çevrimiçi sistemlerde, uygulanabilirliği daha karışık ve uzun olan Pearson algoritması yerine Ortalama Eğimli İşbirlikçi Süzgeçleme algoritmasının uygulanması daha uygun ve kabul edilebilir olacaktır. Bundan dolayı bu tezde Ortalama Eğimli İşbirlikçi Süzgeçleme tasarımı kullanılmıştır.

5. ORTALAMA EĞİMLİ İŞBRİLİKÇİ SÜZGEÇLEME UYGULAMASI

5.1 OEİS Uygulaması İçin Kullanılan Veri Kümesi

Bu tezde kullanılan veri kümesi www.informatik.uni-freiburg.de/~cziegler/BX/ adresinde ki kitap ve oy verilerinden oluşan Book-Crossing veri kümesidir.

BX*CSV dosyasında yer alan bilgiler aşağıdaki gibidir:

- 117,000 adet oy verisi
- 271,000 adet kitap
- 278,000 adet kullanıcı bilgisi

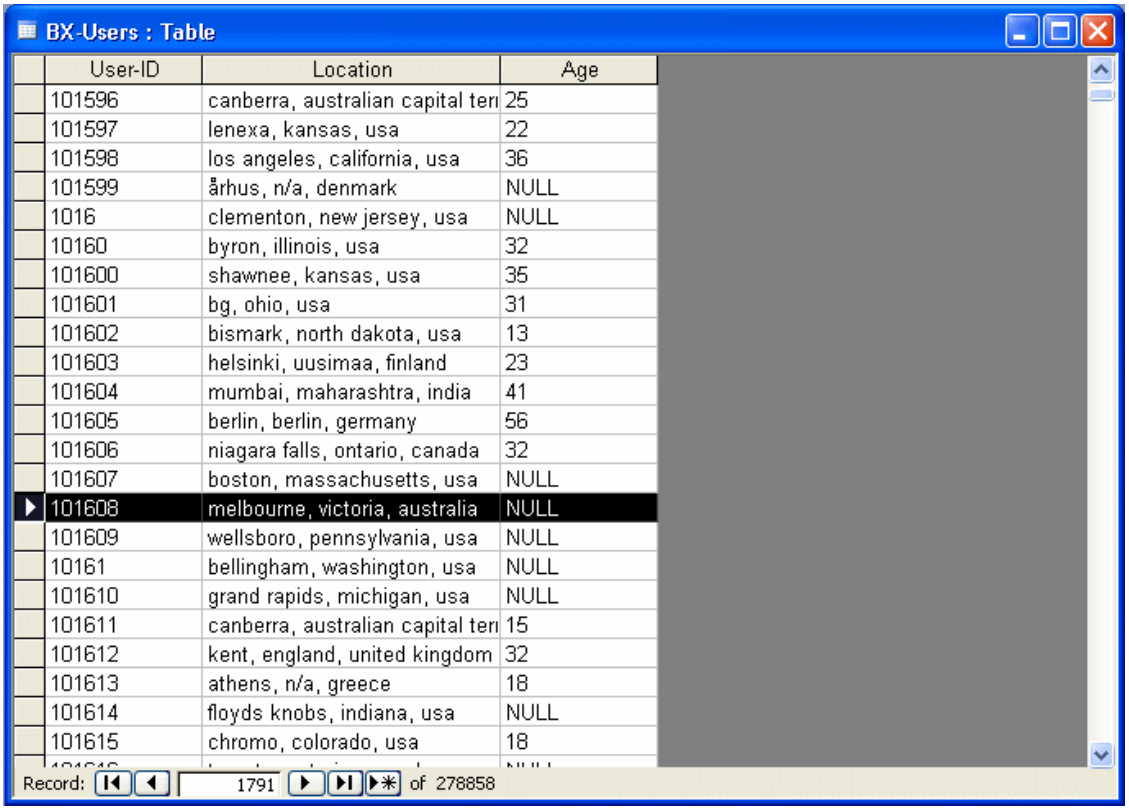
Bu verilerde oy verisi 30,000 adet kullanıcıya ait olduğu için bu tez bu 30,000 adet kullanıcının bilgileri kullanılarak yapılmıştır.

Book-Crossing veri kümesi 3 tablodan oluşmaktadır:

5.1.1 Kullanıcılar

Bu tabloda kullanıcı bilgileri yer almaktadır. Tabloda yer alan kolonlar:

- Kullanıcı Numarası (Birincil Anahtar)
- Yer
- Yaş

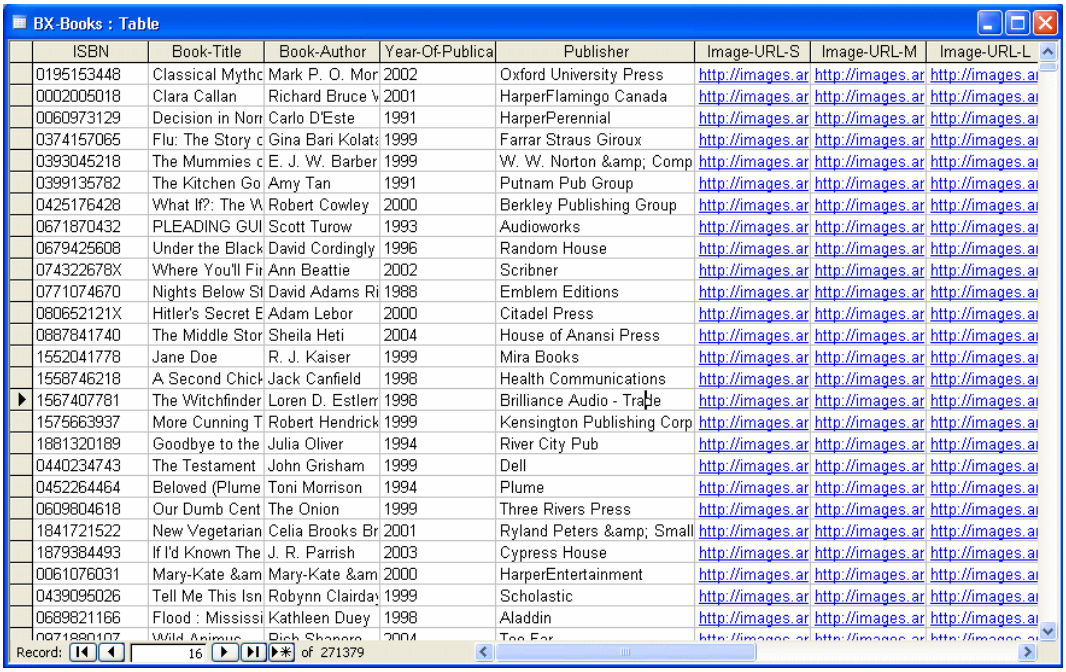


Şekil 5.1 : BX-Crossing veri kümesindeki kullanıcılar tablosu

5.1.2 Kitaplar

Bu tabloda kitap bilgileri yer almaktadır. Kitaplar ISBN numaraları ile tanımlanmıştır. Bu tabloda yer alan kolonlar:

- ISBN (Birincil Anahtar)
- Kitabın Adı
- Kitabın Yazarı
- Yayın Yılı
- Yayınevi
- Kitabın küçük boy resminin URL bağlantısı
- Kitabın orta boy resminin URL bağlantısı
- Kitabın büyük boy resminin URL bağlantısı



BX-Books : Table

ISBN	Book-Title	Book-Author	Year-Of-Publica	Publisher	Image-URL-S	Image-URL-M	Image-URL-L
0195153448	Classical Mythc	Mark P. O. Mor	2002	Oxford University Press	http://images.ar	http://images.ar	http://images.ai
0002005018	Clara Callan	Richard Bruce \	2001	HarperFlamingo Canada	http://images.ar	http://images.ar	http://images.ai
0060973129	Decision in Nor	Carlo D'Este	1991	HarperPerennial	http://images.ar	http://images.ar	http://images.ai
0374157065	Flu: The Story c	Gina Bari Kolat	1999	Farrar Straus Giroux	http://images.ar	http://images.ar	http://images.ai
0393045218	The Mummies c	E. J. W. Barber	1999	W. W. Norton & Comp	http://images.ar	http://images.ar	http://images.ai
0399135782	The Kitchen Go	Amy Tan	1991	Putnam Pub Group	http://images.ar	http://images.ar	http://images.ai
0425176428	What If?: The W	Robert Cowley	2000	Berkley Publishing Group	http://images.ar	http://images.ar	http://images.ai
0671870432	PLEADING GUI	Scott Turow	1993	Audioworks	http://images.ar	http://images.ar	http://images.ai
0679425608	Under the Black	David Cordingly	1996	Random House	http://images.ar	http://images.ar	http://images.ai
074322678X	Where You'll Fir	Ann Beattie	2002	Scribner	http://images.ar	http://images.ar	http://images.ai
0771074670	Nights Below St	David Adams Ri	1988	Emblem Editions	http://images.ar	http://images.ar	http://images.ai
080652121X	Hitler's Secret E	Adam Lebor	2000	Citadel Press	http://images.ar	http://images.ar	http://images.ai
0887841740	The Middle Stor	Sheila Heti	2004	House of Anansi Press	http://images.ar	http://images.ar	http://images.ai
1552041778	Jane Doe	R. J. Kaiser	1999	Mira Books	http://images.ar	http://images.ar	http://images.ai
1558746218	A Second Chick	Jack Canfield	1998	Health Communications	http://images.ar	http://images.ar	http://images.ai
1567407781	The Witchfinder	Loren D. Estlem	1998	Brilliance Audio - Trade	http://images.ar	http://images.ar	http://images.ai
1575663937	More Cunning T	Robert Hendrick	1999	Kensington Publishing Corp	http://images.ar	http://images.ar	http://images.ai
1881320189	Goodbye to the	Julia Oliver	1994	River City Pub	http://images.ar	http://images.ar	http://images.ai
0440234743	The Testament	John Grisham	1999	Dell	http://images.ar	http://images.ar	http://images.ai
0452264464	Beloved (Plume	Toni Morrison	1994	Plume	http://images.ar	http://images.ar	http://images.ai
0609804618	Our Dumb Cent	The Onion	1999	Three Rivers Press	http://images.ar	http://images.ar	http://images.ai
1841721522	New Vegetarian	Celia Brooks Br	2001	Ryland Peters & Small	http://images.ar	http://images.ar	http://images.ai
1879384493	If I'd Known The	J. R. Parrish	2003	Cypress House	http://images.ar	http://images.ar	http://images.ai
0061076031	Mary-Kate &am	Mary-Kate &am	2000	HarperEntertainment	http://images.ar	http://images.ar	http://images.ai
0439095026	Tell Me This Isn	Robynn Clairday	1999	Scholastic	http://images.ar	http://images.ar	http://images.ai
0689821166	Flood : Mississi	Kathleen Duey	1998	Aladdin	http://images.ar	http://images.ar	http://images.ai
0971880107	Wild Animus	Rich Shapero	2004	Too Far	http://images.ar	http://images.ar	http://images.ai

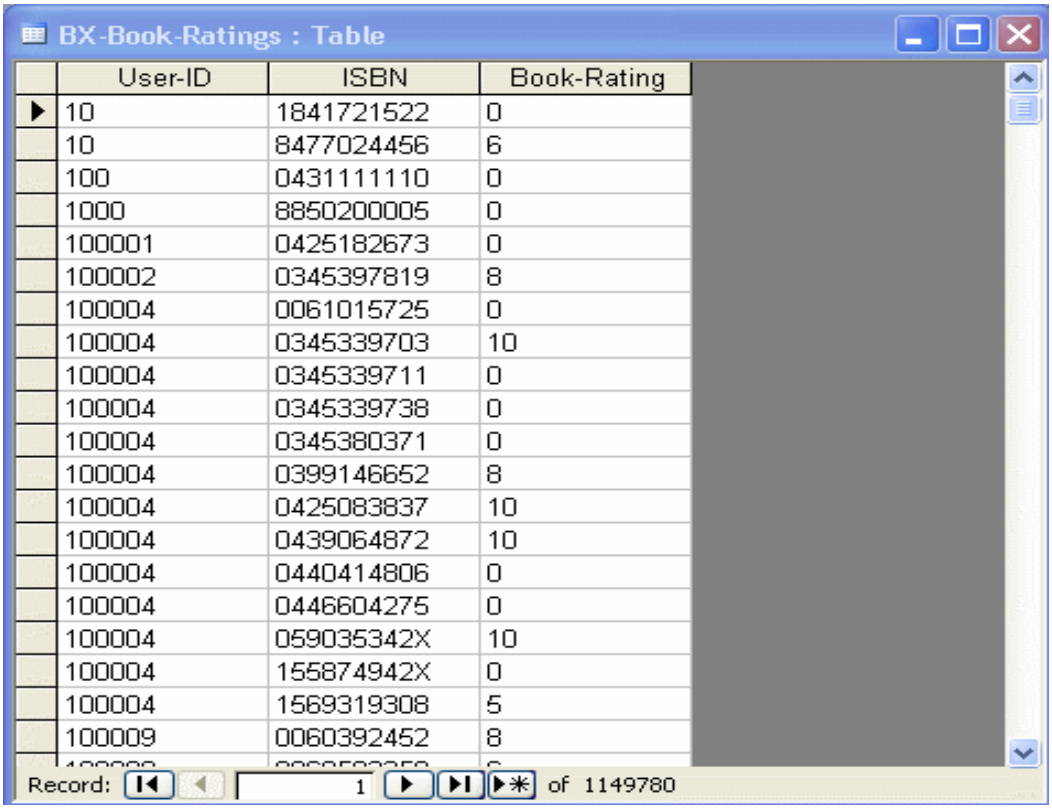
Record: 16 of 271379

Şekil 5.2 : BX-Crossing veri kümesindeki kitaplar tablosu

5.1.3 Kitap oyları

Bu tablo kitap oy bilgilerini içermektedir. Oylar açık veya kapalı olabilir. Oylar 1 ile 10 arasında derecelenmektedir. Yüksek oy yüksek beğeniyi gösterir. Bu tabloda yer alan sütunlar:

- Kullanıcı Numarası (Birincil Anahtar)
- ISBN (Birincil Anahtar)
- Kitap Oyu

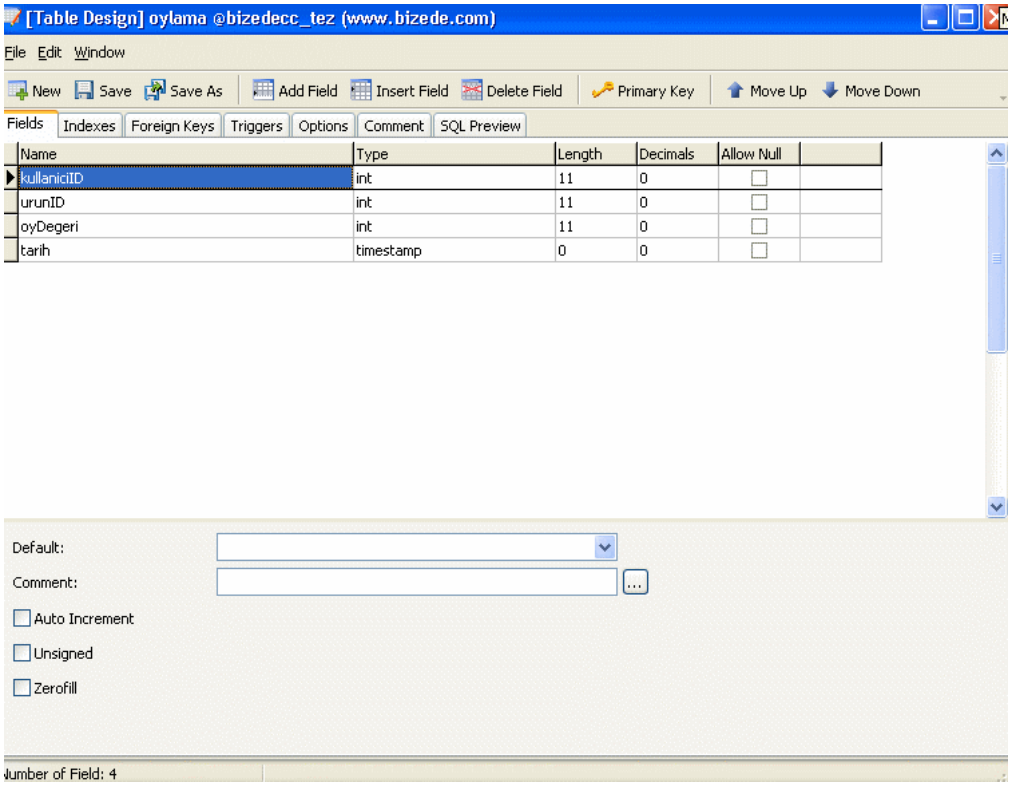


Şekil 5.3 : BX-Crossing veri kümesindeki oylar tablosu

5.2 OEİS Uygulaması

MANTAR uygulaması PHP ve MySQL kullanılarak gerçekleştirildiğinden, OEİS uygulaması da bu sistem üzerinde kurulmuştur.

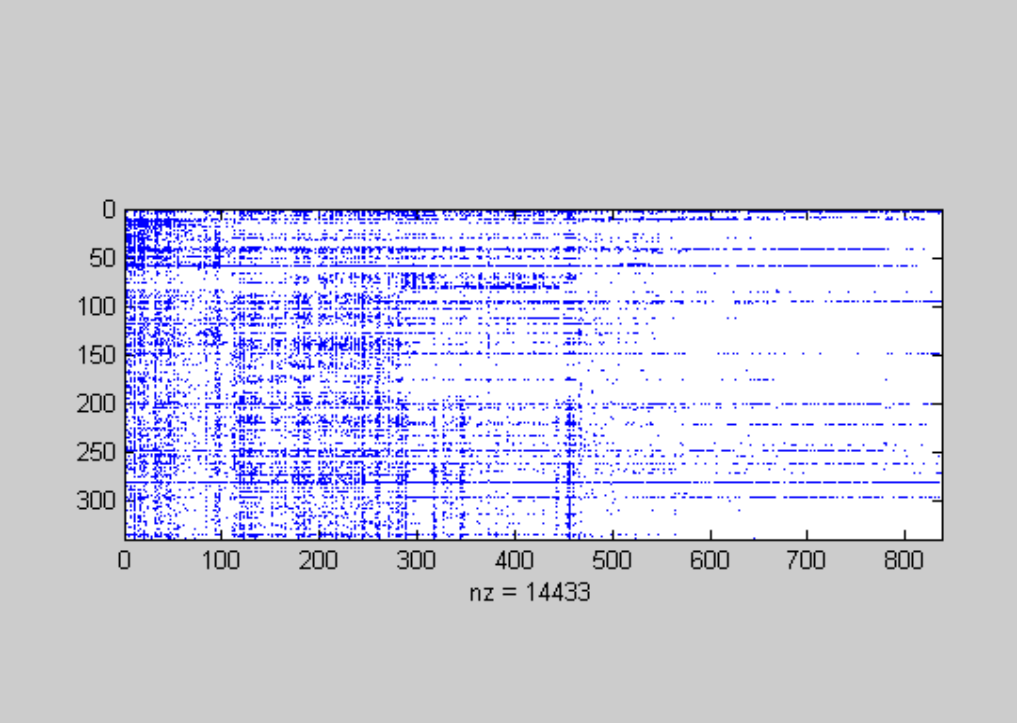
Bölüm 5.1 de kullanılan veri kümesinin çevrimiçi sistemlerde benzer şekillerde kullanılacağını varsayarak oylama tablosu Şekil 5.4 deki gibi oluşturulmuştur.



Şekil 5.4 : OEİS uygulamasındaki oylama tablosu

Oy değeri 1 ile 10 arasında değişen bir rakamdır. kullanıcıID kullancıları numaralarının olduğu başka bir tablo ile ilişkili olup birincil anahtardır (primary key).

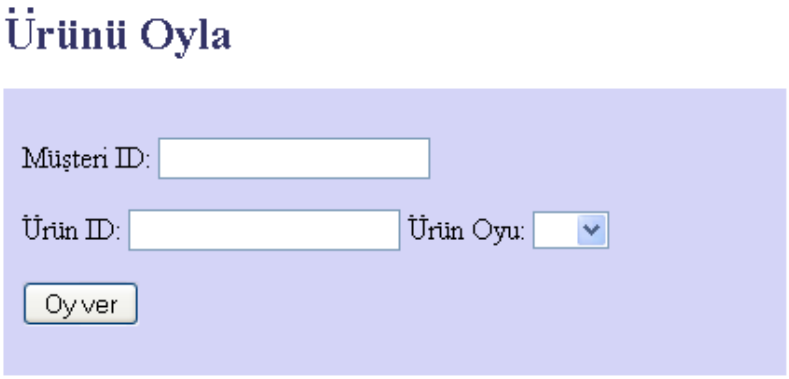
Şekil 5.4 deki tablo Bölüm 4.6 deki çözümlemelerde kullanılmıştır ve tez dosyasında verilen kitap-oy tablosundaki kısıtlı miktardaki veriler ile ilişkilendirilmiştir. BX-Kitap-Oyları tablosu bu yapıda oluşturulmuştur. Bu tabloya ait MySql Tablo oluşturma kodu CD'de verilmiştir.



Şekil 5.5 : Kullanıcılar arasındaki eksik bilgi deseni

Şekil 5.5 kullanıcılar arasındaki eksik bilgi desenini göstermektedir. Bu desendeki sıra uygun kullanıcıları, sütunlar uygun öğeleri göstermektedir. Noktalar ise kullanıcıların öğelere ait oylarını göstermektedir. Sıra ilk 1000 kullanıcıyı, sütun ise ilk 1000 kullanıcının oylarını verdiği öğeleri göstermektedir.

Kullanıcının yeni veya eski bir ürüne ait oy verebilmesi için www.bizede.com/oyver.php sayfası oluşturulmuştur. Bu ara yüzde kullanıcı ilgilendiği ürünün numarasını girerek o ürüne oyunu verebilir. Her yeni bir oylama girildiğinde oylama tablosu güncellenmektedir.



Şekil 5.6 : www.bizede.com/oyver.php sayfasındaki ürün oylama ara yüzü

Bölüm 4.5 de açıklanan OEİS ve Bölüm 4.7 de gözlemlenen çözümleme sonuçlarına bağlı olarak OEİS uygulamasını gerçekleştirmek için öncelikle TEİS yapısı oluşturulmuştur.

TEİS iki öğenin ortalama oylarının birbirinden çıkartılması ile yani çoğunluk farklarına dayalı hesaplanmaktadır. TEİS'e ait bir örnek aşağıdaki çizelgede yer almaktadır.

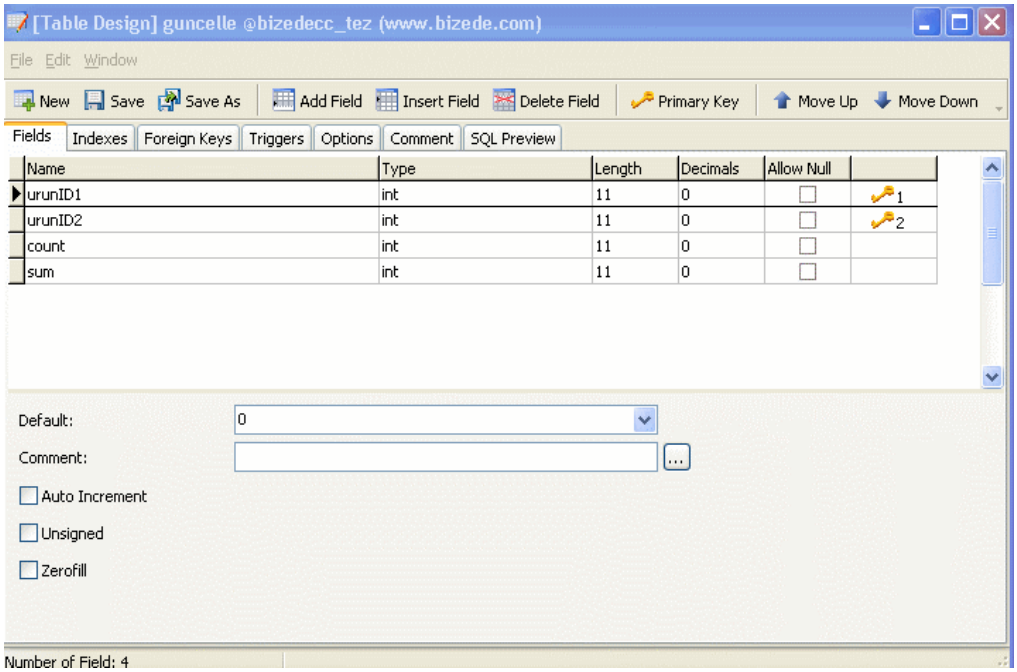
Çizelge 5.1 : TEİS çoğunluk farklarına bir örnek

Number	**Item1 Rate**	**Item2 Rate**
1	8	3
2	6	5
3	7	7
4	6	7
5	6	4
6	4	6
7	8	5
8	5	7
9	9	6
10	6	5
SUM	**65**	**55**

Çizelge 5.1 de görüldüğü üzere 1. öğenin ortalaması 65, 2.öğenin ortalaması 55 dir. Toplam 10 tane oylama olduğu için 1. ve 2. öğe arasındaki ortalama fark 1 dir. Bu sonuca göre 1.öğe 2.öğeden ortalamada 1 oy daha fazladır.

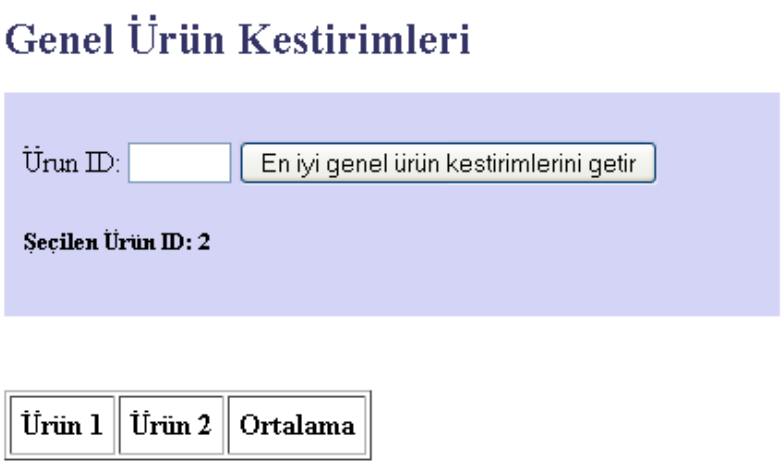
Çizelge 5.1 e öğe matrisi denir. Şekil 5.7 de görüldüğü gibi bu matris herhangi bir öğe çiftine ait oyların COUNT (adet) ve SUM (toplam) değerleri hesaplanarak Güncelleme tablosuna kayıt edilmesi ile oluşur. Güncelleme tablosuna ait MySql Tablo güncelleme kodu CD'sinde verilmiştir.

Her seferinde yeni bir kullanıcı oy verdiğinde güncelleme tablosu yenilenir. Oy güncelleme PHP kodu CD'de verilmiştir.



Şekil 5.7 : OEİS uygulamasındaki güncelleme tablosu

Tavsiyeler listesini ortaya çıkarmak için ise kullanıcının bir ürünü nasıl değerlendireceği göz önüne alınması gerektiğinden bir kestirim işlevi oluşturulması gerekmektedir. Bu işlev OEİS modeli uygulanarak oluşturulmuştur. Bu oluşturulan kestirim listesine ait ara yüz www.bizede.com/genel.php sayfasında yer almaktadır. Bu ara yüzde kullanıcı ilgili ürün numarasını girdikten sonra bu ürünün diğer ürünler ile ilgili ortalamaları verilmektedir.



Şekil 5.8 : Genel ürün tavsiyeler listesi

Bölüm 4 de açıklandığı gibi, öneri algoritmaları, öğelerin oyları arasında düzlemsel bir ilişki var olduğunu düşünerek başlar. Bu yüzden, genel olarak j öğesine ait tercihi, i öğesinin tercihine bağlı olarak kestiririz. Bu durumda algoritma önemli ön işlemler içermektedir ki bunlar öğe/öğe oylama değerleri arasındaki farkları hesaplamaktadır. Aşağıda bu algoritmaya ait sözde program verilmiştir:

her *i* öğesi için
  her *j* öğesi için
    her *i* ve *j* ye oy veren *u* kullanıcısı için
      u kullanıcısının *i* ve *j* için verdiği oyların farkını ortalamaya **ekle**

Aşağıda ise öneri sistemine ait algoritma verilmiştir:

u kullanıcısının oylama yapmadığı **her** *i* öğesi için
  u kullanıcısının oylama yaptığı **her** *j* öğesi için
    i ve *j* arasındaki oylama farklarının **ortalamasını bul**
    farkı *u* kullanıcısının *j* için yaptığı oylama değerine **ekle**
    bunu hesaplanan ortalamaya **ekle**
bu ortalamalar ile sıralanmış öğelerin en başına **dön**

Yukarıdaki öğe/öğe fark algoritması ve de öneri sistemi algoritması kullanılarak www.bizede.com/kisisel.php sayfası oluşturulmuştur. Bu ara yüzde kullanıcı numarası ve istenilen ürün numarası girilerek kişiye ait o ürünle ilgili oy kestirimi verilmektedir.

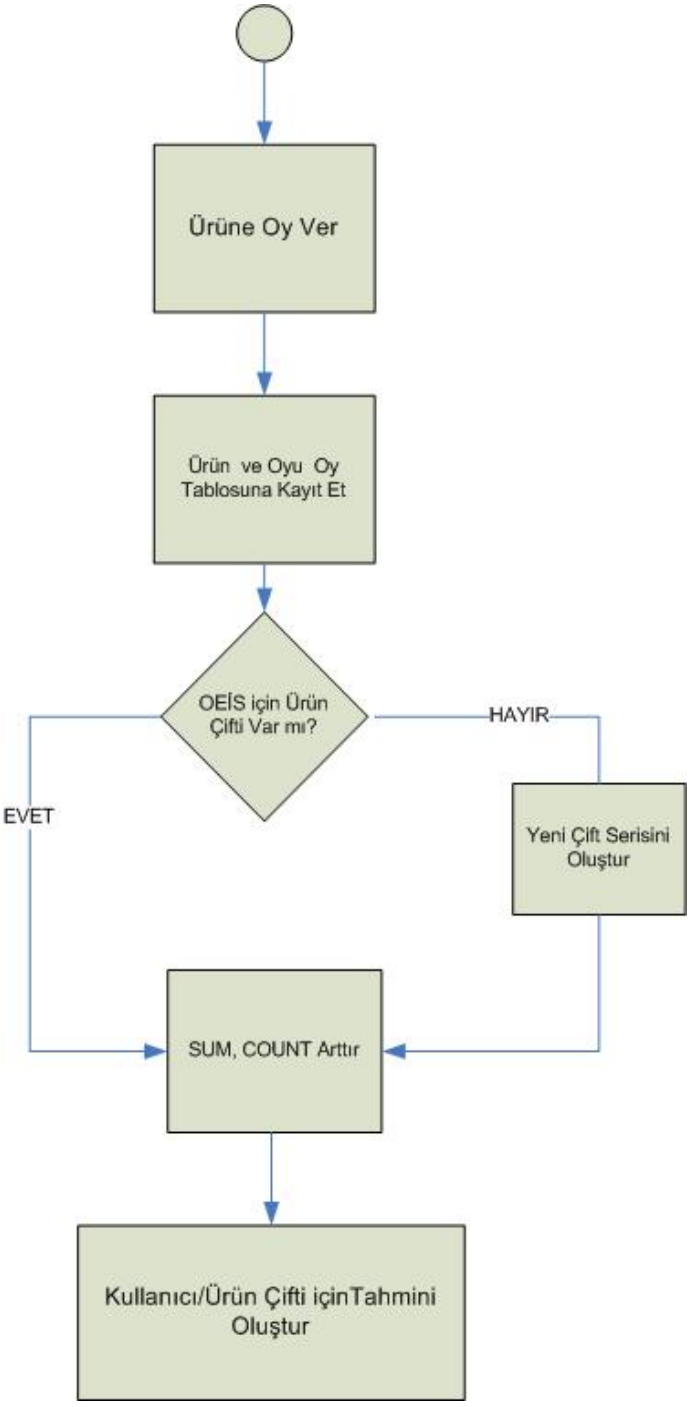


Şekil 5.9 : Ortalama eğimli süzgeçleme oy kestirimi

Ortalama Eğimli İşbirlikçi Süzgeçleme sisteminin uygulamasının akış şeması Şekil 5.10 da verilmiştir.

Uygulama kullanıcın ürüne oy vermesi ile başlar. Herhangi bir kullanıcı bir ürüne oy verdiği zaman, kullanıcı, oy verilen ürün ve oy bilgileri oy veri tabanına eklenir. Oy veri tabanına ekleme işlemi gerçekleştikten sonra, Ortalama Eğimli İşbirlikçi Süzgeçleme yapısına uygun oluşturulmuş kestirim çifti veri tabanında sorgulanır. Eğer oy verilen ürün için daha önceden oy verilmiş ise, bütün ilgili ürün çifteri veri tabanında mevcut olduğundan sadece ilgili ürün çiftinin oy sayısı arttırılır ve ortalama eğilimli işbirlikçi süzgeçleme farkları hesaplanarak veri tabanı yeni değerlerle güncellenir. Eğer oy verilen ürün ile ilgili ürün çiftleri mevcut değilse, kestirim veri tabanı bütün yeni ürün ve eski ürün çiftleri ile güncellenir.

Bu işlemler tamamlandıktan sonra kullanıcın en çok tercih edeceği veya var olan herhangi bir ürün için kullanıcının ne kadar oy vereceği sistem tarafından gösterilir.



Şekil 5.10 : OEİS uygulaması akış şeması

6. SONUÇ VE ÖNERİLER

Çoklu site içerik yönetim sistemleri farklı içerik yönetim sistemleri kullanılarak uygulanabilir. Joomla, yapısı, açık kaynak kod kurulumu ve kullanılabilirliği itibarı ile kuramsal ve uygulama olarak onlarca siteyi tek bir kurulum ile yönetme olasılığını sağlar. Ancak kurulum adedi tanımlanacak alan adı ve eklenecek içerik yazılarına bağlı olarak belirli sayıda sitenin tek bir kurulum üzerinden yönetilebilmesi veri tabanına ulaşım ve daha sonraki bölümlerde uygulanan içerik süzgeçleme sistemlerinin verimli çalışabilmesi açısından önemlidir. Önerilen çoklu site içerik yönetim sistemi her bir sitede 10 yazı ve 100 ürün olduğu varsayılarak en fazla 20 siteyi yönetebilecek kapasiteye sahip olduğu gözlemlenmiştir. Bu işlem veri tabanında 20 x 10 = 200 yazı ve 20 x 100 = 2000 ürün ve ilgili bilgi girişleri yapılarak bir program ile düzenlenmiştir. Çoklu içerik yönetim sistemi kullanılarak oluşturulan 4 sitenin kurulum süresi eklenen 10ar tane yazı ile beraber 23 saat sürmüştür. Bu sonuç, 4 sitenin teker teker kurulduğu ve 10ar tane yazının eklendiği 48 saatlik kurulum süresine göre %52 verimlilik kazancı olacağını göstermiştir. Çoklu site içerik yönetimi sistemi kullanılarak daha fazla site kurulduğunda toplam harcanacak süre daha az olduğundan verimlilik %90lara kadar çıkabilecektir. İleride yapılabilecek çalışmalarda çoklu içerik site yönetim sistemlerinin veri tabanı ulaşım hızları sınanarak kurulabilecek en iyi site sayısı belirlenebilir. Büyük veri tabanlı ve çok kullanıcılı çoklu site içerik yönetim sistemlerinde sunulan yapı uygulanabilir ve başarıma eriştirilebilinir. Kullanılan algoritmalar, oylama bilgileri çözümlenerek, kullanıcılar öbeklere ayrılarak ve kullanıcı geri beslemeleri göz önüne alınarak bu tezde sunulmuş olan çoklu site içerik yönetimi ve süzgeçleme yöntemleri geliştirilebilinir ve daha büyük ve esnek sistemler oluşturulabilinir.

Kullanıcılara sunulan bilgi miktarı ve çeşitliliği artıkça, kullanıcıların ilgilenecekleri bilgi ve ürünleri bulmalarını sağlayacak yöntemlerde artacaktır. Akıllı algoritmik tasarım ve yapılar kullanıcı özelliklerini öğrenip, bu sorunu çözümlemede yardımcı olabilirler. Bu yapılar işbirlikçi, içerik tabanlı ve kişisel süzgeçleme yöntemlerinin ayrı ayrı veya birlikte karma kullanılması ile kuramsal olarak sağlanabilir. Çevrimiçi uygulamalarda bu modeller sistemlere bütünleştirilebilinir ancak verimlilik ve tezde belirtilen kıstaslar göz önüne alındığında Ortalama Eğimli İşbirlikçi Süzgeçleme en uygun süzgeçlerden olduğu sonucuna varılmıştır. Seçilen kıstaslar çevrimiçi uygulamalar için zorunlu olması gereken özellikler için belirlenen kıstaslardır.

Çeşitli işbirlikçi süzgeçlerin ne kadar başarılı oldukları ölçülmüş ve uygulama kolaylıkları karşılaştırılmıştır. Ölçüm işlemi ve kestirim hatalarını hesaplamak için en yaygın olarak kullanılan Ortalama Hata Ölçütü (OHÖ) kullanılmıştır. Kitap ve oy veri tabanları üzerinde uyguladığımız Tek Eğimli İşbirlikçi Süzgeçleme ve Ortalama Eğimli İşbirlikçi Süzgeçleme algoritmaları sırası ile 0.169 ve 0.179 OHÖ hata değerleri vermiştir. Bu Ortalama Hata Ölçütü hata değerleri Kullanıcı Başına Ortalama Süzgeçleme ve Kişiselleştirilmiş İşbirlikçi Süzgeçleme algoritmalarının hata değerlerinden, 0.223 ve 0.211'den daha iyi ve 0.167 hata değerli Pearson Süzgeçleme algoritması kadar iyi sonuçlar vermiştir. OHÖ hesaplanırken, sınama kümesindeki bütün değerlendirmelerden her seferinde oylar saklanmış ve bütün verilmiş 60,000 oy ve 50,000 oy verilmemiş ürün dikkate alınmıştır. Ortalama Eğimli İşbirlikçi Süzgeçleme diğer tekil süzgeçleme yöntemleri ile karşılaştırıldığında, daha başarılı sonuçlar verdiği görülmektedir. Bu sonuçlar doğrultusunda, Ortalama Eğimli İşbirlikçi Süzgeçleme (OEİS) Çoklu İçerik Yönetim Sistemi'ne (ÇİY) kurulup uygulanmıştır. OEİS ile ilgili 3 farklı PHP kodu, 1 MySQL veri tabanı, 2 tablo ve yeni veriler oluşturularak çevrimiçi ÇİY üzerinde uygulama çalıştırılıp sonuçlar beklendiği üzere gözlemlenmiştir.

.

Büyük veritabanlarının kullanıldığı, kullanıcıların oyları dışında kişisel bilgileri ve tercihlerini belirttiği veriler kullanılarak, Ortalama Eğimli İşbirlikçi Süzgeçleme ile beraber başka süzgeçleme yöntemleri ile oluşturulan yeni karma algoritmalar oluşturulabilinir. Yeni oluşturulacak karma algoritmalar çok daha kesin sonuçlar verebilir ve çevrimiçi sistemlerde de uygulanabilir. Elde edilen sonuçlar Ortalama Hata Ölçütü değerleri ile karşılaştırılabilir. Bu tezde kullanılan kitap ve oy bilgilerine uygulandığında 0.160'dan düşük olması durumda daha iyi sonuçlar elde edilmiş olur. Bu tarz bir uygulama için başarımın kabul edilebilir olması ve tasarımın üç veya dört farklı süzgeçleme algoritmasının ortak ve verimli çalışabilecek şekilde geliştirilmesi ile mümkün olabilir.

KAYNAKLAR

[1] **Amrani, M. Y. E., S. Delisle, and I. Biskri ,** 2001. Coping with Information Retrieval Problems on the Web, *Towards Personal Web Weaver `Agents* . In: IC-AI'01. pp. 1225–1231.

[2] **A. Popescul, L. H. Ungar, D. M. Pennock, andS. Lawrence,** 2001 Probabilistic models for unified collaborative and content-based recommendation insparse-data environments. *In Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence.*

[3] **Dai W., 2008.** Technology innovation for SME users, *IEEE International Conference on Service Operations and Logistics, and Informatics* , Melbourne, October 12-15.

[4] **Swierad D., Zabierowski W., Napieralski A.,** 2009. Universal Content Management system with Ajax Technology**,** *10th International Conference - The Experience of Designing and Application of,* Lviv-Polyana, February 24-28.

[5] **Suna M.,** 2002. Business Benefits of Content Management, *Masters Thesis, Department of Industrial Engineering and Managementof,* Lappeenranta University of Technology Lappeenranta, Finland.

[6] **Url-1 <** *http://www.sanayi.gov.tr*>, alındığı tarih 28.09.2009.

[7]**Alekseev N.A., Bondarenko V.Y.,** 2009. Comparative analysis of web-site content management systems, *19th International Crimean Conference,* Sevastopol, September 14-18.

[8] **Giri K. K., Nirgude K. R.,** 2009. Open Source Content Management Software: A Comparative Analysis, *7th International CALIBER-2009*, Ahmedabad, February 25-27.

[9] **Url-2** < *http://www.joomla.org*>, alındığı tarih 12.10.2009.

[10] **Url-3** < *http://www.drupal.org*>, alındığı tarih 15.10.2009.

[11] **Url-4** < *http://www.plone.org*>, alındığı tarih 20.10.2009.

[12] **Deshpande M., Karypis G.,** 2001**,** Item-Based Top-N Recommendation Algorithms, Proceedings *of the 10th international conference on World Wide Web*, Hong Kong.

[13] **Url-5** <*http://en.wikipedia.org/wiki/Content_filtering*>, alındığı tarih 20.12.2009.

[14] **Burke R.,** 2002. Hybrid Recommender Systems, *User Modeling and User-Adapted Interaction, Volume 12, Issue4,* Fullerton, USA, November 2002.

[15] **Schafer J. B.,** 2005, A Dynamic User-Interface for a Meta-Recommendation System, *A Workshop on the Next Stage of Recommender Systems Research*, San Diego, California, USA, January 2005.

[16] **Url-6** <*www.springerlink.com/content/*>, alındıgı tarih 12.01.2010.

[17] **Url-7** <*http://www.netflixprize.com//index*>, alındıgı tarih 15.02.2010.

[18] **Pazzani1 M. J., Billsus D.,** 2007. Content-based Recommendation Systems, *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, May 2007.

[19] **Melville P., Mooney R., Nagarajan R.J.,** 2002. Content-Boosted Collaborative Filtering for Improved Recommendations, *Proceedings of the 18th National Conference on Aritificial Intelligence (AAAI-2002)*, July 2002, Edmonton, Canada.

[20] **Salton G,**2000. Automatic Text Processing. *Addison-Wesley* (1989).

[21] **Balabanovic M.,** (1997), An Adaptive Web Page Recommendation Service, *First International Conference on Autonomous Agents*, Marina del Rey,CA.

[22] **Claypool M., Gokhale A., Miranda T., Murnikov P., Netes D. and Sartin M.,** 1999. Combining content-based and collaborative filters in an online newspaper, *ACM SIGIR Workshop on Recommender Systems*, Berkeley, CA, August 19, 1999.

[23] **Billsus D. and Pazzani M. J.,** 1998, Learning collaborative information filters, *15th International Conference on Machine Learning*. 46-54.

[24] **Hofmann T.,** 2004. "Latent semantic models for collaborative filtering," *ACM Transactions on Information Systems.* 22(1):p.89-115.

[25] **Pazzani M. J.,**1999, A Framework for Collaborative, Content-Based and Demographic Filtering, *Artificial Intelligence Review*, 13, 393 - 408 .

[26] **Weiss S.M. and Indurkhya N.,**2001,Lightweight collaborative filtering method for binary encoded data, Principles *of Data mining and Knowledge Discovery*, 484-491.

[27] **Breese J. S., Heckerman D., and Kadie C.,** 1998, Empirical analysis of predictive algorithms for collaborative filtering, *In Fourteenth Conference on Uncertainty in AI,*Morgan Kaufmann,July 1998.

[28] **Sarwar B. M., Karypis G., Konstan J. A., and Riedl J.,2001**, Itembased collaborative filtering recommender algorithms, *Proceedings of the 10th international conference on World Wide Web*, p.285-295, May 01-05, 2001, Hong Kong, Hong Kong

[29] **Herlocker J., Konstan J., Borchers A., and Riedl J.,** 1999, An Algorithmic Framework for Performing Collaborative Filtering, *Research and Development in Information Retrieval.*

[30] **Canny J.,**2002, Collaborative Filtering with privacy via Factor Analysis, *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, p.238-245.

EKLER

Ek programlar CD de verilmiştir.

ÖZGEÇMİŞ

Ad Soyad:	**Zübeyde Deniz ALAY**
Doğum Yeri ve Tarihi:	**Bornova, Ocak 22, 1979**
Adres:	**190 Columbus Pl, Cliffside Pk, NJ, 07010**
Lisans Üniversite:	**İTÜ**

1979 Bornova, İzmir doğumlu olan Zübeyde Deniz ALAY; lise öğrenimini Haydarpaşa Lisesi'nde tamamlamıştır. 1997 yılında lise mezuniyeti sonrasında İstanbul Teknik Üniversitesi Bilgisayar Mühendisliği Bölümünü kazanmış ve 2002 yılında bu bölümden lisans derecesini almıştır. 2004 yılında Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Yüksek Lisans Programı'na kayıt olmuş ve Mayıs 2010 itibariyle yüksek lisans tezini sunmaktadır.