

# Google Bağımlılığı

**İbrahim C. Arkut<sup>1</sup>**

Yönetim Bilişim Sistemleri  
Girne Amerikan Üniversitesi, KKTC

**Refik C. Arkut<sup>2</sup>**

Bilgisayar Mühendisliği  
Kadir Has Üniversitesi, Türkiye

## ➤ Giriş

Geçen yıl inet'02'de internet davranışı ve Nash dengesini ele almıştık; bu yıl da geleneği bozmayıp alışlagelmişliğin dışında Internet'in kullanıcı açısından neye *bağımlı* olduğunu veya olacağını irdeleyeceğiz. Bu konuyu ele alırken kendisini her geçen gün biraz daha etkin biçimde gösteren internet arama motorlarını ve bunlar içerisinde 1998'de ortaya çıkmasına karşın bizler için vazgeçilmez olan Google<sup>3</sup> arama motorunu seçtik. Google'ın başarı öyküsü ve geri planda ne gibi özelliklere dayandığını aktarırken en azından bu çalışmanın yapılmasının dürtüsü Google ile New York Times gazetesinin arasındaki bir bilişim olayı sürtüşmesi ve durduk yerde yazılım devi Microsoft'un Google'u boy hedefi (esaslı rakip) göstermesi olmuştur. Çalışmanın sonucunda tartışmaya açık şu soru ele alınmıştır: *Bilgi internet'te aranırken acaba isteğimizin dışında bizi yönlendiriyor mu?*

## ➤ Google Farkı

Google daha fazla siteyi daha süratli tarar ve en fazla ilgili sonuçları kullanıcıya ulaştırır. Bunu nasıl başardığını anlamak için kullanılan yazılım ve donanım teknolojisine göz atmak gerekir. Kısaca aramanın başarılı ve süratli olması 'arama-algoritmasının' etkinliği ve bir birlerine bağlı birçok kişisel bilgisayarın oluşturduğu çok süratli arama motoruna bağlıdır. Yazılım Google kurucuları<sup>4</sup> Larry Page ve Sergey Brin (Stanford) tarafından geliştirilen web sayfalarını derecelendiren PageRank<sup>TM</sup> ve bunun ağ üzerinde çalıştırılmasına dayanmaktadır. PageRank<sup>TM</sup> sürekli olarak Google yazılımcıları tarafından geliştirilmekte (ince ayar) ve web aramanın temel unsuru olmaktadır. PageRank web'in demokratik karakterinine dayanmakta ve onun yoğun link yapısını herhangi sayfanın değerininin bulunmasında kullanmaktadır.

---

<sup>1</sup> iarkut@gau.edu.tr

<sup>2</sup> rca@ttnet.net.tr

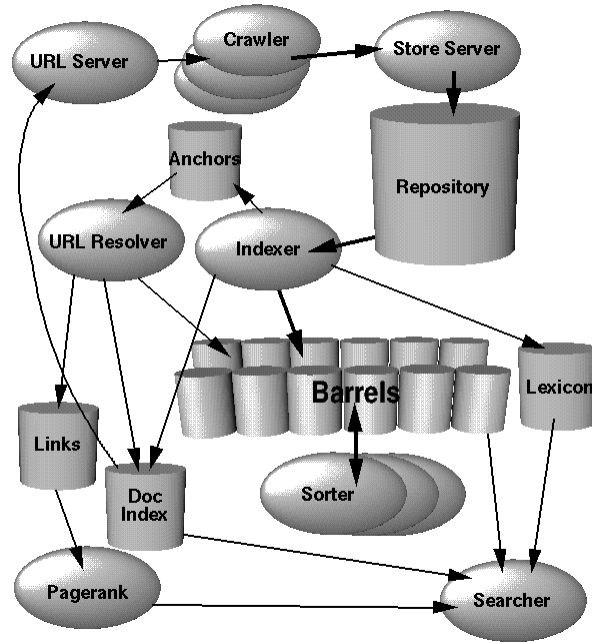
<sup>3</sup> *American Express* kredi kart için 'Don't leave home without it' sloganı Google için 'Don't surf without it' çağrıştırıyor.

<sup>4</sup> Hayırlı öğrenciler Stanford'taki hocaları R. Motwani, J. Ullman ve T. Winogradı unutmadılar.

Aslında Google A sayfasından B sayfasına bir linki bir *oy* olarak algılar ( A sayfası ile B sayfası için) ve aynı zamanda oy'un olduğu sayfanın analizini yapar. Bu şekilde oy büyüklüklerine göre derecelendirilmiş siteler yüksek PageRank değeri alırlar ve her seferinde Google aramasında hatırlanırlar. Başka bir deyişle sizin sayfanızın önemi başka sayfalardan sizin linklerinize ne kadar refererans yapıldığına bağlıdır. Bunun yanında Google karmaşık yazı-uydurma tekniklerini de kullanarak aramanızdaki sonuçların önemli ve ilgili olmasını sağlar. Google'ın karmaşık ve otomatik arama metodların dıştan değiştirilmesinin oldukça zorlaştırmaktadır. Bu konuya daha ilerde tekrar değineceğiz. Burada Google'ın arama sonucuna uygun reklamları ekrana gönderdiğini fakat para karşılığı PageRank yükseltmesinin (sanal şike) mümkün olmadığını veya çok zor olduğunu beyan etmektedir. Bu bakımdan Google arama sonuçları ilgili yüksek-kaliteli web-sitelerini bulmada basit ve adil bir yöntem olarak algılanabilir. Bizlerin yorumu Google ile web'te bir içerik standardının oluşacağı kaçınılmazdır. Örneğin bazı ağ yöneticileri ve web-site yöneticileri Google sügeçinden geçemedikleri için yakınmaktadır.

Google arama motoru genel olarak aşağıdaki dört unsuru göz önüne alarak aranan kelime ile bilgileri sıralar (Şekil 1):

- Aranması istenen anahtar kelimelerin geçtiği (uyduğu) tüm sayfaları bulur.
- Sayfadaki faktörlere göre derecelendirme yapılır.
- İlgili yazı içindeki kilit (önemli kelime)lere göre hesaplama yapılır.
- Son olarak sonuçlar PageRank'a göre ayarlanır.



Şekil. 1 Google Arama Motoru Mimari Yapısı<sup>5</sup>

<sup>5</sup> Google kurucuları Sergey Brin ve Lawrence Page'in ilk makalesinden:  
<http://www.db.stanford.edu/backrub/google.htm>

PageRank hesaplaması ve daha detaylı açıklamalar makalenin sonunda ilgili linklerde verilmiştir. Genel olarak web sayfaları aralarında ve içerisindeki linklerin oluşturduğu yönlendirilmiş çoklu-grafta düğüm derecelerinin bulunması ve ağırlıklandırılmasına dayanır.

Google'ın gücünü Data Recovery Group başkanı P. Ahern göklere çıkarmakta ve Google tarafından sayfası derecelendirmeyen sayfanın Yahoo tarafından listlenmediğini dahi söyleyebilmektedir. Ahren devamla Google olmadan iş yapmanın zorluklarına değinmekte ve şirketi Data Recovery Google dışına aldığı zaman işlerinin yüzde 30 azaldığını söylemektedir. Örneğin işiniz telefon trafiğine bağlı ise ve Google dışına alınmış iseniz önce telefon trafiğinde azalma ve domino etkisi ile de işlernizde düşüş yaşanır. Dotcom krizi yaşanırken Google gücünü artıran web kuruluşlarından biri olmuş ve bazı web-sürfçülerine göre web'in ta kendisi olarak algılanmıştır. Bu nedenle Google'un listeleme stratejilerinin adil olmadığını iddia eden bazı şirketler Google karşı dava açmışlardır.

Google'ın aramadaki gücü, basitliği ve güvenilirliği önde giden Yahoo, America Online ve EarthLink gibi portalleri ve Internet sağlayıcıları Google ile işbirliği yapmalarına yönlendirmiş, kendi arama neticeleri ile Google arama neticelerini birleştirmişlerdir. Yakın geçmişte yapılan bir analize göre ziyaretçilerin sitede kalma sürelerine göre Google ortalama ayda 15 milyon ziyaretçi saat elde ederken Yahoo 6 milyonda kalmıştır (Şekil 2). Web pazarlama danışmanlık firması WebGuerrilla'ya göre bu durum Yahoo'yu Google ile koşulları açıklanmayan bir anlaşmaya itmiştir.

Bu bölümde son olarak Google ve Yahoo arama motorlarına birbiriyle ilgili (anlamlı) iki ve üç İngilizce kelimedenden oluşan 10 adet aynı terimler test giriş yapılarak ve biri hariç tümünde Google arama motorunun daha fazla cevabı daha süratli olarak verdiği gözlemlenmiştir.

## ➤ Bilgi Dışında Bilginin Kullanılması

Bu kısımda Google arama motoru teknolojisi ve doğrudan bilişim hukukunu ilgilendiren iki bilişim olayını inceleyeceğiz: Sırayla New York Times gazetesinin paralı servisinin Google üzerinden bedava elde edilmesi diğeri ise Google yönetiminin kriterlerine göre veya gereksiz bazı kuruluşları listelerinden atması.

Çoğu zaman web'te ancak kayıtlı kullanıcılara açık bilgilere (dosya) erişmek için yasal diğer yollardan bir tanesi de arama motoru üzerinden gitmektir. Bunlardan bir tanesi de New York Times gazetesinin paralı servislerinden olan arşivlerine erişmenin Google üzerinden mümkün olması. Bedavasız varken neden paralılarına gidilsin? Bu işlem Google'ın ön bellekte sakladığı bilgilere erişmekle mümkün olmakta ve Google'u bir çok yayımevi ile karşı karşıya getirmektedir. New York Times eski arşivlere ait linkleri kayıt sayfasına yönlendirmeye çalışmakta buna paralel olarak Google ise bu istenmeyen durumu önleyen yazılımlar geliştirmekte fakat Google'ın yeni özellikleri ile yine de istenmeyen durumlar oluşabilmektedir.

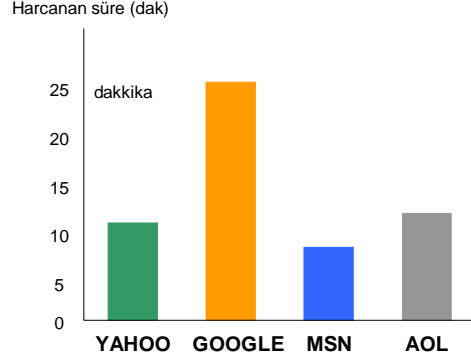
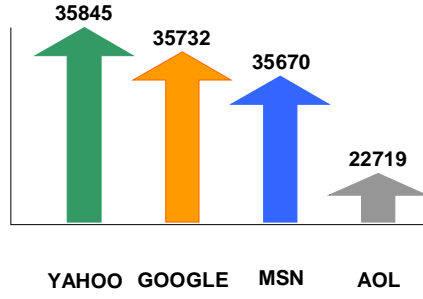
## ➤ Etik Değerler

Google'u özellikle, önemli ve vazgeçilmez yapan kendilerinin kuruluş felsefesi olan yeni servis ve özellikleri yaratıcılıkla deneyip uygulayabilmesidir. Bu kuruluş kültürü her gün milyonlarca ziyaretçi çekmekte ve diğer ilgili sitelere bilinmeyen arama formülleri ile yönlendirmekte, kendisini web'te en güçlü ve popüler şirket haline getirmektedir. Google'un önbellek kullanımı ana Web teknoloji çerçevesinde çetrefelli hukuksal problem oluşturmaktadır. Şu sorunun cevabı temel oluşturmakta: Her hangi Web sayfasının geçici olsa bile ne zaman kopyalanması kabul edilebilir bir olgudur?

Google 1997 yılında diğer arama motorlarından farklı ve kendisine has biçimde oluşturduğu web sayfalarının sayısal kopyalarını kullanıcılarına sunar. Bu şekilde Google tarafından indexlenmiş hemen hemen her sayfa web'teki orijinal yerinde olsun veya olmasın doğrudan Google sunucusundan elde edilebilir. Google diğer arama motorlarının aksine ön bellekte sürekli olarak web kayıtlarını tutmamakta, geçersiz linkleri webte tarayarak önbellekten atmaya elden gelince süratli yapmaktadır. Fakat araştırmacılar için hala Google önbellek sayfaları bir veri hazinesi gibi görülüp veri keşfetmek ve almak için araştırılmaktadır. Güvenlik konularını da içeren bazı durumlarda Google yetersiz kaldığı saptanmıştır. Bu durumdan endişe duyan bazı web-sitesi operatörleri Google'un sitelerinden sayısal kayıt yapmasını engellemek için bazı sayfaları özel kod ile koruma altına alma yoluna gitmişlerdir. Bunun dışındaki diğer bir başka iddia Google ön bellek yolu ile elde edilen bilgiler söz konusu siteye trafiği düşürmekte ve hatta telif hakkı ihlallerine neden olmasıdır. Başka bir durum web-yayın evlerinin düzeltilmiş bilgiler kendi web sayfalarına varken web'te eski ve yanlış bilgilerin bulunabilmesidir. Bu nedenle hukuk adamları yaslar açısından konunun yakın gelecekte mahkemede görüşülmesi ve karara varılması gerektiği söylemektedirler. Diğer bir görüş ise Internet Archive'ini örnek göstererek Internet bulunan bir bilginin ortalama 100 gün olduğunu sonuç itibarıyla eski bilgilerin bir şekilde web'ten silineceğini söylemektedirler.

## ➤ Google Yaklaşımı

Google gün geçtikçe daha etkili olduğunun ve listelerdeki içeriğin bir şekilde başını ağrıtabileceği Google.fr ve Google.de deki bazı aşırı fikir ve düşünceleri besleyen web-sitelerini listeden çıkarmakla ortaya çıkmıştır. Fakat çıkarılan siteler diğer taraftan Google.com'dan elde edilmeye devam etmektedir. Bu yönde Harvard Üniversitesinden B. Edelman ve J. Zittrain'nin raporu 2.5 milyar indekslenmiş sayfa ve karşılaştırma sonucu 113 sitenin (aşırı ırkçı) listenin dışına alındığına değinmektedir. Yakın geçmişte Çin'nin de Internet kullanıcılarına Google'a erişimi kaldırmasından sonra Google'un içerik açısından hukuksal tartışma konusu olmaya devam edeceğini göstermektedir. Google yaptığı açıklamada her hangi bir web-sitesini endeksleme veya liste dışına almakta kendini serbest görmektedir. Bizlere göre bu durum biraz bir pul koleksiyoncusunun topladığı pullar üzerindeki resimlerden sorumlu tutulmasına benziyor.



Şekil 2.

## ➤ Diğer Kaynaklar

Arama motorları için başka bir kaynak sayısal kütüphanelerdir. Bunlardan Old Dominion Üniversitesi'nde bulunan ve birçok yeni DL servisi veren araştırma grubuna dikkatinizi çekmek isteriz. Projeye ABD'deki NSF,NASA,ONR ve daha birçok kuruluş destek vermektedir. DL projesinin kapsadığı arşivler aşağıda verilmiştir.

- [ARC](#)
- [Technical Report Interchange - TRI](#)
- [An OAI based NCSTRL](#)
- [INTEROP](#)
- [ARCHON](#)
- [NASA-CASI Digital Library Extension](#)
- [KEPLER](#)
- [DP9](#)
- [COMOPT](#)
- [DEAN](#)
- [RAPID VISUAL OAI TOOL](#)
- [DOG\(Dienst/Open Archive Gateway\)](#)

Yukardaki arşivlerden DP9 açık kaynak gateway genel arama motorlarına OAI'a (Open Archive Initiatives) uygun endeksleme yapılabilmektedir. DP9 bu işlemi URL istendiği zaman deposundaki kayıtlardan uygun URL ile ve OAI sorgulamasına dönüştürerek yapar. Bunun sonucunda arama motorunun OAI tamamlanmış bilgi deposunda bulunan 'deep Web'i endeksleme imkanı verir. Örneğin ilk yazarın arama motorları tarafından bulunamayan mathprints.com'daki serverdeki makaleleri DP9 tarafından dönüştürülerek Google'un endekslenmesi ile başka bir formatta elde edilebilmektedir. Şekil 3 de detaylar gösterilmiştir.

## ➤ Biraz Teknoloji

Google arama motorunun başarısına karşın WWW’i ortaya atanlardan Massachusetts Institute of Technology (MIT)’den Tim Berners-Lee biraz da internet kullanıcılarının içine düştükleri içerik iletiminde web sayfasının açılmasındaki bekleme sorununu aşmak düşüncesiyle meslekdaşı tanınmış bilgisayar bilimcisi ve ağ teorisyeni Tom Leighton’a akademik çözüm için baş vurur. Tom Leighton ve öğrencileri web trafik sıkışıklığı probleminin çözümünün uygulamalı matematik ve algoritmaları ile mümkün olabileceğini görür ve bu yönde daha sonra kurulacak olan şirketin adı ile anılan ‘Akamai Çözümünü’ geliştirir. 1999 yılında geliştirilen çözümlerle İnternet içeriği iletiminde ‘World Wide Wait’e son veren dünyadaki ilk Web trafiğine bağlı özellikleri kullanan Akamai ortaya çıkmış olur. Bu gün dünyada 60 ülkede Akamai çözümünü kullanan 15000 üzerinde sunucu bulunmaktadır. Bunlar arasında Google ile performans açığını kapatmaya çalışan Yahoo arama motorunu da gösterebiliriz. O halde Tom Leighton’un Akamai çözümü nedir?

## ➤ Akamai Çözümü<sup>6</sup>

Akamai uç-çözümü son-kilometrede (the last mile problem) bulunan İnternet kullanıcılarına web içeriğini İnternet darboğazlarını atlayarak (gecikme ve trafik sıkışıklığı) doğrudan en yakın Akamai sunucusu üzerinden iletmeye dayanır. Bu yöntem merkezi dağıtık sistemlere alternatif uç-iletimi denilir. Bu modelde her web-sitesinin içeriği İnternetin uçunda bulunan sunuculardan veya merkezdeki Akamai sunucusundan elde edilebilir.

Yahoo açısından Akamai çözümünün Google’a göre hala performans açısından geride kalmasından şu sonucu çıkarabiliriz. Ağ üzerine yayılmış akıllı çözümler ağ düğümü içinde (sunucuda) geliştirilen akıllı çözüm yöntemini (PageRank) alt edemez. Fakat ikisinin beraber kullanılması en iyi performansın elde edilmesini sağlar. Dikkat edilirse Akamai’nin de yaklaşımında da webte iş yapan şirketleri ve ISP’leri kendilerine bağımlı yapma düşüncesi yatmaktadır.

## ➤ Sonuç

Yüz milyarlarca link içeren web ormanı gezginleri için kimi zaman bir pusula kimi zaman veri madenciliğinin güçlü aleti olan arama motorları, özellikle Google, her geçen gün biraz daha vazgeçilmez olmaktadır. Bağımlılık yaratan her olgu iyi veya kötü karakterli olsun riskler taşır. Başta sorduğumuz soru: *‘Bilgi internet’te aranırken acaba isteğimizin dışında bizi yönlendiriyor mu?’* gündeme gelmektedir. Böyle bir yazıyı kaleme almamız bunu en azından bizler için doğrulamaktadır.

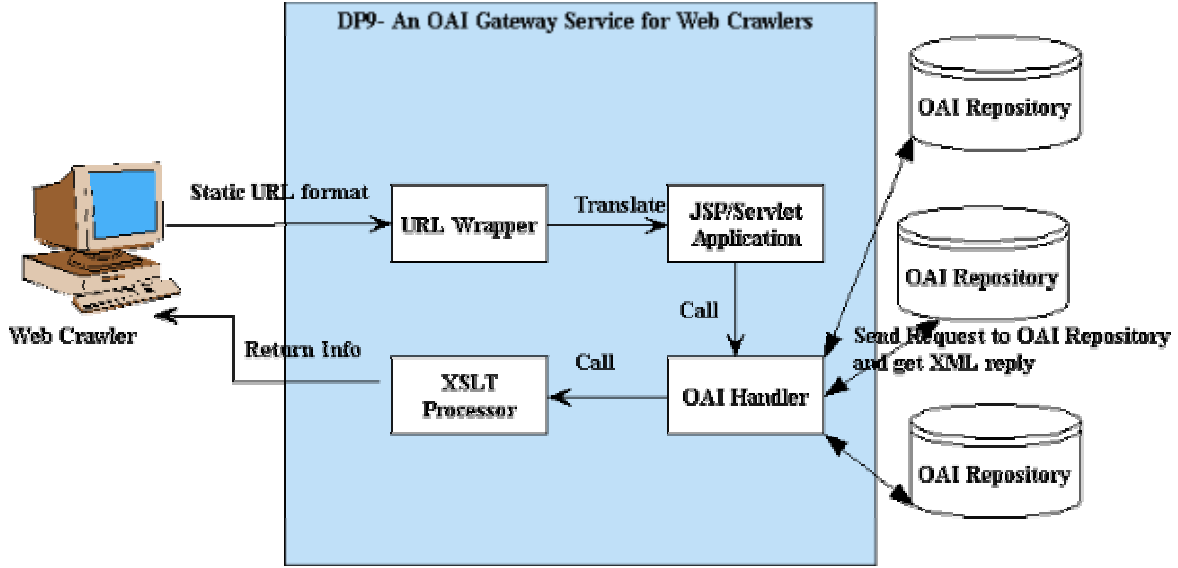
Yazarlar frenlenemez bu gidiş karşısında başta eğitimde karşımıza çıkmakta olan *Google* her işi halleder veya onsuz hiç olmaz yaklaşımının tehlikesine dikkat çekmektedirler.

---

<sup>6</sup> Akamai ve ilgili bilgilere dayalı lisans-üstü dersin tüm notlarına MIT’de başlatılan Open Course Ware (OCW) la ilgili linkten erişebilirsiniz ( <http://ocw.mit.edu> ).

## ➤ Kaynakça

- [1] <http://www.google.com>
- [2] <http://www.yahoo.com>
- [3] <http://ocw.mit.edu>
- [4] <http://www.supportforums.org> (C. Ridings ve M. Shishigin, PageRank Uncovered, 2002)
- [5] <http://www.ipcom.com/papers/programk/>
- [6] <http://www.pcd.stanford.edu/> page



Şekil 2. OAI Gateway Servis Yapısı

```

<?xml version="1.0" encoding="UTF-8" ?>
- <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
    http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2003-07-31T08:50:34+01:00</responseDate>
  <request identifier="oai:MathPreprints:Pure_mathematics/0201019" verb="GetRecord"
    metadataPrefix="oai_dc">http://www.mathpreprints.com/math/OAI/index.htm</request>
- <GetRecord>
- <record>
- <header>
  <identifier>oai:MathPreprints:Pure_mathematics/0201019</identifier>
  <datestamp>2002-01-21</datestamp>
  <setSpec>Pure</setSpec>
  </header>
- <metadata>
- <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
    instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
    http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:creator>Ibrahim Cahit</dc:creator>
  <dc:title>On Zero-Rotatability of Small Graceful Trees: Caterpillars</dc:title>
  <dc:description>The well-known conjecture of Ringel-Kotzig or better known graceful tree conjecture
    has entered to its fourth decade. As today there exists no serious attempts toward its resolution, so
    it is worthwhile to revisit some old concepts. In this paper we have investigated small ( i.e., trees
    with diameter  $\leq 4$ ) graceful trees which are caterpillars, based on the property of rotatability.
    A graceful tree is called rotatable if it is possible to assign the smallest vertex label to an arbitrary
    vertex.</dc:description>
  <dc:subject>Discrete Mathematics and Combinatorics</dc:subject>
  <dc:date>2002-01-21</dc:date>
  <dc:type>text</dc:type>
  <dc:identifier>http://www.mathpreprints.com/Pure_mathematics/0201019</dc:identifier>
  </oai_dc:dc>
  </metadata>
- <about>
- <oai_dc:dc xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
    instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
    http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:publisher>Elsevier</dc:publisher>
  </oai_dc:dc>
  </about>
  </record>
</GetRecord>
</OAI-PMH>

```

OAI Header	
Identifier	oai:MathPreprints:Pure_mathematics/0201019
Datestamp	2002-01-21

Dublin Core Metadata	
Creator	Ibrahim Cahit
Title	On Zero-Rotatability of Small Graceful Trees: Caterpillars
Description	The well-known conjecture of Ringel-Kotzig or better known graceful tree conjecture has entered to its fourth decade. As today there exists no serious attempts toward its resolution, so it is worthwhile to revisit some old concepts. In this paper we have investigated small ( i.e., trees with diameter $\leq 4$ ) graceful trees which are caterpillars, based on the property of rotatability. A graceful tree is called rotatable if it is possible to assign the smallest vertex label to an arbitrary vertex.
Subject	Discrete Mathematics and Combinatorics
Date	2002-01-21
Type	text
Identifier	http://www.mathpreprints.com/Pure_mathematics/0201019

Şekil 3.