

HTML, XML ve WEB SERVİSLERİNİN İNTERNET SUNUCULARI ÜZERİNDEKİ ETKİSİNİN İNCELENMESİ

Erdoğan UZUN, Yılmaz KILIÇASLAN, Erdem UÇAR

Trakya Üniversitesi, Bilgisayar Mühendisliği
Tel: 0(284) 235 39 85, e-Mail: erdinc@trakya.edu.tr

Özet: WWW (World Wide Web) kavramının ortaya çıkışı ile birlikte bilginin ve görselliğin saklandığı HTML (HyperText Markup Language) işaretleme dili İnternetin temelini oluşturmuştur. HTML'in veriyi göstermedeki yetersizlikleri nedeniyle XML (Extensible Markup Language) işaretleme dili İnternet dünyasındaki yerini almaya başlamıştır. XML ile birlikte Web Servisleri kavramı doğmuştur. Bu çalışmanın amacı İnternet üzerinde bilginin paylaşımında kullanılacak HTML, XML ve Web Servislerinin İnternet sunucularına etkisini incelemektir. Bu etki incelenirken dilbilim uygulamalarında için kullandığımız İnternet tabanlı sözlük uygulamaları ve geliştirdiğimiz sözlük uygulamasından yararlanılacaktır.

Anahtar Kelimeler: XML, Web Servisleri, İnternet Sunucuları, Performans

Examining the Effects of HTML, XML and Web Services on Internet Servers

Abstract: Since the emergence of the concept of WWW, the HTML markup language that encodes visual or non-visual information has become the base of the Internet. Because of the inefficiency of HTML to show data, the XML markup language has begun to take a place in the domain of Internet. With XML, the concept of Web services came to existence. The aim of this study is to examine the effect of HTML, XML and Web Services on Internet servers. When examining this effect, we will make use of some dictionaries available on the Internet and similar dictionary which we developed to use in linguistic applications.

Key Words: XML, Web Services, Internet Servers, Performance

Giriş

İnternetteki veri statik veya dinamik olarak ya da popülerliği her geçen gün artan Web servislerini kullanarak İnternet sunucuları üzerinden paylaşımına sunulmaktadır. Statik kullanım, HTML olarak verinin sunucu üzerinde tutulmasıdır. Dinamik kullanım ise PHP gibi bir programlama dili veya ASP.Net gibi bir programlama çerçevesi sayesinde veritabanı gibi sunucu taraflı kaynaklardan çekilen verilerin HTML veya XML dokümanı halinde üretilmesidir. Web servisleri ise XML verileri üretmek ve bu düzenli veri üzerinde işlem yapabilmek için geliştirilmiştir. Web dünyasındaki rotayı belirleyen ticari şirketlerden biri olan Microsoft'a göre .NET geliştirildikten sonraki 10 yıl boyunca Web Servisleri İnternet dünyasına damgasını vuracaktır. (MacManus, 2004)

HTML işaretleme dilinin amacı HTML etiketleri kullanarak Web tarayıcısının veriyi görsel bir biçimde gösterebilmesini sağlamaktır. HTML içinde hem veri, hem de görsellik için etiketlerin saklanması sebebiyle bu veriyi ayırtmak giderek sorun olmaya başlamıştır. Bu noktada SGML dilinin bir uzantısı olan XML işaretleme dili ortaya çıkmıştır.

XML'in işaretleme dilini HTML betik dilinden ayıran en önemli farklardan biri; XML sadece verinin tanımlanması için kullanılmasıdır. XML verinin gösterim şekliyle ilgilenmez. XML'de, HTML'de olduğu gibi etiketlerin bilinmesine gerek yoktur. Etiketler kişi tarafından belirlenebildiği için XML genişletilebilir bir özellik taşır.

Web Servisleri, İnternet üzerindeki istemcilerin başvurabileceği o servise yönelik fonksiyonları barındıran, çok yönlü ve merkezleşmiş bir ünedir (MacManusR., 2004). Web servisleri HTTP (Hyper Text Transfer Protocol) üzerinden gidecek XML veri parçalarını kullansa da, bunların istemcilerin işleyebileceği ve yorumlayabileceği bir hale getirilmeleri gerekir. Bu noktada, Web Servisleri için önemli ve gerekli temel kavramlardan birisi olan SOAP (Simple Object Access Protocol) ortaya çıkar.¹ SOAP etiketlerinin oluşturulması ve çözümlenmesinin, aslında hem istemci hem de İnternet sunucusu üzerinde ek bir maliyeti vardır.

Bu çalışmanın amacı, HTML, XML ve Web Servislerinin sunucu üzerindeki etkisinin incelenmesidir. Bu etki incelenirken dilbilim çalışmaları için kullandığımız İnternet tabanlı sözlük uygulamalarından yararlanılacaktır.

¹ Bu protokoller, dünyada internetin gelişimini belirleyen World Wide Web Consortium (Web Standartlarını Belirleyen Birlik) tarafından düzenlenmektedir.

İlgili Çalışmalar

XML tabanlı çalışmalarda bir uygulama geliştirilirken normalizasyon işlemi çok önemlidir. Sözlük uygulamasında normalizasyon işlemi dikkate alınmıştır. J.Kamps ve Ark., (2004,2005) XML açısından normalizasyonun önemi üzerinde durmuştur.

Web Servisleri konusunda çeşitli çalışmalar vardır. Cai (2002) yaptığı çalışmada Web Servisleri için alternatif kodlama yöntemleri üzerinde durmuştur. Bu çalışmada kodlama yöntemlerinin sonucu bant genişliğine bindirdiği yükü incelemiştir. Chui (2002), bilimsel hesaplamada SOAP'ın performansını araştırmıştır. Bu performans ölçümlerinde bant genişliği ve hafıza göz önüne alınmıştır. Menasce ve Almeida (2002) çalışmalarında Web servislerinin kapasitesinin planlanmasını tanıtmışlardır. Ayrıca sunucuların performansı üzerine çeşitli çalışmaları vardır. Tian ve Ark. (2004) yaptığı çalışmada sonucu üzerindeki bant genişliği ve sunucunun CPU performansı üzerinde çalışmıştır.

Berger ve Ark.,(2002) sözlükte XML ve XSL (Extensible Stylesheet Language Family)'ın kullanılması etiketlenmesi üzerinde durmuştur.

HTML, XML, Web Servisleri ve İnternet Sunucuları

HTML ve XML

HTML İnternet üzerinde Web sayfası oluşturmak için kullanılan bir betik dilidir. HTML dosyalarının aktarımı için HTTP kullanılır. XML, HTML'e çok benzemekle birlikte HTML'in amacı verinin nasıl görüntüleneceği iken, XML'in amacı ise verinin içeriği ile ilgilidir. XML işaretleri önceden tanımlı değildir. Kişiyeye özel olarak bu işaretler belirlenebilir.

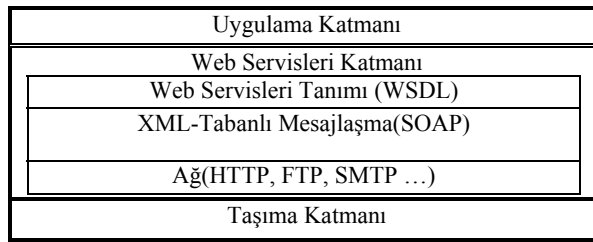
XML, İnternet üzerinde genellikle veri değişimi için kullanılmaktadır. XML, yeni dillerin yaratılması için olanaklar sağladığı için hızlı bir şekilde yaygınlaşmaktadır. XML, hiyerarşik bir yapıya sahiptir. Bu hiyerarşik yapı sayesinde XML'deki veriye HTML'e göre daha kolay şekilde erişilebilir. Veriye daha kolay erişilebildiği için XML'deki arama işlemi çok daha hızlı şekilde yapılabilir. XML, rahatça okunabilen ve anlaşılabilen bir yapıya sahiptir.

XML'de en çok dikkat edilecek noktalardan biri etiket seçimi ve hiyerarşidir. (Kamps Ve Ark., 2004-2005) Etiket seçimi ve hiyerarşik yapı uygulama geliştirilen kişiye bırakıldığı için işlemci performansı ve bant genişliği açısından sorunlarla karşılaşılabilir. (Menasce ve Almeida, 2002) Uygulamayı geliştiren kişinin etiket seçimi ve hiyerarşik yapının kurulmasında dikkat etmesi gerekmektedir.

Web Servisleri

Web Servisleri, XML kullanılarak tanımlanmış fonksiyonları ve arayüzleri kullanan ve URI tarafından tanımlanan bir yazılım sistemidir. Bir Web servisi diğer XML tabanlı Web servisleri ile İnternet protokolü üzerinden XML tabanlı olarak haberleşebilir.

Web Servislerinin mimarisi ve Web uygulamalarındaki yeri W3C (World Wide Web Consortium) tarafından tanımlanmıştır. Bu mimariye göre Web Servisleri, uygulama katmanı ile taşıma Katmanı arasında yer alır (Şekil 1.). Web Servisleri katmanında ise çeşitli protokoller bulunmaktadır. Web Servislerinin tanımlanması için WSDL(Web Services Description Language), XML-tabanlı mesajlaşma için SOAP ve HTTP, FTP ve SMTP gibi ağ iletişim protokolleri kullanılır. WDSL ve SOAP, XML tabanlıdır.

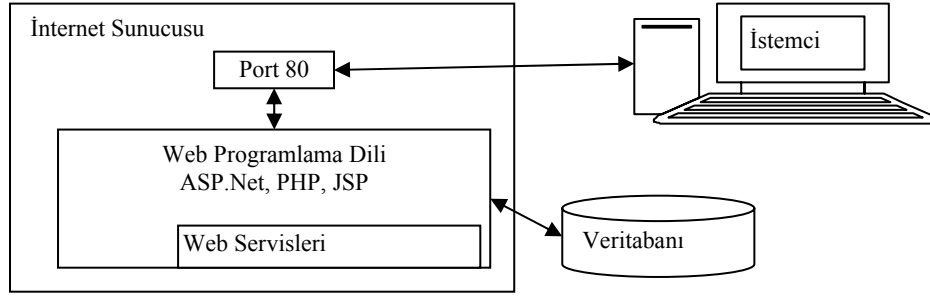


Şekil 1. Web Servis katmanları

Bir Web servisi uygulamasında XML veri ile birlikte her seferinde WSDL ve SOAP mesajları da sunucu tarafından üretilir. Sunucu tarafında üretilen bu mesajlar istemci tarafında ayrıca çözümlenmektedir. XML veri yanında, WSDL ve SOAP tanımlamalarının sunucuda üretilmesi ve istemcide çözülmesi, esasında üretilen veri miktarı arttığı için ek bir maliyet oluşturur.

İnternet Sunucuları

İnternet sunucularının görevi 80. porttan aldıkları HTTP verisini yorumlayıp, istemciye uygun HTML veya XML verisini göndermektir. Ayrıca ASP.Net ve PHP gibi Web programlama dilleri ile dinamik HTML ve XML verileri üretilebilir. Özellikle veritabanı kullanan uygulamalar için, Web programlama dilleri kullanmak gereklidir. Web Servisleri de HTTP arayüzünü kullanarak haberleşmektedir. (Şekil 2.)



Şekil2. İnternet Sunucusu Mimarisi

XML Tabanlı Bir Sözlük Uygulaması

İnternet üzerinde HTML tabanlı çeşitli sözlük uygulamaları vardır. Bilgisayarlı dilbilim çalışmalarının en önemli parçalarından biri sözlüktür. Bilgisayarlı dilbilim çalışmaları için İnternet üzerindeki HTML tabanlı sözlüklere erişim teknikleri kullanılarak veriyi otomatik bir şekilde elde etmek mümkündür.

Bu Çalışmanın Doğuşu:

Dilbilim çalışmalarında TDK (Türk Dil Kurumu <http://www.tdk.org.tr/>), MikroBeta tarafından geliştirilmiş Türkçe Sözlük(<http://www.turkcesozluk.org/>) ve internetten ücretsiz olarak erişilebilen çeviri sözlüğü olarak İngilizce-Türkçe sözlük(<http://www.sozluk.net/>) gibi çeşitli HTML tabanlı sözlüklerden yararlanılmıştır.

HTML dokümanları ayrıştırma tekniklerini kullanarak istenilen veriler HTML doküman içinden çekilebilir. Fakat küçük bir veri parçasına ulaşmak için bile, gereksiz HTML etiketlerinden ve HTML dokümanı ile birlikte gelen resim dosyalarından kurtulmak gerekmektedir. Ayrıca gereksiz etiketler ve HTML dokümanı ile birlikte gelen resimler gereksiz bir ağ trafiği oluşturmaktadır. Bu sebeplerden dolayı XML tabanlı bir sözlük çalışmasına gerek duyuldu.

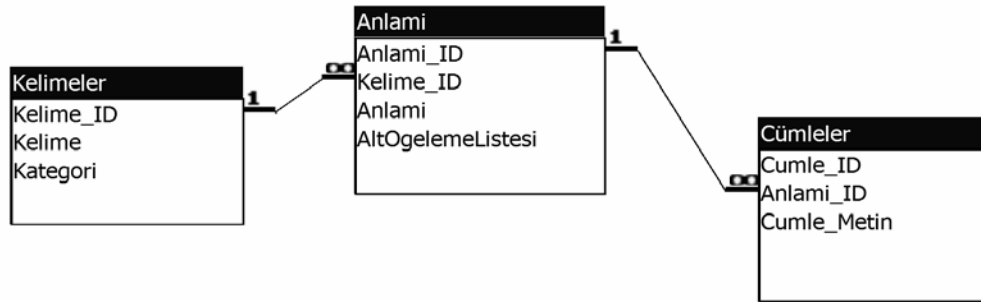
XML Tabanlı Sözlük Çalışması

HTML'in oluşturduğu ağ yükünü hafifletmenin en güzel yollarından biri XML'dir. Fakat XML tabanlı bir sözlük uygulaması ne yazık ki bulunmamaktadır. Bu sebepten dolayı XML tabanlı bir sözlük uygulaması geliştirmeye karar verilmiştir.

XML tabanlı bir çalışma iki şekilde oluşturulabilir. Birincisi XML dosyası üreten dinamik bir Web uygulaması geliştirmek; ikincisi ise XML tabanlı bir Web servisi geliştirmektir.

Veritabanı Yapısı

Dinamik Web uygulamalarında sözlük bilgilerini veritabanında tutmak gereklidir. Sözlük uygulaması için aşağıdaki ilişkileri içeren bir sözlük tasarlanmıştır (Şekil 3).



Şekil 3. Sözlük Veritabanı İlişkileri

Bu çalışmada birbirinden farklı sözcük anlamlarını örnek tümcelerle destekleyecek bir hiyerarşi kurulmuştur. Şekil 3'teki "Kelimeler" tablosunda bir kelimenin isim, fiil, sıfat, zamir durumları "Kategori" alanında tutulur. "Anlami" tablosundaki "AltOgelemeListesi" alanı Türkçedeki alt ögeleme listeleri anlama göre değiştiği için bu bölüme alınmıştır. Alt ögeleme listesi cümlelerin alabileceği öğeleri ve bu öğelerin hal durumlarını içermektedir. (Brent, 1994)

XML verinin elde edilmesi

XML veri iki şekilde elde edilebilir. Birincisi ASP.Net veya PHP gibi Web programlama dilleri ile sorgulanan tablolardan aşağıdaki gibi bir XML çıktısı üretilebilir. İkinci yaklaşım ise Web servisleri üzerinde bu XML çıktısını üretmektir.

```

<kelimeler kelimeid = "20014" kelime = "gelmek" kategori="fiil">
  <Anlami Anlami = "Bir yere gitmek, ulaşmak, varmak" AltOgeleme = "e,den">
    <Cümleler> Gurbetten gelmişim yorgunum, hancı.</Cümleler>
  </Anlami>
<Anlami Anlami = "Geriye dönmek" AltOgeleme = "e,den">
  <Cümleler> adamı Ödemiş'ten aldım geldim</Cümleler>
  </Anlami>
<Translation> come </Translation>
</kelimeler>

```

Tablo 1. Sözlük Uygulamasının Bir XML Çıktısı Örneği

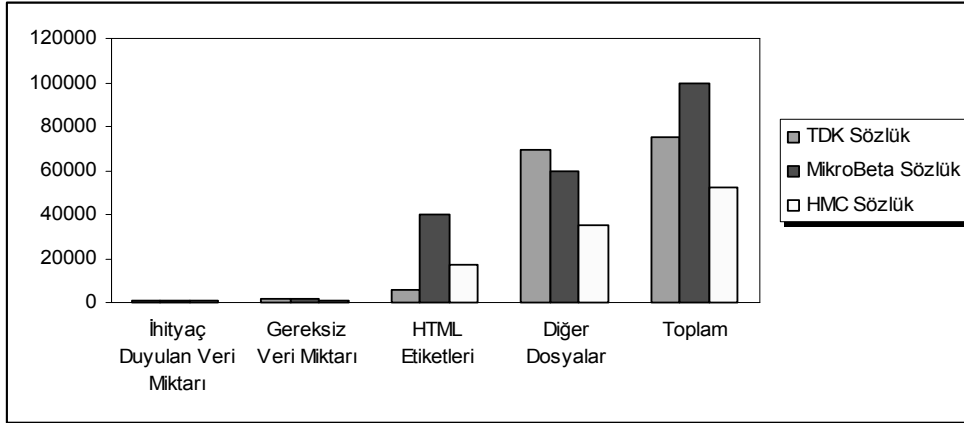
XML verinin XML veya Web Servisi Kullanarak İnternete Sunulması

Yukarıda elde edilen XML veri iki şekilde İnternette paylaşımına açılabilir. Birinci yöntem sadece XML belgesinin üretilip istemcilere gönderilmesidir. İkinci yöntem Web Servislerini kullanarak SOAP üzerinden göndermektir.(Tian ve Ark., 2003)

HTML, XML ve Web Servisleri Uygulama Üzerinde Değerlendirilmesi

Çeşitli HTML tabanlı sözlük uygulamalarının karşılaştırılması

Bir web sayfası HTML etiketlerinden, HTML Etiketleri içindeki verilerden ve resim-javascript gibi diğer dosyalarından oluşur. Bu bölümde mevcut uygulamaların içerdiği HTML etiket miktarı, veri miktarı ve diğer dosyalarının boyutu üzerinde bilgi verilecektir. Veri miktarı, ihtiyaç duyulan veri miktarı ve gereksiz veri miktarı olmak üzere ikiye ayrılmıştır. İhtiyaç duyulan veri miktarından kasıt XML belge için gerekli veri miktarı anlamına gelir. Aşağıda bu dosyaların belirtilen kriterlere göre ortalama ne kadar yer kapladıkları hakkında bilgi verilmektedir.(Şekil 4.)



Şekil 4. TDK, MikroBeta ve sozluk.net HTML tabanlı sözlük uygulamaları

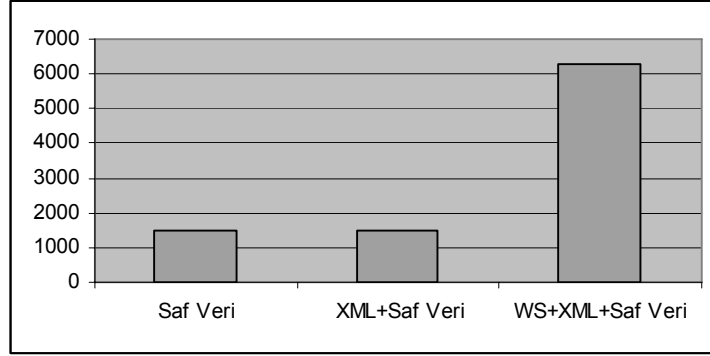
Yukarıdaki tablodaki sonuçlara göre aslında ihtiyaç duyulan veri miktarının 1000 bayt civarında olduğu saptanmıştır. Ama bir Web sayfası transfer edildiğinde ortalama 75000 bayt civarında bir dosya transferi söz konusudur.

HTML Etiketleri yerine XML Etiketlerinin Kullanılması

Bir XML dokümanı XML etiketlerinden ve saf veriden oluşur. Sözlük uygulamamız, hem TDK hem de sozluk.net verilerini içerip yaklaşık olarak 1500 bayt civarında veri içermektedir. Uygulamamızda XML etiketleri eklendiğinde dosya boyutunun yaklaşık iki kat artarak 3000 bayt düzeyine çıktığı görülmüştür. Fakat bu noktada seçilen etiket uzunluklarının ve hiyerarşik düzene göre bu oranların değişebileceği göz önüne alınmalıdır.

Web Servislerini Kullanılması

Web Servisleri kullanılması ile birlikte SOAP ve WDSL gibi ek XML etiketleri devreye girer. Bunun sonucu olarak veri miktarı yaklaşık olarak 3300 baytlık bir artış gösterir. (Tian ve Ark., 2004) (Şekil 5.)



Şekil 5. XML ve Web Servislerinin(WS) dosya boyutuna etkisi

Tartışma ve Sonuç

Bu çalışmada, XML ve Web Servisleri gibi Web dünyasındaki değişime yol açan uygulamalar incelendi. HTML'ye göre farklılıkları ortaya konuldu. Mevcut dinamik Web sayfalarının adaptasyonu hakkında bilgi verildi. Bu adaptasyon sırasında XML ve Web Servisleri gibi ortaya çıkacak iki alternatif incelendi. Bu çalışmadan elde edilen sonuçlar şunlardır:

HTML bilgiye erişim yöntemleri ile HTML içindeki veriyi çıkarmak karmaşık ve maliyetli bir işlemdir. XML ve Web Servis yöntemleri ile dosya boyutu küçüldüğü için bilgiye erişim hızlıdır ve böylece maliyet düşürülmüş olur.

XML yapısı sayesinde HTML dosyalarının kullandığı geniş bant genişliklerinden kurtulabiliriz. Böylece sunucuların yükünü hafifletmiş oluruz.

XML ve Web Servisleri tercih edilirken Web Servisinin oluşturduğu ek etiketler göz önüne alınmalıdır. Özellikle sırf veri üzerine olan basit uygulamalar için sade bir XML dokümanı yeterlidir.

Sade bir XML dokümanı Web programlama dilleri ile kolaylıkla üretilebilir. Bir uygulamacı bir sitedeki verilere ulaşmak istediğinde geliştirilen bir uygulama tarafından üretilen XML dokümanını kullanabilir. XML dokümanı içindeki veriye ulaşmak HTML göre çok daha kolaydır.

Bundan sonraki çalışmalarımızda XML'in görsellik tarafını tamamlayacak XSL ve HTML sayfaları içine gömülen AJAX (Asynchronous JavaScript and XML) inceleneyecektir. Ayrıca dilbilim çalışmalarımız için sözlük modülümüzün geliştirilmesine devam edilecektir.

Kaynaklar

1. Berger D., Reitter D., Stede M., XML/XSL in the dictionary: the case of discourse markers, International Conference On Computational Linguistics, Proceedings of the 2nd workshop on NLP and XML - Volume 17, Sayfa 1-8, 2002
2. Brent M. R., Surface cues and robust inference as a basis for the early acquisition of subcategorization frames, *Lingua*, Volume 92, Sayfa 433-470, 1994
3. Cai, M., A Comparison of Alternative Encoding Mechanisms for Web services. 13th International Conference on Database and Expert Systems Applications, Aix en Provence, France, 2002
4. Chiu K., Investigating the Limits of SOAP Performance for Scientific Computing. IEEE International Symposium on High Performance Distributed Computing, Edinburgh, Scotland, 2002
5. Kamps J., Rijke M., and Sigurbjörnsson B., Length normalization in XML retrieval, BNAIC-2004: Proceedings of the 16th Belgium-Netherlands Conference on Artificial Intelligence, pages 369-370, 2004
6. Kamps J., Rijke M., and Sigurbjörnsson B., Jaap Kamps, Maarten de Rijke, and Börkur Sigurbjörnsson. The importance of length normalization for XML retrieval, *Information Retrieval*, 8:631-654, 2005
7. MacManusR., The Evolution of Corporate Web Sites, 28 Nisan 2004, http://www.digital-Web.com/articles/the_evolution_of_corporate_web_sites/
8. Menasce D., Almeida V., Capacity Planning for Web services, Prentice Hall, Upper, Saddle River, NJ, 2002
9. Tian M., Voigt T., Naumowicz T., Ritter H., Schiller J., Performance Impact of Web Services on Internet Servers. International Conference on Parallel and Distributed Computing and Systems (PDCS 2003), Marina Del Rey, USA, Kasım 2003
10. World Wide Web Consortium, (Web Standartlarını Belirleyen Birlik) <http://www.w3.org/>,