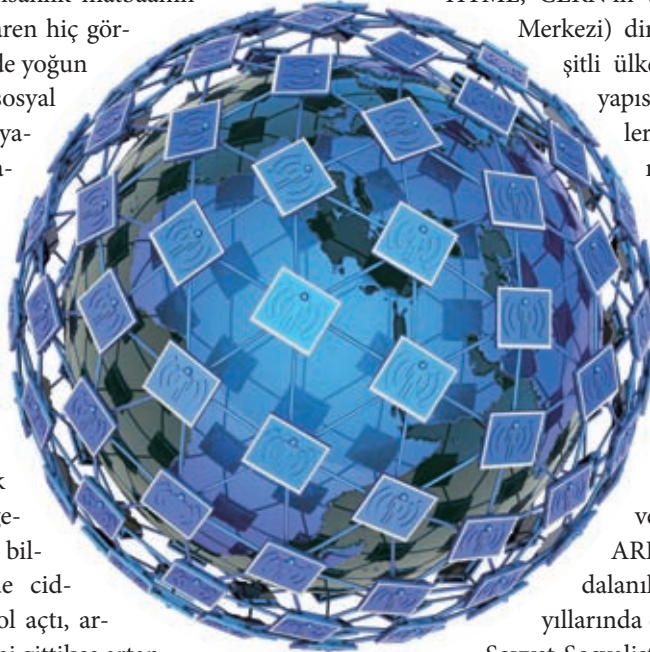


Yeni bilgi modelleme ve programlama felsefesiyle Semantik Web



Web'in günlük hayatımıza girmesiyle, günümüzde insanoğlunun üretmekte olduğu bilgi miktarı ve bu bilgilerin karmaşıklık derecesi insanlık tarihinde görülmemiş boyutlara ulaştı, üstelik bilişim teknolojileri de bu sürece hayli hazırlıksız yakalandı. Son yıllarda hayli hızlı bir şekilde gelişen anlamsal Web teknolojileriyle birlikte önümüzdeki yıllarda yeni nesil bir Web'in doğacağı ve bu yeni nesil Web'in (Semantik Web) günümüzde süregelen bilgi kaosuna son vermekte çok önemli bir rol oynayacağı iddia ediliyor. Semantik Web ile birlikte gerçekten tünelin ucunda ışık görünecek mi? Semantik Web'in önündeki engeller neler? Gerçekten hayata geçirilebilecek mi yoksa bir hayal olmaktan öteye gidemeyecek mi? Bu yazımızda hem Web hem de yapay zekâ dünyasına kısa bir yolculuk yaparak bu sorulara cevap bulmaya çalışacağız.

1970'li yıllardan itibaren bilgisayar alanında adım adım geliştirilen teknolojilerin, doksanlı yıllarda Tim Berners-Lee tarafından geliştirilen ve insanlığın hizmetine sunulan Web'le buluşmasıyla birlikte insanlık matbaanın icadından itibaren hiç görmediği derecede yoğun ekonomik ve sosyal değişiklikler yaşamaya başladı ve buna paralel olarak da bir bilgi patlamasıyla karşı karşıya kaldı. İlk önceleri hayli sevindirici bir gelişme olarak görülen bu gelişme zamanla bilgi yönetiminde ciddi sorunlara yol açtı, ardından da yerini gittikçe artan bir hayal kırıklığına bırakmaya başladı. Sonuç bugün ortada. Günümüzde üretilen bilgilerin çoğunluğunun kaderi internet deryasında birbirlerinden kopuk ve izole bir şekilde unutulmaya terk edilmek oldu.



Web'in Doğuşu

World Wide Web'in belkemiğini HTTP protokolü (*Hypertext Transfer Protocol*) ile HTML (*Hyper Text Markup Language*) oluşturuyor. HTTP ve HTML, CERN'in (Avrupa Nükleer Araştırma Merkezi) direktifleri doğrultusunda, çeşitli ülkelerde bulunan ve farklı ağ yapısına sahip CERN temsilciliklerinde çalışan bilim insanlarının birbirleriyle problem-sizce bilgi alışverişinde bulunabilmesi için Tim Berners-Lee tarafından 90'lı yılların başında geliştirildi. 1990'lı yılların ortalarına doğru CERN tarafından Web'in kullanımının bütün insanlığın hizmetine sunulması kararlaştırıldı ve bu yapılırken büyük ölçüde ARPANET'in altyapısından faydalanıldı (ARPANET soğuk savaş yıllarında özellikle uzay çalışmalarında Sovyet Sosyalist Cumhuriyetler Birliği'nden geri kalmak istemeyen ABD'nin, genelde birbirinden farklı ağ yapılarına sahip Amerikan üniversitelerinin altyapılarını birleştirerek üniversitelerarası bilgi alışverişini mümkün kılmak isteğiyle oluşturulmuştu).

Web'in ilk nesli Web 1.0 (1995-2000) yalnızca HTML belgelerin yer alabildiği "donuk" bir yapıya sahipken, Web 2.0 (2000-2010) ile birlikte kullanıcılarının da aktif olarak katılabildiği etkileşimli ve insan odaklı bir platform doğdu. Web 2.0 sayesinde günümüzün Facebook, Twitter, YouTube gibi en popüler ve önemli kitlesel iletişim araçları doğdu ve bu süreç dünyamıza kelimenin tam anlamıyla yeni bir dinamizm getirdi.

Web 1.0	1995-2000 (Belge odaklı)
Web 2.0	2000-2010 (Etkileşimli ve insan odaklı)
Semantik Web	2010-2020 (Bilgi ve bilgisayar odaklı)

Web 2.0'nin Problemleri

Getirdiği bütün dinamizme rağmen Web 2.0 hâlihazırda bir çok problemi de bünyesinde barındırıyor:

a. Google, Yahoo! gibi anahtar kelime bazında arama yapan güçlü arama motorları dahi artık istenilen sonuçları vermekte zorlanıyor.

b. Web 2.0'da bulunan bilgilerin çoğunluğu metinsel kaynaklı ve yalnızca insanlar tarafından anlaşılabilir bir yapıya sahip.

c. Bilgilerin büyük bir kısmının metinsel kaynaklı olması, bu bilgilerin anlamlandırılıp, bilgisayarlar tarafından "anlaşılmasını" ve aralarında ilişki kurulmasını engelliyor.

d. Aralarında ilişki kurulamayan bilgilerden, otomatik yeni bilgi çıkarsaması imkânsız hale geliyor.

e. Web'in içeriklerinin bilgisayar tarafından anlaşılabilmesi, Web'i büyük bir hızla hemen hemen hiç bir kontrolün ve dolayısıyla sanal güvenliğin bulunmadığı bir ortam haline dönüştürüyor.

Sonuç olarak Web'in günümüzdeki hacmi ve büyüme hızı dikkate alındığında içeriğinin sırf insanlar tarafından değil, aynı zamanda bilgisayarlar tarafından da "anlaşılmaya" başlanması giderek bir zorunluluk haline alıyor. Nitekim Web'in çok yakın bir gelecekte bu tip problemlerle karşılaşacağı en başta Web'in mucidi Tim Berners-Lee ve bir grup başka bilim insanı tarafından daha 2000'li yılların başında öngörüldü ve bu-

na çözüm olarak da içeriğin bilgisayarlar tarafından da anlaşılabilirdiği yeni nesil bir Web düşünüldü: Semantik Web.

Semantik Web Uygulama Örnekleri

Yukarıda da belirtildiği gibi Semantik Web bilgi ve bilgisayar odaklı bir yapıya sahip olacak (bu özelliğinden dolayı Semantik Web'in diğer bir adı da -pek kullanılmamakla birlikte- *Web of Data*'dir). Semantik Web'in yapısındaki bu özellikler kullanımı açısından da insanlığa yepyeni ufuklar açıyor. Her ne kadar tasarlanmakta olan Semantik Web uygulamalarının çoğunluğu daha geliştirme aşamasında olsa da bir kısmı şimdiden hayata geçirildi:

1. Semantik Web ile Web 2.0 arasındaki en önemli fark, arama motorları sistemlerinde fark edilecek. Bir anlamsal arama motoru kullanan kullanıcı, kendisini ilgilendiren konuda bazı anahtar kelimeler girmek yerine sorusunu sisteme doğrudan yöneltebilecek. Örneğin en son dünya futbol şampiyonunun hangi takım olduğunu öğrenmek istediğinizi düşünün. Bu günümüz Web'inin geleneksel arama motorlarında dünya şampiyonu, futbol gibi anahtar kelimeler girerek bulmaktan başka bir şansınız yok gibi. Alacağımız cevap ise önceden bellidir: 1 saniyede yaklaşık 1.000.000 cevap... (Ancak şansınız varsa, gelen ilk 20-30 sonuç arasında muhtemelen aradığınız cevaba dair bir ipucu yakalarsınız.)

Semantik Web'in bize sunacağı anlamsal bir arama motorunda ise anahtar kelimelere, her şeyden önce şansa yer olmayacak. Böyle bir arama motorunda kullanıcı tarafından doğrudan sorulmak istenen soru girilecek ve doğru cevap alınacak:

Örnek:

Sorgu: En son dünya futbol şampiyonu hangi takımdır?

Cevap: İspanya

2. Bir anlamsal arama motoru, yöneltilen soruda geçen eş anlamlı ifadeleri tespit edecek yeteneğe de sahip olacak.

Örnek:

Sorgu: Safari turlarıyla ilgili kapsamlı bilgi istiyorum

Cevap: Afrika Safari turu 1, Afrika Safari turu 2, ..., Afrika Safari turu n

Sistemin kullanıcının sorusuna verdiği cevaptan da anlaşılacağı üzere, sistem Web 2.0'dakinden çok farklı olarak sorudaki "Safari" ile "turlar" arasındaki ilişkiyi anlayacak ve cevap olarak yalnızca kullanıcının ilgilendiği "Safari turlarını" getirecek, Safari sözcüğüyle ilişkili araba modellerini, internet tarayıcısı gibi ürünleri ise otomatik olarak cevap kümesinden eleyecek.

3. Anlamsal teknolojilerin kullanımıyla tost makinesinden buzdolabına kadar her ev aleti "akıllanacak", hatta ihtiyaç halinde birbirleriyle uyumlu bir şekilde çalışacak. Bu şekilde dünyamız "akıllı ev"ler çağına girecek, örneğin sütün bittiğini fark eden buzdolabınız süpermarketten süt ısmarlayabilecek.

4. Kullanıcı ilgilendiği kavramlar arasındaki bağlantıları görsel olarak da tespit edebilecek. Yukarıdaki bahsedilen üç örnekten farklı olarak bu örnek hâlihazırda DBpedia projesi kapsamında (RelFinder) büyük ölçüde hayata geçirildi ve tüm internet kullanıcılarına açık. DBpedia projesiyle ilgili ayrıntılı bilgi için lütfen bir sonraki bölümü (Linked Data) inceleyin. DBpedia projesi kapsamında geliştirilen RelFinder ile RDF tabanlı bilgiler arasındaki ilişkiler görsel olarak da incelenebilir.

Linked Data

Semantik Web'in bütün bu yeteneklere sahip olmasının ardında W3C tarafından geliştirilen Linked Data kavramı var. Linked Data sayesinde her bir bilginin belirli bir anlama sahip olacak şekilde modellenmesi, daha sonra da modellenmiş bu bilgilerin birbirleriyle ilişkilendirilerek birbirine "bağlanması", böylece gelecekte bütün Web'in küresel ölçekte "akıllı" bir veri tabanına dönüştürülmesi tasarlanıyor.





W3C, Web'in mucidi Tim Berners-Lee tarafından 1994 yılında

kurulmuş. Başlıca görevi Web'den en yüksek verimin alınması için gerekli düzenlemelerin yapılmasını sağlamak, ilgili standartları düzenlemek ve gerektiğinde de ihtiyaç duyulan teknolojileri bizzat yaratmak.

W3C tarafından standart haline getirilmiş, önemli bazı günümüz teknolojileri şunlar: XML, HTML, XHTML, RDF, RDF-S, OWL, RIF ve SPARQL.

DBpedia Projesi

Bu amaç için tasarlanan ilk başarılı projelerden biri, iki Alman üniversitesi (Freie Universität Berlin ve Universität Leipzig) ve OpenLink Software firması tarafından şimdiden başarıyla hayata geçirildi: DBpedia



DBpedia projesinin özü, İnternet ansiklopedisi Wikipedia'daki metinsel bilgilerin çoğunluğunun Semantik Web tabanlı algoritmaların kullanımıyla otomatik olarak RDF formatındaki bilgilere dönüştürülüp Linked Data olarak yayınlanması, yeni nesil anlamsal sorgulama lisanı SPARQL ile sorgulamalara hazır hale getirilmesi. Bu proje sayesinde Wikipedia'da bulunan bilgilerin büyük bir kısmı kolaylıkla sorgulanabiliyor ve hatta -daha önce de belirtildiği gibi- bilgiler arasındaki bağlantılar RelFinder gibi RDF tabanlı araçlar üzerinden görsel olarak incelenebiliyor. DBpedia bu tür projelerden yalnızca biri. Özellikle Avrupa ve Amerika'da bir çok kurumda ve üniversitede hâlihazırda CIA World Factbook, GeoNames gibi benzer başka projeler de yürütülüyor, bunlar belirli bir "olgunluk" aşamasına eriştikten sonra DBpedia gibi diğer RDF tabanlı bilgi kümeleriyle birbirlerine "bağlanıyor". Bilgilerin Web'de Linked Data olarak yayınlanabilmesi için ilk önce RDF formatına dönüştürülmesi şart. RDF formatı anlamsal teknolojilerin ana formatını oluşturuyor. RDF, RDF-S ve OWL'in ortak-

laşa kullanımıyla, Semantik Web uygulamalarının kalbini oluşturan, yüksek derecede açıklayıcılık gücüne sahip ontolojiler modelleniyor.

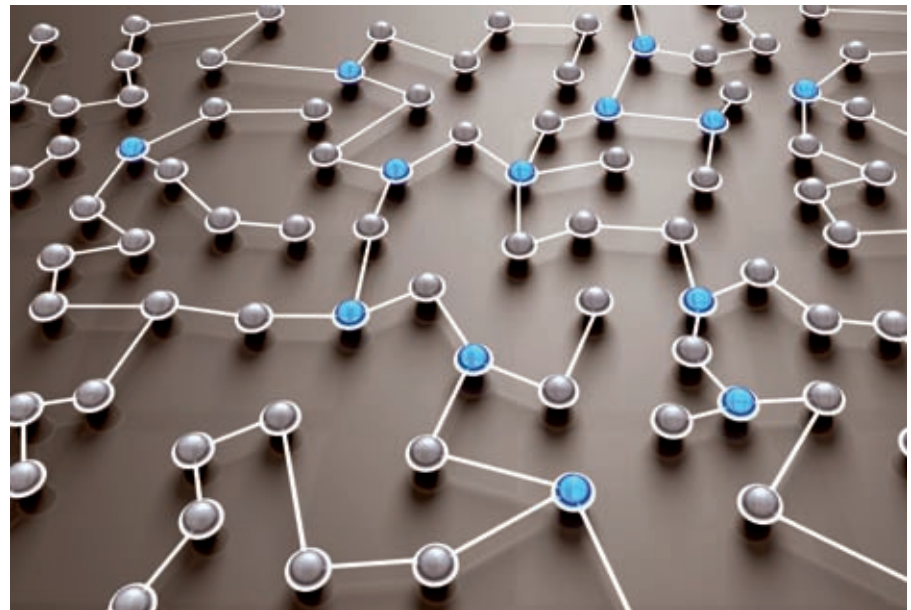
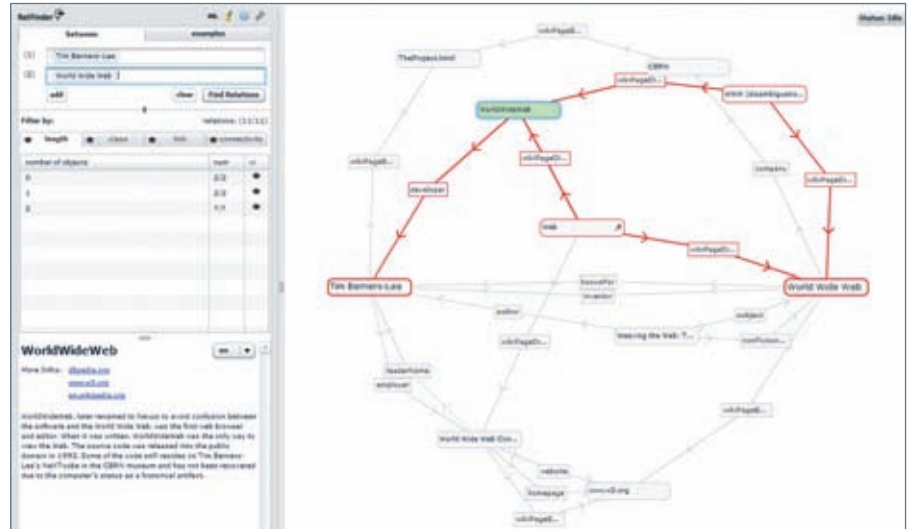
DBpedia türündeki projeler sayesinde Semantik Web aslında daha bugünden bugünkü Web'e paralel olarak "inşa" edilmeye başlandı.

Sonuç

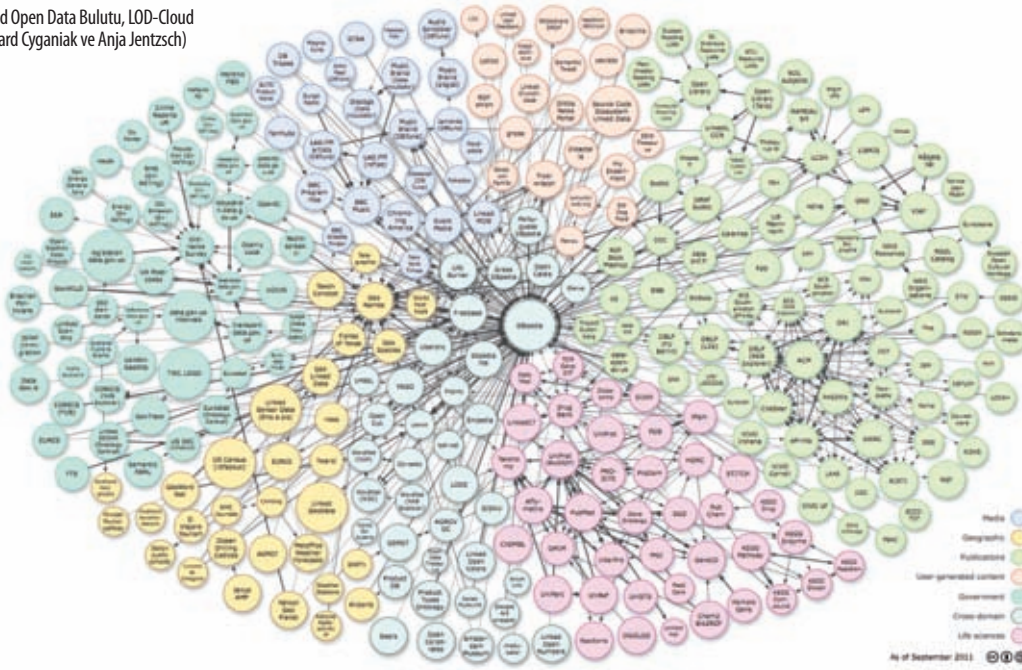
Semantik Web ile birlikte Web, içeriğini yalnızca insanların anladığı insan odaklı bir ortam olmaktan çıkacak, içeriği aynı zamanda bilgisayar tarafından da anlaşılabilir, bilgi ve bilgisayar odaklı, küresel ölçekte akıllı bir veritabanına dönüş-

cek. Bu hedefin önündeki en büyük engel ise hâlihazırda Web 2.0 ortamında bulunan bilgilerin büyük bir kısmının henüz RDF formatına dönüştürülmemiş olması. Web'de bulunan bilgiler ne kadar kısa zamanda RDF formatına çevrilip yayınlanırsa, o kadar kısa zamanda Semantik Web çağına girilecek.

Başta Google, Yahoo! ve Halka olmak üzere başlıca arama motoru sunucuları, anlamsal teknolojilerin sunduğu imkânlardan yararlanmak üzere yoğun çalışmalar yapıyor. Sonuçta ortaya çıkan sistemler, daha şimdiden Web 2.0 gibi açık ortamlarda bulunan, RDF formatında olmayan bilgileri bile doğru bir şekilde yorumlayacak güce erişmeye başladı.



Linked Open Data Bulutu, LOD-Cloud
(Richard Cyganiak ve Anja Jentzsch)

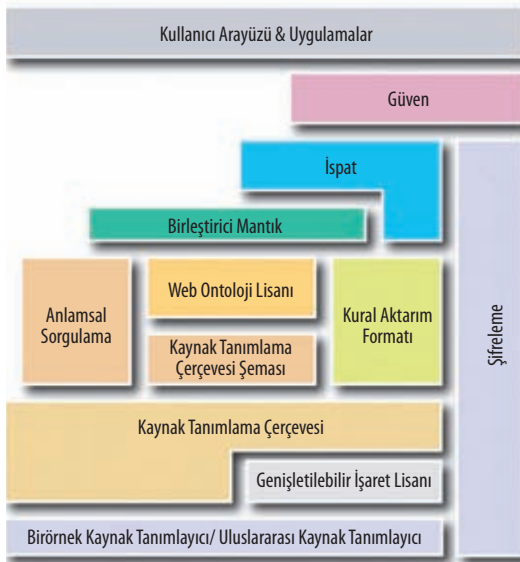


Fakat özellikle kurumsal alan gibi “kapalı” alanlarda, firmaların ve kurumların ellerindeki bilgileri şimdiden doğrudan RDF formatına çevirmesiyle Semantik Web teknolojilerinden daha bugünden büyük ölçüde faydalanılmaya başlayacağı ve dolayısıyla söz konusu firmalar ve kurumların rakiplerine karşı büyük avantajlar elde edeceği açık (W3C tarafından geliştirilmiş olan Semantik Web teknolojileri bunun için gerekli olgunluğa erişti). Tıp ve bioformatik alanında da yıllardan beri anlamsal teknolojiler ile geliştirilen çok başarılı uygulamalar var. DBpedia Web gibi “açık” bir alanda başarılı bir şekilde geliştirilen dünya çapındaki Semantik

Web projelerinden yalnızca biri. Sevindirici diğer bir gelişme de Semantik Web alanındaki açık kod kaynaklı projelerin sayısının günden güne artması. Artık çok yakın bir gelecekte Semantik Web teknolojilerinin günlük yaşamımızın hemen her alanına girmeye başlayacağı kesin. Ayrıca gerek yazımızın önceki bölümlerinde sıralanan sebeplerden gerekse Semantik Web teknolojilerinin beraberinde getirdiği yepyeni bilgi modelleme ve programlama felsefesinden dolayı Semantik Web teknolojileri yalnızca yeni nesil bir Web ve bilgi teknolojisi olarak değil aynı zamanda stratejik bir teknoloji olarak da görülmeli.



Börtçin Ege, Viyana Teknik Üniversitesi Bilgisayar Mühendisliği Fakültesi, Bilgisayar Mühendisliği Bölümü'nü bitirdikten sonra, yüksek lisans öğrenimini de 2005 yılında yine Viyana Teknik Üniversitesi'nde tamamladı. Yüksek lisans çalışması kapsamında Siemens-Almanya için birbiriyle bilgi alışverişinde bulunabilen iki kişisel veritabanı modelleyerek programladı. Yurtdışında bulunduğu süre zarfında özellikle Commerzbank, Siemens-Almanya ve Ericsson-Almanya gibi kuruluşlarda çalıştı. Şu anda Hacettepe Üniversitesi'nde Semantik Web üzerine doktora öğrenimi görüyor. Ayrıca çeşitli firma ve kurumlara Semantik Web teknolojileri konusunda danışmanlık yapıyor.



Semantik Web Katmanları

BAŞLICA SEMANTİK WEB UYGULAMALARI

Orade 11g
<http://www.orade.com/technetwork/database/options/semantic-tech/index.html>
Orade ilişkisel veritabanı ve RDF tabanlı anlamsal bir veritabanı

Sesame
http://en.wikipedia.org/wiki/Sesame_%28framework%29
Açık kod kaynaklı bir RDF tabanlı anlamsal veritabanı

Virtuoso
<http://virtuoso.openlinksw.com/>
Openlink firması tarafından geliştirilen RDF tabanlı anlamsal bir veritabanı

PoolParty
<http://poolparty.punkt.at/tr/>
Semantic Web Company tarafından geliştirilen çok yönlü bir kurumsal bilgi yönetim aracı

Kaynaklar

The World Wide Web Consortium (W3C),
<http://w3.org>
Berners-Lee, T., Hendl, J, ve Lassila, O.,
“The Semantic Web”, *Scientific American*,
17 Mayıs 2001.
Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R. ve Hellmann, S.,
“DBpedia - A Crystallization Point for the Web of Data, Web Semantics: Science, Services and Agents on the World Wide Web”, Cilt 7, Sayı 3, s. 154-165, Eylül 2009.

DBpedia SPARQL Benchmark
<http://blog.aksw.org/2011/dbpedia-sparql-benchmark-paper-wins-isw2011-best-paper-award/>
Semantic Web uygulamalarının performanslarını ölçmek için Leipzig Üniversitesi tarafından geliştirilmiş bir Benchmark

LIMES
<http://aksw.org/Projects/LIMES>
Semantic Web'de link keşiflerinde kullanılan bir framework. Yine Leipzig Üniversitesi tarafından geliştirilmiştir.

Protégé
<http://protege.stanford.edu/>
Stanford Üniversitesi tarafından geliştirilmekte olan açık kaynak kodlu bir ontoloji editörü

Jena
<http://jena.sourceforge.net/>
Java ile Semantic Web uygulamalarını programlanmasını için geliştirilmiş olan bir açık kaynak kodlu bir framework

Blumauer A., Pellegrini T., *Social Semantic Web*, Springer Yayınları, 2009.
Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z. DBpedia: A Nucleus for a Web of Open Data. 6th International Semantic Web Conference (ISWC), Korea 2007 Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. <http://lod-cloud.net/>
DBpedia project, <http://dbpedia.org>
RelFinder: Interactiv Relationship Discovery in RDF, <http://www.visualdataweb.org/relfinder.php>