

PAPER DETAILS

TITLE: Arsiv Belgelerinin Aranmasi Üzerine Anlamsal Arama Botu Gelistirilmesi ve Hukuki Terimler Üzerinden Degerlendirilmesi

AUTHORS: Mehmet ERTUGRUL, Demet ÇELİK ULUSOY

PAGES: 56-79

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/1535612>



Bilgi Yönetimi Dergisi

Cilt: 4 Sayı: 1 Yıl: 2021

<https://dergipark.org.tr/tr/pub/by>



Hakemli Makaleler

Araştırma Makalesi

Makale Bilgisi

Gönderildiği tarih: 26.01. 2021

Kabul tarihi: 29.04. 2021

Yayınlanma tarihi: 30.06. 2021

Article Info

Date submitted: 26.01. 2021

Date accepted: 29.04. 2021

Date published: 30.06. 2021

Anahtar sözcükler

Ontoloji; Anlamsal Ağ; Yapay Zekâ Kuralları; Akıllı Arama Web Botları; Arşiv Doküman Yönetim Sistemleri

Keywords

Ontology, Semantic Web, Artificial Intelligence Rules, Smart Search Web Bots, Archive Document Management Systems.

DOI numarası

10.33721/by.868918

ORCID

0000-0002-5833-4092 (1)

0000-0002-5542-9467 (2)

Arşiv Belgelerinin Aranması Üzerine Anlamsal Arama Botu Geliştirilmesi ve Hukuki Terimler Üzerinden Değerlendirilmesi

*Developing Semantic Search Bot for Searching Archive
Documents and Evaluating Using Legal Terms*

Mehmet ERTUĞRUL

Vodafone Mobile Operations LTD. Department of Information Technology,
Yazılım Uzmanı, mehmet.ertugrul@vodafone.com

Demet ÇELİK ULUSOY

Doğu Akdeniz Üniversitesi Hukuk Fakültesi Öğretim Üyesi,
demet.celik@emu.edu.tr

Öz

Arşivlerde tutulan belgeler tarihi öneme sahip olup; hukuki, siyasi, bilimsel, idari vb. nitelikte, maddi ve/veya manevi değeri olan ulusal ve hatta uluslararası niteliği olan veriler içerir. Günümüzde, arşivlerdeki belgeler, geleneksel belge arama zorlukları nedeniyle, dijital ortama aktarılarak, dijital arşiv katalogları, diğer adıyla arşiv yönetim sistemleri, oluşturulmuştur. Dijital kataloglar sayesinde, belgelerin yıpranmadan saklanması ve düzenlenmesi ve istenen bilgilere hızlı erişim için gerekli belgenin bulunması nispeten kolay hale geldi. Ayrıca, gerekli bir belgeyi talep eden diğer kuruluşlarla anında paylaşılabilmesi zaman ve usul ekonomisi sağlamış ve belge yönetimindeki sıkıntıları azaltmıştır. Ancak günümüz dijital kataloglar, Web 2.0 teknolojisi ile geliştirilmiş, sorgu tabanlı çalışan yazılımlardır. Bu tür sistemlerde ihtiyaç duyulan veriyi içeren belgenin taramasında söz dizimsel arama yapılmaktadır. Bu nedenle, ilgili belgenin bulunmasında ıskalamar olabilmektedir. Özellikle, hukuki kavram içeren belge aramalarında, hem daha çok sayıda tarihsel nitelikli arşiv belgesi bulunması, hem de eski lügatlerin halen kullanılabilmesi nedeniyle, sistemden dönen belge sonuçları daha karmaşık veya ciddi belge ıskalamarına neden olabilmekte ve bu durum, kullanıcıları ilgili belge bulmada zorlamaktadır. Oysa gelişen yapay zekâ teknolojisi sayesinde, bu katalogları akıllıca işletmek ve mevcut yapısını geliştirmek mümkündür. Anlamsal Web, meta veri depolamayı kolaylaştırarak ve bilgi alma sürecinin kalitesini artıran, anlamsal yetenekler ekleyerek mevcut dijital katalogları güçlendirebilir. Bu makalede, ontoloji ile biçimlendirilmiş olan arşiv belgeleri üzerinde anlamsal arama yapabilen bir yazılım botu sunulmaktadır. Bot, kural tabanlı çalışan anlamsal arama motoru olup, dijital kataloglar üzerinde yapılan veri taramasında kullanıcıların arama kelimelerini girdi kullanır. Bot, ontolojideki tüm kavramları, kavramsal bağları, türemiş kelime özelliği, eş anlamlı özelliği, tekil-çoğul ilişkisi vb. yapıları girdi kelimeler ile tarar ve aramayı zenginleştirir. Değerlendirme sürecinde, eski lügat dahil ederek hukuki terimler üzerinden bot ile çeşitli belge arama çalışmaları yapılmıştır. Daha sonra, söz dizimsel arama yapan başka iki dijital katalog üzerinde aynı terimler belge aramada tekrarlanmıştır, böylece; dönen sistem sonuçları kıyaslanmıştır. Elde edilen sonuçlar umut verici olup, bot'un mevcut kataloglar üzerinde uygulanabilirliğini, etkililiğini ve verimliliğini göstermektedir.

Abstract

Documents kept in the national archive are of historical importance and are the heritage of a nation. These documents can be legal, political, scientific, administrative etc., have material and / or moral value, may contain a variety of sensitive and important personal / national data. Today, documents in national

archives have been transferred to digital environment due to traditional document search difficulties, and digital archive catalogs, also known as archive management systems, have been established.

Thanks to digital catalogs, it has become relatively easy to store and organize documents without fraying and to find the necessary document for quick access to the desired information. In addition, the ability to instantly share a required document with other demanding organizations has provided time and procedure economy and reduced document management difficulties. However, these digital catalogs have been developed with today's Web 2.0 technology and are query-based systems. When searching for documents on such systems, syntactic search is performed. This situation may cause misses in searching and finding relevant documents. Especially in document searches with legal concepts, since there are more historical archive documents and old dictionaries can still be used, documents returned from the system may cause more complex results or serious document misses, which makes it difficult for users to find relevant documents. However, thanks to the developing artificial intelligence technology, it is possible to operate and develop these catalogs intelligently. Semantic Web technologies can enhance existing digital archive catalogs by facilitating metadata storage and adding semantic capabilities that improve the quality of the information retrieval process. In this article, a software bot capable of semantic search on archive documents which are previously formatted with ontology, is presented. Bot is a rule-based semantic search engine and uses users' search words as input for required data search on digital catalogs. Bot searches all the concepts, conceptual links, derived word features, synonym features, singular-plural relations etc. in ontology according to the entered keywords and enriches the current search process. During the evaluation phase, various document search studies were carried out on the bot, including the old dictionary and using legal terms. Next, the same terms are repeated in the document search on two other digital catalogs that perform syntactic search, so that returning system results were compared. The results of the sample document search studies applied are promising and show the applicability, effectiveness and efficiency of the proposed bot in existing archive catalogs.

1. Giriş

Arşivlerde dijitalleştirme, kurumlara ait rapor, form, evrak, yazışma, sözleşme, fatura, faks, film vb. kağıt gibi taşıyıcı ortamlarda bulunan belge ve dokümanların yüksek çözünürlüklü tarayıcı cihazlarla taranarak sayısallaştırılması ve bilgisayar sistemlerine istenilen dosya formatında kaydedilmesi akla gelir. Dijitalleştirmede çeşitli dosya biçimleri (.pdf, .jpeg, .tiff vb.) olup, bu dokümanların arana-bilirlik, küçük dosya boyutları ve dönüştürüle-bilirlik açısından en çok tercih edilen dosya biçimleri olarak kabul edilmiştir. Dijital ortama aktarılmış her bir doküman/belge arşiv yönetim sistemi veri tabanı üzerinde indeks bilgileri ve envanter, dosya, klasör numarası, içerik vb. meta verileri ile tutulmaktadır. Meta veriler, kısaca veri hakkında veri olup, eldeki verinin bir veya daha fazla özelliği hakkında bilgi sağlayan tanımlayıcı verilerdir. Amaç, veriler hakkındaki temel bilgileri özetlemek ve verilere kolay erişim sağlamaktır. Örneğin, belge tarihi, belge başlık, belge oluşturma saati ve tarihi. İlgili kurum/kuruluşlar veya araştırmacılar, dijital kataloğlara ait bu meta-veriler sayesinde, istenilen tarihi belgelere veya verilere, geleneksel arama metotlarına göre, nispeten daha hızlı bir şekilde erişebilmektedir. Hepsinden önemlisi bu tarihi verilerin asıllarının kaybolmasının önüne geçerek, daha az yıpranması ve sonraki kuşaklara aktarılmasını mümkün kılmaktadır. Bu tür tarihi belgeler, bir ülke için miras niteliğinde olup, oldukça önem arz etmektedir.

Dijital arşivlerde belge arama süreci en önemli hizmetlerden biridir. Günümüz Web teknolojisi (Web 2.0) (O'reilly, 2009, s.1) ile geliştirilen ve içerik yönetimi sağlayan yazılım çözümlerinde, bilginin anlamının makineler tarafından anlaşılmasına ve bilgiler arasında anlamsal bağlar kurulması sağlanamamaktadır. Teknolojik olarak mevcut Web ortamında sunulan bir sayfanın veya içeriğin bilgisayarlar tarafından anlaşılabilir olması zordur. Bu nedenle, Tim Berners-Lee ve arkadaşları, Web ortamında bulunan dosya veya belgelerin, sadece insanların algıladığı farklı biçimlerdeki milyonlarca bilgi dosyasını bilgisayarlar tarafından işlenerek anlamlandırılması ve istenilen bilgiye ulaşabilmesini sağlayan Anlamsal Web (AW) diğer adıyla Semantic Web, yaklaşımını geliştirmişlerdir (Berners-Lee, Hendler, & Lassila, 2001, s.1). AW'nin temel katkısı Ontoloji Bilgi Tabanı (OBT) (Gruber, 2009), kısaca Ontoloji, Web ortamında bulunan farklı yapıdaki kavramsal bilgilerin ve aralarındaki anlamsal bağların bilgisayarlar tarafından anlaşılır hale getirilmesini sağlayan bir yapıdır.

Ertuğrul ve Güneş çalışmalarında (Ertugrul & Gunes, 2017), Arşiv Belge Yönetim Sistem'lerine ait içerik terminolojisini bir ontoloji formatında toplamış, ayrıca resim formatında arşivlenmiş binlerce belgelerin otomatik olarak bir arşiv yönetim sistemi ara yüzünden meta veriye dönüştürülmesini sağlamışlardır. Sistem ontolojisini geliştirmek için Web Ontology Language (OWL) dili kullanmışlardır (McGuinness & Van Harmelen, 2004). Burada, Protégé ontoloji editörü (Noy et al. 2001, p.1)

kullanılarak oluşturulan ontoloji, OWL 2.0 versiyonunda (Hitzler et al. 2009) geliştirilmiştir. Bu ontoloji, Arşiv Yönetim Sistemi Ontolojisi (AYSO) olarak adlandırılmış, yeni belge kayıt modülünün kalıp ontolojisi olarak düşünülmüştür. Arşiv yetkilisi arşive atacağı her belgeyi, kayıt modüle bağlı bir ara yüz üzerinden kaydeder. Sistem AYSO kalıbını kullanarak o belge için arkada otomatik olarak özel bir “belge ontolojisi” dokümanı üretilmesini sağlar ve bu da sistem veri tabanında saklanır. Amaç, kullanıcıların ilerideki olası belge aramaları için her üretilen bu belge ontolojileri üzerinden anlamsal arama yapabilmektir. Bunun için, Anlamsal Arama Motoru (Bot) kullanılabilir. Ertuğrul ve Güneş çalışmalarında, bahsi geçen bu süreçleri detaylı olarak (Ertugrul & Gunes, 2017, s. 89-101) sunmuşlardır.

Bu çalışmada, yukarıda bahsi geçen AYSO ontolojine bağlı bir bot önerilmiştir. Bot, mevcut dijital kataloglara entegre olabilecek yapıda düşünülmüş ve mevcut sistemler ile genel çalışma prensibinden bu makalede bahsedilmiştir. Bot’un değerlendirme çalışmalarında, belge arama ara yüzü üzerinden, genellikle Arapça ve Farsça’nın baskın olduğu hukuk dili (Altay, 2002, s.22) göz önüne alınarak bazı hukuki terimlerle belge arama çalışmaları yapılmıştır. Bu deneysel belge aramalarında, hukuk terimlerinin seçilmesinin nedeni, kullanıcıların Farsça veya Arapça kökenli olan hukuk diline özgü kelimeleri aramalarda sıklıkla kullanmasıdır. Bugün hukuk dilinde büyük çoğunluğu Arapça kalmış olması dilin en önemli özelliği olarak gösterilir (Altay, 2002, s.23). Türk hukuk dilinde eski ve dolayısıyla Türkçe’nin olağan tamlama yapısına uymayan kalıpları Arapça ve Farsça tamlamalar oluşturmaktadır. Nitekim Türkçe hukuk dilinde gündelik yaşamda çok az kullanılan veya hiç kullanılmayan eski sözcükler halen önemli bir yer tutmaktadır (Altay, 2002, s.18,30). Zira hukuk alanında kullanılan terimlerin Arapça veya Farsça kökeninden kaynaklanarak kendine özgü bir niteliği haiz olduğu bilinmektedir. Bu nedenle, saha çalışması olarak hukuk alanı seçilmiş ve makalenin kapsamı sınırlandırılmıştır. Nitekim hukuki kavramların hâlâ Osmanlı Türkçesi ağırlıklı olması sebebiyle Arapça ve Farsça kökenli deyim, terkip ve nitelemelerin çokluğudur. Bu yapı üzerine bina edilen hukuk terminolojisi bugün de yaygın olarak kullanılmaktadır. Gerek güncel belgelerde ve gerekse arşiv malzemesi olmuş olanlarda belge türlerinden işlemlerine, doğdukları fonksiyon ve vakalar, hatta belge içeriğinde, Osmanlı Türkçesi ifadeler fazlaca yer almaktadır. Özellikle tarihsel nitelikteki hukuk belgelerinin aranmasında; geri dönen belge sayısının fazla olabilmesi, anlamsal bakımdan hedef belgeye isabet oranının az olabilmesi, dönen belge sonuçlarının karmaşık olabilmesi nedeniyle, arayıcının hedef belgeye erişmek için dönen sonuçlar üzerinde ileri aramalar yapmasına neden olmaktadır. Bu karmaşıklığa bir de hukuk alanına ait terimlerin eş anlamlısı olan sayısız terimlerin varlığını da hesaba kattığımızda (eski lügatleri içeren terimler olduğu da düşünüldüğünde) arayıcının aradığı belgeleri ıskalması oldukça ihtimaldir.

Bu çalışmada, geliştirilen sistemin mevcut başka sistemler ile değerlendirilmesi için, aynı sorgular Web 2.0 ile yazılmış ve söz dizimsel arama çalıştıran başka iki dijital belge katalogu üzerinde tekrarlanmış ve dönen belge sayısına bağlı kıyaslama çalışmaları gerçekleştirilmiştir. Bu çalışmada, geliştirilen Bot, Java dili kullanılarak gerçekleştirilmiştir. Java yazılımı ile geliştirilen bot ‘un, anlamsal kural tabanı AYSO üzerinde yerleşik olup, Semantic Web Rule Language (SWRL) (Horrocks et al. 2004, s.3-30) teknolojisi ile oluşturulmuştur.

Bu makalenin katkıları ilgili bölümler bazında sırasıyla şu şekilde düzenlenmiştir: önerilen sistemin geliştirilmesinde kullanılan teknolojiler ve algoritmaları tartışmak; sistemin ana modüllerini ve modüller aracılığıyla işlenen verilerini kısaca sunmak; sistem ontolojisini ve bot tarafından belge aramasında gerekli olan anlamsal kuralları kısaca açıklamak; sistemin botunun değerlendirmesini, kullanıcıların belge arama ara yüzlerini, birkaç örnek belge arama çalışmalarını ve sonuçlarını tartışmak; arşiv hizmetlerinde kanunlara göre kişisel/ ulusal verilere erişimin önemini tartışmak; Web 2.0 ile yazılmış başka bir dijital katalog üzerinde aynı sorgular denenerek, karşılaştırma sonuçlarını ortaya koyarak sonuçta araştırma çalışması hakkında bir sonuç sağlamaktır.

2. Materyal ve Metot

Kullanıcının bildiği ve ulaşmayı hedeflediği belirli bir belge bilinmiyorsa, bunun yerine ulaşmaya çalıştığı veriyi ona verecek belgeleri tespit etmek mümkündür. Literatürde bu arama daha çok bilgisayarlı arama olarak yer almaktadır (Rose & Levinson, 2004). Kullanıcı bir konu hakkında bilgi arar ve yapılan aramanın temel amacının algılanması da ayrıca mümkündür. Böylece ne tür belge aradığı anlaşıldığında

arama alanı daralmış olacak ve daha kolay bilgi ve dokümana erişim sağlayabilecektir (Levene, 2011). Milyonlarca belge üzerinde sürdürülen bir aramada, kullanıcının amacının algılanması, hedef veriyi içeren benzer belgeleri içeren mevcut dosyalar ve klasörlerin önceden belirlenmesini sağlar. Benzer verileri içeren fakat gerçekte hedef olmayan belgeleri içeren binlerce dosya veya klasör, böylece gereksiz yere taranmayacak ve aramaya dahil edilmeyecektir. Böylece, milyonlarca belge içinde, ilgili dosya veya klasörün önceden belirlenebilmesi işlem hızına katkı sağlayacaktır.

Yapay zekâya dayalı anlamsal belge arama konusunda çeşitli çalışmalar yapılmış ve akıllı arama sistemlerinde avantaj sağlayan anlamsal aramada fonksiyonel yetkinlikler tartışılmıştır (Preethi & Devi, 2012, s.18; Domingue, Fensel, & Hendler, 2011, s.659; Madhu, Govardhan, & Rajinikanth, 2011, s. 3-7). Bu fonksiyonel yetkinliklerin, Arşiv Belge Yönetim Sistemleri'ndeki anlamsal tabanlı belge arama süreci bakımından avantaj olduğu, benzer çalışmalarla da ortaya konulmuştur (MUNIN Project, 2015; Ahonen & Hyvonen, 2009; Hyvönen et al. 2008).

Bu çalışmada, AW teknolojisi kullanılarak, kullanıcının sisteme girdiği anahtar kelimeler işlenmiş ve arşiv belgelerinin içeriğini algılamaya yönelik, kullanıcı sorgusunun anlamsal bağlılığı olan başka terimlerle zenginleştirilmesi hedeflenmiştir. Halbuki geleneksel aramada, kullanıcı anahtar kelimeleri baz alınır ve tüm belgelerde girilen kelimeler içeriyor ya da içermiyor olduğuna bakılır. Anlamsal arama bir belgenin içeriğini anlamaya yönelik ne tür bir bilgi içerdiği, kimin tarafından oluşturulduğu, dönemi, vb. bilgilerin yanı sıra, o belgenin içeriğinde dair anahtar kelimelerinin de anlamını algılayarak, arama alanını genişletmektedir. Anlamsal arama, anlam biliminden faydalanarak, belge aramada en yakın ilgili dokümana ulaştırmayı hedefler. Sadece arşivde saklanan dosya numarası, tarih, başlık vb. etiket bilgisi içeren belgelerin bir listesini bulmanın aksine, ontolojideki kavramsal anlamlarına (ör. türemiş kelime özelliğine, eş anlamlı özelliğine, tekil-çoğul ilişkisine vb. niteliklere göre) aramasını sürdürebilir.

Bu çalışmanın ekibinde, bir yazılım uzmanı sistemin uçtan uca geliştirilmesinden ve bir hukukçu-da Web 2.0 ile geliştirilmiş çevrim içi kütüphane ve milli arşivlerde kataloglanan arşiv belgelerinin etiket yapısının çıkarılmasından, hukuki belgelerdeki kavramsal içerik tespitinden, Arapça veya Farsça kökenli hukuk dili dâhil ederek, hukuki terimler üzerinden bot ile çeşitli belge arama çalışmalarının yapılmasından sorumlu olmuştur. Çalışmanın temel işlev adımları, iki kısımda incelenebilir:

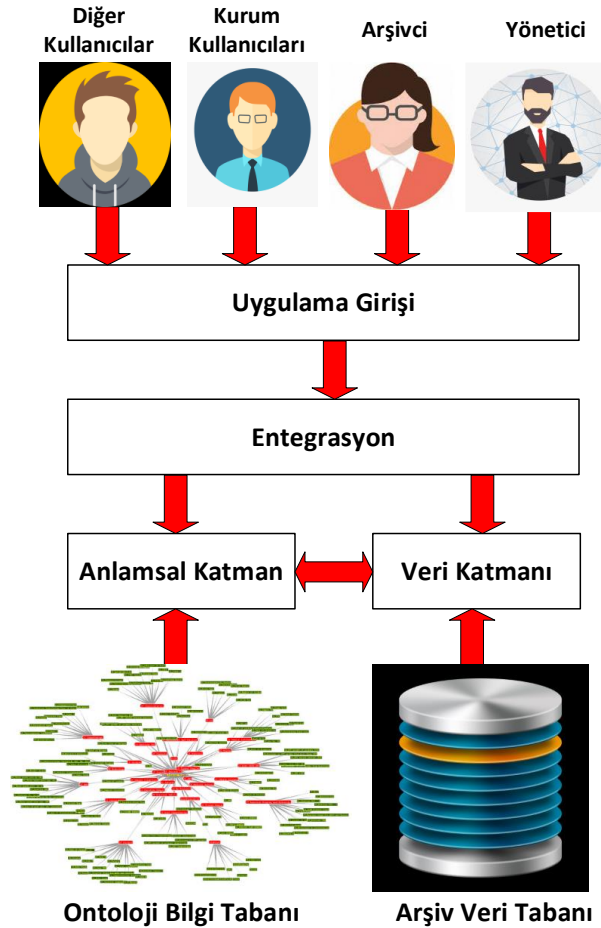
Hazır ontolojinin kullanılması süreci: Ertuğrul ve Güneş çalışmasında (Ertugrul & Gunes, 2017), Arşiv Yönetim Sistemi Ontolojisi (AYSO) geliştirilmesi ve her yeni kaydedilen belgenin bir arşiv yönetim sistemi üzerinden AYSO kalıbı kullanılarak otomatik olarak birer ontolojiye dönüştürülmesini sağlamışlardır. Bu çalışmada, önerilen anlamsal belge arama bot'u ile hâlihazırda üretilmiş ve veri tabanında saklanan bu belge ontolojileri üzerinden anlamsal belge aramanın yürütülmesi hedeflenmiştir. İki çalışma birbirini tamamlayan nitelikte düşünülmüştür. Ertuğrul ve Güneş çalışmasında yapılan katkılar:

- Dijital Arşiv Yönetim Sistemlerine bağlı olarak; klasör, dosya, belge ve sayfa kavramlarına ait fişler/etiketlerin belirlenmesi,
- AYSO içindeki kavram ya da sınıf yapılarının (Ontology Classes) oluşturulması,
- AYSO'nun içinde belirlenen sınıf yapılarının arasındaki nesne ve veri türü özelliklerin (Object Type - OTP ve Data Type properties - DTP), önceden belirlenmiş fişler/etiketlere göre kurgulanması ve yaratılması,
- AYSO'nin içinde belirlenen sınıf yapılarının yapılarına ait veri elemanlarının (Ontology Individuals) oluşturulması,
- Ontoloji üzerinde kelime eşleştirmeye yönelik anlamsal kural tabanının geliştirilmesi,
- Yeni belge kayıt için "belge ontolojisi" üretme modülünün geliştirilmesi,
- AYSO ve belge ontolojisi üretme modülünün doğrulaması ve testlerinin yapılması süreçleri sağlanmıştır.

Bu çalışmanın katkısı: Temel katkı Anlamsal Arama Motoru (Bot) geliştirilmesi ve değerlendirilmesi süreçleridir. Java¹ dili kullanılarak Netbeans 8.0² platformu ile bot'un geliştirilmesi ve bot'un belge arama performans değerlendirme çalışmaları yürütülmüştür. Bot'un geliştirilmesindeki işlenen temel adımlar şöyledir:

- Arama bot 'unun terimler üzerinden anlamsal eşleşme mekanizmasının tasarlanması,
- Arama bot 'unun Web arama servislerinin geliştirilmesi,
- Web servislerinden ASYO bilgi tabanına kullanıcı sorgularının aktarılması ve AYSO üzerindeki SWRL kurallarını çalıştırarak bilgi çıkarımın sağlanması,
- Bot için bir ara yüzü tasarlanarak test veri tabanını üzerinde çeşitli arama sorguları ile sistemin değerlendirilmesi,
- Test ve revizyonların tamamlanması adımlarını içermektedir.

Ertuğrul ve Güneş çalışmasında (Ertugrul & Gunes, 2017) bahsi geçen AYSO'nun ve üretilen her belge ontolojisinin yapısı, önerilen bot'un çalışması mekanizmasını anlamak için önemli olduğundan, bu kısımlara dair ait detaylar bu makalede kısa olarak tanımlanmıştır. Bu çalışmada daha çok, ikinci süreçten anlamsal arama bot 'undan bahsedilmiştir. Anlamsal Arama Bot'u, sistemin en önemli modülü olup, mevcut dijital katalog sistemlerine arka planda bir Web servis gibi entegre olacak biçimde düşünülmüştür. Mevcut dijital katalog sistemleri, yetkili veya yetkisiz kullanıcılar tarafından, kısıtlı ya da kısıtsız olarak rollere göre kullanılmaktadır.



Şekil 1. Anlamsal Tabanlı Dijital Arşiv Sisteme Genel Bakış

Şekil 1'de, sistemin belge arama süreci, kullanıcı tipleri ve arama sürecine dahil edilen tüm modüller sunulmuştur. Önerilen Anlamsal Arama Bot'u ile, mevcut dijital katalogların belge arama web ara yüzüne arkadan entegrasyonu sayesinde, dijital katalogların kendi arama modülleri ile belgelerine daha

¹ <http://java.sun.com/products/archive/j2se/6u7/index.html>

² <https://netbeans.org/downloads/8.0/>

akıllıca erişimi hedeflenmiştir. Kurum kullanıcıları, arşivciler, yönetici ve Web üzerinden belgelere erişmek isteyen diğer tüzel kişiler vb. kullanıcılar sistemdeki kayıtlı yetki seviyelerine göre sistemi kullanabilir ve erişime yetkili oldukları belgelerin içeriğini görüntüleyebilirler. Bot, kullanıcıların aradıkları belgeleri AW teknolojisi sayesinde, akıllı bir web hizmeti gibi (Şekil 1’de görüldüğü üzere; “Semantic & Data Layer” kısımları üzerinden sistemin veri tabanı ve ontolojisine girilen kullanıcı sorgusunun taşınması ve işlenmesiyle) yapabilmesine olanak sağlayacaktır.

Bu çalışmada, Java’dan ASYO bilgi tabanına bağlantı kurmak, ontolojideki kavramsal bilgilerin okunmasını, güncellenmesi, yeni bilgi eklenmesi veya silinmesini gibi işlemler, OWL API (Ontology Parser) (Horridge & Bechhofer, 2011) kütüphanesi kullanılarak yapılmıştır. Ayrıca SWRL kuralları, bir belge içindeki geçen kavramsal tanımları, kullanıcının girdiği anahtar kelimeler ile eşleştirerek ilgililerini belirlemek amaçlı oluşturulmuştur. Anlamsal Arama Bot’u belge anahtar kelimeleri ile kullanıcı girdi anahtar kelimeleri ontoloji üzerinde kavramsal yakınlıklarına bakarak, her belge için bir benzerlik skoru hesaplamaktadır. Sorguda ilgili bulunan tüm belgeler için ayrı ayrı bu skor hesaplanır ve kullanıcıya büyükten küçüğe doğru liste halinde sunulur. Bu sayede, kullanıcıların belge ararken sisteme girdiği anahtar kelimeler, ontolojideki kavramlar ve özellikler/ilişkiler ile eşleştirilir ve arşiv veri tabanındaki ilgili belge(lerin) bulunması sağlanır. Bahsi geçen eşleştirme sürecinde, kullanılan metotlar genel olarak:

- Ontoloji ve SWRL kuralları üzerinden, anlamsal eşleştirme (Semantic Matchmaking) ve çıkarsama,
- Jaro-Winkler algoritması³ (Winkler, 1999; Winkler & Thibadeau, 1991) üzerinden, kelimeler arasında benzerlik skoru hesaplanır.

Jaro-Winkler algoritması, kullanıcı arama kelimeleri/kavramları $R = \{R_1, R_2 \dots R_m\}$ kümesi, belge $B = \{B_1, B_2 \dots B_n\}$ anahtar kelimeleri arasında uygulanır. Bu kümeler, sistem botunun eşleştirme fonksiyonuna dizin parametre olarak girer. Sonuçta, o belge için total bir benzerlik skoru aşağıdaki denklemlere göre hesaplanır. Kartezyen çarpıma alınan her iki küme elemanlarından oluşan her kelime ikilisi için (R_i, B_j) , benzerlik skoru Denklem 1’e göre hesaplanır (Winkler, 1999; Winkler & Thibadeau, 1991). Kelime ikililerinden en yüksek eşleşme skoru oluşturan, genel toplama katılır ve o iki kelime hafızadan çıkarılır. Aynı işlem geri kalan tüm kelime ikililerine uygulanır. Tüm kelime ikililerinin eşleşmesi bitince, genel toplamın ortalaması alınır, Denklem 2 de gösterilmiştir.

İki kelime (s_1, s_2) arasındaki Jaro benzerlik mesafesi d_j olsun;

$$Similarity2Words(R_i, B_j) = d_j = \begin{cases} 0, & \text{if } m = 0 \\ \frac{1}{3} \left(\frac{m}{s_1} + \frac{m}{s_2} + \frac{m-t}{m} \right) |a|, & \text{otherwise} \end{cases} \quad (1)$$

m : iki kelime arasında eşleşen karakterlerin sayısı,

t : iki kelime arasında yer değişikliği görülen karakterlerin sayısının yarısı.

$SimilarityScore(R, B)$

$$= avg \left(\sum_{j=1}^{\max(n_R, n_B)} [\max_{j,i} \{Similarity2Words(R_i, B_j)\}, delete R_i, B_j] \right) \quad (2)$$

3. Sistem Mimarisi ve Kullanımı

Daha önce belirtildiği gibi, sistemin ana teknik bölümleri şunlardır: (1) Arşiv Yönetim Sistemi Ontolojisi (AYSO), (2) Anlamsal Web Kuralı Bilgi tabanı ve (3) Anlamsal Arama Motoru (Bot). Sistemin genel mimarisi Şekil 2’de sunulmuştur.

³ http://en.wikipedia.org/wiki/Jaro%E2%80%93Winkler_distance



Şekil 2. SWRL Kuralları Aracılığıyla Sistemin Botu Tarafından İşlenen Girdiler ve Döndürülen Çıktılar

Sistemin çalışma prensibi ve kullanılan metodolojiler, aşağıda işlem sırasıyla kısaca anlatılmıştır ve Şekil 2’deki numaralara işaret edilmiştir.

- (1) Kullanıcı belge arama ara yüzüne bağlanır ve belge aramak için arama kelimelerini girer.
- (2) Girilen arama kelimelerinin zenginleştirilmek üzere, sistemin Anlamsal Arama Motoru (Bot), AYSO’ya bağlanır ve AYSO üzerinde bulunan tüm SWRL kuralları çalıştırılır.
- (3) AYSO’da bu arama kelimelerine işaret eden diğer mevcut kavramlar ve veri /nesne türü özellikleri (ör. türemiş kelime özelliği, eş anlamlı özelliği, tekil-çoğul ilişkisi vb. bağlı kelimeler) tespit edilir ve kullanıcı arama kelimelerini zenginleştirir.
- (4) Bot, veri tabanında bulunan daha önceden yaratılmış tüm belge ontoloji dosyalarına (B000000X.owl) bağlanır ve zenginleştirilmiş arama kelime kümesini, her belge ontolojisindeki anahtar kelime nesne özelliğine bağlı kelimeler ile eşleştirir. Kelime eşleştirmesi esnasında Jaro-Winkler algoritması (Winkler, 1999; Winkler & Thibaudeau, 1991) kullanılır. Bu algoritmanın girdileri, zenginleştirilmiş yeni arama kelime kümesi ile her bir belge ontolojisinde kayıtlı “b_keywords” nesne türü özelliğine bağlı tüm anahtar kelimelerdir. O belgeye ait tüm kelimeler ile girilen arama kelimeleri, bu algoritmaya birer parametre olarak girer ve her kelime eşleşme işlemi için bir skor hesaplanır. Daha sonra o belgeye bağlı tüm anahtar kelimelerin skorları toplanarak, o belge için genel bir skor değeri hesaplar.
- (5) Sistem, veri tabanındaki belge ontolojilerine ait metin dosyalarını taradıktan sonra, her belge için elde edilen skora göre büyükten küçüğe göre sıralanır ve kullanıcının ara yüzüne listeler.

AYSO ontolojisi, belge ontolojilerinin otomatik yaratılması, kural tabanının çalıştırılması ve her belge için Jaro-Winkler benzerlik skorunun hesaplanması aşağıda bölümlerde detaylı sunulmuştur.

3. 1. Arşiv Yönetim Sistemi Ontolojisi (AYSO) ve Belge Ontolojileri B00000X.owl

Protégé (Protégé Ontology Editor, 1999), ontoloji oluşturmak için açık kaynaklı bir yazılımdır ve sistemimizin ontolojisini geliştirmede kullanılmıştır. Sistem, iki farklı ontoloji yapısı içermektedir:

owl:class,	54	owl: Thing, arama_sonuc, kurum, lokasyon, bolge, ilce, sehir, ulke,
rdfs:subClassOf		osmanlica_sozluk, user_keywords, vgm_arshiv, belge_yonetim,
		belge_alt_tur, belge_arshiv_sekli, belge_kategory, belge_tur, erisim_haklari,

		klasor, dosya, belge, sayfa, vakif_defter_turleri, yayin_bilgisi, belge_ortam, dil, kisi, editor, personel, yazar, tarihi_donem, yayinci
ObjectProperty	17	arama_sonucu_deger, b_altTuru, b_arsivSekli, b_bolge, b_dili, b_dosyasiKodNo, b_duzenleyenKurum, b_editoru, b_erisimHakki, b_eseri, b_excellSiraNo, b_geldigiKurum, b_ilce, b_kategory, b_kayitEdenPersonelNo, b_klasorununKodNo, b_ortami, b_sehir, b_tarihiDonemi, b_turu, b_ulke, b_vakifDefterTuru, b_yayincisi, b_yayinlamistir, b_yazari, d_klasorununKodNo, k_belgeyiIcerir, k_dosyayiIcerir
DatatypeProperty	22	b_arkaKapagi, b_dijitalizeTarihi, b_duzenleyenKisi, b_ilkSayfasi, b_kayitNo, b_keywords, b_klasorunAdi, b_onKapagi, b_orjinalKayitNo, b_sayfaAdeti, b_tarihAraligi, b_tarihi, b_url, b_vakifaGelisTarihi, b_yeniKlasorKodNo, d_acilisTarihi, d_belgeAdeti, d_gomlekSiraNo, d_keywords, d_mikroFormNo, d_orijinalNo, d_tasnifKodu, d_varakAdeti, esanlamlisi_kelime, s_byteDegeri, s_sayfaNo, s_x_pixel, s_y_pixel, turemis_kelime, u_keywords
NamedIndividual	1347	Arama motoru çalıştıkça değişkenlik göstermektedir.
SWRL Kural Tabanı	7	Rule 1: belge(?b), osmanlica_sozluk(?o), user_keywords(?t), b_keywords(?b, ?t), turemis_kelime(?o, ?t) -> arama_sonucu_deger(?b, ilgili) Rule 2: belge(?b), user_keywords(?uk), b_keywords(?b, ?uk) -> arama_sonucu_deger(?b, ilgili) Rule 3: belge(?b), osmanlica_sozluk(?o), user_keywords(?uk), b_keywords(?b, ?b_kyw), esanlamlisi_kelime(?o, ?b_kyw), turemis_kelime(?o, ?term), u_keywords(?uk, ?term) -> arama_sonucu_deger(?b, ilgili) Rule 4: belge(?b), osmanlica_sozluk(?o), user_keywords(?uk), b_keywords(?b, ?b_kyw), esanlamlisi_kelime(?o, ?term), turemis_kelime(?o, ?b_kyw), u_keywords(?uk, ?term) -> arama_sonucu_deger(?b, ilgili) Rule 5: belge(?b), osmanlica_sozluk(?o), user_keywords(?uk), b_keywords(?b, ?t), turemis_kelime(?o, ?t), u_keywords(?uk, ?t) -> arama_sonucu_deger(?b, ilgili) Rule 6: belge(?b), osmanlica_sozluk(?o), user_keywords(?uk), b_keywords(?b, ?e), esanlamlisi_kelime(?o, ?e), u_keywords(?uk, ?e) -> arama_sonucu_deger(?b, ilgili) Rule 7: belge(?b), osmanlica_sozluk(?o), user_keywords(?e), b_keywords(?b, ?e), esanlamlisi_kelime(?o, ?e) -> arama_sonucu_deger(?b, ilgili)

Çizelge 1: AYSO Ontolojisinin Büyüklüğü ve İçerdiği Kavramlar Hakkındaki Veriler

(1) AYSO, diğer adıyla “Arsiv_Ontology. owl”: Sistemin ontolojisi yaratılırken, arşiv belgelerinin genel yapısı incelenmiş ve bu belgelerin içeriğine işaret eden anahtar kelimeler, yapısal özelliklerine işaret eden dosya ve klasör detayları, yazarı, kayıt eden kişi, biçimsel formatı, sayfa sayısı, bağlı olduğu diğer kaynaklar, tarihi dönemi vb. yapıların belge aramalarında önemli veriler olduğu anlaşılmıştır. Bu nedenle, sistemin ontolojisinde bu yapılarla yer verilmiş, ontoloji kavramları (ör., “belge”, “yazar”), veri özelliği (ör., “belge tarihi”), nesne özelliği (ör., “belge_dili”) vb. ontoloji elemanları olarak, OWL dili kullanılarak yaratılmıştır. Aynı zamanda, AYSO, her yeni bir belge kaydında o belgeye ait yaratılan belge ontolojisi (“B000000X.owl” metin dosyası) için bir kalıp ontolojidir. Bu nedenle, bu ontolojiyi sistemin “Upper Ontolojisi” olarak düşünebiliriz. Diğer deyişle, üst düzey ontolojisi, üst model veya temel ontolojisi olarak da düşünebiliriz. Sistemde sadece bir adet üst model ontoloji bulunmaktadır. Bu üst model ontolojide, arşivlenmiş belgelerin genelinin aranmasında ve yönetiminde etkin olan ilgili tüm alan ilgili kavramlar birer sınıf veya alt sınıfları olarak tanımlanmıştır. Sınıflar arasında kurulan bağlar, özellik tanımları olarak yer almıştır. Bu çalışmada, her belge öreni için bir adet belge ontolojisi saklanmaktadır. Bu belge ontolojilerinde tutulan terimler, üst model ontolojimizdeki terimlere işaret eder ve belge ontolojilerindeki tüm sınıfların süper sınıfları veya üst kümeleri olarak bağlar oluşturur. Üst model ontolojisinde, sınıflar arası çeşitli nesne ve veri türü özellik bağları tanımlanmıştır. Örneğin,

“belge (owl:Class) → b_tarihiDonemi (OTP)→ tarihi_donem (owl:Class)”, veya “belge (owl:Class) → d_gomlekSiraNo (DTP)→ Integer (owl:xsd^^Int)”, vb).

Çizelge 1’de AYSO üzerinde yaratılan tüm kavramsal sınıfları, alt sınıfları, nesne türü özellikleri, veri türü özellikleri, veri öğeleri (OWL individuals), SWRL kuralları verilmiştir. Sonuç olarak, AYSO sistemin kalıp ontolojisi olup; 31 temel sınıf ve alt sınıf, 30 veri türü özellik, 28 nesne türünde özellik, yaklaşık 1347 veri öğesi (OWL: NamedIndividual) ve 7 adet SWRL kuralı içermektedir.

(2) Belge Ontolojileri (B000000X.owl): Belgelere ait üretilen belge ontolojilerinin amacı ise AYSO’dan farklı olup, ileride belge arama sırasında kullanılmak üzere veri tabanında depolanmaktadır. Arşiv personeli tarafından, dijital arşive aktarılacak olan her yeni belge için, o belgeye özgün “B000000X.owl” belge ontolojisi bir metin dosyası gibi üretilir. Daha önce de belirtildiği gibi, belge ontolojileri üretilirken, AYSO bir kalıp olarak kullanılır. Arşivdeki milyonlarca belge için milyonlarca belge ontolojisi yaratılabilir. Sonunda, o belge için, B000000X.owl adında bir ontoloji dosyası yaratılarak veri tabanına aktarılır.

Belge ontolojileri, yeni belge kaydı tanımlama ara yüzü kullanılarak, belgenin verileri bir defaya mahsus olmak suretiyle dijital kataloğa girildiğinde otomatik üretilir. Sistem, o anda o yeni belgenin kavramsal içeriğini ve yapısal özelliklerini makinalarca anlaşılan bir formatta, diğer deyişle OWL 2 formatına, dönüştürür. Böylece, belge üzerinde çeşitli özelliklere göre ileride anlamsal aramalar yapılabilir. Sistem tarafından otomatik olarak o an da üretilen “B0000001.owl” belge ontolojisine ait OWL 2 dosyası Çizelge 2’de gösterilmiştir. Belge ontolojisi, çeşitli nesne ve veri türü özellik tanımların içermektedir (ör. b_keyword → “Türkiye Dış İlişkiler Antlaşmalar”, b_sayfaAdeti → “565”, b_tarihi → “01/01/1982” vb).

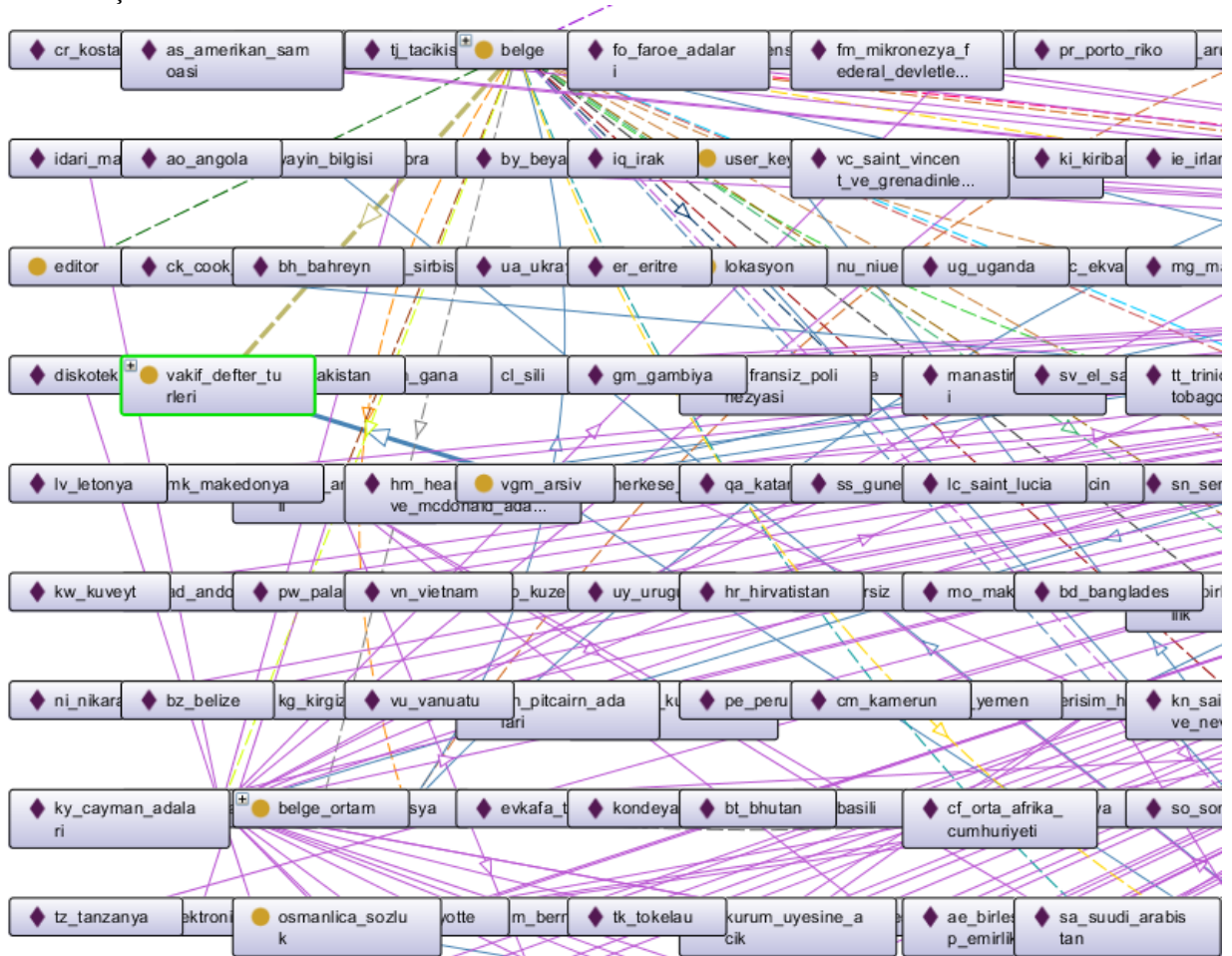
“B0000001.OWL”: Sisteme kaydedilen yeni bir belgenin ontolojisi.

1	<?xml version="1.0"?>	53	<ObjectPropertyAssertion>
2	<Ontology xml:base=".../ArsivOntology#">	54	<ObjectProperty IRI="#b_ortami"/>
3	<ClassAssertion>	55	<NamedIndividual IRI="#B0000001"/>
4	<Class IRI="#belge"/>	56	<NamedIndividual IRI="#dijital"/>
5	<NamedIndividual IRI="#B0000001"/>	57	</ObjectPropertyAssertion>
6	</ClassAssertion>	58	<DataPropertyAssertion>
7	<ObjectPropertyAssertion>	59	<DataProperty IRI="#b_keywords"/>
8	<ObjectProperty IRI="#b_erisimHakki"/>	60	<NamedIndividual IRI="#B0000001"/>
9	<NamedIndividual IRI="#B0000001"/>	61	<Literal datatypeIRI="#string">Türkiye Dış
10	<NamedIndividual IRI="#herkese_acik"/>	62	İlişkiler Antlaşmalar</Literal>
11	</ObjectPropertyAssertion>	63	</DataPropertyAssertion>
12	<ObjectPropertyAssertion>	64	<DataPropertyAssertion>
13	<ObjectProperty IRI="#b_geldigiKurum"/>	65	<DataProperty IRI="#b_keywords"/>
14	<NamedIndividual IRI="#B0000001"/>	66	<NamedIndividual IRI="#B0000001"/>
15	<NamedIndividual	67	<Literal
16	IRI="#vakiflar_genel_mudurlugu_merkez_kutuphanesi"/>	68	datatypeIRI="#string">Antlaşmalar</Literal>
17	</ObjectPropertyAssertion>	69	</DataPropertyAssertion>
18	<ObjectPropertyAssertion>	70	<DataPropertyAssertion>
19	<ObjectProperty IRI="#b_kayitEdenPersonelNo"/>	71	<DataProperty IRI="#b_keywords"/>
20	<NamedIndividual IRI="#B0000001"/>	72	<NamedIndividual IRI="#B0000001"/>
21	<NamedIndividual IRI="#vp00135"/>	73	<Literal datatypeIRI="#string">Türkiye Dış
22	</ObjectPropertyAssertion>	74	İlişkiler </Literal>
23	<ObjectPropertyAssertion>	75	</DataPropertyAssertion>
24	<ObjectProperty IRI="#b_klasorununKodNo"/>	76	<DataPropertyAssertion>
25	<NamedIndividual IRI="#B0000001"/>	77	<DataProperty IRI="#b_keywords"/>
26	<NamedIndividual IRI="#113"/>	78	<NamedIndividual IRI="#B0000001"/>
27	</ObjectPropertyAssertion>	79	<Literal datatypeIRI="#string">Türkiye Dış
28	<ObjectPropertyAssertion>	80	İlişkiler ve Ortadoğu</Literal>
29	<ObjectProperty IRI="#b_ortami"/>	81	</DataPropertyAssertion>
30	<NamedIndividual IRI="#B0000001"/>	82	<DataPropertyAssertion>
31	<NamedIndividual IRI="#dijital"/>	83	<DataProperty IRI="#b_sayfaAdeti"/>
32	</ObjectPropertyAssertion>	84	<NamedIndividual IRI="#B0000001"/>
33	<ObjectPropertyAssertion>	85	<Literal datatypeIRI="#integer">565</Literal>
34	<ObjectProperty IRI="#b_sehir"/>	86	</DataPropertyAssertion>
35	<NamedIndividual IRI="#B0000001"/>	87	<DataPropertyAssertion>
36	<NamedIndividual IRI="#ankara"/>	88	<DataProperty IRI="#b_tarihi"/>
37	</ObjectPropertyAssertion>	89	<NamedIndividual IRI="#B0000001"/>
38	<ObjectPropertyAssertion>	90	<Literal
39	<ObjectProperty IRI="#b_tarihiDonemi"/>	91	datatypeIRI="#string">01/01/1982</Literal>
40	<NamedIndividual IRI="#B0000001"/>	92	</DataPropertyAssertion>
41	<NamedIndividual IRI="#tbmm_tc"/>	93	<DataPropertyAssertion>
42	</ObjectPropertyAssertion>	94	<DataProperty IRI="#b_vakifaGelisTarihi"/>

43	<ObjectPropertyAssertion>	95	<NamedIndividual IRI="#B0000001"/>
44	<ObjectProperty IRI="#b_turu"/>	96	<Literal
45	<NamedIndividual IRI="#B0000001"/>	97	datatypeIRI="#string">14/02/2017</Literal>
46	<NamedIndividual IRI="#kitap"/>	98	</DataPropertyAssertion>
47	</ObjectPropertyAssertion>	99	<DataPropertyAssertion>
48	<ObjectPropertyAssertion>	100	<DataProperty IRI="#b_yeniKlasorKodNo"/>
49	<ObjectProperty IRI="#b_yazari"/>	101	<NamedIndividual IRI="#B0000001"/>
50	<NamedIndividual IRI="#B0000001"/>	102	<Literal
51	<NamedIndividual IRI="#resat_yazici"/>	103	datatypeIRI="#string">k2016_113</Literal>
52	</ObjectPropertyAssertion>	104	</DataPropertyAssertion>
		105	<DataPropertyAssertion>
			<DataProperty IRI="#d_acilisTarihi"/>
			<NamedIndividual IRI="#B0000001"/>
			<Literal
			datatypeIRI="#string">01/02/2010</Literal>
			</DataPropertyAssertion>

Çizelge 2. "B0000001" Belgesinin ("B0000001.owl") Ontolojisinin OWL 2 Formundan Bir Kesit

AYSO bilgi tabanın Protégé ontoloji editörü Onto Graph tabından elde edilen görüntüsü aşağıda Şekil 3'de verilmiştir.



Şekil 3. AYSO Bilgi Tabanın Protégé Ontoloji Editörü OntoGraph Görüntüsü

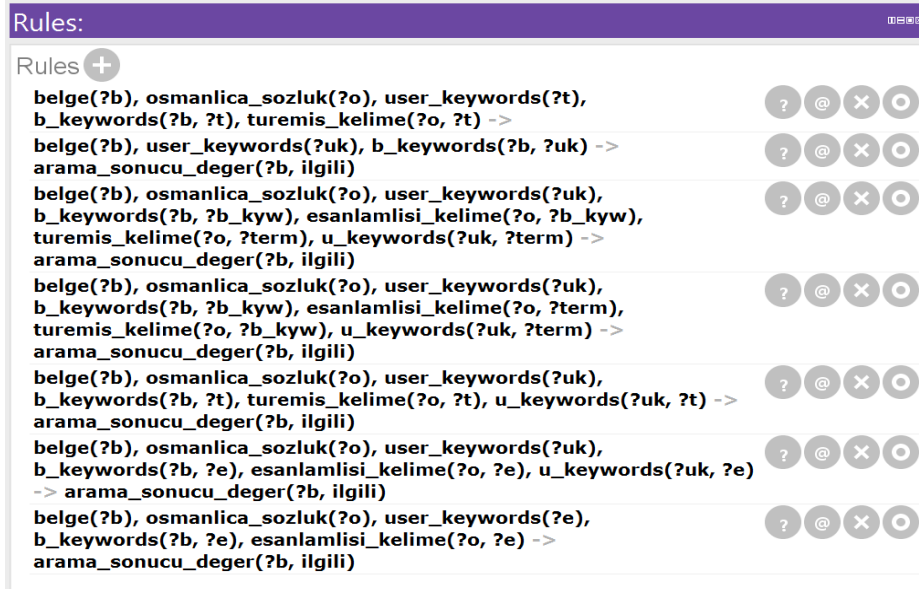
AYSO'ya bağlı belge ontolojilerinin otomatik bir ara yüz üzerinden oluşturulması hakkındaki detaylı bilgi Ertugrul & Gunes'in (2017) çalışmalarında verilmiştir. Bu nedenle arşiv belgelerinin birer belge ontolojisinin yaratılması ve ilgili ara yüz çalışmalarına bu makalede değinilmemiştir. Bu süreçte, dijital arşiv kataloğunda yer alan her arşiv belgesinin ayrıca bir adet belge ontolojisinin de sistemin veri tabanında yerleşik olduğu varsayılmıştır. Bu makalemizin gelecek bölümlerinde, önceden yaratılmış hali hazırda veri tabanında yerleşik belge ontolojileri üzerinde anlamsal aramaların nasıl yürütüldüğü ve sistemden dönen cevaplar tartışılmıştır.

3. 2. Anlamsal Web Kuralı Bilgi Tabanı

Bu çalışmada, Semantic Web Rule Language (SWRL) (Horrocks et al. 2004), sistemin anlamsal kural tabanını geliştirmek için kullanılmıştır ve daha güçlü tümdengelimle muhakeme yetenekleri sağlar. Anlamsal Sorgu Geliştirilmiş Web Kural Dili (SQWRL) (O'Connor & Das, 2009), SWRL'ye dayalı olan ifade edici bir sorgu dilidir. SQWRL, anlamsal kuralları sorgulamak için kullanılır. AYSO'daki SWRL kuralları sayesinde, çıkarsanan kelimelere (örneğin; eş anlamlı özelliği: “hürriyet, serbestlik, erkinlik, serbesti, hürlük, bağımsızlık”) erişmek için girdi olarak anahtar kelimeler (örneğin; “özgürlük”) ontolojiye sürülür. Daha sonra, sistem zenginleştirilmiş girdi kelimeleri kullanarak dijital kataloglardaki belgelerin OWL metin dosyalarını tarar. AYSO üzerinde şimdilik 7 adet SWRL kuralı oluşturulmuştur. Milyonlarca belge için bu kurallar genel olup, AYSO'daki kavramsal bağları, türemiş kelime özelliği, eş anlamlı özelliği, tekil-çoğul ilişkisi vb. yapıları girdi kelimeler aracılığıyla tarar ve kullanıcı arama kelimelerini zenginleştirir. Şekil 4'de, sistem çıkarım botunun çalıştığı SWRL kuralları ve bu kurallarda işlenen girdileri ve döndürülen çıktıları göstermektedir. Örneğin, **Kural 2**'yi incelersek: “belge (?b), user_keywords (?uk), b_keywords (?b, ?uk) → arama_sonucu_değer (?b, ilgili)”.

- belge (?b): ?b değişkeni bir belgeyi temsil edip, eğer
- user_keywords (?uk): ?uk bir kullanıcı anahtar kelimesi ise, eğer
- b_keywords (?b, ?uk): ?b belgesi kullanıcının bu ?uk kelimesini içeriyor ise, bu durumda,
- arama_sonucu_değer (?b, ilgili): ?b belgesi için sonuç “ilgili” dir.

Kullanıcının girdiği kelimeleri zenginleştirmek amacıyla, AYSO'da alan ilgili tutulan kelimeler ile kullanıcı girdi anahtar kelimeleri arasında benzerlik bakılır. Bu benzerlikler çeşitli anlamsal bağlara bakılarak yapılır: Osmanlıca dan türemiş, eş anlamlısı, aynısı, gibi özellikler taranır. Sonuç belge ilgili veya ilgisiz karardır. İlgisiz belgeler gelecek adıma dahil edilmez. Fakat ilgili bulunan belgeler gelecek adım olan benzerlik skoru hesaplaması yapılan servise gönderilir. Amaç, ilgili bulunan her belgenin ne oranda hedef veriye yakın olduğunu hesaplamaktır. Bu nedenle, bu serviste ilgili bulunan her belge için bir benzerlik skoru hesaplanır.



Şekil 4. Protégé Editörü Üzerinde SWRL Kurallarının Görünümü

4. Değerlendirme—Bot Örnek Belge Arama Çalışmaları

Bu bölümde, sistemin belge arama ara yüzünde, örnek bir sorgulama ve sorgu sonucunda sistemden dönen çıkarsama sonuçlarının kullanıcıya sunulması anlatılmıştır. Bir arama başlatıldığında, sistem arşivindeki belgelere ait olan OWL dosyalarına girer ve yukarıdaki SWRL çıkarsama kurallarını çalıştırır ve sonucunda “ilgili” ya da “ilgisiz” çıktısını üretir. Buradaki amaç, ilgisiz belirlenen belgelere gereksiz yere benzerlik skoru hesaplatmamak ve sistemi yavaşlatmamaktır. İlgili sonucu dönen tüm

belgeler benzerlik skolarması yapıldıktan sonra, kullanıcıya sonuçlar ve belge detayları liste olarak sunulur.

Kurum üyesi, bu listeden istediği belgenin; “kayitEdenPersonelNo”, “tasnifKodu”, “belgeYeniDosyaNo”, “belgeYeniKlasorKodNo”, “orjinalKayitNo”, “belgeKlasorunAdi”, “tarihAraligi”, “dijitalizeTarihi”, “gomlekSiraNo”, “mikroFormNo”, “url”, “vakifaGelisTarihi”, “yayinlanmaTarihi”, “b_keywords” vb. diğer tüm bağlı verisine ulaşabilecektir. Ayrıca,

- Belgenin geldiği kurum,
- Belgeyi düzenleyen kurum,
- Belge tarih aralığı,
- Belge tarihi dönemi,
- Belge dili,
- Belge kategorisi,
- Belge türü,
- Belge alt türü,
- Belgenin yazarı, vb. diğer özellikler üzerinden de belge araması yapılabilir.

Örneğin, “Belgeler içinde antlaşma kelimesi içeren ve Osmanlıca dilinde olan ve Vakıflar Genel Müdürlüğü Merkez Kütüphanesinden gelen vb.” bir sorgu düşünülebiliriz. Bu durumda sistem, “antlaşma” kelimesi içeren belgeleri tarama ile başlayacaktır. Eğer hiçbir belge “antlaşma” kelimesini içermiyorsa, bu durumda, “antlaşma” kelimesinin türemiş kelimeleri ve aynı zamanda eş anlamlısı kelimeleri (aşağıdaki kelimeler- bkz. “antlaşma” kelimesi için türemiş kelimeleri ve aynı zamanda eş anlamlısı kelimeler⁴), tüm belgelerin ontoloji dosyalarında ilgili etikete bakarak eşleşenlerin belge numaralarını saklayacaktır. Aramada AYSO içindeki tüm SWRL kuralları çalıştırılmış ve “ilgili” dönen belge numaraları bulunmuştur. Fakat henüz bu adımda arama tamamlanmamıştır. Yukarıda bahsi geçen diğer özelliklere göre ikinci bir arama başlatılacaktır.

```

—   turemis_kelime "muaheze"^^xsd:string
—   turemis_kelime "muahezeler"^^xsd:string
—   turemis_kelime "antlaşmalar"^^xsd:string
—   esanlamlisi_kelime "antlaşma"^^xsd:string
—   esanlamlisi_kelime "muahede"^^xsd:string
—   turemis_kelime "muahedenname"^^xsd:string
—   turemis_kelime "muahed"^^xsd:string
—   esanlamlisi_kelime "muahedeler"^^xsd:string
—   turemis_kelime "muahede"^^xsd:string
—   turemis_kelime "muahedei ticari"^^xsd:string
—   esanlamlisi_kelime "muahedat"^^xsd:string
—   turemis_kelime "muahat"^^xsd:string
—   turemis_kelime "muahezenname"^^xsd:string
—   turemis_kelime "muahedei ittifakiyye"^^xsd:string
—   esanlamlisi_kelime "anlaşma"^^xsd:string
—   turemis_kelime "muahhar"^^xsd:string
—   turemis_kelime "anlaşmalar"^^xsd:string

```

Bu ikinci aramada, AYSO içindeki SWRL’ler kullanılmamıştır. Bunun yerine, iki kelime arasındaki benzerlik için bir skor hesaplayan Jaro-Winkler (Winkler, 1999; Winkler & Thibaudau, 1991) yaklaşımı kullanılmıştır. Örneğimize göre, SWRL kuralları sayesinde “ilgili” sonuç döndüren birinci aramadan dönen belgelerin, ontoloji dosyalarına girilip, sadece Osmanlıca dilinde olan ve Vakıflar Genel Müdürlüğü Merkez Kütüphanesi’nden gelen belgeler ayrıştırılacaktır.

Aşağıda, sistemin AYSO dosyasında bulunan 7 adet SWRL kurallarının çalıştırılmasını gösteren bir örnek verilmiştir (Şekil 5). Örnekte kullanıcı “antlaşma” kelimesini belge sorgulama ara yüzüne

⁴ <http://www.nedirnedemek.com/muahede-nedir-muahede-ne-demek>

girmiştir ve aramayı başlatarak bu ontoloji içindeki yapay zekâ SWRL kural setini, tüm belge tanımlarında “b_keywords” nesne türü özellik bağlantılarında aratmıştır. Özellik yapıları, Belge → “b_keywords” → “Terim” mantığında, yani Özne-Nesne-Yüklem üçlemesiyle kurulmuş, ontolojide bağlı olan tüm bu bağlar araştırılmaktadır. Örneğimize göre, herhangi bir belge içinde “b_keywords” bağlantısında; antlaşma kelimesi aranır veya bulunmazsa, eş anlamlısı aranır ve yine bulunmazsa, türemiş kelimeleri taranır. Tüm bu yapısal kelime bağlantıları ontoloji içinde bulunan “osmanlica_sozluk” sınıfı içinde önceden tanımlanmış olmalıdır.

Örnekle açıklamak gerekirse; “antlaşma” kelimesi, Osmanlıca sözlük içinde “o2” bağlantısına bağlanmıştır. “o2” kavramının birçok türemiş ve eş anlamlı kelimesi sözlükte tanımlanmıştır. “o2 → esanlamlisi_kelime → muahede” gibi (Şekil 5’de o2 kelimesinin tüm türemiş ve eş anlamlı kelimeler, kırmızı dikkörtgen içinde ontolojideki hali verilmiştir. Şimdilik sadece, türemiş kelimeler ve eş anlamlı kelimeler sözlüğe eklenmiştir. Ayrıca, benzer bağlantılara ontoloji üzerinde yer verilebilir (örneğin; tümleşik kelimeler). Bu çalışma kapsamında, ontolojinin büyümesi zamanla olabileceğinden, sadece türemiş kelimeler ve eş anlamlı kelimelere yer verilmiştir.

The screenshot displays the AYSO ontology editor interface. The top navigation bar includes tabs for 'Active Ontology', 'Entities', 'Classes', 'Object Properties', 'Data Properties', 'Individuals by class', 'OntoGraf', 'OWL/XML', and 'Instances: o2'. The left sidebar shows the 'Class hierarchy: osmanlica_sozluk' with a tree structure including 'owl:Thing', 'arama_sonuc', 'kurum', 'lokasyon', 'osmanlica_sozluk', 'user_keywords', 'vgm_arsiv', 'belge_yonetim', 'klasor', 'dosya', 'belge', 'vakif_defter_turleri', and 'yayin_bilgisi'. The main area is divided into three sections: 'Instances: o2' showing 'o1', 'o2', and 'o3'; 'Rules' containing several SWRL rules; and 'Property assertions: o2' which is highlighted with a red box. This section lists various data property assertions for 'o2', such as 'turemis_kelime', 'esanlamlisi_kelime', and 'muahede', each with a corresponding string value in XML format.

Şekil 5. AYSO Üzerinde Osmanlıca Sözlükteki Tanımlı Kelimeler İçin Türemiş ve Eş Anlamlı Bağlar Kurulması

T.C. BAŞBAKANLIK VAKIFLAR GENEL MÜDÜRLÜĞÜ

KAYIT YAPAN PERSONEL NO: VP00135

TANIMLANACAK BELGE NO: 8

BELGENİN TARİHİ: Belirsiz

DİJİTALLEŞTİRME TARİHİ: Belirsiz

VAKİFA GELİŞ TARİHİ: Belirsiz

BELGENİN GELDİĞİ KURUM: Belirsiz

BELGEYİ DÜZENLEYEN KURUM: Belirsiz

BELGEYİ DÜZENLEYEN KİŞİ: Belirsiz

BELGE TARİH ARALIĞI: XXXX-XXXX

TARİHİ DÖNEMİ: Belirsiz

BELGE DİL: Belirsiz

BELGE KATEGORİSİ: Belirsiz

BELGE TÜRÜ: Belirsiz

BELGE ALT TÜRÜ: Belirsiz

BELGE DEFTER TÜRÜ: Belirsiz

BELGE ÜLKE: Belirsiz

BELGE BÖLGE: Belirsiz

BELGE ŞEHİR: Belirsiz

BELGE İLÇE: Belirsiz

BELGENİN YAZARI: Belirsiz

BELGE EDITÖRÜ: Belirsiz

BELGENİN YAYINCISI: Belirsiz

SAYFA ADETİ: 0

BELGE ÖN KAPAK: Belirsiz

BELGE ARKA KAPAK: Belirsiz

BELGE İLK SAYFA: Belirsiz

BELGE ERİŞİM HAKKI: Belirsiz

BELGE ARŞİV ŞEKLİ: Belirsiz

BELGE ORTAMI: Belirsiz

BELGE ANAHTAR KELİMELER

BELGE ÖN RESİM

BELGE KLASÖR KOD NO (SıraNo): 0

BELGE KLASÖR YENİ NO: K

BELGE KLASÖR ADI: Belirsiz

BELGE ORJİNAL NO: Belirsiz

BELGE LİNK URL: Belirsiz

DOSYA TASNİF KODU: VGM

BELGE DOSYA NO: D

DOSYA GÖMLEK SIRA NO: Belirsiz

DOSYADAKİ BELGE ADETİ: 0

DOSYADAKİ VARAK ADETİ: 0

DOSYA AÇILIŞ TARİHİ: Belirsiz

DOSYA ORJİNAL NO: Belirsiz

MİKRO FROM NO: Belirsiz

DOSYA ANAHTAR KELİMELER

Şekil 6. Kullanıcı Arama İşlemi İçin Arama Motorunda Dosya Arama Butonuna Basıyor

Şekil 6'da kullanıcı dosya arama butonuna (4.ncü butona) bastığında, arama işlemi başlatılıp, imleç belge anahtar kelime kutusuna çekilir.

Şekil 7. Kullanıcı Anahtar Kelime ve Aradığı Belgenin Geldiği Kurum Bilgisine Yönelik Arama Başlatıyor

Kullanıcı burada “Antlaşma” kelimesini girmiş olsun. Ayrıca belgenin geldiği kurum bilgisine “Vakıflar Genel Müd Merkez Kütüphanesi” girmiş olsun (Şekil 7).

Bot kullanıcıdan aldığı “Antlaşma” arama kelimesini ontolojiye aktarıyor. Bunu yapmasından kasıt, sistemin SWRL kural tabanı arama kelimelerini zenginleştirmek için çalıştırmasıdır. Eğer bu kelime hiçbir belge ontolojisinde yok ise, sistem AYSO’daki Osmanlıca sözlüğündeki kelimeleri taramaktadır. Bazı hukuki belgelerde halen Osmanlıca kelimeler kullanılabildiğinden, sistem Osmanlıca sözlük içinde bu kelimenin bağlı olduğu türemiş kelimeleri veya varsa eş anlamlısı kelimeleri, kısaca; tüm bağlı kelimeleri belge ontolojilerinde “b_keywords” bağlantılarında aramaya başlar. Belgelerde, en az bir tane bile bu kelimelerden içeriyorsa, gelecek adımda Jaro-Winkler benzerlik skor puanı hesaplanmak üzere o belgenin numarası kayıt edilir. Bu noktaya kadar, tüm belgelere aynı işlem uygulanır. Bu noktada aramanın ilk kısmı tamamlanmıştır. Çünkü kullanıcı başka özellikleri de hesaba katarak arama işlemi istemiş ve aradığı belgenin geldiği kurum bilgisini de “Vakıflar Genel Müd Merkez Kütüphanesi” olarak sorguya girmiştir. Bu durumda, herhangi bir belgenin içinde bu tür kurum adları, belge tarih aralığı formatı, tarihi dönem adları, belge dili, belge kategorisi, belge yazarı vb. özellik tanımları kişiden kişiye farklı hatalar farklı yazım kısaltmaları içerebilir. Bu durum arama sonucunu etkileyebilir. Sistem bu noktada yeniden, kural çalıştırmak yerine, iki kelime veya cümle arasında benzerlik skoru hesaplayan Jaro-Winkler yaklaşımı çalıştırır. Herhangi bir belgenin herhangi bir özelliğini ontolojiden cümle halinde söküp, kullanıcının girdiği özellik değeri ile kıyaslayarak bir benzerlik skoru oluşturabilmektedir. Böylece, ulaşılan belgeler girilen tüm anahtar kelimelere ve belge yapısal özellik değerlerine bağlı olarak, toplamda benzerlik oranı en büyük olan, kullanıcının aradığına en yakın belge niteliğindedir.

Örneğimizden devam edersek; Şekil 7’de belge arama butonuna (5.nci butona) basıldığında, kullanıcıya bir mesaj kutusu gelmektedir. Şekil 8’de görüldüğü üzere, ilk değerler belge numaralarını göstermektedir. Daha sonra kullanıcının girdiği arama kelimeleri ve diğer özellik tanımlarından (ör. belgenin geldiği kurum) bilgisine yönelik arama sonuçları gösterilmektedir.

Şekil 8. İlgili Bulunan Belgeler ve Jaro-Winkler Yaklaşımı ile Benzerlik Skorlarının Hesaplanması

Şekil 9. Kullanıcı, İlgili Belgeyi Olarak B0000008 Nolu Seçmiş ve Forma Yükleme İçin OK Butonuna Basmıştır

Örneğin, kullanıcı, aradığı belgeye ait, geldiği kurum bilgisini “Vakıflar Genel Müd Merkez Kütüphanesi” diye girmiştir. Yine, ontolojide tanımlı olan, B0000010 nolu belgenin geldiği kurum bilgisini “Vakıflar Genel Mudurlugu Merkez Kutuphanesi” olarak tanımlıdır. Bu iki kurum tanımları arasında benzerlik skoru 0.906 olarak hesaplanmış ve Şekil 9’da gelen mesajdaki ilk satırda gösterilmiştir. Aynı durum, B0000008 nolu belgede de görülmektedir (Şekil 10).

Şekil 10. Kullanıcıya, B0000008 Nolu Belgeye Ait Detaylı Sonuçlar Raporlanmıştır

Yaklaşık 11 adet belge “Antlaşma” kelimesini içerdiğinden “ilgili” bulunmuş (Şekil 8) fakat geldiği kurum bilgisinde farklı eşleşme skorları göstermiştir. Bu noktada, en çok benzeyenlerden birinin kullanıcı tarafından seçilmesi beklenmektedir.

The screenshot shows the VAKIFLAR GENEL MÜDÜRLÜĞÜ web interface. The header includes the logo and name of the institution. The main content area is divided into several sections:

- KAYIT YAPAN PERSONEL NO:** vp00135
- TANIMLANACAK BELGE NO:** B0000008
- BELGENİN TARİHİ:** 01/01/1965
- DİJİTALLEŞTİRME TARİHİ:** 08/07/2015
- VAKIFA GELİŞ TARİHİ:** 11/07/2015
- BELGENİN GELDİĞİ KURUM:** _mudurlugu_merkez_kutuphanesi
- BELGEYİ DÜZENLEYEN KURUM:** turk_tarih_kurumu
- BELGEYİ DÜZENLEYEN KİŞİ:** ismail_hakki_uzuncarsili
- BELGE TARİH ARALIĞI:** XXXX-XXXX
- TARİHİ DÖNEMİ:** tbmm_tc
- BELGE DİL:** turkce
- BELGE KATEGORİSİ:** idari_mali
- BELGE TÜRÜ:** kitap
- BELGE ALT TÜRÜ:** diger_kitaplar
- BELGE DEFTER TÜRÜ:** Belirsiz
- BELGE ÜLKE:** tr_turkiye
- BELGE BÖLGE:** belirsiz
- BELGE ŞEHİR:** ankara
- BELGE İLÇE:** belirsiz
- BELGENİN YAZARI:** ismail_soysal
- BELGENİN YAYINCISI:** turk_tarih_kurumu_yayinlari
- SAYFA ADETİ:** 471
- BELGE ÖN KAPAK:** applebcinstaller.log
- BELGE ARKA KAPAK:** sehir.txt
- BELGE İLK SAYFA:** belirsiz
- BELGE ERIŞİM HAKKI:** kurum_uyesine_acik
- BELGE ARŞİV ŞEKLİ:** dijital_resim
- BELGE ORTAMI:** dijital

BELGE ANAHTAR KELİMELER: muahezename, siyasi antlaşmalar, türkiye dış ilişkiler

BELGE ÖN RESİM: (Image of a document with a blue background)

BELGE KLASÖR KOD NO (SıraNo): 0003529

BELGE KLASÖR YENİ NO: k327.56/soy

BELGE KLASÖR ADI: türkiye dış ilişkiler

BELGE ORJİNAL NO: 956.101/uzu

BELGE LİNK URL: .vgm.gov.tr/yordambt/yordam.php?dilsecim=0

DOSYA TASNİF KODU: vgm

BELGE DOSYA NO: d327.56/soy

DOSYA GÖMLEK SIRA NO: 0002758

DOSYADAKİ BELGE ADETİ: 62

DOSYADAKİ VARAK ADETİ: 56

DOSYA AÇILIŞ TARİHİ: 11/03/2014

DOSYA ORJİNAL NO: 0003529

MİKRO FROM NO: iv. cilt, 1. kısım

DOSYA ANAHTAR KELİMELER: siyasi antlaşmalar

Şekil 11. Kullanıcıya Seçtiği İlgili Belgenin Ontoloji Bilgi Tabanındaki Tüm Bilgileri Sunulmuştur

Şekil 9’da gösterildiği üzere, B0000008 nolu belge seçilmiş ve Şekil 11’deki gibi ara yüze yüklenmiş olsun. Şekil 11’de görüldüğü üzere, B0000008 nolu belgenin arama kelimeleri, “Muahezename, Siyasi antlaşmalar, Türkiye dış ilişkiler” olarak vakıf personeli tarafından tanımlanmıştır. Fakat, kullanıcı sisteme sadece “Antlaşma” kelimesini ve ayrıca geldiği kurum bilgisine de “Vakıflar Genel Müd Merkez Kütüphanesi” kelimelerini girmiştir. Sistemin SWRL kural tabanında, “Antlaşma” kelimesi ile “Muahezename” kelimesi arasındaki “turemis_kelime” nesne özelliği bağlantısından dolayı, sonuç “ilgili” olarak bulunmuştur. Aynı zamanda, Şekil 11’de görüldüğü üzere, ontolojide tanımlı olan, B0000008 nolu belgenin geldiği kurum bilgisini “Vakıflar Genel Mudurlugu Merkez Kutuphanesi” olarak girilmesine rağmen, arama motorumuz, kullanıcının girdiği kurum bilgisi olan “Vakıflar Genel Müd Merkez Kütüphanesi” terimle kıyasladığında büyük oranda benzerlik bulduğundan bu belgeyi 5.nci sırada kullanıcıya sunmuştur.

Yukardaki benzer örnekler test çalışmalarında tekrar edilerek, sorgu modülüne; “antlaşma”, “antlaşmalar”, “anlaşma”, “anlaşmalar”, “muahede” vb. diğer kelimeler de ayrıca girilmiş ve aynı sonuçları elde edilmiştir. Fakat sisteme, “anlaşmazlıklar” kelimesi girildiğinde, yukarıda “ilgili” olarak elde edilen belgelere rastlanmamıştır.

5. Arşiv Hizmetlerinde Kanunlara Göre Kişisel/ Ulusal Verilere Erişimin Önemi

Bu araştırmanın deneysel çalışmalarında, devlet vakıfları arşivi ve milli arşivlerde kataloglanmış belgelerde eski lügatıda dâhil ederek ve hukuki terimler kullanılarak bot ile çeşitli belge arama çalışmaları yapılmıştır. Yapılan belge arama işlemlerinde hem anlamsal hem de söz dizimsel açıdan kullanıcı aramaları irdelenmiştir. Bilindiği üzere, devlet vakıfları arşivi ve milli arşivlerin dijital kataloglarında hukuka ilişkin çok sayıda tarihsel/bilimsel, eski ve/veya yeni çok sayıda belgeler yer almaktadır. Bu niteliğe haiz, çok eski yazıtlar, parşömene, kâğıt dışı malzeme üzerine hatta taş, deri, vb. kâğıt belge olmayan cisimler üzerine yazılmış çeşitli belgeler de olabilir. Günümüz dijital katalogları üzerinde yapılan, hukuki nitelikteki belge aramalarında, dönen belge sayısının oldukça fazla olması nedeniyle, kullanıcıları ne yazık ki sistemden dönen sonuç belgeler üzerinde ayıklama yapmaya zorlamakta, bu da ciddi zaman kaybına neden olabilmektedir. Çıkan sonuçların fazla sayıda olması araştırmacının amacına ulaşması bakımından da engelleyici olabilmektedir. Bu nedenle, anlamsal irdeleme ile arama alanının daraltılması sağlanmalıdır. Diğer bir sorun da, dönen sonuç belgelerdeki hukuk disiplinin kullandığı eski lügat terimlerini, günümüz sistemleri tarafından anlamsal bakımdan irdelenememesidir.

Hukuk alanında birbiri yerine kullanılan kavramsal bağlı kelimeler (ör. türemiş kelime, eş anlamlı, tekil-çoğul ilişkisi vb. bağlı kelimeler), araştırmacının tek bir kelime üzerinden değil, ayrı ayrı tüm bağlı kelimeleri aramaya katmasını zorunlu kılmaktadır. Bu durum, yukarıda bahsi geçen arama sonuçlarına ilişkin kullanıcı zorluklarının tekrar tekrar yaşanmasına sebep olmaktadır.

Hukuk dili gündelik dilde kullanılmayan sözcükleri barındırdığından ve kendine özgü bir dil yapısı olduğundan dolayı, çoğu zaman hedef veriye ait belgelere erişim için dile bağımlı bu terimleri ve dil yapısını bilmek gerekir. Bazen de hedef veriyi içeren ayrıca sorgudaki zaman kısıdına uyan hedef belgelere erişmek için farklı yapıda sorgulara ihtiyaç duyulabilir. Bu durumda araştırmacının, içerik ve/veya yapısal olmak üzere iki farklı boyutta arama yapması gerekebilir. Örneğin, içerik bakımından aramaya ihtiyaç duyulursa; “temel hak ve özgürlüklere ilişkin uluslararası antlaşmaların” taranması örnek verilebilir. Diğerisi ise, yapısal bakımdan aramaya ihtiyaç duyulursa (ör. zamana bağlı): “son 10 yılda iç hukuka geçirilmiş temel hak ve özgürlüklere ilişkin uluslararası antlaşmaların” taranması örnek verilebilir. Zira temel hak ve özgürlüklere ilişkin uluslararası anlaşmalar üzerinden yapılacak içerik bakımından aramalarda, güncelliğini yitirmiş, yürürlükten kalkmış anlaşmalara da sorgu sonucunda ulaşılabilir. Bu gibi durumlarda, araştırmacının arama işleminde başa dönmesine ve zaman kaybına neden olabilir. Ancak, eğer ilgili belge, ilk kaydedildiğinde dönemsel etiket bilgisi ile girilmişse, o belgenin yürürlükte olup/olmadığı ayırımı da yapılabilir ve yapısal arama niteliğinde bu zorluk da aşılabılır. Önerilen bot sayesinde hem kavramsal hem de yapısal arama yapmak mümkündür. Bot’un deneysel çalışmalarında, hukuki nitelikte sahip ve önemli bir kelime olan “anlaşma” terimi seçilmiştir. Bu hukuki terimin eş anlamlısı olarak hem Türkçe dilinde geçen yeni kullanımları, hem de eski hukuk lügatında olabilecek ilgili terimler de aramaya dahil edilmiştir. Dönen belge sonuçları sayısal bakımdan incelenmiştir. Bu deneysel çalışmalar ve değerlendirme sonuçları gelecek bölümde sunulmuştur.

6. Kıyaslama —Diğer Dijital Arşiv Kataloğu Üzerinde Yapılan Aramalar ve Değerlendirme Sonuçları

Bu bölümde, bazı eski/yeni hukuk lügatında olan terimler üzerinden belge sorgulamaları yapılarak sistem botunun değerlendirilmesi anlatılmıştır. Geliştirilen botun, Vakıflar Genel Müdürlüğü Merkez Kütüphanesi⁵ ile Türkiye Cumhuriyeti Cumhurbaşkanlığı Devlet Arşivleri Başkanlığı’na⁶ ait dijital arşiv kataloğunun arama motoruyla, önceki bölümde sunulan benzer sorgulara verdiği yanıtlar ve sonuç belgeler sayısal bakımdan irdelenmiştir. Vakıflar Genel Müdürlüğü Merkez Kütüphanesinde

⁵Vakıflar Genel Müdürlüğü Merkez Kütüphanesinin Arama Motoru, toplam belge sayısı: 34396, <https://kutuphane.vgm.gov.tr/yordambt/yordam.php>

⁶Türkiye Cumhuriyeti Cumhurbaşkanlığı Devlet Arşivleri Başkanlığı Arama Motoru, toplam belge sayısı belirtilmemiş, <https://katalog.devletarsivleri.gov.tr/Sayfalar/Arama/Arama.aspx>

kullanıcılar için bir ara yüz bulunmakta olup ve kütüphane bünyesinde mevcut kaynakların sorgulanması bu ara yüz üzerinden mümkün olmuştur.

Vakıflar Genel Müdürlüğü Merkez Kütüphanesi arama motorunda; sadece “Bölüm (Seçilen Kütüphane)”, “Tür”, “Alt Tür”, “Şekil”, “Ortam”, “Dil”, “Yeniler”, “Görsel”, “Kurum Yayını”, ve “Aranan Anahtar Kelimeler” vb. bilgilerle arama yapılabilir. Şekil 12’de görüldüğü üzere, Vakıflar Genel Müdürlüğü Merkez Kütüphanesi’nin arama motoru üzerinde yaptığımız bir sorgu sonucu gösterilmiştir. Sorgu yapılırken Kütüphane sekmesinde “Vakıflar Genel Müdürlüğü Merkez Kütüphanesi” seçilmiş ve arama kutusuna “anlaşma” kelimesi girilerek sorgu başlatılmıştır. Şekil 12’den anlaşıldığı gibi “anlaşma” kelimesini içeren belgeler liste şeklinde gösterilmiştir. Toplam 9 belge bulunmuştur.

Vakıflar Genel Müdürlüğü Kütüphaneleri

Vakıflar Genel Müd

Bölüm

Tür

Alt Tür

Şekil

Ortam

Dil

Yeniler

Görsel

Kurum Yayını

Q anlaşma

Tümü

+

BUL

Temizle

1



Türkiye İslam ülkeleri; anlaşmalar ve mevzuat /

Vakıflar Genel Müdürlüğü Merkez Kütüphanesi
0003807
341 YAZ 1982 k.1
Rafta

2



Türkiye Cumhuriyeti'nin taraf olduğu ikili ve çok taraflı kültür anlaşmaları ve kültürel değişim programları.

Vakıflar Genel Müdürlüğü Merkez Kütüphanesi
0004794
301.202 TÜR 1984 k.1
Rafta

3



Osmanlı tarihi :Karlofça Anlaşmasından XVIII. Yüzyılın Sonlarına Kadar /

Vakıflar Genel Müdürlüğü Merkez Kütüphanesi
0001766
956.07 UZU 1956 k.1/c.4/1
Rafta

4



Türkiye İslam ülkeleri; anlaşmalar ve mevzuat /

Vakıflar Genel Müdürlüğü Merkez Kütüphanesi
0002931
341 YAZ 1982 k.1
Rafta

5



Şark meselesi / , Raif Karadağ

Vakıflar Genel Müdürlüğü Merkez Kütüphanesi
0001331
322.42 KAR 1971 k.1
Rafta

6



Türk İnkılabı Tarihi :Trablusgarp ve Balkan Savaşları Osmanlı Asyasının Paylaşılması için anlaşmalar /

Vakıflar Genel Müdürlüğü Merkez Kütüphanesi
0006940
956 BAY 1991 k.1/c.2/2
Rafta

7



Türk İnkılabı Tarihi :Trablusgarp ve Balkan Savaşları Osmanlı Asyasının Paylaşılması için Anlaşmalar /

Vakıflar Genel Müdürlüğü Merkez Kütüphanesi
0006943
956 BAY 1991 k.1/c.2/3
Rafta

8



Türk İnkılabı Tarihi : Cilt II Trablusgarp ve Balkan Savaşları Osmanlı Asyasının Paylaşılması için Anlaşmalar Kısım:1 1911 başından Balkan Savaşına kadar /

Vakıflar Genel Müdürlüğü Merkez Kütüphanesi
0006946
956 BAY 1991 k.1/c.2/1
Rafta

9



Osmanlı tarihi :Karlofça Anlaşmasından XVIII. Yüzyılın Sonlarına Kadar /

Vakıflar Genel Müdürlüğü Merkez Kütüphanesi
0008091
956.07 UZU 1995 k.2/c.4/1
Rafta

Şekil 12. Merkez Kütüphanesinde “Anlaşma” Kelimesinin Sorgulama Sonuçları

Şekil 13’ de gösterildiği gibi Vakıflar Genel Müdürlüğü Merkez Kütüphanesi’nin arama motorunda yaptığımız diğer bir sorgu sonucu gösterilmiştir. Bu sefer de sorgu yapılırken “Kütüphane” sekmesinde yine “Vakıflar Genel Müdürlüğü Merkez Kütüphanesi” seçilmiş ve arama kutusuna bu sefer “anlaşmalar” kelimesi girilmiştir. Bilindiği gibi “anlaşma” kelimesi “anlaşmalar” kelimesinin tekil halidir. Fakat sorgu sonucunu incelediğimizde, toplam 6 (altı) belge bulunmuş; sonuç sayısı düşmüştür.

Vakıflar Genel Müdürlüğü Kütüphaneleri

Vakıflar Genel Müd Bölüm Tür Alt Tür Şekil Ortam Dil Yeniler Görsel Kurum Yayını

anlaşmalar Tümü + BUL Temizle

1		Türkiye İslam ülkeleri; anlaşmalar ve mevzuat /	Vakıflar Genel Müdürlüğü Merkez Kütüphanesi 0003807 341 YAZ 1982 k.1 Rafta
2		Türkiye Cumhuriyeti'nin taraf olduğu ikili ve çok taraflı kültür anlaşmaları ve kültürel değişim programları.	Vakıflar Genel Müdürlüğü Merkez Kütüphanesi 0004794 301.202 TÜR 1984 k.1 Rafta
3		Türkiye İslam ülkeleri; anlaşmalar ve mevzuat /	Vakıflar Genel Müdürlüğü Merkez Kütüphanesi 0002931 341 YAZ 1982 k.1 Rafta
4		Türk İnkılabı Tarihi :Trablusgarp ve Balkan Savaşları Osmanlı Asyasının Paylaşılması için anlaşmalar /	Vakıflar Genel Müdürlüğü Merkez Kütüphanesi 0006940 956 BAY 1991 k.1/c.2/2 Rafta
5		Türk İnkılabı Tarihi :Trablusgarp ve Balkan Savaşları Osmanlı Asyasının Paylaşılması için Anlaşmalar /	Vakıflar Genel Müdürlüğü Merkez Kütüphanesi 0006943 956 BAY 1991 k.1/c.2/2 Rafta Detay için tıklayın
6		Türk İnkılabı Tarihi : Cilt II Trablusgarp ve Balkan Savaşları Osmanlı Asyasının Paylaşılması için Anlaşmalar Kısım:1 1911 başından Balkan Savaşına kadar /	Vakıflar Genel Müdürlüğü Merkez Kütüphanesi 0006946 956 BAY 1991 k.1/c.2/1 Rafta

Şekil 13. Merkez Kütüphanesinde “Anlaşmalar” Kelimesinin Sorgulama Sonuçları

Aynı zamanda taranan bu kelimelerin, anlamsal bakımdan bağlı olduğu diğer kelimeleri de kaçırmamak ve sistemin doğru değerlendirilmesi için sorgumuza bu sefer; “antlaşma” ve “antlaşmalar” kelimeleri ile devam edilmiştir. Çünkü, “anlaşma” kelimesi; uluslararası politikada tarafların belli konularda benzer görüşleri paylaştıkları, birlikte hareket edecekleri anlamındadır. Gizli veya açık olabilir. Davranışsal birliği ifade edebileceği gibi yazılı şekilde de olabilir.

Diğer yandan, “antlaşma” kelimesi ise; uluslararası yapılan görüşmeler sonucunda kaleme alınan bir metnin imzalanmasıdır. Antlaşmalar, bu metni imzalayan devletlerin ilgili kurullarından kabul gördükten sonra yürürlüğe girer. İki devlet arasında olabileceği gibi çok taraflı da olabilir. Sonuç olarak, bu kelimeler birbirlerine anlamsal ve fonksiyonel bakımdan bağlı olan kelimelerdir. Şekil 14’de gösterildiği gibi yeni bir sorgu daha yapılmış, bu sefer “antlaşma” kelimesi sorgulanmış ve bu kelimeyi içeren ilgili belgeler toplam 14 (on dört) adet bulunmuştur.

Vakıflar Genel Müdürlüğü Kütüphaneleri

Vakıflar Genel Müd Bölüm Tür Alt Tür Şekil Ortam Dil Yeniler Görsel Kurum Yayını

antlaşma Tümü + BUL Temizle

0006197	0006198	0006199	0006200	0006201	0002758	0003529	0000643
Muahadat mecmuası 3	Muahadat mecmuası 5	Muahadat mecmuası 4	Muahadat mecmuası 1	Muahadat mecmuası 2	Osmanlı tarihi :Karlofça	Türkiyenin dış münasebetleriyle	Türkiye ve Nato
0004799	0001646	0006474	0006491	0006589	0006851		
Tarihçeleri ve açıklamaları ile	Mısakı milli ve Lozan sulh	Tarihçeleri ve açıklamaları ile	1774 Küçük Kaynarca	Türkiye ve mavera-ı Kafkasya	Türkiye ve mavera-ı Kafkasya		

Türkiye ve mavera-ı Kafkasya ilişkileri içerisinde Trabzon ve Batum konferansları ve antlaşmaları (1917-1918) Şahin, Enis ; Enis Şahin

ISBN-İSSN: 9751615437

Yayın Bilgisi : Ankara: Türk Tarih Kurumu Yayınları , 2002

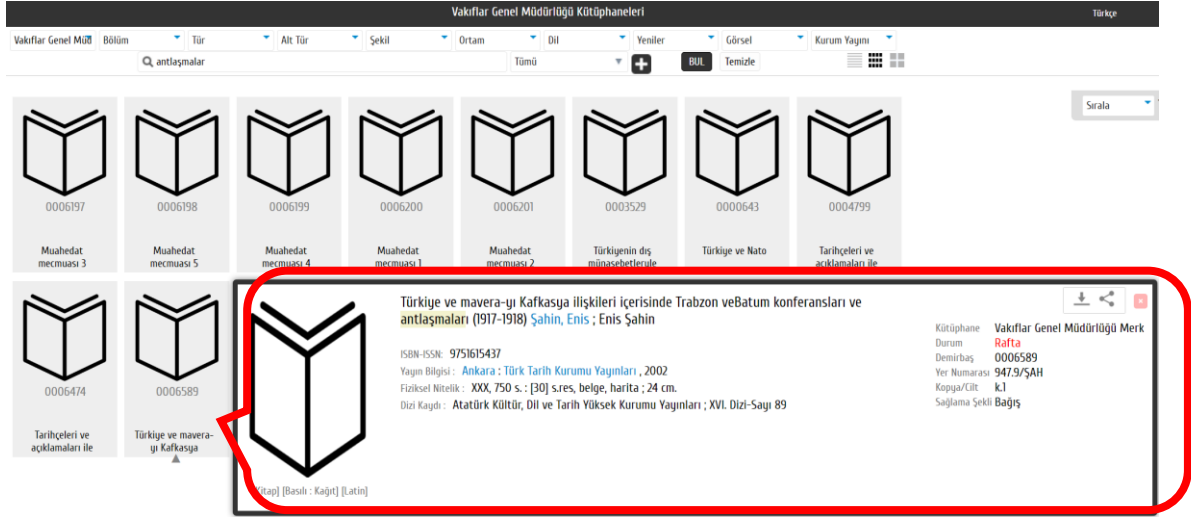
Fiziksel Nitelik : XXX, 750 s. : [30] s.res, belge, harita ; 24 cm.

Dizi Kaydı : Atatürk Kültür, Dil ve Tarih Yüksek Kurumu Yayınları ; XVI. Dizi-Sayı 89

Kütüphane Vakıflar Genel Müdürlüğü Merk
Durum Rafta
Demirbaş 0006589
Yer Numarası 947.9/ŞAH
Kopga/Cilt k.1
Sağlama Şekli Bağış

Şekil 14. Merkez Kütüphanesinde “Antlaşma” Kelimesinin Sorgulama Sonuçları

Son olarak, Şekil 15’de görüldüğü gibi sorgu yapılırken “Kütüphane” sekmesinde yine “Vakıflar Genel Müdürlüğü Merkez Kütüphanesi” seçilmiş ve bu sefer arama kutusuna “antlaşmalar” kelimesi girilmiştir. Bilindiği üzere, “antlaşma” kelimesi “antlaşmalar” kelimesinin tekil halidir. Her sorguda farklı sonuçlar ve belgeler çıkmasının sebebi; bu arama motorunun söz dizimsel arama yapmasından kaynaklanmaktadır. Kelimelerin; anlamlarına, türemiş kelime özelliğine, eş anlamlı özelliğine, tekil-çoğul gibi anlam ilişkilerine bakmadan arama yapmaktadır.



Şekil 15. Merkez Kütüphanesinde “Antlaşmalar” Kelimesinin Sorgulama Sonuçları

Aranan Terimler	T.C. Cumhurbaşkanlığı Devlet Arşivleri Başkanlığı Arama Motoru Sonuçları	Vakıflar Genel Müdürlüğü Merkez Kütüphanesi Arama Motoru Sonuçları	Sistem Bot 'unun Sonuçları
Anlaşma	15221	15	11
Anlaşmalar	15221	8	11
Antlaşma	3497	16	11
Antlaşmalar	3497	11	11
Sözleşme	6148	5	11
Sözleşmeler	6148	0	11
Akit	131	0	11
Akitler	131	0	11
Mukavele	3647	5	11
Mukaveleler	3772	0	11
Kontrat	1651	0	11
Kontratlar	1651	0	11
Muaheze	80	0	11
Muahezeler	26	0	11
Muahed	4	11	11
Muahede	2145	5	11
Muahedename	258	2	11
Muahede-i Ticari	11	0	11
Muahedeler	2931	2	11
Muahedat	564	8	11
Muahat	1	0	11
Muahezename	0	0	11
Muahede-i İttifakiyye	1	0	11
Muahhar	48	0	11

Çizelge 3. T.C. Cumhurbaşkanlığı Devlet Arşivleri Başkanlığı, Vakıflar Genel Müdürlüğü Merkez Kütüphanesi, Anlamsal Arama Bot 'unun Hukuki Terimler ile Sorgulanması ve İlgili Bulunan Belge Sayıları

T.C. Cumhurbaşkanlığı Devlet Arşivleri Başkanlığı'na ait belge sayısı milyonlar iken, Vakıflar Genel Müdürlüğü Merkez Kütüphanesi'ne ait toplam belge sayısı 34396 adettir. Bizim sistemimizde, test çalışmaları için, yukarıdaki iki kütüphaneden rastgele seçerek sistemin yeni belge kayıt modüle bağlı

ara yüz üzerinden manuel giriş ile oluşturduğumuz toplam belge ontolojileri sayısı 80 adettir. Bu belge ontolojilerinin her birinde kayıtlı “b_keywords” nesne türü özelliğine bağlı tüm anahtar kelimeler, sistem botu tarafından kullanıcı girdi kelimesi ile eş anlamlı, türemiş veya tekil-çoğul bağıntılarıyla kıyaslanır. Tüm belge ontoloji dosyalarında “b_keywords” etiketi taranır, kavramsal boyutta eşleşenlerin belge numaraları sistem botu tarafından saklanır, böylece hangi eş anlamlı, türemiş veya tekil-çoğul bağlantılı terim girilirse girilsin hep aynı sayıda belge sonuç olarak döndürmektedir (bkz. son kolon, Çizelge 3). Sistem botu daha çok belge üzerinde denense de aynı hareketle davranacaktır. Sistemin diğer sistemlere entegrasyonu çalışmaları halen devam etmektedir. T.C. Cumhurbaşkanlığı Devlet Arşivleri Başkanlığı arama motoru sonuçlarına bakılırsa tekil-çoğul ilişkisine bağlı güncel kelimeler de aynı sonucu döndürmektedir. Bu durumda hassasiyetin iyi olduğu anlaşılmaktadır. Fakat eski lügat terimler kullanıldığında, özellikle tekil-çoğul ilişkilerinde, dönen sonuçlar çok değişkenlik göstermektedir (bkz. Çizelge 3). Bu durum da bazı belgelerin gözden kaçırıldığına işaret etmektedir. T.C. Cumhurbaşkanlığı Devlet Arşivindeki milyonlarca belge arasından dönen sonuçların, Vakıflar Genel Müdürlüğü Merkez Kütüphanesi’ne ait arama motorunun döndürdüğü sonuçlara nispeten daha iyi olduğunu belirtebiliriz. Nedeni ise, “Anlaşma, Anlaşmalar” veya “Antlaşma, Antlaşmalar” gibi tekil-çoğul ilişkisi ile arama kelimeleri girildiğinde, T.C. Cumhurbaşkanlığı Devlet Arşivi’nin arama motorunun daha kararlı bir yapıya sahip olduğu Çizelge 3’ün birinci ve ikinci kolonlarında görülebilmektedir.

Önceki bölümde de anlatıldığı gibi, anlamsal arama botu üzerinde aynı sorgu kelimesi çalıştırılmıştır. Sorguya girilen “antlaşma” arama kelimesinin; anlamlarına, türemiş kelime özelliğine, eş anlamlı özelliğine, tekil-çoğul gibi anlam ilişkilerine göre bot aramasını yürütülmüştür. Sonuçta, arama terimine göre veri setinde gerçekte ilgili olan belgeler kaçırılmamış ve dönen ilgili belge sayısı hep aynı kalmıştır. Aynı zamanda, Vakıflar Genel Müdürlüğü Merkez Kütüphanesine nazaran daha fazla ilgili belge ele geçirilirken, T.C. Cumhurbaşkanlığı Devlet Arşivleri Başkanlığı arama motoruna nazaran bot tarafından daha az sayıda ilgili belge bulunmuştur. Bunun sebebi ise, T.C. Cumhurbaşkanlığı Devlet Arşivleri Başkanlığının dijital kataloğunun ciddi olarak (milyonlarca) daha fazla belge içermesinden kaynaklanmaktadır. Ayrıca, önerilen bot’un aksine, devlet arşivleri başkanlığı kataloğundan dönen belgeler için bir skorlama işlemi yapılmamakta ve en yakın belgenin dönen sonuçlar arasında kullanıcı tarafından bulunmasını gerektirmektedir.

7. Sonuçlar

Milli arşivlerde bulunan tarihsel öneme sahip milyonlarca adet belgenin; hızlı ve güvenilir şekilde elektronik ortama aktarılması, elektronik ortamda yönetilmesi, diğer kurumlar ile paylaşılması için Dijital Arşiv Yönetim Sistemleri kullanılır. Bu sistemlerde, milyonlarca belge içinden istenilen belgenin en kısa zamanda doğru ve akıllı bir üslupla bulunabilmesi önemli bir süreçtir.

Bu tür arşiv yönetim sistemlerin modellenmesinde, Web 2.0 teknolojisinin kısıtları nedeniyle, genelde söz dizimsel arama yapılabilir. Anlamsal bazda içerik tabanlı arama yapabilmek için, bilgi tabanlarına ihtiyaç duyulmaktadır. İçeriklerin kavramlaştırılması ve kavramlar arası anlamsal bağlarının/ özelliklerin önceden tanımlanması ancak Ontoloji Bilgi Tabanları ile yapılabilir. Bu nedenle, bu çalışmada; OBT’nin ana teknolojisi olan Anlamsal Web teknolojisi kullanılmış, dijital ortama taşınmış tarihi değeri olan belge/belgelerin veya ilgili verilerin bir anlamsal arama botu üzerinden, kullanıcılar tarafından kolaylıkla aranmasına olanak sağlayan bir sistem ortaya konulmuştur. Sistemin ana teknik bölümleri şunlardır: (1) Arşiv Yönetim Sistemi Ontolojisi (AYSO), (2) Anlamsal Web Kuralı Bilgi tabanı ve (3) Anlamsal Arama Motoru (Bot). Bu makalede, önerilen sistem sayesinde, her bir arşiv belgesine ait içeriğin anlamsal etiketlenmeleri sayesinde, OWL formatında metin dosyası olarak saklanabilmesi ve anlamsal arama bot’ları ile bu belge ontolojileri üzerinde anlamsal aramalar yapılabilceği gösterilmiştir.

Ayrıca, sistemin en önemli parçası olan Anlamsal Arama Bot’unun değerlendirilmesinde, eski lügat dahil edilerek hukuki terimler üzerinden, çeşitli belge sorgulamaları yapılmıştır. Bot, kullanıcının girdiği arama kelimesinin ontolojideki kavramsal yapılarına; türemiş kelime özelliğine, eş anlamlı özelliğine, tekil-çoğul gibi anlam ilişkilerine göre arama setini otomatik olarak genişletir. Kıyaslama çalışmalarında, söz dizimsel arama yapan başka iki dijital katalog üzerinde aynı terimler belge aramada

tekrarlanmıştır, böylece; dönen belge sayıları kıyaslanmıştır. Elde edilen sonuçlar umut verici olup, bot'un mevcut kataloglar üzerinde uygulanabilirliğini, etkililiğini ve verimliliğini göstermektedir.

Etik Standartlar ile Uyumluluk

Çıkar Çatışması: Yazar herhangi bir çıkar çatışmasının olmadığını beyan eder.

Etik Kurul İzni: Bu çalışma için etik kurul iznine gerek yoktur.

Yazar Katkı Beyanı: Yazarlar makale için eşit oranda katkıda bulunduklarını beyan ederler.

Finansal Destek: Yoktur.

Kaynakça

- Ahonen, E. ve Hyvonen, E. (2009, September). Publishing Historical Texts on the Semantic Web a Case Study. In 2009 IEEE International Conference on Semantic Computing (pp. 167-173). IEEE.
- Altay, A. (2002). İngilizce ve Türkçe Hukuk Dillerinin Özellikleri: Karşılaştırmalı Bir Çalışma. *Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi*, 2(19), 13-32.
- Berners-Lee, T., Hendler, J. ve Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 34-43.
- Domingue, J., Fensel, D. ve Hendler, J. A. (Eds.). (2011). Handbook of Semantic Web Technologies. Springer Science & Business Media.
- Ertugrul, M., & Gunes, A. (2017). Arşiv Yönetim Sistemleri için Ontoloji Bilgi Tabanı Geliştirilmesi. *Aksaray University Journal of Science and Engineering*, 1(2), 85-109.
- Gruber, T. (2009). Ontology. *Encyclopedia of Database Systems*, 1, 1963-1965.
- Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F. ve Rudolph, S. (2009). OWL 2 Web Ontology Language Primer. *W3C Recommendation*, 27(1), 123.
- Horridge, M. ve Bechhofer, S. (2011). The Owl Api: A Java Api for Owl Ontologies. *Semantic Web*, 2(1), 11-21.
- Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosz, B. ve Dean, M. (2004). SWRL: A Semantic Web Rule Language Combining OWL and RuleML. *W3C Member Submission*, 21(79), 1-31.
- Hyvönen, E., Viljanen, K., Tuominen, J. ve Seppälä, K. (2008). Building a National Semantic Web Ontology and Ontology Service Infrastructure—the FinnONTO Approach. In *European Semantic Web Conference* (pp. 95-109). Springer, Berlin, Heidelberg.
- Kişisel Verilerin Korunması Kanunu. (2016, 24 Mart). Resmi Gazete (6698). Erişim adresi: <https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=6698&MevzuatTur=1&MevzuatTertip=5>
- Levene, M. (2011). An Introduction to Search Engines and Web Navigation. John Wiley & Sons.
- Madhu, G., Govardhan, D. A. ve Rajinikanth, D. T. (2011). Intelligent Semantic Web Search Engines: A Brief Survey. *arXiv preprint arXiv:1102.0831*.
- McGuinness, D. L. ve Van Harmelen, F. (2004). OWL Web Ontology Language Overview. *W3C Recommendation*, 10(10), 2004.
- Noy, N. F., Sintek, M., Decker, S., Crubézy, M., Fergerson, R. W. ve Musen, M. A. (2001). Creating Semantic Web Contents with Protege-2000. *IEEE Intelligent Systems*, 16(2), 60-71.
- O'Connor, M. J. ve Das, A. K. (October). SQWRL: a Query Language for OWL. In *OWLED*, 2009(529).
- O'Reilly, T. (2009). What is Web 2.0. O'Reilly Media, Inc.
- Preethi, N. ve Devi, T. (2012). Case and Relation (CARE) Based Page Rank Algorithm for Semantic Web Search Engines. *International Journal of Computer Science Issues (IJCSI)*, 9(3), 329.

- Protégé Ontology Editor. (1999). Stanford University School of Medicine: <http://protege.stanford.edu/>
- Rose, D. E. ve Levinson, D. (2004, May). Understanding user goals in web search. *In Proceedings of the 13th International Conference on World Wide Web* (pp. 13-19).
- The MUNIN Project-An overview, MUNIN Final Event. (2015). Dipl.-Wirtsch.-Ing. University MUNIN Project Coordinator: Hans-Christoph Burmeister, Hamburg, Germany. Erişim adresi: <http://www.unmanned-ship.org/munin/wp-content/uploads/2015/06/MUNIN-Final-Event-B-1-CML-The-MUNIN-Project-An-Overview.pdf>.
- Winkler, W. E. (1999). The State of Record Linkage and Current Research Problems. *In Statistical Research Division*, US Census Bureau.
- Winkler, W. E. ve Thibaudeau, Y. (1991). An Application of the Fellegi-Sunter Model of record Linkage to the 1990 US Decennial Census. Washington, DC: US Bureau of the Census.