



Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü
Bilgi ve Belge Yönetimi Anabilim Dalı

**İÇERİK TABANLI ATIF ANALİZİ MODELİ TASARIMI: TÜRKÇE
ATIFLAR İÇİN METİN KATEGORİZASYONUNA DAYALI BİR
UYGULAMA**

Zehra Taşkın

Doktora Tezi

Ankara, 2017

İÇERİK TABANLI ATIF ANALİZİ MODELİ TASARIMI: TÜRKÇE ATIFLAR İÇİN METİN
KATEGORİZASYONUNA DAYALI BİR UYGULAMA

Zehra Taşkın

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü

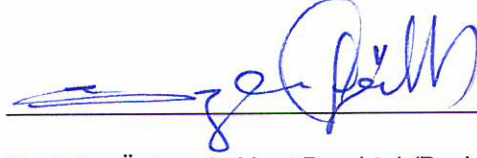
Bilgi ve Belge Yönetimi Anabilim Dalı

Doktora Tezi

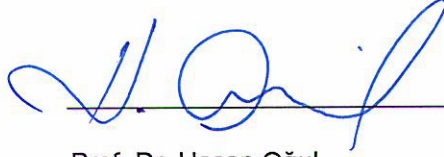
Ankara, 2017

KABUL VE ONAY

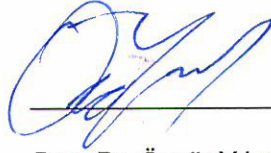
Zehra Taşkın tarafından hazırlanan "İçerik Tabanlı Atıf Analizi Modeli Tasarımı: Türkçe Atıflar İçin Metin Kategorizasyonuna Dayalı Bir Uygulama" başlıklı bu çalışma, 15.05.2017 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Doktora Tezi olarak kabul edilmiştir.



Prof. Dr. Özlem Gökkurt Demirtel (Başkan)



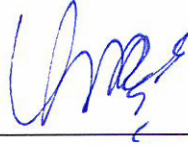
Prof. Dr. Hasan Oğul



Doç. Dr. Özgür Yılmazel



Doç. Dr. İrem Soydal



Doç. Dr. Umut Al (Danışman)

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. Sibel Bozbeyoğlu
Enstitü Müdürü

BİLDİRİM

Hazırladığım tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kâğıt ve elektronik kopyalarının Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- Tezimin/Raporumun tamamı her yerden erişime açılabilir.
- Tezim/Raporum sadece Hacettepe Üniversitesi yerleşkelerinden erişime açılabilir.
- Tezimin/Raporumun yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

15.05.2017



Zehra Taşkın

TEŞEKKÜR

Bu çalışmanın fikir olarak ortaya çıkmasından sonuçlanmasına kadar geçen sürede destek olan pek çok kişi ve kuruluşa teşekkür borçluyum. Bu tezin her aşamasında destek olan, tezcanlı bir yapıya sahip olan beni gerektiğinde frenleyip gerektiğinde ise harekete geçmem için motive eden; motivasyon kaynaklarımı çok iyi bilen sevgili danışmanım Doç. Dr. Umut Al'a sonsuz teşekkür ederim. Hem yüksek lisans hem de doktora tezimde kendisi ile çalışabildiğim için çok şanslıyım.

Öte yandan bu araştırma Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından 115K440 proje numarası ile desteklenmiştir. TÜBİTAK, Hacettepe Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi ve Proje ekibinde görev alan herkese, özellikle Proje bursiyerleri Ayşe Esra Özkan Çelik, Güleda Doğan, İpek Şencan, Elçin Keleş ve Kardelen Aktaş'a bu çalışmaya yaptıkları katkı için sonsuz teşekkürler.

Tezim süresince tez izleme komitelerimde yer alarak yapıcı eleştirileri ile tezin her zaman daha ileri bir noktaya gelmesi için çabalayan hocalarım Doç. Dr. İrem Soydal ve Prof. Dr. Hasan Oğul'a teşekkür ederim. Teze başlangıç aşamasında bu işin altından kalkabilir miyim diye kara kara düşünürken Eskişehir'deki ofisinde beni ağırlayan, veri toplama ve analiz konusunda fikir veren ve savunma sınavımda da yer alarak katkısını farklı bir boyuta taşıyan Doç. Dr. Özgür Yılmazel'e ve yine savunma sınavımda yer alarak değerli görüşlerini ileten Prof. Dr. Özlem Demirtel'e sonsuz teşekkürler.

Mutlu bir insansanız etrafınız iyi dostlarla çevrili demektir. Can dostum güzel insan Güleda Doğan'a sahip olduğu güzel kalbi ve eşsiz dostluğu için, her soruma anında cevap verdiği için ve danışmanımın bir sayfadan daha uzun teşekkür edilmez diye yaptığı uyarılar sebebi ile uzun uzadıya sayamadığım pek çok sebepten dolayı teşekkür ederim. Tezim için yaptığı onca hızlandırıcı şeye ek ve hatta ondan daha çok eşsiz dostluğu ve sohbeti için güzel dostum Ayşe Esra Özkan Çelik'e; her başım sıkıştığında yanımda olan, beni benden daha iyi tanıyan canım Şule Turan'a çok teşekkür ederim. Dostlarınız olmadan yaptığınız hiçbir şeyin anlamı yok. İyi ki varsınız. Bizi oğlan anneleri kurtaracak!

Bu tezin tamamlanması sürecinde desteğini gördüğüm kocaman aileme çok teşekkür ederim. Sabırsızlıkla bu tezin bitmesini bekleyen tüm geniş aileme bu süreçteki tüm destek ve fedakârlıklarından dolayı teşekkürler.

Son olarak hayattaki tek önceliğim çekirdek ailemin bu tezde katkısı büyük. Değişiyorsam, geliyorsam en önemlisi kocaman gülebiliyorsam hayatımdaki iki dev adamın katkısı sayesinde. Sevgili eşim Abdülkadir Taşkın ve güler yüzlü oğlum Emir Taşkın'a bu süreçte hep yanımda oldukları, hep destek verdikleri ve günlerimi güzelleştirdikleri için teşekkür ederim. Hep söylenen "... olmasaydı bu tez bitmezdi" tanımlamasındaki boşluklara sizin adınızı yazıyorum.

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kâğıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

- **Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.**

(Bu seçenekle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etseniz ve kütüphane bu talebinizi yerine getirirse bile, teziniz arama motorlarının önbelleklerinde kalmaya devam edebilecektir)

- **Tezimin/Raporumun tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.**

(Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisi alınabilir)

- **Tezimin/Raporumun..... tarihine kadar erişime açılmasını istemiyorum ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.**

- **Serbest Seçenek/Yazarın Seçimi**



15/05/2017

ETİK BEYAN

Bu alıřmadaki bütn bilgi ve belgeleri akademik kurallar erevesinde elde ettiđimi, grsel, iřitsel ve yazılı tm bilgi ve sonuları bilimsel ahlak kurallarına uygun olarak sunduđumu, kullandıđım verilerde herhangi bir tahrifat yapmadıđımı, yararlandıđım kaynaklara bilimsel normlara uygun olarak atıfta bulunduđumu, tezimin kaynak gsterilen durumlar dıřında zgn olduđunu, Do. Dr. Umut Al danıřmanlıđında tarafımdan retildiđini ve Hacettepe niversitesi Sosyal Bilimler Enstits Tez Yazım Ynergesine gre yazıldıđını beyan ederim.



Zehra Tařkın

ÖZET

TAŞKIN, Zehra. *İçerik Tabanlı Atıf Analizi Modeli Tasarımı: Türkçe Atıflar İçin Metin Kategorizasyonuna Dayalı Bir Uygulama*, Doktora Tezi, Ankara, 2017.

Araştırma/cı performanslarının ölçümünde önemli unsurlardan biri alınan atıf sayılarının değerlendirilmesidir. Aldıkları atıfların sayısına göre akademisyenler teşvik almakta, yükseltilmekte veya ödüllendirilmektedirler. Başlangıçta temel amacı birbirleri ile ilişkili yayınları saptamak olan atıfların kullanım amacı da yüklenen bu misyon ile birlikte değişmeye başlamış, bu değişim de etik olmayan bazı uygulamalar ile atıf sayılarının manipüle edilmesine kadar uzanmıştır. Gelenen noktada atıfların niceliksel olarak değerlendirilmesinin yanında içeriksel analizinin yapılması gerekliliği ortaya çıkmaya başlamıştır.

Bu çalışmada Türkçe atıflar için içerik tabanlı bir atıf analizi modeli tasarlanması hedeflenmiştir. Bu hedef doğrultusunda Türkçe kütüphanecilik ve bilgibilim literatüründe yayımlanmış 423 hakemli makale derinlemesine incelenmiş ve analizler gerçekleştirilmiştir. İçerik tabanlı atıf analizi modeli yaratılabilmesi için öncelikle makalelere ait üst veri, kaynakça ve tam metin bilgileri toplanmış, bu makalelere ait 12.881 kaynakça ve 101.019 cümle çalışma için oluşturulan veri tabanında depolanmıştır. Bu işlemin ardından atıflar için taksonomik türler belirlenmiş ve bu makaleler içindeki atıf cümleleri yürütülen uzlaşmalı etiketleme süreci ile sınıflara ayrılmıştır. Son aşamada ise Weka yazılımı kullanılarak sınıflamanın başarımı test edilmiş ve bu başarı oranları göz önüne alınarak içerik tabanlı atıf analizi modeli sunulmuştur.

Çalışmada atıflar; anlamı açısından, yapıma amacı açısından, şekli açısından ve dizilimi açısından atıflar olarak dört temel taksonomik kategoriye ayrılmış, ardından her bir tür kendi içinde alt kategorilere bölünmüştür. Bu alt kategoriler; anlamı açısından atıflar için pozitif, negatif ve nötr; yapıma amacı açısından atıflar için literatür, tanım yapma, yöntem, veriye atıf ve veri doğrulama; şekli açısından atıflar için ise yazar adı anılarak, toplu ve tırnak içinde atıf şeklinde belirlenmiştir. Dizilimi açısından atıflarda ise atıfın yapıldığı bölüm, kaç kez ve kaç farklı bölümde atıf yapıldığı dikkate alınmıştır.

Atıfların makinece sınıflandırılmasında kelime ön işleme yöntemi olarak 1-2 gram sözcük parçalayıcısı seçilmiş ve dur sözcükleri korunarak uygulama gerçekleştirilmiştir. Bunun temel sebebi dur sözcüklerinin atıf sınıflarını belirlemede önemli yere sahip olmasıdır. Kelime ön işleme sürecinin ardından sınıflamanın başarımı Weka yazılımı ile test edilmiş ve üç sınıf için de %90'ın üzerinde başarı elde edilmiştir.

Anlamı ve yapıma amacı açısından atıfların makinece sınıflandırılmasında Naive Bayes Multinomial algoritması kullanılmış ve anlamı açısından atıflarda %96,5; yapıma amacı açısından atıflarda %90,4'lük başarımlar elde edilmiştir. Anlamı açısından atıfların sınıflandırılmasında en düşük başarımların negatif atıfların saptanması konusunda olduğu görülmüştür. Bu durum yazarların daha üstü kapalı bir dille negatif atıf yaptığı savını doğrulamaktadır. Literatürde yapılan çalışmalar doğal dil işlemenin diğer uygulamaları analize eklendiğinde (duygu sözlüğü ve gövdeleme gibi) başarımların arttığını göstermektedir. Bu sebeple gelecek çalışmalarda çeşitli analizlerin eklenmesi ile başarımlar artırılabilir. Yapıma amacı açısından atıflarda en yüksek başarımlar veri doğrulama atıfları için saptanırken, en düşük başarımlar yöntem ve tanım yapma atıfları için elde edilmiştir. Bunun temel nedeninin de yöntem açıklarken yapılan tanımlar olduğu düşünülmektedir. Şekli açısından atıfların sınıflandırılmasında Random Forests algoritması kullanılmış ve algoritma %92 oranında başarılı bir sınıflama yapabilmektedir. En yüksek başarımlar yazar adı anılarak yapılan atıflar için, en düşük başarımlar ise tımağ içinde belirtilen atıflar için hesaplanmıştır.

Türkçe kütüphanecilik ve bilginbilim literatüründe atıfların %85'inin giriş ve literatür değerlendirmesi bölümlerinde yapıldığı, negatif atıflar ile veri doğrulama atıflarının genellikle bulgular ve sonuç kısımlarında yer aldığı, yazar adlarının en sık sonuç kısmında yapılan atıflarda anıldığı gibi sonuçlara ulaşılmıştır. Bunun yanında kaynakçalarda yer alan künyelerin %67'sine metin içinde yalnızca bir kez atıf yapıldığı ve %6'sına metin içinde hiç yer verilmediği saptanmıştır. Ayrıca metin içinde yer alan göndermelerin %1'i de kaynakçada yer almamaktadır. Bu durum yazarlar ve editörlerin atıf verirken daha dikkatli olması gerekliliğini ortaya koymaktadır.

Çalışma sonunda ortaya koyulan içerik tabanlı atıf analizi modeli ile araştırmacılar, editörler ve yönetici/karar vericiler için atıf değerlendirmelerinde dikkate alınması gereken temel noktalar belirlenmiş, bilimsel iletişim sürecindeki her bir rolün üzerine düşen görevler tanımlanmıştır. Bu noktada en önemli unsur tüm atıfların eşit olmadığına süreçteki herkes tarafından bilinmesidir. Söz konusu bilinç oluştuğunda atıflar üzerinden yapılacak manipülasyonlar da en aza indirilebilecektir.

Anahtar Sözcükler

İçerik tabanlı atıf analizi, doğal dil işleme, metin kategorizasyonu, Weka, atıf sınıflandırması, Türkçe kütüphanecilik ve bilginbilim literatürü

ABSTRACT

TAŞKIN, Zehra. *Designing a Model for Content-Based Citation Analysis: An Application for Turkish Citations Based on Text Categorization*, PhD Dissertation, Ankara, 2017.

One of the important components of measuring research/er performance is to evaluate the number of citations received. Academics receive incentives, promotions or rewards contingent to the number of these citations. Although the initial purpose for the recording of citations was to determine the publications related to one another, the use of citations also changed as a result of these new contingents. This situation sometimes brings about unethical practices such as the manipulation of the number of citations. Consequently, there emerged the necessity of conducting content analysis of citations in addition to quantitative evaluations.

The main aim of this study is to design a content-based citation analysis model for Turkish citations. For this end, 423-refereed articles, which have been published in library and information science literature in Turkey, are thoroughly examined. Firstly, all metadata, references, and full-texts of the articles are stored in a database to create a content-based citation analysis model. A total of 12,881 references and 101,019 sentences have been stored in this database. Then, the main taxonomic categories have been determined and citation sentences have been classified into these categories by tagging them with inter-annotator agreement process. At the last stage, the performance of the classification has been tested by using Weka software and a content-based citation analysis model is presented considering these performance ratios.

In this study, citations are divided into four main categories: citations in terms of meanings, citations in terms of purposes, citations in terms of shapes, and citations in terms of arrays. Then, each category is divided into sub-categories. The sub-categories are positive, negative, and neutral citations for meaning; literature, definition, method, data and data validation for purpose; mentioning author name, multiple citations in single sentence, and citation using direct quotations for shape. In evaluating the citations in terms of arrays, the sections of citations (introduction, method etc.), the number of use, and the number of citations in different sections in the texts have been considered.

In the categorization of citations by the machine, 1-2 gram word tokenizer has been chosen as the word preprocessing method and the application is run with the stop words

preserved. The main reason for this is that the stop words have importance in the determination of citation classes. Following the word preprocessing, the performance of the classification has been tested with the Weka software and over 90% performance is achieved for all three main categories.

Naive Bayes Multinomial algorithm is used to classify citations in terms of meanings (performance ratio is 96.5%) and purposes (performance ratio is 90.4%). It has been found that the lowest achievement in classifying citations in terms of meaning is in determining negative citations. This finding confirms the argument that authors make negative citations with more allusive words. Studies in the literature show that when different applications of the natural language processing are added to the analysis (such as sentiment dictionary or parsing), the performance can be improved. Success rates may be increased by adding various analyses in future studies. According to the results of citation purpose classification, the best performance has been determined for data validation citations, while the lowest performance has been detected for method and definition citations. The main reason for this is thought to be the definitions made when explaining the method. Random Forests algorithm has been used for the classification of citations in terms of shapes and the algorithm has been able to classify citations with the success rate of 92%. Highest achievement has been determined for citations with author names, while the lowest performance has been calculated for the citations indicated in quotation marks.

The results show that in the Turkish library and information science literature citation sentences are generally placed in introduction and literature review sections (85%), and negative and data validation citations are seen in the findings and conclusions sections. Additionally, citations by using the name of cited authors are generally found in conclusion sections. It is determined that 67% of the references are cited only once in the texts, and 6% are not cited in the texts at all. In addition, 1% of the citations in the texts are not found in the reference lists. This suggests that writers and editors should be more careful when citing and editing the papers.

In this study, the fundamental points to be taken into consideration during citation evaluation processes by researchers, editors, and managers/decision makers are presented by content-based citation analysis model. With this model, the tasks ideally assigned to each role in the scholarly communication process are also defined. The most important issue at this point is a realization on the part of all parties involved that the

meaning of a citation is not the same in every case. Once such an awareness is in place, it may be possible to minimize the manipulations done through the citations.

Keywords

Content-based citation analysis, natural language processing, text categorization, Weka, citation classification, Turkish library and information science literature

İÇİNDEKİLER

Kabul ve Onay	I
Bildirim.....	II
Teşekkür.....	III
Yayımlama ve Fikri Mülkiyet Hakları Beyanı	IV
Etik Beyan	V
Özet.....	VI
Abstract	VIII
İçindekiler	XI
Tablolar Dizini.....	XIV
Şekiller Dizini.....	XV
1. Bölüm: Giriş	1
1.1. Konunun Önemi	1
1.2. Araştırmanın Amacı ve Hipotezi	6
1.3. Araştırma Soruları	6
1.4. Araştırmanın Kapsamı	7
1.5. Araştırmanın Düzeni	7
1.6. Kaynaklar.....	8
2. Bölüm: Araştırmacı Değerlendirme Ölçütü Olarak Atıflar ve İçerik Tabanlı Atıf Analizi	
Çalışmaları	9
2.1. Giriş	9
2.2. Atıf ve Atıf Analizi Kavramı.....	9
2.2.1. Atıf Analizleri Üzerine Tartışmalar	10
2.2.2. Türkiye’de Atıfların Önemi	16
2.3. Atıf Analizi Türleri.....	18
2.3.1. Niceliksel Atıf Değerlendirme Çalışmaları	19
2.3.2. İçerik Tabanlı Atıf Değerlendirme Çalışmaları.....	19
2.3.2.1. Semantik (Anlamsal) Çalışmalar	20
2.3.2.2. Sentaktik (Sözdizimsel) Çalışmalar	24
2.4. Türkiye’de Atıf Değerlendirme Çalışmaları	26
3. Bölüm: Doğal Dil İşlemede Metin Kategorizasyonu	27
3.1. Giriş	27
3.2. Doğal Dil İşleme.....	27
3.2.1. Doğal Dil İşleme Süreçleri	28

3.2.1.1. Veri Toplama.....	29
3.2.1.2. Veri Ön İşleme	29
3.2.2. Doğal Dil İşleme Uygulamaları.....	30
3.3. Metin Kategorizasyonu.....	31
3.3.1. Metin Kategorizasyonunda Verilerin Hazırlanması.....	32
3.3.2. Metin Kategorizasyonunda Teknik ve Algoritmalar	32
3.3.3. Metin Kategorizasyonunda Performans Değerlendirme	33
3.3.4. Metin Kategorizasyonu Yazılımları Ve Weka	35
4. Bölüm: Yöntem ve Teknik.....	37
4.1. Giriş	37
4.2. Veri Setinin Tanımlanması	38
4.3. Verilerin Toplanması ve Analize Uygun Hale Getirilmesi	38
4.4. Veri Giriş Arayüzünün Tasarlanması	39
4.4.1. Üst Veri Bilgilerinin Tanımlanması	40
4.4.2. Kaynakça Bilgilerinin Toplanması	41
4.4.3. Bölüm Bilgilerinin ve Yapısal Tam Metinlerin Toplanması	42
4.5. Atıf Sınıflarının Belirlenmesi	44
4.5.1. Anlamı Açısından Atıflar	45
4.5.2. Yapılma Amacı Açısından Atıflar	46
4.5.3. Veriliş Şekli Açısından Atıflar	47
4.5.4. Dizilimi Açısından Atıflar	48
4.6. Etiketleme Süreci	48
4.6.1 Etiketleme Arayüzünün Tasarımı.....	48
4.6.2. Etiketleme Eğitimi.....	51
4.6.3. Uzlaşmalı Etiketleme Süreci	51
4.7. Atıfların Sınıflandırılmasında Doğal Dil İşleme Yöntemlerinin Uygulanması ...	51
4.7.1. Tüm Atıf Sınıflarına Ait Derlemlerin ve Arff Dosyalarının Oluşturulması	52
4.7.2. Kelime Ön İşleme Süreci	52
4.7.3. Algoritmaların Uygulanması.....	53
4.7.4. Algoritma Performanslarının Değerlendirilmesi.....	54
4.7.4.1. Metodolojik Değerlendirme.....	54
4.7.4.2. Sayısal Değerlendirme	54
5. Bölüm: Bulgular ve Değerlendirme	55
5.1. Giriş	55
5.2. Bulgular.....	55

5.2.1. Anlamı Açısından (Pozitif, Negatif ve Nötr) Atıfların Sınıflandırılması	55
5.2.2. Yapılma Amacı Açısından Atıfların Sınıflandırılması	59
5.1.3. Yapılma Şekli Açısından Atıfların Sınıflandırılması	64
5.1.4. Dizilimi Açısından Atıfların Sınıflandırılması	66
5.1.4.1. Atıf Yapılan Bölümlerin Değerlendirilmesi.....	66
5.1.4.2. Metin İçinde Atıf Yapılma Sayılarının ve Atıf Yapılan Bölümlerin Değerlendirilmesi	69
5.2. Değerlendirme ve İçerik Tabanlı Atıf Analizi Modeli	70
6. Bölüm: Sonuç ve Öneriler	74
6.1. Sonuç	74
6.2. Öneriler	75
6.3. Gelecek Çalışmalar.....	82
7. Bölüm: Kaynakça.....	83

TABLolar DİZİNİ

Tablo 1. Atıf sınıflama modeli örneđi.....	24
Tablo 2. Metin kategorizasyonunda kullanılan teknik ve algoritmalar	33
Tablo 3. Tahmin karışıklık matrisi	34
Tablo 4. Yapılma amacına göre atıf türleri, tanımlamalar ve örnekler	47
Tablo 5. Yapılma amacı açısından atıflar için analiz sonuçları ve karışıklık matrisi	64

ŞEKİLLER DİZİNİ

Şekil 1. Veri işleme süreci.....	29
Şekil 2. Veri ön işleme aşamaları.....	30
Şekil 3. ROC eğrisi örneği.....	35
Şekil 4. İçerik tabanlı atıf analizi modeli tasarımı süreci	37
Şekil 5. İlişkisel veri tabanının varlık-ilişki şeması	39
Şekil 6. Veri giriş arayüzü	40
Şekil 7. Örnek üst veri girişi	41
Şekil 8. Veri girişi sonrası künyelerin gösterimi	42
Şekil 9. Bir makalenin temel bilimsel bölümleri ve IMRAD yapısı	43
Şekil 10. Bölümlere göre tam metin verisi giriş ekranı.....	44
Şekil 11. Taksonomik atıf sınıfları	45
Şekil 12. Etiketleme arayüzünün genel görünümü	50
Şekil 13. Türkçe atıflar için makine öğrenmesi süreci.....	52
Şekil 14. Anlamına göre atıfların sınıflandırılması ve makine tarafından saptanması süreci	56
Şekil 15. Yapılma amacına göre atıfların sınıflandırılması ve makine tarafından saptanması süreci	61
Şekil 16. Yapılma amacı açısından temel atıf sınıfları.....	63
Şekil 17. Yapılma şekli açısından temel atıf sınıfları	65
Şekil 18. Dizilimi açısından atıfların bölümlere dağılımını gösterir infografik	67
Şekil 19. Kaynakçalarda yer alan referansların metin içinde geçme sıklıkları.....	69
Şekil 20. Türkçe atıflar için içerik tabanlı atıf analizi modeli.....	70
Şekil 21. Bilimsel iletişim sürecindeki dört farklı rol için atıf yapma, değerlendirme, uygulama önerileri.....	76

1. BÖLÜM: GİRİŞ

1.1. KONUNUN ÖNEMİ

Bilimsel yayınlar arařtırmacıların arařtırma bulgularını sundukları ortam olmalarından dolayı önemlidir. Bu bağlamda çalışmalar arasında kurulmuş ilişkiler arařtırmacılar için kritik öneme sahiptir (Shum, 1998, s. 19). Yani bilimsel bilginin arařtırma çevrelerinde yayılmasında ilişkilerin önemli bir yeri bulunmaktadır. Bu ilişkileri sağlayan en temel unsur atıflardır (Teufel, 1999, s. 33). Yazarlar yayınlarında atıfları; örnekleme, ayrıntılandırma veya eleştirme amaçlı olarak kullanırlar. Yazar, atıf yaptığı kaynağın kendi çalışması ile ilgili olduğunu ve okuyucusunun ilgisini çekebileceğini düşünür. Bu sayede yayınlar arasında atıflarla sağlanan bir ilişki kurulumu (Oppenheim, 1996, s. 155). Bu açıdan bakıldığında atıf yapmanın temel işlevinin atıf yapan ile yapılan yayın arasında bir bağ kurmak ve ikisi arasında ilişkiyi sağlamak olduğu vurgulanmaktadır (Smith, 1981, s. 84). Atıfların mecazi olarak akademik başarının içinde dondurulmuş ayak izlerine benzediği ve bu sayede fikirler geçidine tanıklık ettikleri belirtilmektedir (Cronin, 1981, s. 16). Bu ayak izleri sayesinde yönlendirmelerin açığa çıkarılması, daha önce o yoldan geçenlerle ilgili bilgi sağlanması ve amaca yönelik ipuçlarının saptanması mümkün olmaktadır. Yani atıflar sayesinde insan bilgisinin gelişim gösterdiği, yenilik sürecine maddi katkı sağlandığı ve eğer atıflar düzgün bir şekilde sıralandıysa arařtırmacılara güçlü ve çok yönlü bir araç sunulduğu iddia edilmektedir. Bilimsel bilginin gelişimi; keşfetme, değerlendirme, sağlamlaştırma ve yeniden değerlendirme sürecinin tamamıdır (Cronin, 1981, s. 20). Bu nedenle bilimsel arařtırmalarda tekrarlanan deneyimler büyük öneme sahiptir. Deneyimlerin tekrarlanabilmesi, yayınlar arasında atıflar aracılığı ile sağlanmış ilişkiler ağı ile mümkün olmaktadır.

Atıfların performans değerlendirmelerinde kullanılması konusu uzun yıllar tartışmalara sahne olmuştur. Literatürde atıf değerlendirmeleri için iki farklı görüş bulunmaktadır (Bornmann ve Daniel, 2008, s. 46). Bir görüşe göre (örn.: Cole, 2000; Van Raan, 2004) atıf analizleri ödüllendirme, Nobel adaylarının belirlenmesi, birimlerin prestij ve imajlarının saptanması, akademik sıralamalar, akran değerlendirmeleri ve yükselmeler gibi konularda kullanılmaya oldukça uygun bir araçtır. Diğer tarafta yer alan yazarlara göre ise (örn.: Cozzens, 1985; Woolgar, 1991) atıf analizleri beraberlerinde getirdikleri çeşitli kuşkular sebebi ile arařtırma ve arařtırmacı değerlendirmelerinde kullanılamazlar. Bu gerekçelerden bazıları Bornmann ve Daniel (2008, s. 46) tarafından şu şekilde sıralanmaktadır:

- *Zaman ile ilgili faktörler:* Yıllar içinde yaşanan yayın sayısındaki devasa artış nedeniyle güncel yayınlar eskilerine nazaran daha fazla atıf almaktadırlar. Ancak, bu durum eski yayınların kötü olduğu anlamına değil, güncel yayınların daha fazla göz önünde olduğu anlamına gelmektedir. Diğer taraftan geçmişte çok atıf almış yazılar günümüzde de atıf almaya devam etmektedir. Bu sebeple bir bilimsel yayının yayımlanma tarihi aldığı atıf sayısını etkilemektedir.
- *Disiplin temelli faktörler:* Atıf yapma motivasyonları disiplinlere, hatta bir disiplinin alt disiplinlerine göre farklılık gösterebilmektedir. Bunun yanında bazı disiplinlerde yapılan yayın sayılarının diğer disiplinlere göre fazla olması da atıf sayılarını etkileyen faktörlerden biridir. Dolayısıyla bu unsur göz ardı edilirse hata yapılması kaçınılmazdır.
- *Dergi temelli faktörler:* Makalelerin yayımlandıkları dergiler, bu dergilerin görünürlükleri, uluslararası nitelikleri ve prestijleri gibi dergi temelli faktörler alınan atıf sayısında önemli rol oynamaktadır.
- *Makale temelli faktörler:* Makalelerin türü (araştırma makalesi, derleme, mektup vb.), çok yazarlı olup olmadığı, uzunluğu veya kaynakçasında fazla referansa sahip olması aldıkları atıf sayısını etkileyen faktörlerdir.
- *Yazar/okuyucu temelli faktörler:* Makalelerin dili, kültür bariyerleri, yazarlar arasında kurulan ağlar ve bu ağların dışında kalan yazarların daha az atıf alması gibi sebepler makalelerin aldıkları atıf sayılarını etkilemektedir.
- *Yayınlara erişilebilirlikleri:* Günümüzde açık erişim ile ilgili önemli gelişmeler kaydedilse de hâla makalelerin tam metinlerine erişim ile ilgili sorunlar bulunmaktadır. Bu sorunlar nedeniyle erişilemeyen yayınların aldıkları atıfların sayısı da haliyle diğer yayınlara oranla daha düşük olmaktadır.
- *Teknik problemler:* Mevcut yayınlarda yapılmış atıf hataları, atıf dizinlerindeki standardizasyon ile ilgili sorunlar veya dizinleme problemleri yazarların iz sürerek literatür değerlendirmesi yapmasını zorlaştırmakta, bu da bazı yayınların aldıkları atıf sayılarını etkilemektedir.

Yukarıda sıralanan tüm bu faktörler dikkate alınarak atıf temelli değerlendirmelerin iyileştirilmesini sağlamak mümkündür. Ancak bu aşamada önemli olan “yazarlar neden atıf yapar?” sorusunun doğru bir şekilde cevaplanabilmesidir. Literatürde atıf yapma motivasyonları üzerine çok çeşitli sınıflamalar olsa da en bilinen sınıflama atıf dizinlerinin yaratıcısı Garfield (1970, s. 82) tarafından yapılmıştır. Bu sınıflamaya göre yazarlar:

- *Konunun öncülerine saygı göstermek,*
- *İlişkili yayınlara kredi vermek,*
- *Yöntem ve kullanılan araçları açıklamak,*
- *Literatür arka plan bilgisini sağlamak,*

- *Kendinin ya da başkalarının çalışmalarını düzeltmek,*
- *Önceki çalışmaları kritik etmek,*
- *İddiaları kanıtlamak,*
- *Araştırmacıları gelecekteki çalışmalar konusunda uyarmak,*
- *Atıf almayan, kötü ünü olan veya kötü dizinlenmiş yayınlara öncülük ederek ortaya çıkarmak,*
- *Verileri doğrulamak,*
- *Bir konuyu, kavramı ya da kuramı tartışan orijinal yayınları kimliklendirmek,*
- *Bir fikri ya da çalışmayı reddetmek/onaylamamak,*
- *Başkalarının iddialarına itiraz etmek amaçları ile atıf yapmaktadırlar.*

Sayılan sebepler her ne kadar atıfların neden yapıldığı sorusuna cevap verse de, çeşitli faktörler nedeniyle en doğru yayınlara atıf yapılamazken; ilgisiz yayınlara atıf yapma potansiyeli bulunmakta (Smith, 1981, s. 84), yazarlar gelişigüzel şekilde kaynaklara atıf yapabilmektedirler (Kochen, 1974, s. 74). Bu nedenle mevcut beklenti en ilgili yayınların anılarak atıf zincirinin kuvvetlendirilmesi yönünde olsa da, bu durumda yayının konusu ile doğrudan ilgili kaynaklara atıf yapılmazken, görece daha az ilgili yayınlara atıf yapılabilmektedir. Price (1986, s. 58) yazarlarda var olan bu bireysel atıf verme şekillerini dikkate alarak tüm yazarların aynı tutarlılık, tamlık ve dürüstlükte kaynaklara atıf yapamayacaklarını belirtmektedir. Price'ın yaptığı bu değerlendirmenin doğruluğu günümüzdeki uygulamalarda daha net şekilde görülebilmektedir. Günümüzde bazı yazarlar stratejik düşünerek çeşitli yazılara atıf yapmakta ya da yapmamakta; dergi editör veya hakemleri makalelerden talep ettikleri zoraki atıf (coercive citation) eklemeleri ile kendilerinin ya da kurumlarının atıf sayılarını etkileyebilmektedirler (COPE, 2012). Bu gibi uygulamalar ile atıf sayıları kolaylıkla manipüle edilebilmektedir. Bu tür manipülasyonların en bilinen örneklerinden birinin *Web of Science*'ta dizinlenen Türk dergileri arasında yer alan *Energy Education Science and Technology* dergisi tarafından yapıldığı ortaya çıkarılmıştır. Derginin kendine atıf oranının %94 olması ile birlikte ortaya çıkan şüpheler sonucu dergide yapılan yayınlar incelenmiş ve bu yayınların büyük bir kısmının aslında aynı dergiden ya da aynı editörün yönettiği "kardeş" dergiden yapıldığı saptanmıştır (Kaplan, 2014; Öztürk, 2012). Sözü edilen dergideki atıf uygulamaları ile ilgili çeşitli yazılar literatürde yer bulmuş (Al ve Soydal, 2012; Öztürk, 2013) ve bu durumun *Web of Science* tarafından saptanması üzerine dergi 2013 yılında dizinden çıkarılmıştır. Derginin dizinden çıkarılmasından bir yıl önce TÜBİTAK tarafından verilen "Uluslararası Bilimsel Yayınları Teşvik Programı" kapsamında fen bilimleri alanında beşinci (163 başvuru - 36.226 TL); sosyal bilimler alanında ise açık ara farkla birinci sırada (289 başvuru - 151.624 TL) en fazla teşvik verilen dergi olması (Kaplan, 2013) atıfların yönetici ve karar vericiler açısından ne denli önemli olduğunu göstermesi bakımından anlamlıdır. Bu örnek, dergiler ve bu dergilerde var olan atıf manipülasyonları ile ilgili literatürde karşılaşılan tek örnek değildir.

Jeofizik alanında bir derginin baş editörü ve çeşitli başka jeofizik dergilerinde hakemlik rollerine sahip olan bir akademisyenin atıf manipülasyonu yaptığı saptanmış ve bulgular raporlanmıştır (Davis, 2017). “Atıf karteli mi yoksa editör sahtekarlığı mı?” şeklinde yorumlanan olay sonrasında sözü edilen editör görevinden istifa etmiş, bu editör hakkında soruşturma başlatılmıştır (Oransky, 2017). Atıflar araştırma, araştırmacı veya kurumların değerlendirilmesinde önemli bir ölçüt olarak görülmeye devam ettiği sürece bu gibi haberlere rastlanması şaşırtıcı olmayacaktır.

Atıf sayma ile ilgili geçmişten günümüze çeşitli sorun, manipülasyon ve eleştirilerin literatürde yer buluyor olmasına rağmen hâla sıklıkla tercih ediliyor olmasının en temel sebebi bu sistem ile yönetici ve karar vericilerin yazarlarla doğrudan iletişime girmeden (anket ve görüşme gibi) değerlendirme yapabiliyor olmalarıdır (MacRoberts ve MacRoberts, 1996, s. 435; Smith, 1981, s. 84). Bu bağlamda atıf değerlendirmeleri için “kutsal bilgi kaynağı” varsayılan atıf veri tabanlarına (*Web of Science*, *Scopus* gibi) başvurulmaktadır. Bu veri tabanları kullanılarak yapılan değerlendirmeler aracılığı ile en çok atıf alan yazarlar etkili yazarlar olarak sunulmakta,¹ atıf sayısına göre teşvikler verilmekte ve bilim adamları ödüllendirilmektedir (Lerner ve Wulf, 2007, s. 634; Miller, Coble ve Lusk, 2013, s. 520). Bibliyometrik yöntemler kullanılarak alanların en önemli yazarları, kurumları ve ülkeleri belirlenebilmekte, bilimsel alanlar haritalandırılabilenekte, ortak atıf analizleri gerçekleştirilmekte ve bilimin etkisi değerlendirilebilmektedir.² Ancak atıf sayıları yalnızca atıf yapan yazar o makaleyi gerçekten kullandığı ve kullanım bir yarara (kalite, değer ya da etki) dönüştüğünde bir gösterge olarak kabul edilebilmelidir (Smith, 1981, s. 87).

Son yıllarda atıfları sayarak araştırmacıların değerlendirilmesi ile ilgili olarak pek çok yayın yapılmakta ve bu durumun önüne geçilmeye çalışılmaktadır. Yapılan bilimsel çalışmaları değerlendirirken ya da araştırmacılara teşvik verilirken niceliksel değerlendirmeler yapılmasının ötesine geçilmesi gerekmektedir. Araştırmacıların

¹ Essential Science Indicators (http://wokinfo.com/products_tools/analytical/essentialscienceindicators/), Highly Cited Researchers (<http://highlycited.com/>) ve ScienceWatch (<http://archive.sciencewatch.com/>) gibi web sitelerinde yayın ve atıf sayılarına göre etkili yazarlar, kurumlar ve ülkelerin sıralamaları sunulmaktadır.

² Türkiye’de de akademik yükselme ve teşviklerde atıf sayısı önemli bir gösterge olarak kullanılmaktadır. Örneğin, Hacettepe Üniversitesinde yayınlarına yüksek dış atıf almış yazarlar Bilimsel Araştırma Projeleri Koordinasyon Birimi Yüksek Atıf Başarı Desteği kapsamında yurtdışı konferans seyahat hakkı kazanmaktadır (Hacettepe Üniversitesi..., 2015). Ayrıca akademik kadro tayinleri ve yükselmelerde atıf dizinlerinde yer alan makalelerin aldığı atıf sayıları -disiplinlere ve üniversitelere göre farklılık göstermekle beraber- önemli bir yer tutmaktadır (Öğretim Üyeliğine Yükseltilme..., 1982). Devlet üniversitelerinde çalışan akademik personele verilen “Akademik Teşvik Ödeneği”nde atıflar için ayrı bir bölüm yer almakta, her bir atıfa kadro derecesi ve atıf yapılan makalenin yazar sayısı doğrultusunda puan verilmektedir (Akademik, 2016)

değerlendirilmesinde çalışmalarının alana ne kadar katkı sağladığına bakılması gerektiği vurgulanırken, bunun bir yolunun da yazarların nerede ya da kaç tane yazdıklarının değerlendirilmesinin bir kenara bırakılarak ne yazdıklarına odaklanılması olduğu aktarılmaktadır (Al ve Soydal, 2014, s. 40).

Asıl işlevi yayınlar arasında ilişki kurmak olarak belirtilen atıfların kullanım amaçlarının yıllar içinde değişmesi ile atıf sayma yöneticiler ve karar vericiler için elzem, araştırmacılar için ise bir zorunluluk haline gelmiştir. Bilim alanında gelişme katetmiş ülkelerde (A.B.D. gibi) performanslar çalışmaların özgünlüğü, alana katkısı veya önemi dikkate alınarak jüri kararı ile değerlendirilirken, bilim alanında önemli gelişme kat edememiş ülkeler araştırmacılarına atıf dizinlerinde yer alan yayınları ve atıfları üzerinden teşvik vermekte ya da akademik kadro tayini gerçekleştirmektedirler (Tonta, 2014, s. 2). Atıf dizinlerinin yaratıcıları da dâhil pek çok araştırmacı atıf dizinlerinin bu amaçla kullanılmasının anlamlı olmadığını vurgulamıştır. Çünkü atıf dizinlerinin bu denli yoğun kullanılmaya başlanması ile yazarlar formalite gereği atıflar yapmakta veya orijinalini hiçbir zaman görmedikleri yayınlara atıf yapmaktadırlar. Özellikle Türkiye’de atıf dizinlerine giren Türkiye menşeli yayın sayısının artması ile birlikte atıflar konusunda da önemli sıkıntılar yaşandığı görülmüştür. Türkiye menşeli araştırmacılar ve Türkçe dergiler atıf sayılarını şişirmek amacı ile etik olmayan çeşitli uygulamalar gerçekleştirmekte (Kaplan, 2014), bu da bilimin ilerlememesine neden olmaktadır. Uluslararası literatürde atıf sayma uygulamasının değiştirilmesi gerektiği ile ilgili 1950’lerden itibaren pek çok yayın yapılmış ve son yıllarda içerik tabanlı atıf analizi kavramı geliştirilmiştir. Bu analizler sayesinde atıfların farklı şekilde değerlendirilmesi mümkün olmakta, formalite gereği yapılan atıflar ile negatif atıflar saptanabilmekte, ayrıca bilim dallarının atıf yapma motivasyonları belirlenebilmektedir. Ulusal literatürde ise atıf saymanın yerini alabilecek bir içerik tabanlı atıf analizi çalışması yapılmamıştır. Niteliksel atıf göstergeleri belirlemeyi amaçlayan tek tez çalışmasında (Tunç, 2012) ise uygulanması hayli zor olan ve teknoloji temeline sahip olmayan bir yol önerilmiştir. Çalışmamızın önemi hem Türkçe atıflar için semantik ve sentaktik yapıları çözümleyebilecek bir araç tasarlanması, hem de atıfların taksonomik türlerinin belirlenebilecek olmasıdır. Bu çalışma sonrasında atıf saymanın yerini alabilecek bir değerlendirme modelinin önerilmesi ile tüm atıfların eşit olarak değerlendirilmesinin önüne geçilmesi mümkün olabilecektir.

1.2. ARAŞTIRMANIN AMACI VE HİPOTEZİ

Günümüzde akademik ödüllendirme, teşvikler ve yükselmelere ilişkin değerlendirmelerin niceliksel veriler kullanılarak yapılması pek çok problemi de beraberinde getirmektedir. Sayıca fazla atıf sahibi olan yazarlar bu atıfların özelliklerine bakılmaksızın başarılı yazarlar olarak nitelendirilmektedirler. Bu nedenle de bilimsel yayınlarda yapılan atıfların sayısı gün geçtikçe artmakta, ancak bu atıflara yönelik herhangi bir içeriksel değerlendirme yapılmamaktadır.

Bu araştırma “tüm atıflar eşit değildir” hipotezi etrafında şekillendirilmiş ve bu hipotez doğrultusunda temel amaç “Türkçe atıflara yönelik semantik ve sentaktik yapıları çözümleyebilecek bir makine tasarlayarak araştırma değerlendirmeleri için içerik temelli bir değerlendirme modeli sunmak” olarak belirlenmiştir. Bu bağlamda Türkçe metinler incelenerek atıflar için çeşitli taksonomik kategoriler oluşturulmuş ve büyük yoğunluklu metinlerden bu kategorilerin otomatik olarak saptanmasına yönelik makine öğrenmesi süreci işletilmiştir. Türkiye’de içerik temelli atıf analizi konusunda daha önce yapılmış herhangi bir makine öğrenmesi modeli bulunmadığından bu araştırmanın sonuçlarının bilim ve teknoloji konusunda karar vericiler ve yöneticilere kolaylık sağlayacağı düşünülmektedir.

1.3. ARAŞTIRMA SORULARI

Bu çalışma kapsamında cevaplanması öngörülen temel araştırma sorusu “Türkçe atıflar için otomatik metin kategorizasyonu (automatic text categorization) tekniği kullanılarak semantik ve sentaktik temelli bir içerik tabanlı atıf analizi modeli yaratılması mümkün müdür?” olarak belirlenmiştir. Diğer araştırma soruları ise şunlardır:

- Türkçe atıfların taksonomik kategorizasyonu nasıldır ve bu atıflar için bir sınıflama şeması yaratılması mümkün müdür?
- Türkçe yayınların bilimsel bölümlerinde (giriş, yöntem, bulgular vb.) yapılmış atıfların taksonomik kategorileri arasında herhangi bir farklılık var mıdır?
- Türkçe yayınlarda pozitif, negatif ve nötr atıfların yeri nedir ve bu atıflarda ayırt etmeyi kolaylaştıracak bir dil kullanımı mevcut mudur?
- Bu çalışmanın sonuçları kullanılarak Türkçe metinler içinden atıfların türlerini saptayabilecek bir makine öğrenmesi modeli yaratılması mümkün müdür?

1.4. ARAŞTIRMANIN KAPSAMI

Çalışma kapsamında 2000-2015 yılları arasında *Bilgi Dünyası* dergisinde yayımlanmış 172 ve 1996-2015 yılları arasında *Türk Kütüphaneciliği* dergisinde yayımlanmış 251 hakemli Türkçe makale derinlemesine incelenmiştir.³ *Bilgi Dünyası*'nda yayımlanmış 23, *Türk Kütüphaneciliği*'nde yayımlanmış altı İngilizce makale araştırma kapsamının Türkçe literatür ile sınırlı olmasından dolayı araştırmaya dâhil edilmemiştir.

Çalışmada *Türk Kütüphaneciliği* ve *Bilgi Dünyası* dergilerinde yayımlanmış toplam 423 makalenin tüm üst verileri ve tam metinleri hazırlanan bir arayüz aracılığı ile erişilebilir hale getirilmiş, ardından otomatik metin kategorizasyonunu uygulayabilmek için etiketleme süreci işletilmiştir. Araştırmada kullanılan yöntem ve teknikler ile ilgili detaylı bilgi dördüncü bölümde sunulmaktadır.

1.5. ARAŞTIRMANIN DÜZENİ

Bu araştırma temel olarak altı ana bölüme ayrılmıştır.

İlk bölümde çalışmanın önemi, araştırma soruları, kapsamı ve düzeni bulunmaktadır. Bu bölümün temel amacı araştırmanın neden yapıldığının ve ne gibi sonuçlara ulaşılmışının hedeflendiğinin detaylı olarak ortaya koyulmasıdır.

Konunun önemi bölümünde de aktarıldığı gibi atıflar günümüzde pek çok farklı nedenle kullanılmaktadırlar. Bu bağlamda atıfların öneminin ve kullanım alanlarının tanımlanması büyük önem taşımaktadır. Bu çalışmanın ikinci bölümünde araştırma/cı değerlendirme ölçütü olarak kullanılan atıflar ile ilgili temel bilgiler verilmektedir. Nicel ve nitel atıf değerlendirme çalışmaları, atıf sınıflandırma sistemleri ve doğal dil işleme uygulamalarının atıf analizlerinde kullanımı ile ilgili bilgiler yine bu bölümde yer almaktadır. Ayrıca çalışma kapsamında kullanılan atıf sınıflandırma sistemi bu bölümde detaylıca aktarılan arka plan çalışmalar aracılığı ile belirlenmiştir.

³ *Bilgi Dünyası* dergisi yayın hayatına 2000 yılında başlamıştır. Çalışma kapsamında dergide 2015 yılı sonuna kadar yayımlanmış tüm hakemli makaleler değerlendirilmiştir. *Türk Kütüphaneciliği dergisi* ise 1952 yılında başladığı yayın hayatını 1996 yılından itibaren hakemli makaleler yayımlayarak sürdürmüştür. Çalışma kapsamında dergide 1996 yılından başlayarak 2015 yılı sonuna kadar yayımlanmış tüm hakemli makaleler değerlendirilmiştir.

Türkçe kütüphanecilik ve bilgilim literatüründe otomatik metin kategorizasyonu sık tercih edilen bir yöntem değildir. Ancak bu yöntem pek çok disiplinde yoğun bir şekilde kullanılmaktadır. Bu çalışmanın üçüncü bölümünde doğal dil işleme otomatik metin kategorizasyonu tekniğinin tanımı, kullanım alanları, verilerin hazırlanması ve metin kategorizasyonu teknikleri gibi konular detaylı olarak anlatılmaktadır.

Araştırmanın çeşitli aşamalarında birbirinden farklı yöntem ve teknikler uygulanmış ve çeşitli uygulamalar geliştirilmiştir. Bu nedenle her bir aşamada izlenen yolun tanımlanması büyük önem taşımaktadır. Verilerin toplanması aşamasından başlayarak, veri tabanının tasarımı, verilerin etiketlenmesi, metin kategorizasyonunun uygulanması gibi her aşamanın işletilmesi ile ilgili detaylı bilgiler çalışmanın dördüncü bölümünde sunulmaktadır.

Çalışmanın beşinci bölümünün ilk kısmında elde edilen bulgular çeşitli infografikler aracılığı ile sunulmakta ve yorumlanmaktadır. İkinci kısımda ise bu bulgulardan hareketle atıflar için bir değerlendirme modeli önerilmektedir. Çalışmanın son bölümü olan altıncı bölümde ise sonuç ve öneriler sunulmakta, gelecekte yapılabilecek çalışmalara yer verilmektedir.

1.6. KAYNAKLAR

Çalışma ile ilişkili yayınlara ulaşmak amacı ile çeşitli basılı ve elektronik kaynaklar kullanılmıştır. Atıflar ile ilgili yayınlar genellikle aşağıdaki dergilerde yayımlandığından bu çalışmanın odağını temel olarak aşağıdaki dergilerde yayımlanan makaleler oluşturmaktadır.

- *Journal of the Association for Information Science and Technology*
- *Journal of Informetrics*
- *Scientometrics*
- *Social Studies of Science*

Tezin yazımında Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü “Tez Yazım Yönergesi” kullanılmıştır.⁴

⁴ http://www.sosyaltbilimler.hacettepe.edu.tr/belgeler/Tez_ve_Rapor_Yazim_Yonergesi.pdf

2. BÖLÜM: ARAŞTIRMACI DEĞERLENDİRME ÖLÇÜTÜ OLARAK ATIFLAR VE İÇERİK TABANLI ATIF ANALİZİ ÇALIŞMALARI

2.1. GİRİŞ

“Konunun Önemi” bölümünde de belirtildiği gibi başlangıçta temel amacı yayınlar arasında ilişki kurmak olan atıfların kullanım alanları yıllar içinde farklılık göstermiş ve araştırmacıların değerlendirilmesinde kullanılmaya başlanmıştır. Bu sebeple son yıllarda akademik çevrelerce niceliksel açıdan değerlendirilen atıflar ile araştırmacılara ödül, teşvik veya yükselme verilip verilemeyeceği tartışılmaktadır. İkinci bölümde atıfların tanımı, kullanım amaçları, yazarların atıf yapma motivasyonları, nicel ve nitel atıf değerlendirme teknikleri gibi çeşitli konular derinlemesine incelenmektedir. Bunun yanında son yıllarda pek çok araştırmaya da konu olan ve bu tezin de temelini oluşturan içerik tabanlı atıf analizleri ile ilgili detaylı incelemeler de yine bu bölümde sunulmaktadır.

2.2. ATIF VE ATIF ANALİZİ KAVRAMI

Türk Dil Kurumu'nun Güncel Türkçe Sözlüğü'nde atıf “yöneltme, çevirme”, “ilişkili bulma” ve “gönderme” olarak tanımlanmaktadır (TDK, 2015). Oxford Sözlüğü'nde kökeni Ortaçağ İngilizcesinde kullanılan ve Latince “citare” sözcüğüne dayanan “citation” sözcüğünün ilk anlamı “genellikle bilimsel çalışma içinde bir kitap, evrak ya da yazardan alıntı yapmak ya da kaynak göstermek” olarak tanımlanmıştır (Oxford Dictionaries, 2015). Yazarlar kendi güncel çalışmaları ile bilimsel literatürün engin arşivlerinde yer alan önceki çalışmalar arasında belirgin bağlantıları atıflar aracılığı ile kurarlar (Garfield, 1997). Bu belirgin bağlantılar bir yazarın yayımlanmış bir teori, metot ya da veriyi kullanışlı görmesi ve kaynak göstermesi ile oluşturulur.

Kavram olarak atıf yapma Rönesans döneminde, matbaanın bulunmasından sonra kullanılmaya başlanmıştır (Al ve Tonta, 2004, s. 21). Başlangıçta konunun öncülerine saygı göstermek, araştırmalar arasında bağ kurmak, yöntemi ve kullanılan araçları açıklamak, literatür arka planı sağlamak, önceki çalışmaları kritik etmek, iddiaları kanıtlamak, bir fikri doğrulamak ya da yanlışlamak veya verileri doğrulamak gibi amaçlarla yapılan atıflar günümüzde araştırmacıların değerlendirilmesi amacı ile kullanılmaya başlanmıştır (Al, 2008, s. 19; Garfield, 1970, s. 82; Weinstock, 1971, s. 191). Bunun yanında günümüzde atıflar çalışmaları için iz süren veya yükselme/teşvik başvurusunda

bulunacak olan arařtırmacılar, editörler, hakemler ve yönetici/karar vericiler tarafından birbirinden farklı amaçlarla yoğun olarak kullanılmaktadırlar.

Yayınlarla ait atıf bilgileri çeřitli veri tabanlarında tutulmaktadır. Clarivate Analytics firması tarafından sunulan *Web of Science* ve Elsevier yayınevini oluşturduđu *Scopus* veri tabanları bu atıf kaynaklarının en bilinen örnekleridir (Clarivate Analytics, 2017; Elsevier, 2015). Dünya çapında bilinen performans deđerlendirmeleri ve sıralama sistemlerinde bu iki veri tabanından yararlanılarak sonuçlar sunulmakta ve arařtırma/cılar deđerlendirilmektedir. Örneđin, CWTS Leiden Rankings ve URAP (University Ranking for Academic Performance) yaptıkları sıralamalarda *Web of Science* veri tabanını kullanırlarken (CWTS Leiden Ranking, 2015; URAP, 2015), SciMago ve QS World University Rankings *Scopus* veri tabanını kullanmaktadır (QS, 2015; SJR, 2015). Bunun yanında dünya çapında sıklıkla başvuru ve yayın/atıf sayılarına göre etkili yazarlar, kurumlar ve ülkeleri sunan *Essential Science Indicators* (Thomson Reuters, 2014), *Highly Cited Researchers* (Thomson Reuters, 2015) ve *ScienceWatch* (ScienceWatch, 2015a) gibi web sayfaları da veri kaynađı olarak bu veri tabanlarını kullanmaktadırlar. Sözü geçen veri tabanları aracılıđı ile yapılan deđerlendirmeler neticesinde en çok atıf alan yazarlar etkili yazarlar olarak etiketlenmekte, alınan atıf sayılarına göre teřvikler verilmekte ve bilim adamları ödüllendirilmektedir (Lerner ve Wulf, 2007, s. 634; Miller, Coble ve Lusk, 2013, s. 520). Atıf analizleri sonucu elde edilen bulgular ile alanların en önemli yazarları, kurumları ve ülkeleri belirlenebilmektedir. Tüm bu çalıřmalardan elde edilen sonuçların yorumlanması ile de bilimin etkisi deđerlendirilmektedir.

Atıf analizleri arařtırmacıların deđerlendirilmesi amacı ile ilk kez kullanılmaya bařlandığında, arařtırma yöneticileri atıfların neyin ölçüsü olduđunu ve bunu nasıl ölçtüđu konusunda kafa karıřıklıđı yařamıřlardır (Cozzens, 1989, s. 437). Atıflar ile ilgili söz konusu tartıřmalar ile ilgili genel çerçeve "Atıf Analizleri Üzerine Tartıřmalar" bařlıđı altında detaylı olarak aktarılmaktadır.

2.2.1. Atıf Analizleri Üzerine Tartıřmalar

Literatürde atıf saymak kaliteyi, etkiyi, önemi, etkinliđi, kullanıřlılıđı, görünürlüđu veya diđer konuları ölçebilir mi sorularına pek çok çalıřmada cevap aranmıřtır. Bunun yanında zamanlamanın, makalenin yayımlandıđı derginin, alan farklılıklarının, kendine atıfların, yazarların deneyim sürelerinin, disiplinlerarası alan örüntülerinin, yayınların dilinin, grup atıflarının ve negatif atıfların etkilerinin ne olabileceđi konusu çeřitli çalıřmalarda tartıřılmıřtır

(Cozzens, 1989, s. 437-438). Bibliyometri alanında çalışanların yukarıda sıralanan sorulara cevap verebilmeleri halinde atıf analizlerinin anlamlı olabileceği ve gelecekteki çalışmalarla geliştirilebileceği vurgulanmaktadır (Cozzens, 1989, s. 437). Atıfların performans değerlendirme amacıyla kullanılıyor olması bilim alanlarındaki gelişmelerin resminin yanıltıcı şekilde sunulması sebebi ile çeşitli tartışmaları da beraberinde getirmiştir (Moravcsik ve Murugesan, 1975, s. 86). Atıf dizinlerinin yaratıcısı Garfield, sadece atıfların bir yazarı değerlendirmek amacı ile kullanımının akıllıca olmadığını vurgulamış, atıf sıklığının bir yazarın çalışmasının önemini ölçmekten çok araştırma faaliyetinin bir ölçüsü olduğunu savunmuştur. Garfield'a göre atıf sıklığı ile yazarlar değil, o yazarlara ait eserlerin etkisinin ölçülmesi mümkün olabilmektedir. Bu nedenle de yazar değerlendirmede atıf sıklıkları ile birlikte diğer ölçümlerin de kullanılması gerekmektedir (Garfield, 1973, s. 407). Bir diğer görüş ise (Goudsmith, 1974, s. 28) atıf sayma işinin bilimsel performansın ölçülmesi, ödüllendirme veya kadro tayinlerinde kullanılmaya başlanmasının araştırmacılarda daha fazla sayıda atıf yaparak daha fazla atıf alma ya da rakibe atıf yapmazken arkadaşlara atıf yapma mantığına dönüşebildiğini belirtmektedir. Bir başka çalışmada (Ziman, 1968, s. 58) benzer şekilde pek çok atıfın kibarlıktan ya da politikaya uymak gerektiğinden verildiği; bu sebeple etkili bir gösterge olarak kabul edilemeyeceği görüşü savunulmaktadır. Oppenheim da (1996, s. 156) bu konunun üzerinde önemle durarak yazarların çalışmalarında ne kadar çok kaynağa atıf yaparlarsa o kadar çok atıf aldıklarını doğrulamıştır. Sonuç olarak bu gibi durumların yayınlarda yapılan atıfların sayılarını artıracakları ancak çalışmalara fazla katkı sağlamayacakları öngörülmüştür. Bu öngörünün gerçekleşip gerçekleşmediğini kestirmek amacı ile küçük bir uygulama yapıldığında bu durum net olarak anlaşılmaktadır. Örneğin, 26 Şubat 2017 tarihinde *Web of Science* veri tabanından "bibliometrics" konu sorgusu ile elde edilen kayıtların 1975-1980 yılları arasında yapılmış olanlarının kaynakçalarında yer alan referansların ortalaması 8 iken (ortanca 8) 2011-2016 yılları arasında yayımlanan makalelerde bu sayının 37 (ortanca 31) olduğu sonucuna ulaşılmıştır. Bu sonuç literatürde yer alan ve önemli dergilerin değerlendirildiği bir çalışma ile de desteklenmektedir (Nicolaisen ve Frandsen, 2015, s. 19-27). Bu çalışmaya göre 1992 yılında dergilerin kaynakçalarında yer alan toplam referans sayısı 190.732 iken 2012 yılında bu sayı 241.682'ye ulaşmıştır.⁵ Bu durum yıllar içinde yapılan yayın sayısının artışı ile bağlantılı olsa da, elde edilen bulgular yıllar içinde kaynakçada yer verilen referans sayılarının arttığı öngörüsünü desteklemekte ve endişeleri haklı çıkarmaktadır. Bu endişeler konusunda Vinkler (1994, s. 499) olaya farklı bir noktadan yaklaşmış ve elbette sayıların veri, verilerin

⁵ Çalışmada analiz edilen dergilerin ilgili yıllarda kaç makale yayımladığına ilişkin bilgi bulunmamaktadır.

de gösterge olarak kabul edilemeyeceğini, ancak ilgili veri, uygun yöntemler ve göstergeler olmaksızın bilimetrimin sürdürülmesinin de mümkün olamayacağını savunmuştur.

Atıf sayısı üzerinden yapılan değerlendirmeler veri tabanlarının sunduğu sayısız kolaylık sebebi ile yöneticiler ve karar vericiler tarafından sıklıkla kullanılmaktadırlar (Al, 2008, s. 19). Bir yayının hiç atıf almamış olmasının o yayını değersiz yapmamasına rağmen atıf analizleri üzerine çalışan bazı yazarlar Nobel ödülü almış bilim adamları gibi bilim dünyasında katkı sahibi olan yazarların genellikle çok sayıda atıf aldıklarını iddia etmektedirler (Garfield ve Malin, 1968; Garfield, 1986, s. 186; Sher ve Garfield, 1983). Thomson Reuters firması bu bağlantıyı göz önüne alarak her yıl Nobel ödülü alması muhtemel fizik, tıp, ekonomi ve kimya alanında çalışan bilim adamlarını "Citation Laureates" başlığı altında listelemektedir (King, 2015). 2002 yılından beri bu listeler aracılığı ile bilimsel atıf verileri kullanılarak Nobel alacak kişiler tahmin edilmeye çalışılmaktadır. Şimdiye kadar sözü geçen dört alanda ödül almış toplam 131 bilim adamından 37'si (%28) tahmin edilebilmiştir (ScienceWatch, 2015b). 2004, 2006 ve 2015 yıllarında ise tahminlerinde başarıya ulaşamamışlardır. Literatürde yapılmış bir çalışmada (Gingras ve Wallace, 2010, s. 411) bibliyometrik analiz araçları ile Nobel Ödüllerini tahmin etmenin neredeyse imkânsız olduğu sonucuna ulaşılsa da, %28'lik bir tahmin oranı her 10 ödülde 3'ünün tahmin edilebildiği anlamına geldiğinden yayın sayısının devasa boyutlara ulaştığı günümüzde yadsınamaz bir başarıdır. Bu durum Nobel Ödüllerinde atıf sayısının etkili olabileceğini gösterse de, tahminde tek etkenin atıflar olmadığı da açıktır.

Bibliyometri literatürü 1970'lerin sonu ve 1980'li yıllarda atıf analizlerinin anlamlılığı ile ilgili pek çok çalışmaya sahne olmuş ve olumsuz, kendine (self-citation) ve yöntemsel kaynaklara yapılan atıflar üzerinde eleştiriler yapılmıştır. 1979 yılında hazırlanan makalede (Garfield, 1979) atıf analizlerinde negatif ve kendine atıfların çok büyük etkisi olmadığı savunulmuştur. Bunun yanında negatif atıfların bilimsel yayınlarda son derece nadir görüldüğü ve bunun atıf analizlerini etkilemediği belirtilmiştir (Carter, 1974). Bir diğer iddia ise negatif atıfların da anlamlı atıflar olduğu, çünkü bilimin eleştiri ile geliştiğidir (Garfield, 1979, s. 362). Bazı görüşlere göre eleştiri alabilecek kadar hatalı yayınlar da bilimsel literatüre yaptıkları katkı dolayısıyla değerlidir (Cole ve Cole, 1971, s. 26). Buna örnek olarak Nobel almış yazarların Nobel öncesi makalelerinin olumsuz eleştiriler alarak çeşitli dergiler tarafından reddedildiği gösterilmektedir. Aynı çalışmada yöntemsel makalelerin (methodological papers) diğer makalelere göre daha fazla atıf aldığı varsayımı üzerinde durulmuş, ancak bu görüşün geçerli olmadığı, bazı yöntem makaleleri çok sayıda atıf alırken çoğu yöntem makalesinin (%73) ise hiç atıf almadığı

belirtilmiştir (Garfield, 1979, s. 363). Makalenin iddiasına göre atıf analizleri bilim çevreleri tarafından tanımlanmamış olan etkiyi ölçemez. Bilim çevreleri tarafından tanımlanma da atıflar aracılığı ile sağlanabilmektedir. Atıfların değerlendirilmesi konusunda oldukça ses getiren bu makaleye yönelik karşıt görüş içeren bir çalışma Chubin (1980) tarafından kaleme alınmış ve özellikle “bilim çevreleri tarafından tanımlanma” işini hangi bilim çevrelerinin yapacağı konusu tartışmaya açılmıştır. Atıf saymanın anlamı/anlamsızlığı üzerine çalışmalar günümüzde de yoğun bir şekilde yapılmaya devam etmektedir.

1980’li yıllarda atıf dizinlerinde çoğunlukla Anglo-Sakson dünyasından İngilizce yazılmış dergiler yer aldığından üçüncü dünya ülkelerini bu dizinler üzerinden değerlendirmenin anlamlı olmadığı iddia edilmiştir (Arunachalam ve Manorama, 1988, s. 394, 406). Ancak, atıf dizinlerinin yıllar içinde kullanım amaçlarının değişmesi ve ticari amaçla yürütülen bölgesel genişleme stratejisi (Testa, 2008) ile birlikte geçmişteki değerlendirme zorluğunun getirdiği problemler günümüzde boyut değiştirmiştir. Atıf dizinlerinden yararlanılarak kolaylıkla elde edilebilen sayısal bilgiler Türkiye gibi ülkelerde akademik performansı değerlendirmek ya da üniversitelerin bilim camiasındaki yerini belirlemek amacı ile kullanılmaya başlanmıştır. Atıf dizinlerine dayanan bu listelerin araştırmacıların ya da onların yayınlarının kalitesini ölçmek için kullanılması otoriteler tarafından “ölümcül günah” şeklinde nitelendirilmesine rağmen, Türkiye gibi ülkeler bu sayılara dayalı yöntemlerin ölümcül cazibesine kapılmaya devam etmektedirler (Tonta, 2014, s. 16-17; Van Raan, 2005).

Atıf analizi çalışmalarının bilimi ölçmek için yetersiz olduğu pek çok çalışma ile dile getirilmiştir. Örneğin, yazarların çalışma yapacakları alan ile ilgili rastgele seçilmiş bazı yayınları okudukları ve bu yayınlara atıf yaptıkları iddia edilmiştir (MacRoberts ve MacRoberts, 1996, s. 436). Kaynakçalarda yapılan hatalardan hareketle atıflar için bir model geliştiren çalışmada (Simkin ve Roychowdhury, 2003, s. 269) yazarların atıf yaptıkları yayınların yalnızca %20’sinin orijinalini gördükleri belirtilmektedirler. Aynı yazarlar son yıllarda karşılaştırmalı çalışmaların atıfları sayma üzerine yoğunlaştığını, ancak bilim adamlarının çoğunun atıf yaptığı kaynağı görmemesi nedeni ile atıf saymanın anlamsız olduğunu vurgulamaktadırlar (Simkin ve Roychowdhury, 2006, s. 172).

Yazarlar atıf verirken başkalarının yaptıkları atıfları kopyalama eğilimindedirler. Bu durum da pek çok bilim dalında sıklıkla kullanılan bir kuram olan, en az çaba ve geçmiş deneyimlerin yardımı ile en iyi çıktılara ulaşma isteğini tanımlayan en az çaba kuramının (Zipf, 1949, s. 1) araştırmacıların günümüzdeki atıf yapma stilleri için de geçerli olduğunu

göstermektedir. Yazarların makaleleri görmeden atıf yapmaları konusunda literatürdeki önemli örneklerden biri Wetterer (2006) tarafından ortaya çıkarılan ve Madeira adasında yaşayan büyük başlı karıncalar hakkında yazılmış Almanca bir makale ile bu makaleye yapılan atıflardır. Bahse konu makaleye atıf yapan bir yazar makale içindeki bilgiyi İngilizce'ye çevirirken hata yapmış, ardından bu Almanca makaleye yapılan atıfların pek çoğunda aynı hata tekrarlanmıştır. Hatalı bilginin yeni çalışmalarda sürekli tekrar edilmesinin en önemli nedeninin, yazarların orijinal metne ulaşmadan hâlihazırda çevirisi yapılmış makaleyi kullanma arzusu olduğu sonucuna ulaşılmıştır.

Yazarların orijinal kaynağı görmemesinin yanında çeşitli diğer sorunlar da bulunmaktadır. Örneğin, önyargılı atıflar, atıf yapmak için ikincil kaynakların tercih edilmesi, konferanslarda ya da diğer başka ortamlarda görüşülen resmi olmayan bilgiye atıf yapılmaması ancak bilimsel çalışmalarda bu sayede elde edilen bilgilere yer verilmesi, alanlara göre atıf yapma motivasyonlarının, zamanın, atıf sayısının farklı olması, kendine atıflar, literatürü reddetme, önyargılı veriler ve teknik sebeplerden dolayı da atıf analizlerinin güvenilirliği sorunlu görülmektedir (MacRoberts ve MacRoberts, 1996, s. 436-438).

Atıf analizleri ile ilgili bir başka yaklaşım da bir atıf almış yayının hiç atıf almamış bir yayına göre atıf alma ihtimalinin daha yüksek olduğu ile ilgilidir. Çünkü akademisyenler ve bilim adamları kendi yayınlarını yaparken buldukları yayınların kaynakçalarını izleyerek atıf yapma eğilimindedirler. Esasında bu eğilim literatürde yeri olan ve "citation pearl growing" (Markey ve Cochrane, 1981, s. 19) olarak adlandırılan bir yaklaşım olsa da, yazarlar atıf yaptıkları kaynakların orijinallerini görmediklerinde yaptıkları atıfın anlamını yitirmesine neden olmaktadır. Bu atıf izleme yolu "en az bir atıf almış bir yayın gelecekte de mutlaka atıf alacaktır" gibi bir çıkarımı da beraberinde getirmektedir. Bu atıf yapma mantığının Matta İncil'i 25'inci ayetinde geçen "çünkü kimde varsa ona daha çok verilecek ve o bolluk içinde olacak. Ama kimde yoksa kendisinde olan da elinden alınacak (unto every one that hath shall be given, and he shall have abundance)" yaklaşımı ile benzerlik gösterdiği belirtilmektedir (Matthew 25:29, 2004; Simkin ve Roychowdhury, 2006, s. 181). Bu ayet temelinde 1968 yılında Matthew Etkisi kavramını ortaya çıkaran Robert Merton; ünü duyulmuş ya da seçkin araştırmacıların çalışmaları benzer olsa bile adı duyulmadık bir araştırmacıya göre daha fazla itibar ve atıf alacağını savunmaktadır (Merton, 1968, s. 59). Merton'a göre, Matthew Etkisi bireysel araştırmacıların kariyerlerine ve ödüllendirilmelerine doğrudan etki etmektedir. Literatürde Matthew Etkisi'ne benzer olarak Stigler Kanunu da geliştirilmiştir. Bu kanuna göre bilimsel buluşların çoğu gerçek kâşifinin adı ile anılmamaktadır (Stigler, 1980, s.

147). Hatta Stigler kendi adı ile geliştirdiği bu kanunun da aslında Matthew Etkisi'ni bulan Robert Merton'un bir hipotezinden yola çıkılarak geliştirildiğini savunmaktadır. Tüm bu nedenlerle atıf saymanın ününün ününe ün kattığı, ünsüzün ise bu ünü sağlamak için ünlü araştırmacılar ile birlikte yayın yapması gerektiği sonucuna ulaşılmış ve bu durumun anlamsızlığı vurgulanmıştır. Matthew Etkisi'nin atıf analizlerine yakıştırıldığını ancak bunun gerçeği yansıtmadığını iddia eden çalışmalar da literatürde yer bulmuştur (Van Raan, 1998, s. 130). Matthew etkisine benzer şekilde "başarı başarıyı besler (success breeds success)" yaklaşımında da literatürde hâlihazırda atıf almakta olan yayınların diğerlerine göre ve yaşlı araştırmacıların gençlere göre avantajlı olduğu vurgulanmaktadır (Cozzens, 1985, s. 149).

Atıf saymayı eleştiren bir çalışmada tüm atıflar eşit değerde olmadığından bu analizlerin güvenilir olmadığı savunulmuştur (Oppenheim, 1996, s. 157). Ayrıca atıf sayma işinin her şeyi sadece sayılara indirgediğinden "etki" ve "kalite" gibi unsurları ölçmeye yetmediği, atıf verilirken yapılan hataların bile sonucu %10-%20 oranında etkileyebileceği belirtilmiştir. Atıf analizlerinin önemli sorunlarından birinin de hatalı yayınlar olduğu vurgulanmıştır. Örneğin, bilim çevreleri tarafından John Darsee isimli yazarın ürettiği 55 yayının yalnızca yedisinin geçerli, 40'ının şaibeli ve sekizinin kesinlikle hileli olduğu ortaya çıkarılmıştır. Ancak bu yazara 1982 ve 1990 yılları arasında toplam 298 yayından atıf yapılmıştır. Yani yazar hatalı yayınlar ile oldukça yüksek miktarda atıf almıştır. Ancak enteresan olan bir diğer nokta da atıf yapan 256 yayının (%86) Darsee'nin makalelerini onaylar ya da yalnızca bu çalışmalardan söz eder nitelikte atıf yapmalarıdır (Oppenheim, 1996, s. 158). Geri çekilmiş (retracted) makalelerde de durum farklı değildir. Geri çekilmiş makaleler ile ilgili yapılan bir içerik tabanlı atıf analizi çalışmasında 2015 yılında en çok atıf yapılan ilk beş geri çekilmiş makaleye yapılan atıflar incelenmiştir (Halevi ve Bar-Ilan, 2016). Bu atıfların çoğunluğunun negatif olması beklentisine karşın atıf yapan yazıların büyük kısmında geri çekme ile ilgili herhangi bir ifade bulunmadığı, hatta pozitif atıf yapıldığı belirtilmiştir. Bunun yanında yazarlar yayıncıların atıf almaya devam eden bu yayınları bilerek web sitelerinden tamamıyla ücretsiz olarak erişilebilir olarak sunduklarını, bu yayınların atıf almalarını sağlayan birer katalizör görevi gördüğünü belirtmektedirler. Geri çekilmiş makaleler ile ilgili bir diğer yazıda da bu gibi çalışmaların geri çekildikten yıllar sonra bile hâla atıf aldıkları vurgulanmaktadır (Al ve Soydal, 2015, s. 32). Araştırma kapsamında incelenen geri çekilmiş makalelerin aldıkları atıfların yarısına yakını (%45) makaleler geri çekildikten sonra yapılmıştır. Bir diğer çalışmada geri çekilmiş makalelerin hâla atıf alıyor olma sebeplerinin açık olmayan yayıncı web siteleri, korsan web sitelerinin

varlığı veya eski sürümlerin web’de yer almaya devam etmesi, yazarların yayının geri çekildiğini saklamaya yönelik tutumları gibi çeşitli etmenlerle açıklanmaya çalışılmış ve bu makalelerin hâla atıf alıyor olmalarının önemli bir probleme işaret ettiği belirtilmiştir (Silva ve Dobránszki, 2017, s. 1653).

Atıf sayma üzerine getirilen tüm bu eleştirilere rağmen atıflara performans değerlendirmede hâla sıklıkla başvurulmaktadır. Bunun temel sebebi atıfların zahmetsizce karar vermede kullanılabilmesi iken, atıf süreçlerinde günden güne artan manipülasyonlar, kopyala yapıştır usulü hiç görülmeden atıf yapılan kaynaklar, gereksiz referanslar ve unutulmuş önemli yayınlar gibi çeşitli sebeplerden bu süreç ister istemez kötü etkilenmekte ve atıfların amaçları dışında kullanılmasına sebep olmaktadır.

2.2.2. Türkiye’de Atıfların Önemi

Türkiye’de yayın ve atıf sayıları akademik yükselme ve teşviklerde önemli bir gösterge olarak kullanılmaktadır. 14 Aralık 2015 tarihinde Bakanlar Kurulu tarafından kararlaştırılan “Akademik Teşvik Ödeneği Yönetmeliği”nin T.C. Resmi Gazete’de yayımlanması ve yürürlüğe girmesi ile araştırmacılara devlet tarafından verilecek teşvikler için şartlar ve teşvik miktarları belirlenmiştir (Akademik, 2015). Yönetmelik 2016 yılında güncellenmiş, bu güncelleme ile atıflardan sağlanacak puanlarda az da olsa değişikliğe gidilmiştir. Akademik teşvik ödeneği puanlarının hesaplanmasında 2017 yılı itibarıyla şu şekilde bir yol izlenmektedir (Akademik, 2016):

- Devlete bağlı tüm yükseköğretim kurumlarında Akademik Teşvik Ödeneği Yönetmeliği’nde belirtilen esasları yerine getirmek üzere en az üç üyeden oluşan akademik teşvik komisyonu kurulur ve bu komisyon Aralık ayının ilk haftasında teşvik takvimini yayımlar. Komisyon üyeleri üç yıllık olarak seçilir.
- Her bir birim (bölümler, dekanlıklar, enstitüler vb.) içinde ön inceleme heyetleri kurulur. Bu ön inceleme heyetlerinin görevi başvuruları komisyona sunmadan önce detaylı incelemelerin yapılmasıdır. Bu kurulların görev süresi de üç yıldır.
- Başvuru sahipleri faaliyetlerine uygun birime başvuru yapmakla yükümlüdür. Yani, Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü’nde görev yapan bir öğretim elemanı başvuru dosyası ile bağlı bulunduğu bölüme başvurur. Bölüm ön inceleme heyeti dosyayı inceler ve bir üst heyete (Edebiyat Fakültesi Dekanlığı) gönderir. Burada da incelemesi tamamlanan dosya Rektörlük Akademik Teşvik Komisyonu’nun onayına sunulur.

- Akademik Teşvik Ödeneği için başvuru dosyaları hazırlanırken dokuz faaliyet alanında (proje, araştırma, yayın, tasarım, sergi, patent, atıf, tebliğ ve ödül) yıl içinde yapılan faaliyetler YÖKSİS üzerinden sisteme kayıt edilir ve raporlandırılır. Her bir faaliyet alanında alınabilecek maksimum puan 30 olarak belirlenmiştir. Bunun yanında bir akademisyenin alabileceği toplam puan 100'ü geçemez.
- Puanlama sisteminde atıfların da önemli yeri bulunmaktadır. Atıflarda uygulama ilkeleri diğer faaliyet alanlarına göre değişiklik gösterebilmektedir. Örneğin, diğer faaliyet alanlarında yazar sırası puanlamaya etki ederken (ilk yazar daha yüksek puan) atıf faaliyet türünün hesaplanmasında yazar sayısı dikkate alınmamaktadır (atıf alan tüm yazarlara eşit puan). Bunun yanında bir eserde aynı yayına birden fazla atıf yapılmış ise bu atıflar tek atıf olarak değerlendirilmektedir. Kendine atıflar teşvik kapsamında değildir.
- Atıfların puana dönüştürülmesinde izlenen yol ise şöyledir:
 - o Atıf faaliyet alanında atıflar sınıflara ayrılmış ve her bir sınıfın oranı belirlenmiştir. Bu oranlara göre; uluslararası kitaplardan alınan atıfların oranı %8, ulusal kitaplar %2, atıf dizinlerinde yer alan dergiler %4, atıf dizinleri dışındaki uluslararası dergiler ile *ESC'*de dizinlenen dergiler %2, hakemli ulusal dergiler %2'dir. Bu oranlar her bir faaliyet türü için belirlenen sabit 30 puan ve kadro derecesine göre belirlenmiş katsayılar ile çarpılarak nihai atıf puanı hesaplanmaktadır. Kadro derecelerine göre belirlenen katsayılar profesör, doçent ve uzmanlar için 1, yardımcı doçentler için 1,5, araştırma görevlisi, öğretim görevlisi ve okutman kadrolarında bulunanlar için 2'dir. Yani bir profesör veya uzman atıf dizinlerinde yer alan dergilerden bir atıf aldığına topladığı puan 2,4 iken; bir araştırma görevlisi aynı dergiden atıf aldığına topladığı puan 4,8'dir.
- Akademik teşvik ödemesi en yüksek devlet memuru brüt aylık tutarının kadro derecesi gözetilerek belirlenen oranlar ile çarpılması ve bu sayının da akademik teşvik puanının yüze bölünmesi sonucu elde edilen rakam ile çarpılması ile hesaplanır. Yani hesaplama formülü şöyledir; Akademik teşvik ödemesi tutarı=En yüksek devlet memuru brüt aylığı×akademik kadro unvanlarına göre belirlenmiş oran×(akademik teşvik puanı/100). Hesaplama profesör için belirlenmiş oran %100, doçent için %90, yardımcı doçent için %80, araştırma görevlisi, öğretim görevlisi, uzman, okutman, çevirici gibi kadrolarda bulunanlar için %70'tir. Bu bağlamda 2017 yılında 100 puan alan bir profesörün alacağı teşvik miktarı 843,76 Lira iken (T.C. Maliye Bakanlığı, 2017) aynı puanlı bir

araştırma görevlisinin alacağı teşvik miktarı 590 Lira'dır. Özetlemek gerekirse puanlama sisteminde bir profesörün 100 puana ulaşması bir araştırma görevlisine göre daha zor iken, bir profesör 100 puana ulaştığında aldığı ödenek de araştırma görevlisine göre daha yüksektir. Sistemde bu şekilde bir dengeleme yapıldığı görülmektedir. Ancak, ilginç bir başka nokta da uzman kadrosunda bulunan akademisyenlerin teşvik puanları profesörler ile benzer şekilde hesaplanırken teşvik ödemelerinde katsayının %70 olarak belirlenmiş olmasıdır. Bu durumda da 100 puanlı bir uzman bir profesör kadar üretim yapmış olsa da alacağı teşvik ödeneği 590 lira olacaktır.

Türkiye'de atıflar için uygulanan bir diğer teşvik kanalı ise araştırmacıların konferans/kongre katılımlarının desteklenmesidir. Örneğin, Hacettepe Üniversitesi'nde *Web of Science*'in *SCI*, *SSCI* ve *A&HCI* dizinlerinde taranan dergilerden yüksek atıf alan araştırmacılar son iki yılda aldıkları atıf sayısı dikkate alınarak "Yüksek Atıf Başarı Desteği" almaya hak kazanırlar (Hacettepe Üniversitesi..., 2015, s. 12). Her bölümün ulaşması gereken atıf sayıları alan özellikleri dikkate alınarak komisyon tarafından önceden belirlenmiş ve araştırmacılara sunulmuştur. Örnek vermek gerekirse Bilgi ve Belge Yönetimi Bölümü gerekli atıf sayısı alt sınırı 5 iken Kimya Mühendisliği Bölümü'nün sınırı 99 olarak belirlenmiştir.

Parasal ve konferans katılım desteklerinin yanında atıf sayılarının dikkate alındığı bir diğer uygulama da akademik kadro tayinleri ve yükselmelerdir. Disiplinlerine ve çalıştıkları üniversitelerin özelliklerine göre araştırmacılar atıf dizinlerinde yer alan makalelerinin aldıkları atıf sayıları üzerinden değerlendirilerek kadro veya yükselme alabilmektedirler (Öğretim Üyeliğine Yükseltme..., 1982).

Sayılan örneklerden de anlaşılacağı üzere atıflar Türkiye'de büyük öneme sahiptir. Atıfların niceliksel olarak değerlendirilmesi ile araştırmacılar desteklenmekte, ya da akademik kariyerlerine devam ettirilmektedirler. Bu nedenle de atıflar ile ilgili mevcut durumun ortaya çıkarılması ve bu yönde farkındalığın artırılması önemlidir.

2.3. ATIF ANALİZİ TÜRLERİ

Literatürde atıf analizleri iki şekilde yapılmaktadır. Bunlardan ilki ve en sık kullanılanı nicel yöntemler iken son yıllarda nitel çalışmalar da yapılmaya başlanmıştır. Nicel yöntemler bibliyometrik çalışmaların başlangıcından beri sıklıkla tercih edilen yöntemler olmasına

rağmen yukarıda da belirtildiği üzere beraberinde getirdiği dezavantajlar sebebi ile yerini içerik tabanlı atıf analizi çalışmalarına bırakmaya başlamıştır. Bu kısımda da çalışmanın temel çerçevesini oluşturan nitel atıf analizi çalışmalarına daha geniş yer ayrılmıştır.

2.3.1. Niceliksel Atıf Değerlendirme Çalışmaları

Literatürde sıklıkla karşılaşılan atıf analizi türü niceliksel atıf analizleridir. Atıf sayıları sadece bilimsel dergilerin ölçüm nesnesi olarak değil bireysel araştırmacılar, araştırma grupları, bölümler ve enstitüler, bilimsel disiplinler ve ülkeler için kullanılabilir bir ölçüttür (Moed, 2005, s. 14). Bibliyometrik çalışmaların yapılmaya başlandığı ilk yıllarda yayınların etkisi veya kalitesinin ölçülmesi için atıfların sayılmasının gerekli olduğu düşünülmüştür (Cole ve Cole, 1971, s. 23).

En sık kullanılan niceliksel atıf analizi teknikleri farklı iki kaynaktan aynı kaynağa atıf yapılması esasına dayanan bibliyografik eşleştirme ve bir kaynaktan farklı iki yayına atıf yapılması esasına dayanan ortak atıftır (Al ve Tonta, 2004, s. 23). Bunun yanında yazarların kendilerine ait geçmiş çalışmalara yaptıkları atıfları tanımlayan kendine atıf (Aksnes, 2005, s. 41) ve yayınların yıllar geçtikçe daha az atıf aldıklarını gösteren literatür eskimesi ve yarı yaşam (Meadows, 1967, s. 30) göstergeleri de niceliksel atıf değerlendirmelerinde sıklıkla kullanılmaktadır.

2.3.2. İçerik Tabanlı Atıf Değerlendirme Çalışmaları

Literatürde atıfların niceliksel olarak sayılmasından ziyade bu atıfların değerlerine göre işlem görmesi gerektiği ile ilgili çalışmalar 1950'lerden beri gerçekleştirilmektedir. Bu çalışmaları belli gruplarda toplamak gerekirse; atıfın metin içindeki yerini dikkate alan sözdizimsel (sentaktik) yaklaşımı esas alan değerlendirmeler, atıf yapılan yayın ile yapılan arasındaki ilişkiyi anlamsal (semantik) olarak değerlendiren çalışmalar (pozitif, negatif, nötr gibi), atıf yapılan kaynağa atıfın sıklığını araştıran çalışmalar (bir makalede kaç kez atıf yapıldığı) ile atıf motivasyonlarının sınıflandırıldığı çalışmalar şeklinde bir gruplandırmaya gitmek mümkündür. Bu çalışmaların hemen hemen hepsinde ilk sorunun "yazarlar neden atıf yaparlar?" olduğu görülmüştür.

Yazarların neden atıf yaptıkları dikkate alarak geliştirilen yeni nesil atıf analizlerine literatürde "içerik tabanlı atıf analizi (content-based citation analysis)" adı verilmiştir (Ding, Zhang, Chambers, Song, Wang ve Zhai, 2014, s. 1820). İçerik tabanlı atıf analizi sadece

sıklık sayarak yayınların etkilerinin ölçümünden çok tam metinlerin incelenerek atıf kapsamının değerlendirilmesi anlamına gelmektedir (Ding, Zhang, Chambers, Song, Wang ve Zhai, 2014, s. 1821). Bu analizleri semantik ve sentaktik yaklaşımlar olarak iki temel grupta incelemek mümkündür. Semantik yaklaşımlar atfın nasıl yapıldığı ile ilgilenirken, sentaktik yaklaşımlar atıf yapılan makaleye metin içinde nerede atıf yapıldığı sorusuna cevap vermek üzere uygulanmaktadır.

30-40 yıl önce içerik tabanlı atıf analizi amaçlı yapılmış çalışmaların örneklem boyutları ve kullandıkları teknikler nedeni ile yazarları tarafından çoğunlukla genele uyarlanamayacakları söylenmiştir. Ancak, günümüzde hesaplamalı dilbilim (computational linguistics) alanının hızla gelişmesi ile birlikte içerik tabanlı atıf analizlerini genele uyarlamak daha kolaylaşmıştır. Günümüzde tam metin dokümanlara açık erişim, büyük yoğunluklu metinleri işleme yeteneği ve çeşitli algoritmaların geliştirilebiliyor olması nedeni ile bu işin daha kolay yapılabileceği ifade edilmektedir (Teufel, 1999, s. 38). Makine öğrenmesi algoritmalarının geliştirilmesi ve pek çok eylemin bilgisayarlara kolaylıkla yaptırılabilmesi ile birlikte, içerik tabanlı atıf analizleri de makine öğrenmesi aracılığı ile gerçekleştirilmeye başlanmıştır.

İçerik tabanlı atıf analizlerinin semantik ve sentaktik yaklaşımlarla yapıldığı çeşitli çalışmalar literatürde yer bulmuştur. Bu analizler, temelinde atıfların niteliksel olarak değerlendirilmesini amaçlayan uygulamalar olan atıf tavsiye sistemleri (Liu, Yan ve Yan, 2013); bilgi madenciliği (Schneider ve Borlund, 2005); bilgi erişim (Aljaber, Stokes, Bailey ve Pei, 2010; Fu ve Aliferis, 2010; Liu, Chen, Ding, Wang, Xu ve Lin, 2014; Ritchie, 2008); duygu analizi (Athar, 2011; Cavalcanti, Prudêncio, Pradhan, Shah ve Pietrobon, 2011; Tandon ve Jain, 2012; Yu, 2013); atıfların sınıflandırılması (Bertin, 2008) ve atıf özetleme (Elkiss, Shen, Fader, Erkan, States ve Radev, 2008; Tandon ve Jain, 2012) gibi amaçlarla kullanılmaktadırlar. Literatürde yer alan içerik tabanlı atıf analizi çalışmalarında farklı yöntem ve teknikler kullanılarak sonuçlara ulaşılmaya çalışılmıştır. Atıflar daha önce de belirtildiği gibi semantik ve sentaktik çalışmalar olmak üzere iki şekilde sınıflandırılmaktadır.

2.3.2.1. Semantik (Anlamsal) Çalışmalar

Çalışmanın ilk kısmında da belirtildiği gibi yazarların atıf yapma motivasyonlarının neler olduğuna ilişkin yapılmış en bilinen sınıflama atıf dizinlerinin yaratıcısı Garfield tarafından geliştirilmiştir. Bu sınıflama dışında literatürde atıfların anlamları dikkate alınarak yapılmış

çeşitli başka sınıflandırmalar da bulunmaktadır. Örneğin, Nanba ve Okumura'ya göre (1999, s. 927) üç tip atıf vardır. Bunlar; diğer araştırmacıların teorilerine dayanan atıflar (Tip B), ilgili çalışmalarla karşılaştırma yapan ve onların problemleri üzerine odaklanan atıflar (Tip C) ile diğer tip atıflardır (Tip O). Çalışmanın yazarlarına göre en önemli atıf tipi C'dir, çünkü bu tipte önceki çalışmalara bir giriş ve bu çalışmaların problemlerine odaklanma söz konusudur. Bir diğer çalışmada Lipetz (1965), atıf yapılan kaynağa bilimsel katkıyı temel alan bir gruplamayı tercih etmiş ve atıf yapılan kaynağa bilimsel katkıyı en temel motivasyon göstergesi olarak kabul etmiştir. Oppenheim ve Renn de (1978, s. 226) atıf yapma nedenlerini yedi gruba ayırarak bu nedenler arasında en önemli ve en çok tercih edilenin tarihsel arka plan bilgisi sağlamak olduğunu vurgulamışlardır. Bonzi, kütüphanecilik ve bilgibilim alanı için dört temel atıf tipi belirlemiştir (Bonzi, 1982, s. 211). Bunlar; metin içinde özel olarak sözü edilmeyen toplu atıflar (literatürdeki çalışmalar göstermektedir ki... gibi), metin içinde özel olarak sözü edilen atıflar (Taşkın'ın da ifade ettiği gibi... gibi), tırnak içinde belirtilen alıntılar ile metin içinde birden fazla tırnak içinde verilmiş alıntılardır. Atıf yapma nedenleri de tıpkı bilgi davranışları gibi bilim dallarına göre farklılık göstermektedir ve bu nedenle de her alana özgü atıf yapma tiplerinin saptanması önemlidir (Cano, 1989). Bonzi'ye (1982, s. 211) göre kütüphanecilik ve bilgibilim alanında en çok tercih edilen tip yazar adı anılarak yapılan atıflar iken, en değerli tip aynı kaynaktan birden çok kez yazar adı anılarak alıntı yapılmasıdır. Literatürde yapılmış bu sınıflamaların en önemli özelliği atıf yapan yazarların yaptıkları atıfta en çok neyi dikkate aldıklarını göstermesidir. Moravcsik ve Murugesan (1975) yaptıkları çalışmada atıfların formalite gereği yapılıp yapılmadığı, tartışma amacı ile mi yoksa onaylama amaçlı mı yapıldığı gibi çeşitli sorulara cevap aramışlardır. Çalışma sonunda atıfların üçte birinin gereksiz olmasına rağmen yayınlarda yer bulduğu, kavramsal atıflardan çok işlevsel atıfların tercih edildiği, atıfların %40'ının formalite gereği yapıldığı, atıfların yedide birinin olumsuz (negational) atıf olduğu sonuçlarına ulaşılmıştır. Çalışma sonunda elde edilen bilgilerden hareketle atıf davranış grupları geliştirilmiştir. Bu gruplar; formalite gereği yapılan ve organik atıflar,⁶ kavramsal ve işlevsel atıflar,⁷ evrimsel ve yakınsal atıflar,⁸ onaylayıcı ve

⁶ Formalite gereği yapılan atıflarda çalışmaya atıf yapılmasına ihtiyaç yoktur ancak yapılmıştır. Organik atıflar ise gerçekten çalışmaya katkı sağladığı için yapılmıştır.

⁷ Kavramsal atıflarda atıf yapılan kaynağın ortaya çıkardığı kurama atıf yapılırken, işlevsel atıflarda atıf yapılan kaynağın kullandığı yöntem, teknik ve araçlara atıf yapılmaktadır.

⁸ Evrimsel atıflar bir önceki çalışmaya alternatif getirirken, yakınsal atıflar bir önceki çalışmayı destekler nitelikte verilmektedir.

tartışmacı atıflar⁹ olarak belirlenmiştir. Bir diğer çalışmada da Moravcsik ve Murugesan'ın genel görüşü geliştirilerek bilimsel etkinin kalitesinin atıf sayısı ile ölçülmesi için bazı değişkenler belirlenmiştir (Chubin ve Moitra, 1975). Bu çalışmanın sonucunda uzun makalelerin kısa makalelere göre iki kat daha fazla atıf aldıkları, mektup türü yayınların da daha uzun süre atıf aldıkları saptanmıştır. Tartışmacı atıfların ise yayımlanmalarının hemen ardından yüksek miktarda atıf aldıkları, ancak zaman içinde popülerliklerini yitirdikleri öne sürülmüştür. Moravcsik ve Murugesan'ın davranış modeli Cano tarafından anket aracılığı ile çözümlenmeye çalışılmış, yazarlar için en önemli atıf modelinin kavramsal ve işlevsel atıflar olduğu doğrulanmıştır (Cano, 1989).

Bazı çalışmalarda da atıflar fonksiyonel olarak gruplandırılmışlardır (Spiegel-Rosing, 1977). Yöntemsel fonksiyon ve genel fonksiyon (literatür arka planı gibi) olarak ayrılan bu gruplarda genellikle tercih edilen fonksiyonun genel fonksiyon olduğu savunulmuştur. Araştırma kapsamında incelenen dergide yayımlanan makalelerin yaptıkları atıfların %80'inin çalışma ile ilgili detay bilgi vermek amacı ile kullanıldığı saptanmıştır. Atıflar değerlendirildiğinde elde edilen diğer sonuçlara göre pozitif atıfların oranı %2,4 iken negatif atıfların oranının %0,4 olduğu; daha önceki çalışmaları doğrular atıfların oranı %0,3 iken, yanlışlar atıfların oranının %0,4 olduğu görülmüştür. Bu bulgular Garfield'in 1979'da öne sürdüğü tezini doğrular nitelikte olsa da, aynı zamanda nötr atıfların önemini ortaya koyması açısından da önemlidir. Nötr atıflara Moravcsik ve Murugesan tarafından kullanılan (1975) "formalite gereği atıf" yakıştırmasını yapmak mümkündür.

Yapılan bir diğer çalışmada da atıf analizlerinde her zaman eleştirilere maruz kalan yazarların kendine atıf yapma motivasyonları değerlendirilmiştir. Sonuç olarak kendine atıf ile diğer atıfların yapılma motivasyonu arasında anlamlı bir farklılık olmadığı sonucuna ulaşılmıştır. Tüm yazarlar hem kendine atıflarda hem de diğer atıflarında temel olarak konu ile alakalı literatürü tanımlamak amacı ile atıf yapmaktadırlar (Bonzi ve Snyder, 1991). Literatürde yer alan ve 1990'lı yılların ortalarına kadar yapılmış pek çok çalışmada atıf yapma nedenleri ve atıf sınıflamaları gerek anketler gerekse yayınların incelenmesi yolu ile saptanmaya çalışılmıştır. Ancak, genellikle elci sistemle yapılan bu çalışmaların pratik olmadığı da tüm çalışmalarda vurgulanmıştır (örn.: Cano, 1989; McCain ve Turner, 1989; Shadish, Tolliver, Gray ve Gupta, 1995; Vinkler, 1987).

⁹ Onaylayıcı atıflar bir önceki çalışmanın bulgularını desteklerken, tartışmacı atıflar eleştiri getirmektedir.

Yapılan bir çalışmada (Ding, Zhang, Chambers, Song, Wang ve Zhai, 2014, s. 1823) atıf analizleri ile ilgili ilk semantik çalışmaların 1950 sonlarındaki atıf yapma motivasyonları ile ilgili ilk çalışmalara dayandığını belirtilmiştir. Çünkü atıf yapma motivasyonları atıflara yüklenen anlamları doğrudan etkilemektedir.

2006 yılında yapılan bir çalışmada ise atıflar için makine öğrenmesi temelli bir sınıflama şeması geliştirilmesi gerekliliği üzerinde durulmuş ve dört temel atıf fonksiyonu üzerinden bir uygulama gerçekleştirilmiştir (Teufel, Siddharthan ve Tidhar, 2006). Bu dört atıf fonksiyonu; atıf yapılan kaynağın zayıflığının açık ifadesi, diğer çalışmalarla iyi ya da kötü yönde karşılaştırma, diğer çalışmaları mevcut çalışmaya uyarlama ya da kullanma ile nötr kategoridir. Atıf fonksiyonlarını ortaya çıkarmak amacı ile makine öğrenmesi süreçlerinde sık kullanılan duygu analizi (sentiment analysis) tekniğini uygulamışlardır. Duygu analizi çalışmalarının temeli herhangi özel bir durum karşısında takınılan olumlu ve olumsuz tutumların büyük yoğunluklu metinler içinden doğal dil işleme aracılığı ile tespit edilmesine dayanır (Nasukawa ve Yi, 2003, s. 70). Literatürde duygu analizi çalışmaları son yıllarda sosyal ağlar üzerinde yoğunlaşsa da, atıfların özelliklerinin saptanması için de uygulanmaktadır. Bir başka duygu analizi çalışmasında da anahtar sözcükler analiz edilerek içerik tabanlı atıf analizi çalışması yürütülmüş, bu aşamada bibliyometrik yaklaşımlar ile doğal dil işleme teknikleri birlikte kullanılmıştır (Small, 2011).

Astronomi makalelerine yönelik gerçekleştirilen semantik temelli bir içerik analizi çalışmasında, bu gibi araştırmalarda en temel problemin atıfın makinece saptanması olduğu belirtilmiş ve bu gibi çalışmalara başlanmadan önce atıfların anlamlandırılmasını ve bağlantılandırılmasını sağlayacak bir içerik analizi yapılması gerekliliği vurgulanmıştır (Kim ve Weber, 2006). Ding ve arkadaşları da içerik tabanlı atıf analizinde yarı otomatik bir yaklaşım söz konusu olduğunu ve atıf yapılan yayınların mevcut yayın ile ilişkisini anlayabilmek için içerik analizinin mutlaka yapılması gerektiğini savunmuşlardır (Ding, Zhang, Chambers, Song, Wang ve Zhai, 2014, s. 1829-1830). İçerik analizinin ardından yapılacak doğal dil işleme çalışmaları daha etkili sonuç verecektir. Kim ve Weber'in çalışmasında yapılan içerik analizi sonrasında saptadıkları en belirgin özellik atıf yapıldığında "they", ve "those" gibi çoğulluk bildiren sözcüklerin kullanılmasıdır. Yazarlar bu sayede atıfların çıkarımını yapmışlar ve bu çıkarımın %96 oranında tutarlı davranmasını sağlamışlardır. Siddharthan ve Teufel (2007) ise bu yaklaşımın zayıflıkları üzerinde durarak daha iyi bir sistemin geliştirilmesi gerektiğini savunmuşlardır. Bu bağlamda bilimsel nitelikleri saptamak üzere iki temel ayırım belirlemişlerdir. Bunlar artgönderim (anaphoric reference) ve altbölüm (subpart)'dür. Artgönderimde makalenin

tamamın atıf yapılırken, diğerinde makalenin sadece bir bölümüne atıf yapılmaktadır. Hazırladıkları makine öğrenmesi modeli bu iki ayrıma göre referansları ağırlıklandırmakta ve ortalama olarak %70 oranında doğru sınıflandırmaktadır.

Psikoloji yayınları üzerinden yürütülen bir çalışmada içerik tabanlı atıf analizlerinin en önemli probleminin en fazla yapılan atıf türünün formalite gereği yapılan atıflar olmasından hareketle niteliksel değerlendirmenin zorlaşması olduğu belirtilmiştir (Krampen, Becker, Wahner ve Montada, 2007). Örneğin, psikoloji alanında yapılan atıfların üçte birinin yayınlara hiçbir katkısı olmayan formalite gereği yapılan atıflar olduğu saptanmıştır. Literatürdeki çalışmalar da en çok yapılan atıf türünün formalite gereği yapılan atıflar olduğunu doğrulamaktadır. Çalışmaların hemen hemen hepsinde formalite gereği atıfların her zaman %40'tan fazla oranlarda olduğu belirtilmektedir (Liu, 1993, s. 372).

Literatürde yer alan güncel bir çalışmada da atıflar taksonomik olarak sınıflandırılmış ve Tablo 1'deki gibi bir sınıflama modeli ortaya çıkarılmıştır (Erikson ve Erlandson, 2014, s. 628).

Tablo 1. Atıf sınıflama modeli örneği

Kanıtlama	Sosyal işbirliği	Ticari işbirliği	Veri
Sınırlandırma	Bilimsel gelenek	Kredi verme	Gözden geçirme
Etkin destek (doğrulama)	Bilimsel öz-ımar	Şahsi tanıtım	Meta analiz
Edilgen destek	Emeklerin telafisi	Atıf takası	Metin çalışması
İlave okuma		Kişisel reklâm (kendine atıf)	
		Atıf rehin verme (pledging)	

Tablo 1'de gösterilen taksonomik sınıflamada atıflar dört temel sınıfta değerlendirilmiş ve her bir sınıf kendi içinde alt sınıflara ayrılmıştır. Bu modelin önemli bir özelliği atıflara yönelik olarak günümüzde gerçekleştirilen etik dışı uygulamaların da sözü edilen sınıflarda yer bulmuş olmasıdır. Örneğin, "ben x'e atıf yapayım, x'te bana yapsın" mantığı ile yapılmış atıflar atıf takası sınıfı altında yer bulmuştur. Literatürdeki en güncel ve farklı sınıflandırmanın Erikson ve Erlandson tarafından yapılan bu sınıflandırma olduğu açıkça görülmektedir.

2.3.2.2 Sentaktik (Sözdizimsel) Çalışmalar

Bir bilimsel yayının giriş, sonuç, bulgular gibi belirli bölümleri vardır ve her bir disiplin için bu bölümler bazı istisnalar hariç benzerlik göstermektedir. Bu bölümlerin ayrılması sayesinde yapılan atıfların algılanan kullanılabilirlik düzeyi hesaplanabilmektedir (Ding, Zhang, Chambers, Song, Wang ve Zhai, 2014, s. 1823). Literatürdeki sentaktik atıf analizi çalışmalarının ilki 1976 yılında yayımlanmıştır (Voos ve Dagaev, 1976). Bu çalışmada merak edilen unsur, atıflara çalışmaların hangi bölümlerinde kaç defa yer verildiğinin hesaplanması ve hangi bölümlerde yapılan atıfların daha anlamlı olduğunun

saptanmasıdır. Çünkü yazarların çıkarımına göre bir yayından birden fazla kez atıf alan makalelerin kalitesinin diğerlerine göre daha yüksek olması söz konusu olabilir. Çalışma sonunda atıfların makalelerdeki yerinin (giriş, sonuç, bulgular gibi) ve kaç kez atıf yapıldığının hesaplanması ile atıfın kalitesinin hesaplanmasının gerçekleştirilebileceği, ancak bunun elci sistemle yapılması nedeni ile uzun zaman alacağı vurgulanmıştır. Bunun yanında çalışma sonunda elde edilen bulguların genele yorumlanamayacağı, ancak gelecekte yapılacak çalışmalar için öngörü sağlayacağı konusunda fikir belirtilmiştir.

Literatürdeki bir diğer çalışmada da bir makalede bir yayına birden fazla kez atıf yapılmışsa, bunun iki makalenin konusal olarak yakın ilişki içinde olduğunu gösterdiği belirtilmiştir (Herlach, 1978, s. 310). Sosyal ve ilgili diğer bilimlerde atıfların rolünü saptamak ve bir sınıflama şeması yaratmak amacı ile yürütülmüş bir çalışmada içinde kütüphanecilik alanının da bulunduğu sekiz kategoride atıfların yeri gözetilerek bir sınıflama sistemi oluşturulmuştur (Peritz, 1983). Çalışma tarafından saptanan en temel nokta disiplinler arasında atıf nitelikleri konusunda farklılıklar olduğudur. Bunun yanında Peritz, makalelerde geçen atıfların ağırlıklandırılması gerekliliğine vurgu yaparak; bir makalede bir yayına birden fazla kez atıf yapılıyorsa bu atıfın diğer atıflardan ayrı tutulması gerektiğine işaret etmiştir (Peritz, 1983, s. 310). Aynı makalede üzerinde durulan bir diğer nokta da atıfların etiketlenmesinin gerekliliği iken, bunu gerçekleştirecek bir bilgi sisteminin o yıllarda henüz tasarlanmadığı belirtilmiştir (Peritz, 1983, s. 303). Sentaktik dizilim ile ilgili bir diğer yayında (Maričić, Spaventi, Pavičić ve Pifat-Mrzljak, 1998, s. 530-540) atıfların değeri için derecelendirme yapılması hedeflenmiştir. Çalışma sonunda elde edilen bulgulara göre çalışmaların bulgular ve yöntem kısımlarında yapılan atıfların en değerli atıflar olduğu belirtilmiştir. Her bir bölümde yapılan atıflara yaptıkları içerik analizi sonuçlarına göre puan veren yazarlar; giriş kısmında yapılan atıflara genel bir bakış yaptıkları ve çoğunda sadece anıldıkları için 15, yöntem ve bulgular kısımlarında yapılan atıflara genellikle araştırmanın temeli bu çalışmalara dayandırıldığı ve genellikle pozitif bir yaklaşım sergilendiği için 30, tartışma ve yorum kısmında yapılan atıflara da iki çalışma arasında bir karşılaştırma yapıldığı için 25 puan verilmesi gerektiğini öngörmüşlerdir. Bu bulguları destekleyen bir başka çalışmada da giriş ve literatür değerlendirme kısmında yapılan atıfların hiçbir şekilde kalite göstergesi olamayacağı vurgulanmıştır (Adedayo, 2015, s. 305).

2.4. TÜRKİYE'DE ATIF DEĞERLENDİRME ÇALIŞMALARI

Türkçe literatürde atıfların sınıflandırılması ya da niteliksel olarak değerlendirilmesine yönelik çok az sayıda yayın yapılmıştır. Bu yayınlardan ilki Türkçe kütüphanecilik ve bilgilendirme literatüründe yer alan dergilerin atıf analizinin yapıldığı bir çalışmada gerçekleştirilen atıf sınıflandırmasıdır (Atılgan, Atakan ve Bulut, 2008, s. 399). Türkçe kütüphanecilik ve bilgilendirme atıflarını iki gruba ayıran çalışmada eğer atıf alıntı ya da bir yargıyı desteklemek amacı ile kullanılmışsa kanıt gösterme; kaynağa yönlendirme şeklinde yapılmışsa yönlendirme atıfı olarak sınıflandırılmıştır. Sözü geçen çalışmada herhangi bir makine öğrenmesi tekniği kullanılmamış, atıflar tek tek değerlendirilerek sınıflama yapılmıştır.

Niteliksel atıf göstergeleri ile ilgili yapılmış tek yayında ise negatif, pozitif ve nötr atıflar dikkate alınarak geliştirilecek yeni bir h-endeksinin gerekliliğinden bahsedilmiştir (Tunç, 2012). Bu çalışmada pozitif, negatif ve nötr atıfların ayırt edilebilmesi için yazarlar tarafından işaretleme yapılmasının gerekliliğinden söz edilmektedir. Çalışmaya göre yazarlar verdikleri pozitif atıfın başına artı (+ Taşkın, 2015 gibi), negatif atıfın başına eksi işareti koyarak (- Taşkın, 2015 gibi) ve nötr atıfta da herhangi bir işaretleme yapmayarak atıfların niteliklerini kendileri belirlemelidirler. Ancak yazarların bazı kişisel ya da farklı sebeplerle atıfları işaretlemek istememeleri, bilerek hatalı işaretleme yapmaları, geçmiş yayınların işaretlenmesi ile ilgili sorunlar, çözümün pratik bir yol sunmaması ve atıf gelenekleri gibi çeşitli sebeplerle bu yöntem uygulanabilir görünmemektedir. Benzer bir öneri 1978 yılında uluslararası literatürde de tartışılmış ancak kullanışlı bir yöntem olmadığı vurgulanmıştır (Small, 1978, s. 329).

Yukarıda sayılan iki çalışma dışında Türkçe için semantik veya sentaktik içerik tabanlı atıf analizi çalışması literatürde yer bulmamıştır. Bu çalışmada Türkçe kütüphanecilik ve bilgilendirme literatürü için hem semantik hem de sentaktik yapıları temel alan bir atıf analizi modeli tasarlanması amaçlanmaktadır. İçerik tabanlı atıf analizleri ile kütüphanecilik ve bilgilendirme literatürünün özellikleri birebir örtüşmektedir. Yapılan bir çalışmada veri türü, miktarı, kullanılan yöntemler ve maliyet açısından içerik tabanlı atıf analizinin kütüphanecilik ve bilgilendirme alanına hız ve kullanışlılık açısından önemli katkılar sağlayacağı aktarılmaktadır (Zang, Ding ve Milojević, 2012, s. 1494). Çalışmamızda geliştirilen içerik tabanlı atıf analizi modeli ile hem zamandan tasarruf sağlanması, hem de performans değerlendirmelerinin etkili bir şekilde yapılabilmesine yönelik öneriler geliştirilebilecektir.

3. BÖLÜM: DOĞAL DİL İŞLEMEDE METİN KATEGORİZASYONU

3.1. GİRİŞ

Doğal dil işlemede kullanılan pek çok farklı yöntem ve teknik (metin kategorizasyonu, özetleme, makine çevirisi gibi) bulunmaktadır. Bu çalışma kapsamında yürütülen içerik tabanlı atıf analizi uygulaması için yöntem metin kategorizasyonu olarak belirlenmiştir. Bu bölümde doğal dil işlemede metin kategorizasyonunun kullanımı, çeşitli algoritmalar ve farklı kategorizasyon teknikleri tüm detayları ile sunulmaktadır.

3.2. DOĞAL DİL İŞLEME

Doğal dil işleme doğal metinler, konuşmalar ya da kullanışlı diğer malzemelerin bilgisayarlar tarafından nasıl kullanılacağına ilişkin anlaşılması ve bilgisayarlara bunların nasıl işletileceğinin öğretilmesidir (Chowdhury, 2003, s. 51). Bir dizi görev veya uygulama için bir insanın kullandığı dile en çok benzeyen dil işleme başarısına ulaşmak ve bu sayede doğal dille oluşturulmuş metinleri bilgisayarlı tekniklerle analiz edebilmek doğal dil işleminin temel amaçlarından biridir (Liddy, 2010, s. 3864).

Çeşitli çalışmalarda doğal dil işlemeyi anlayabilmek için yedi bağımsız düzey belirlenmiştir (Feldman, 1999, s. 62-64; Liddy, 2010, s. 3867-3868). Bu düzeyler aşağıda açıklanmaktadır:

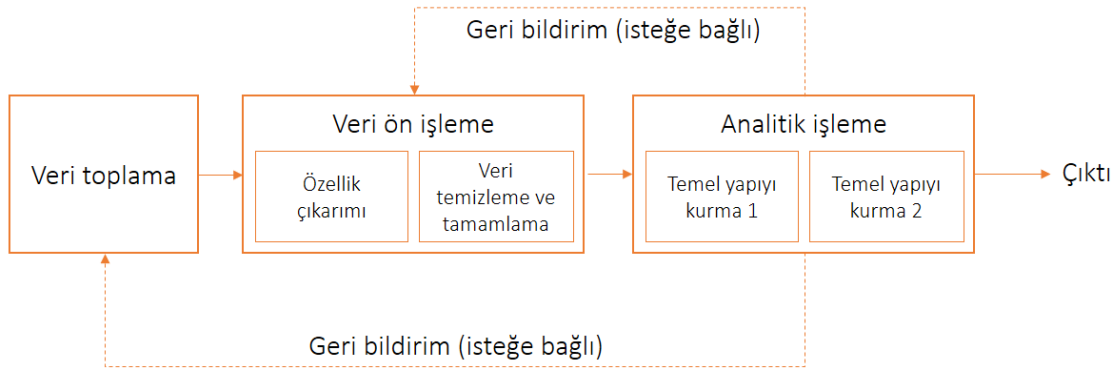
- *Fonetik veya fonolojik düzey:* Bu düzeyde yazılı metinlerden ziyade konuşma dilinin veya telaffuzun anlaşılması üzerine çalışmalar yapılmaktadır. Türkçe dâhil pek çok dilde konuşma dilinin tanımlanması amacı ile çeşitli çalışmalar literatürde yer bulmuştur (Çarkı, Geutner ve Schultz, 2000; Schultz ve Waibel, 2001; Shen, Wu ve Tsai, 2015).
- *Biçimbilimsel (morfolojik) düzey:* Bu düzey kelimelerin morfemlerden (biçimbirim) oluşan çok bileşenli doğasını ortaya çıkarmakla uğraşır. Morfemler sözcüğün anlamının en küçük parçasıdır. Örneğin, sondan eklemeli bir dil olan Türkçe'de morfolojik analizi yapmak üzere çeşitli araçlar geliştirilmiştir (Eryiğit, 2014; Zemberek NLP, 2015). Bu araçlar sayesinde Türkçe yazılmış sözcükler en küçük birimleri olan köklerine ayrılabilen; kökler ve eklerin dilbilimsel özellikleri çıkarılabilmektedir.
- *Sözcüksel (lexical) düzey:* Kelime tabanlı anlamayı temel alan düzeydir. Anlamsal düzeyin en basit safhasını temsil eden bu düzeyde kelimeler işaretçilere (token) ayrılır. Sözcüksel analizlerde bir sözlük (lexicon) kullanmak zorunludur.

- *Sözdizimsel (sentaktik) düzey:* Bu düzeyde dilin özellikleri göz önüne alınarak kelimelerin cümle içindeki yerleri belirlenir. Örneğin, Türkçe için özne, tümleç, yüklem, nesne vb. öğelerin cümle içinde dizilimini ortaya çıkaran araştırmalar sözdizimsel çalışmalara örnektir (Arısoy, Saraçlar, Roark ve Shafran, 2010).
- *Anlamsal (semantik) düzey:* Bu düzeyde kelimelerin sözlük anlamlarının yanı sıra metnin bağlamı içinde söz konusu kelimelerin hangi anlamlarda kullanıldığını da saptamak hedeflenir. Bir kelimenin birden çok anlamı olabileceğinden ötürü bu düzeyin doğal dili işleme ve bu dili makine tarafından anlaşılabilir hale getirmede zor süreçlerden biri olduğu söylenebilir.
- *Söylemsel (discourse) düzey:* Semantik ve sentaktik düzeylerin birlikte kullanıldığı düzeydir. Son yıllarda genellikle büyük miktarda veri içeren haber metinlerini ya da tweetleri analiz etmekte kullanılan bu düzeyde, temel amaç kelime ve cümlelerin hem semantik hem de sentaktik olarak incelenmesi ve özelliklerinin ortaya koyulmasıdır.
- *Yararcı (pragmatic) düzey:* Bu düzey vukuf (knowledge) anlamındaki bilginin bilgi erişim sistemlerine tanımlanması ve bu sayede doğal dilin çok daha detaylı olarak anlaşılması üzerine kuruludur. Örneğin, “The city councilors refused the demonstrators a permit because they feared violence” haber cümlesinde geçen “they” sözcüğünün anlaşılabilmesi için bilgi sistemine ekstra bilgi sağlanması zorunluluğu bulunmaktadır (Liddy, 2010, s. 3868). Bu gibi ekstra bilgi gerektiren doğal dil metinlerinin işlenmesi oldukça zor olduğundan, yararcı düzeyin en karmaşık doğal dil işleme düzeyi olduğunu söylemek yanlış olmayacaktır.

Yapılacak herhangi bir doğal dil işleme çalışmasında yukarıda sıralanan bu düzeylerden hepsinin ya da birkaçının araştırılması zorunludur. Her bir düzey bir önceki düzeyin devamı niteliğindedir. Yani, en basit doğal dil işleme çalışmaları fonetik düzeyde yapılırken, en detaylı çalışmalar yararcı düzeyde gerçekleştirilmektedir.

3.2.1. Doğal Dil İşleme Süreçleri

Doğal dil işlemede bir veri setinin işlenmesi ve analiz edilebilmesi için belirli aşamaların dikkatle uygulanması gerekir. Bu aşamalar sırasıyla veri toplama, veri ön işleme ve analitik işlemdir. Bu işlemlerin birbirleri ile ilişkisi Şekil 1’de sunulmaktadır (Aggarwal, 2015, s. 4).



Şekil 1. Veri işleme süreci

Doğal dil işleme sürecinde başlangıçta problemin doğru bir şekilde tanımlanması, verinin anlaşılması, verilerin hazırlanması (veri entegrasyonu, veri temizleme, veri dönüştürme, veri indirgeme gibi), modelin kurulması ve son olarak kullanılacak yazılım ve algoritmaların belirlenmesi büyük önem taşımaktadır (Akpınar, 2014, s. 77-80).

3.2.1.1. Veri Toplama

Doğal dil işlemede en önemli aşamalardan biri en doğru verilerin etkili bir şekilde toplanabilmesidir. Bunun temel nedeni en başta yapılacak iyi seçimlerin analizin başarımını olumlu yönde etkileyebilecek olmasıdır. Bu nedenle de veri analizine geçilmeden önce araştırma kapsamında yapılacak çalışmalar titizlikle listelenmeli ve analizi yapılacak veriler doğru bir şekilde toplanmalıdır. Eğer büyük veri ile uğraşılıyorsa veri toplama amacı ile özelleştirilmiş bir donanım ya da yazılım kullanılarak verilerin toplanmasını kolaylaştırmak mümkündür. Veri toplama sürecinde toplanan verilerin bir veri tabanı ya da veri ambarında sistematik bir şekilde saklanması önemlidir (Aggarwal, 2015, s. 3).

3.2.1.2. Veri Ön İşleme

Doğal dil işleminin ilerleyen safhalarında model kurma esnasında ortaya çıkacak sorunlar veri ön işleme aşamasına sık sık geri dönülmesine ve verinin yeniden düzenlenmesine sebep olacaktır (Akpınar, 2014, s. 77). Bu nedenle veri ön işleme sürecinde işin özenli bir şekilde yapılması önemlidir. Bu bağlamda literatürde sıklıkla kullanılan ve girdi kalitesi ile çıktı kalitesi arasındaki ilişkiyi tanımlayan “çöpler içeri, çöpler dışarı (garbage in, garbage out)” ifadesinin veri ön işleme aşamasında en temel ilkelere biri olduğunu söylemek yanlış olmayacaktır (Lee, Lu, Ling ve Ko, 1999, s. 751).

Veri ön işleme aşamasına gelmiş verilerin belirli bir kalite düzeyinde olması yapılacak işlemeyi kolaylaştıracaktır. Bu kalite düzeyleri; anlaşılabilirlik, değer katabilirlik, doğruluk, erişilebilirlik, güncellik, güvenlik, inanılabilirlik, nesnellik, tamlık, uygunluk ve yorumlanabilirliktir (Batini, Cappiello, Francalanci ve Maurino, 2009, s. 18). Tüm bu kalite düzeyleri yakalandığında veri ön işlemenin başarımı da artacaktır. Belirli bir kalite düzeyinde olan verilerin veri ön işleme süreçleri Şekil 2'de gösterildiği gibidir (Akpınar, 2014, s. 89).



Şekil 2. Veri ön işleme aşamaları

Veri ön işleme aşamalarının tamamlanmasının ardından veriye uygun doğal dil işleme yönteminin seçilerek analiz safhasına geçilmesi önemlidir. Doğal dil işleme türleri hakkında genel bilgi ilerleyen bölümlerde aktarılmaktadır.

3.2.2. Doğal Dil İşleme Uygulamaları

Doğal dilde oluşturulmuş metinlerin işlenmesinin çeşitli amaçları olduğundan her bir amaca uygun olarak birçok uygulama geliştirilmiştir. Literatürde en sık karşılaşılan doğal dil işleme uygulamaları ve bu uygulamaların temel mantığı şöyledir:

- **Bilgi erişim (information retrieval):** Son yıllarda hızla artan bilgi miktarı ve bu bilginin içinden anlamlı olanlarına erişmeye yönelik ihtiyaç ile birlikte doğal dil işlemenin en önemli uygulamalarından biri anlamlı bilgiye erişim olmuştur. Çeşitli çalışmalarda “document retrieval (belge erişim)” ya da “text retrieval (metin

- erişim)” şeklinde tanımlanan bu uygulamada temel amaç, insanların herhangi bir paragrafın, kitabın ya da büyük yoğunluklu metnin içinden ihtiyaç duydukları herhangi bir kısma erişimlerinin sağlanabilmesidir (Lewis ve Jones, 1996, s. 92).
- *Bilgi çıkarımı (information extraction)*: Bilgi çıkarımı uygulaması büyük yoğunluklu bir metnin içinden temel anahtar unsurların (kişiler, kurumlar, konular, ülkeler gibi) tanımlanması, etiketlenmesi ve çıkarımına dayanır (Liddy, 2010, s. 3871). Bu uygulama yürütülecek doğal dil işleme çalışmasının temelini de oluşturabilir. Zira bazı çalışmalarda metnin içindeki bilginin çıkarımı sonuca ulaştırırken, bazı çalışmalarda bu bilginin çıkarımı sayesinde elde edilen veriler işlenmektedir. Bu nedenle bilgi çıkarımı diğer uygulamalar ile beraber de kullanılabilir (Blake, 2013, s. 129).
 - *Makine çevirisi (machine translation)*: Metinlerin ya da konuşmaların bir dilden diğer dile otomatik olarak çevrilmesini temel alan bu uygulamada farklı kültürlerin kaynaşması ya da eski dillerin yaşatılması amaçlanmaktadır (Manning ve Schütze, 1999, s. 463).
 - *Özetleme (summarization)*: Özetleme sistemleri büyük yoğunluklu metinlerde yer alan cümle ya da paragraflar içinden dilsel veya istatistiksel bazı yöntemleri kullanarak en önemli sözcükleri ya da cümleleri seçmeye ve metni temsil eden anlamlı bir özet çıkarmaya dayanır (Chowdhury, 2003, s. 60).
 - *Metin kategorizasyonu (text categorization - text classification)*: Metin kategorizasyonu uygulamaları resim veya örüntü tanıma, hava durumu tahmini, hastalık tanılama gibi çeşitli amaçlarla kullanılan tahminleyici bir modeldir (Silahtaroglu, 2013, s. 67).

3.3. METİN KATEGORİZASYONU

Sınıflama (kategorizasyon) algoritmaları metinlerin temel özellikleri dikkate alınarak çeşitli sınıflar yaratılmasına dayanır (Blake, 2013, s. 136). Veri sınıflamada temel iki çeşit yöntem bulunmaktadır. Bu yöntemler şunlardır:

- *Gözetimsiz metin kategorizasyonu (unsupervised text categorization)*: Gözetimsiz sınıflamada etiketlenmemiş veya sınıflanmamış veriler üzerinden bilinmeyen bir yapıyı ortaya çıkarmak hedeflenir (Akba, 2014, s. 9). Bu nedenle gözetimsiz metin kategorizasyonu bilgisayarlı tekniklerle veya çeşitli yazılımlarla sınıflama yapılmasına dayanır (Extension, 2013). Uzman desteği olmadan yalnızca bilgisayarlı tekniklerle sınıflama yapılması dilin özelliklerinin ortaya

çıkarılmasındaki güçlüklerden dolayı gözetimli kategorizasyona göre çoğunlukla düşük başarımlara ulaşmaktadır (Aggarwal, 2015, s. 285).

- *Gözetimli metin kategorizasyonu (supervised text categorization)*: Uzman desteği alınarak doğru sınıfların belirlendiği yönteme gözetimli kategorizasyon adı verilmektedir (Blake, 2013, s. 136). Önceden değerleri ve sınıfları bilinen veriler sayesinde geliştirilen yöntemin testinin yapılması ve analizin başarısının ortaya koyulması kolaylaşmaktadır (Akba, 2014, s. 9; Love, 2002, s. 829).

3.3.1. Metin Kategorizasyonunda Verilerin Hazırlanması

Genellikle tüm sınıflama algoritmalarında veriler eğitim (training) ve test olarak adlandırılan iki setten oluşur (Aggarwal, 2015, s. 286). Eğitim setinde her bir kayıt niteliklerin bir kümesini kapsar. Yani eğitim seti ile bir model kurulur. Test setinde amaç eğitim seti ile kurulan modelin doğrulanmasıdır. Bu nedenle yapılacak metin kategorizasyonu çalışmalarında iki veri seti ile çalışılması önemlidir.

Diğer taraftan 3.2.1 şeklinde numaralandırılan bölümde aktarılan veri toplama, ön işleme ve analize hazır hale getirme aşamalarının tamamının metin kategorizasyonu çalışmalarında da dikkatle sürdürülmesi çalışmanın başarımının artırılması açısından büyük önem taşımaktadır.

3.3.2. Metin Kategorizasyonunda Teknik ve Algoritmalar

Metin kategorizasyonu konusunda literatürde çok sayıda teknik ve algoritma bulunmaktadır. En çok kullanılan teknik ve algoritmalar Tablo 2'de gösterildiği gibidir (Aggarwal, 2015, s. 285-343; Silahtaroğlu, 2013, s. 67-136).

Tablo 2'de sıralanan teknik ve algoritmalar dışında da çeşitli algoritmalar (lazy, misc, meta gibi) bulunmaktadır. Metin kategorizasyonu sonunda elde edilecek başarıyı artırmak için bu algoritmalarından en uygun olanları veriye uygulanmalı ve en doğru algoritma kullanılmalıdır. Tüm algoritmalar matematiksel işlemler içerseler de hesaplamaları otomatik yapmayı sağlayacak yazılımlar da geliştirilmiştir. Bu sayede tüm hesaplamalar saniyeler içinde yapılabilmektedir.

Tablo 2. Metin kategorizasyonunda kullanılan teknik ve algoritmalar

Teknik	Algoritmalar
<p>Karar ağaçları: Sınıflandırma için bir ağaç oluşturulur (kök). Daha sonra her bir kayıt bu ağaca uygulanır (yaprak/düğüm). Çıkan sonuca göre kayıtlar sınıflandırılır.</p>	<p>ID3: Entropiden (bilgilerin sayısallaştırılması) yararlanarak ağaçta yer alan değişkenler içinde en ayırt edici olanın bulunması için kullanılır. C4.5 ve C5: ID3 algoritması kayıp verileri hesaba katmaz. C4.5 ve C5 algoritmalarında ise kayıp veriler diğer veri ve değişkenler yardımı ile öngörülür ve bu sayede daha duyarlı ve daha anlamlı kurallar çıkarılabilir. CART: Karar ağacını dallara ayırmada kök ve düğümlere karar verilmesinin yanında belirlenen düğümlerin hangi noktasında düğümün ikiye ayrılacağı da CART algoritmaları ile hesaplanabilir. SLIQ: Hem sayısal hem de kategorik verilerin sınıflandırılmasında kullanılır. Önceki üç algoritma önce derinlik ilkesine göre çalışırken SLIQ önce genişlik ilkesine göre çalışır. Yani aynı anda birden çok yaprağı oluşturabilir. SPRINT: Yapısal olarak SLIQ algoritmasına benzese de, SPRINT algoritmasında her bir değişken için ayrı bir değişken listesi hazırlanır. Bu sayede veri tabanındaki değişken sayısı kadar tablo oluşturulur. Elde edilen ilk listeler de ağacın kökü ile ilişkilendirilir.</p>
<p>İstatistiğe dayalı algoritmalar: Veri işlemede verilerin önceden sınıflara ayrılması aslında gelecekte elde edilen sonuçların tahmini, yani sınıfların tahminidir. Bu nedenle istatistiksel yöntemler kullanılarak sınıflama yapılması mümkündür.</p>	<p>Bayesyen: Hâlihazırda sınıflanmış verilerin özellikleri dikkate alınarak yeni bir verinin mevcut sınıflardan hangisine girme olasılığının olduğunu hesaplayan bir yöntemdir.</p> <p>Regresyon: Herhangi bir değişkenin ile bir veya daha fazla başka değişkenler arasındaki ilişkinin matematiksel bir denklem şeklinde yazılmasıdır. Öncelikle verileri sınıflara bağlı olarak çeşitli bölgelere ayırır, ardından tahmin için formüller üretir.</p> <p>CHAID: Ki-kare testini kullanarak veriler üzerinden bir karar ağacı oluşturan algoritmadır. Bu bakımdan hem istatistiksel teknikleri kullanır, hem de karar ağacı yaratılır.</p>
<p>Mesafeye dayalı algoritmalar: Eldeki verilerin birbirlerine olan uzaklığı veya benzerliği kullanılarak sınıflandırma yapılır.</p>	<p>K- En yakın komşu: Sınıflandırma yapılırken veri tabanındaki her bir kaydın diğer kayıtlara uzaklığı hesaplanır. Bu yöntem sıklıkla coğrafi bilgi sistemlerinin tasarımında ve yapılan hesaplamalarda kullanılır.</p> <p>En küçük mesafe sınıflandırıcısı: Veri tabanındaki her bir sınıfa ait noktanın merkezi ile tahmin edilmesi gereken veri arasındaki mesafe göz önüne alınarak işlem yapılır.</p>
<p>Yapay sinir ağları: Biyolojik sinir ağlarından esinlenerek geliştirilmiş sistem, yapay sinir hücrelerinin birbirleriyle çeşitli şekilde bağlanmasından oluşur ve genel olarak katmanlar şeklinde düzenlenir. Üç temel katman vardır (girdi, gizli ve çıktı).</p>	<p>İleri sürümlü yapay sinir ağları: Girdiler ya da girdileri temsil eden vektörler gizli katmana ulaşmakla yükümlüdür.</p> <p>Hata geriye yayma yöntemi: Doğrusal olmayan sistemlerin tanımlanması, hareketli görüntü analizleri ve ses tanıma programlarında kullanılan ve eğitim düşüm yöntemine dayanan bir yöntemdir.</p>

3.3.3. Metin Kategorizasyonunda Performans Değerlendirme

Doğal dil işlemede verilerin hazırlanması, işlenmesi, sınıflanması ve çeşitli algoritmalar aracılığı ile test edilmesinin ardından yapılacak işlem sınıflamanın etkinliğinin test edilmesidir. Bunun için çeşitli yöntemler bulunmaktadır. Metin kategorizasyonunda başarıyı test etmek üzere metodolojik değerlendirme ve sayısal değerlendirme şeklinde iki çeşit yöntem bulunmaktadır (Aggarwal, 2015, s. 334).

- *Metodolojik değerlendirme*: Bu yöntem verilerin test ve eğitim seti olarak ikiye bölünmesi ile otomatik olarak gerçekleştirilen değerlendirme sürecidir. Eğitim seti üzerinden tanımlanan modelin test setine uygulanması ile kendi içinde tutarlı bir değerlendirme yapılmış olur. Çapraz doğrulama, verileri 10'ar veya 20'şerli gruplara bölme ya da veriyi %66-%33'lük bölümlere ayırmak gibi çeşitli değerlendirme yaklaşımları bulunmaktadır.
- *Sayısal değerlendirme*: Metin kategorizasyonunun başarımının sayısal olarak ölçümü için çeşitli yöntemler ve terimler geliştirilmiştir. En bilinen değerlendirme ölçütleri anma (precision), duyarlılık (recall) ve anma ve duyarlılık skorları kullanılarak hesaplanan *f*-ölçütüdür (*f*-measure). Anma değeri erişilen ilgili belge sayısının derlemdeki toplam ilgili belge sayısına oranını gösterirken, duyarlılık erişilen ilgili belge sayısının erişilen toplam belge sayısına oranını gösterir (Van Rijsbergen, 1979, s. 113-114).

Anma, duyarlılık ve *f*-ölçütü ile ilgili formülleri metin kategorizasyonuna uyarlayıp Tablo 3'te yer alan karışıklık matrisi doğrultusunda örneklendirmek gerekirse (Biricik, 2011, s. 25-26):

Tablo 3. Tahmin karışıklık matrisi

		Gerçek kategori	
		Kategori 1	Kategori 2
Tahmin edilen kategori	Kategori 1	Doğru pozitif (TP)	Yanlış pozitif (FP)
	Kategori 2	Yanlış negatif (FN)	Doğru negatif (TN)

Bir veri kümesinde gerçekte pozitif sınıfta yer alıp, sınıflandırmada da pozitif sınıfta yer alan örneğe doğru pozitif (TP); gerçekte negatif sınıfa ait olup sınıflandırma sonucunda da negatif sınıfta yer alan örneğe doğru negatif (TN) adı verilir. Gerçekte negatif sınıfta yer almasına rağmen pozitif tahmin edilen örnek yanlış pozitif (FP); pozitif sınıfta olmasına rağmen negatif tahmin edilen örnek ise yanlış negatif (FN) olarak tanımlanabilir. Bu bağlamda anma, duyarlılık ve *f*-ölçütü formülleri şöyledir:

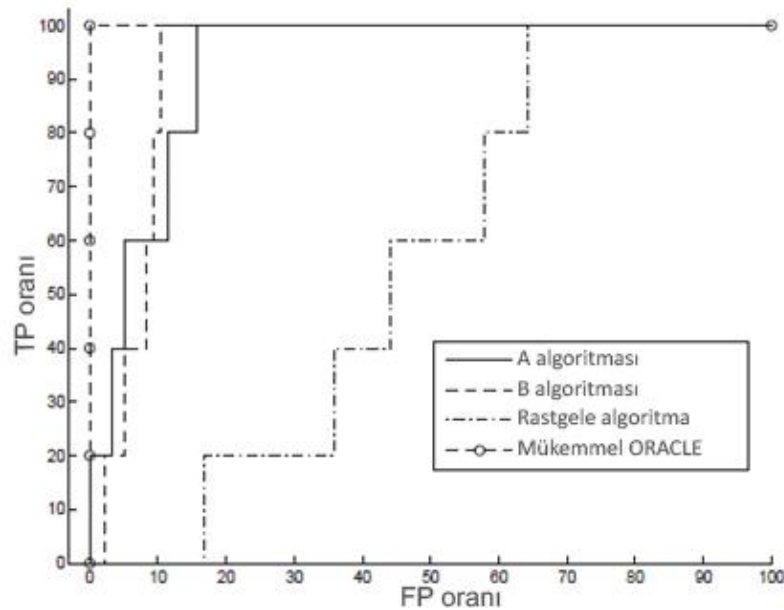
$$\text{Anma} = \frac{TP}{TP+FN} \quad (\text{Formül 1})$$

$$\text{Duyarlılık} = \frac{TP}{TP+FP} \quad (\text{Formül 2})$$

$$f = \frac{2 \times \text{anma} \times \text{duyarlılık}}{(\text{anma} + \text{duyarlılık})} \quad (\text{Formül 3})$$

Literatürde f -ölçütünün hesaplanması için çeşitli formüller bulunmaktadır (f_a , f_2 gibi). Metin kategorizasyonu çalışmalarında anma ve duyarlılık ölçümlerinin ağırlıklı harmonik ortalamasını veren f_1 -ölçütü formülü sıklıkla kullanılmaktadır. Bunun temel sebebi ise kategoriler arası dağılımı homojen olmayan ya da belirli bir kategoride çok az sayıda örnek barındıran veri setlerinde başarımın en iyi f_1 ölçütü ile hesaplanabiliyor olmasıdır (Biricik, 2011, s. 26). Anma ve duyarlılık değerlerinin harmonik ortalamalarının alınmasıyla yapılan sınıflamanın her iki unsur dikkate alınarak değerlendirilmesi mümkün olabilmektedir.

Anma, duyarlılık ve f -ölçütlerinin yanında bir de ROC eğrilerinin kullanımı ile performans değerlendirmesi yapılabilmektedir. Bu eğri ile metin sınıflandırmasının ne kadar güvenilir olduğu görselleştirilerek sunulabilmektedir. ROC eğrisi doğru pozitif örnek sayısının yanlış pozitif örnek sayısına oranını gösterir. ROC eğrisi 0 noktasına yaklaştıkça kategorizasyonun başarıya ulaştığını gösterir. Örnek bir ROC eğrisi Şekil 3'te gösterilmektedir (Aggarwal, 2015, s. 341).



Şekil 3. ROC eğrisi örneği

3.3.4. Metin Kategorizasyonu Yazılımları ve Weka

Metin kategorizasyonunu uygulamak ve sınıflandırma uygulamalarının performanslarını değerlendirmek üzere Weka, KNIME, Angoss Knowledge Studio, Matlab gibi çok çeşitli yazılımlar bulunmaktadır. Silahtaroğlu "Veri madenciliği: Kavram ve Algoritmaları" adlı kitabında bu alanda kullanılabilecek yüze yakın yazılımı listelemiştir (Silahtaroğlu, 2013, s. 259-264). Bu tez çalışmasında açık kaynak kodlu bir yazılım olan ve University of

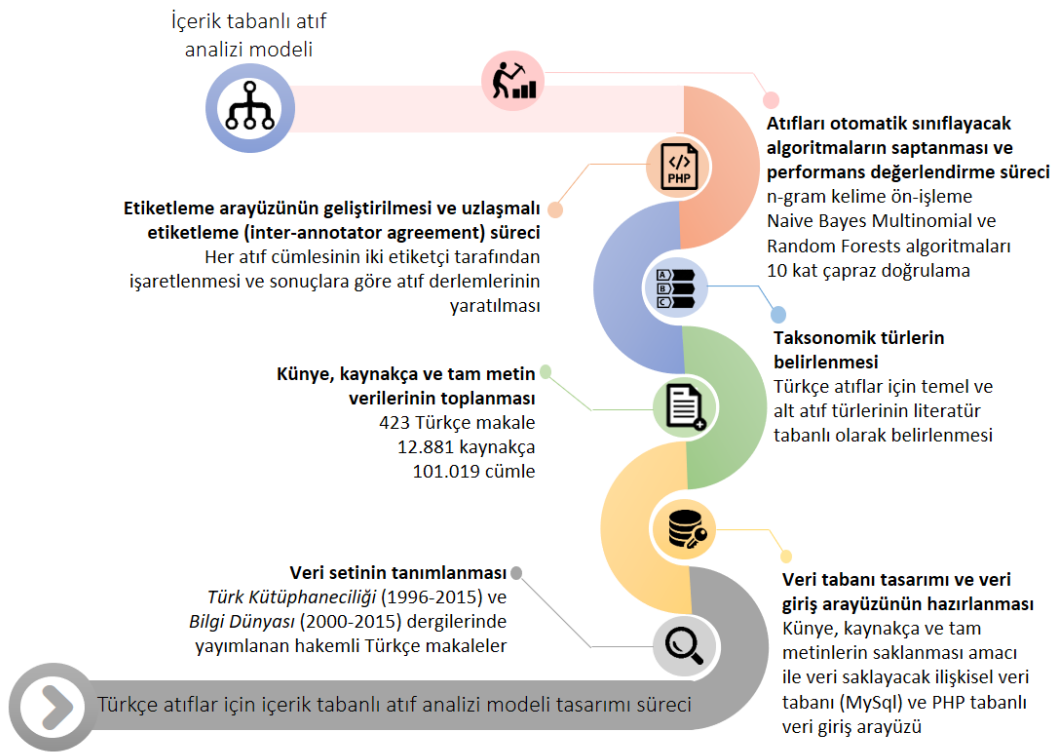
Waikato tarafından geliştirilen Weka kullanılmıştır. Bu çalışma kapsamında araç olarak Weka'nın tercih edilmesinin çeşitli sebepleri vardır. Bunlardan en önemlisi Weka'nın GNU Genel Kamu Lisansı ile geliştirilmiş bir özgür yazılım olması ve bu sayede esnek bir yapı sunmasıdır. Bunun yanında Java programlama dilinin kolaylıkla uygulanabiliyor olması, metin kategorizasyonunda kullanılan kelime ön işleme ve analiz algoritmalarının neredeyse tamamının tek arayüzden sunulması ve kolaylıkla uygulanabilmesi de bu aracın seçilmesini sağlayan sebeplerdendir.

Weka veri madenciliği görevleri için makine öğrenmesi algoritmalarını bünyesinde toplayan bir açık kaynak kodlu yazılımdır. Veri ön işleme, sınıflama, regresyon, kümeleme, birliktelik kuralları ve görselleştirme gibi çeşitli veri işleme aşamalarını gerçekleştirebilen yazılım Java tabanlı olarak çalışmaktadır (Weka, 2016). ARFF dosya formatını kullanan Weka'da sayısal, tarihsel, metinsel veya küme verileri çeşitli notasyonlar kullanılarak tanımlanmakta, bu sayede yazılım veriyi doğru bir şekilde işleyebilmektedir (Şeker, 2016). Yazılımın kullanımı ile ilgili çeşitli kılavuzlar ve videolar literatürde yer almaktadır (Şeker, 2016; Weka Eğitim Serisi, 2015; Witten, Frank ve Hall, 2011).

4. BÖLÜM: YÖNTEM VE TEKNİK

4.1. GİRİŞ

Bu çalışma kapsamında verilerin toplanması, işlenmesi ve değerlendirilmesi süreçlerinde farklı yöntem ve teknikler kullanılmıştır. Bu bağlamda her bir aşamanın tek tek tanımlanması benzer çalışmaların farklı disiplinlerde uygulanmasını kolaylaştıracaktır. Bu bölümde tez çalışmasının her aşaması detaylı olarak anlatılmakta ve uygulamanın ne şekilde yapıldığı açıklanmaktadır. Şekil 4'te araştırmanın temel yapısı sunulmaktadır. Çalışmada yapılan tüm veri toplama ve değerlendirme işlemleri 115K440 Proje numaralı "Türkçe atıflar için içerik tabanlı atıf analizi modeli tasarımı" başlıklı TÜBİTAK SOBAG Projesi ile paralel yürütülmüş ve Proje'de görevli uzman bursiyerler etiketleme aşamasında uzlaşmayı (inter-annotator agreement) sağlamak üzere görevlendirilmiştir.



Şekil 4. İçerik tabanlı atıf analizi modeli tasarımı süreci

4.2. VERİ SETİNİN TANIMLANMASI

Literatürde yer alan atıf değerlendirme çalışmalarında üzerinde sıklıkla durulan konulardan biri alan temelli değerlendirmelerin yapılmasının gerekliliği ile ilgilidir. Nasıl ki her bireyin bilgi arama davranışı farklı ise, alanlara özgü atıf yapma stilleri de birbirinden farklıdır ve bu farklılıklar dikkate alınarak değerlendirme yapılmalıdır (Cano, 1989, s. 284). Bu bağlamda bu çalışma kapsamında *Türk Kütüphaneciliği* ve *Bilgi Dünyası* dergilerinde yayımlanan Türkçe makaleler temel veri setini oluşturmuştur. Her iki dergi de açık erişimli dergiler olduğundan koleksiyonlarına erişimde herhangi bir sıkıntı yaşanmamıştır. Atıf yapma stillerinin benzer olması açısından çalışma kapsamına yalnızca hakemli makaleler alınmış, diğer doküman tipleri değerlendirme dışında bırakılmıştır.

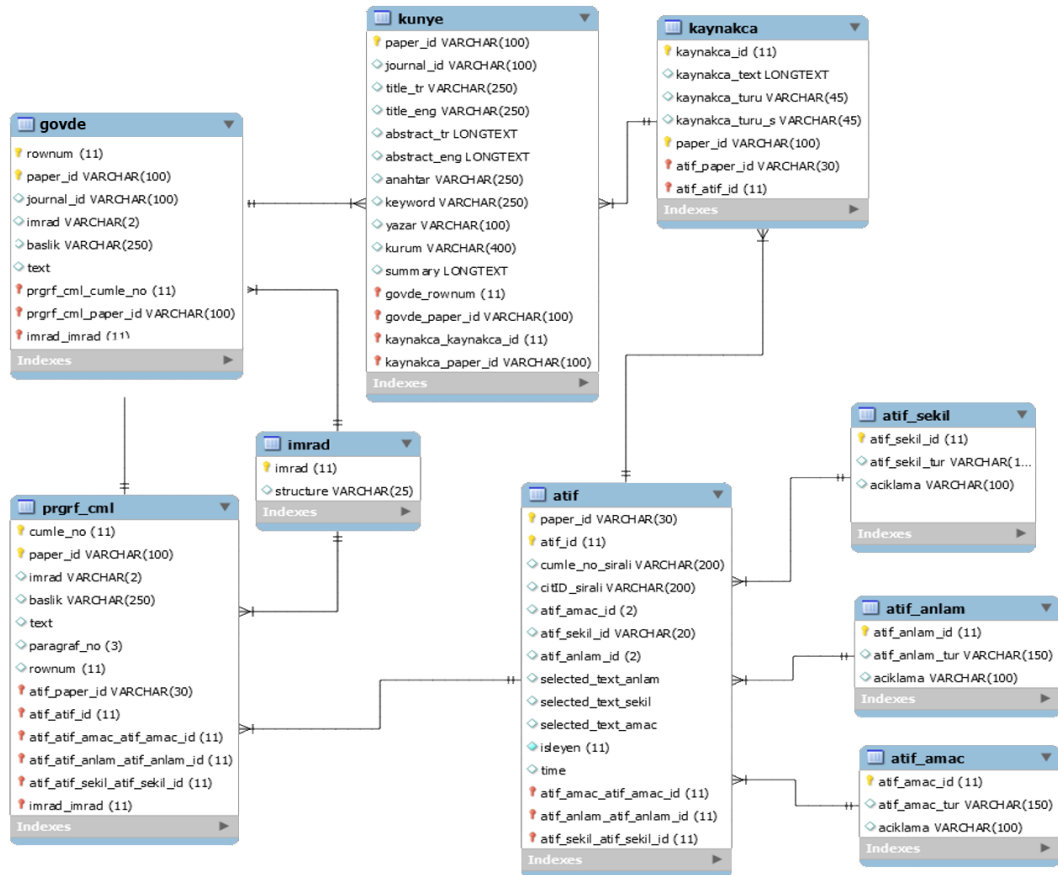
4.3. VERİLERİN TOPLANMASI VE ANALİZE UYGUN HALE GETİRİLMESİ

Çalışma kapsamında *Türk Kütüphaneciliği* ve *Bilgi Dünyası* dergilerinde yayımlanmış tüm hakemli makaleler pdf formatında kaydedilmiştir. İndirilen yayınların arasından OCR (Optical Character Recognition) işlemi yapılmamış dokümanlar için -karakter sorunu yaşanmaması açısından- bir OCR süreci işletilmiştir.

İndirilen tüm yayınlar akıllı birer numara ile kimliklendirilmiştir. Bu kimlik numaralarında *dergi adı+yıl+sayı+no* şeklinde bir yapı gözetilmiştir. Örneğin, TK201031 numaralı makale *Türk Kütüphaneciliği* dergisinde 2010 yılının üçüncü sayısında yayımlanmış ilk makaleyi temsil etmektedir.

Makalelerin pdf kopyalarının indirilmesinin ardından her bir makale .txt ve .doc biçimlerine dönüştürülmüştür. Bu dönüştürmede dil kodlaması olarak UTF-8 seçilmiştir. Bu sayede Türkçe karakterlerde herhangi bir problem yaşanmasının önüne geçilebilmiştir. Dönüştürülen tüm dosyalar veri girişi için tasarlanan arayüzün arkaplan yapısında kullanılmıştır.

Verilerin toplanması için MySql tabanlı ilişkisel bir veri tabanı tasarlanmıştır. Bu ilişkisel veri tabanının varlık-ilişki şeması Şekil 5'te gösterildiği gibidir.



Şekil 5. İlişkisel veri tabanının varlık-ilişki şeması

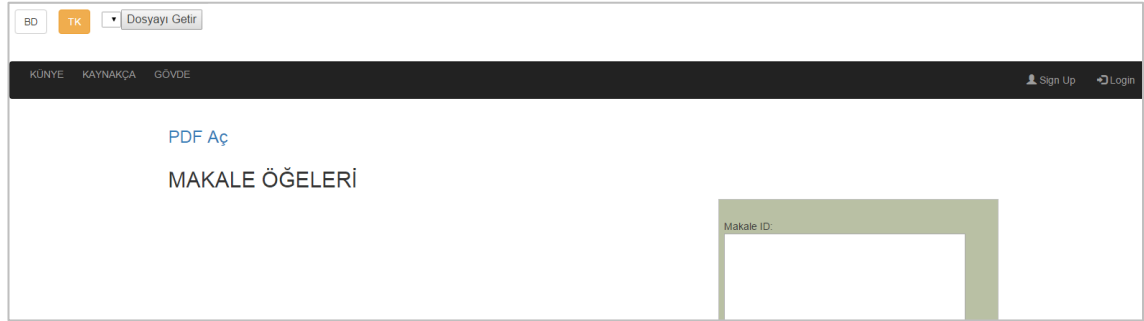
Oluşturulan ilişkisel veri tabanının amacı verileri standart bir yapıda tutmayı sağlamaktır. Makalelere ait temel bilgiler, kaynakçalar ve tam metinlere ait bilgilerin bu veri tabanında tutulması hedeflenmiştir. Veri tabanının oluşturulmasının ardından veri girişini kolaylaştıracak arayüzün tasarlanması aşamasına geçilmiştir.

4.4. VERİ GİRİŞ ARAYÜZÜNÜN TASARLANMASI

Bu çalışma içerisinde makalelere ait tüm metinlerinin elde edilmesi büyük önem taşımaktadır. Bu nedenle temel üst veri bilgilerinin yanında kaynakça ve tam metinlerin de veri tabanında depolanması gerekmektedir. Bu bağlamda hazırlanan arayüzde üç aşamalı bir veri giriş yapısı oluşturulmuştur. Hazırlanan arayüz <http://115k440.hacettepe.edu.tr> adresinden erişilebilir durumdadır. Hâlihazırda tüm künye, kaynakça ve gövde verileri web sayfası üzerinde “Metin ön işleme” sekmesi altında saklanmaktadır.

4.4.1. Üst Veri Bilgilerinin Tanımlanması

Hazırlanan arayüzde ilk aşama makalelerin temel öğelerine ayrılmasıdır. Belirlenen temel öğeler yazar adları, Türkçe ve İngilizce başlıklar ile öz ve anahtar sözcükler, yazarların çalıştıkları kurum bilgileri ve *Türk Kütüphaneciliği* dergisi için uzun İngilizce özetlerdir. Arayüz açıldığında ekrana Şekil 6'da gösterildiği gibi bir görüntü yansıtılmaktadır.



Şekil 6. Veri giriş arayüzü

Veri giriş arayüzünün en üstünde yer alan TK ve BD bağlantıları üzerinde çalışılacak dosyaların seçimi için tasarlanmıştır. Veri girişi yapılırken ilk aşamada üzerinde çalışılacak dosya seçilmelidir. Üzerinde çalışılacak dosyanın seçimi için BD ya da TK sembolünün üzerinde tıkladığında yıllara göre bir açılır kutu belirlemektedir. Bu açılır kutudan yıl seçimi yapıldıktan sonra her yılın altından belirli makaleler seçilebilmektedir.

Bir makale seçildiğinde MySQL sorguları aracılığı ile oluşturulan sistem başlık, yazar adı, özet veya diğer alanları ayrıştırabilmektedir. Bu ayrıştırma kullanılan dil, enter tuşu kullanımı, başlıklar gibi bazı özellikler göz önüne alınarak yapılmaktadır. Örneğin, öz bilgileri başlığın “öz” olduğu yerde başlayıp “anahtar sözcükler”e kadar geçen bölümün işaretlenmesi üzerine planlanmıştır. Ancak gerek dergilerde yayımlanan tüm yazıların aynı standart yapıda olmaması, gerekse metin dosyalarının özelliklerinden kaynaklanacak bazı sorunlar sebebi ile otomatik belirlenen öğelerin insan eliyle doğrulanması amacı doğrultusunda arayüzün sağ tarafında yer alan kutucuklar tasarlanmıştır. Sistem tarafından önerilen öğeler sol kısımda yer alan kutucukların içinde gösterilirken, öğe işaretlemesi yapan kişi her bir öğeyi kolaylıkla sağ tarafa aktarabilmektedir. Tüm kutucuklar düzenlenebilir formatta olduğundan ana metinde yazım, karakter veya herhangi başka bir sorun olsa da veri girişini yapan kişi tarafından düzeltilmesi olanaklı hale getirilmiştir. Örnek ekran Şekil 7’de gösterildiği gibidir.

Hakemli Yazılar / Refereed Papers	'Türk Kütüphaneciliği 24, 1 (2010), 11-32'
MAKALE DERGİ BAŞLIK TR BAŞLIK ING ABS.TR ABS.ING ANAHTAR SÖZ. KEYWORDS YAZARLAR KURUM	Başlık TR: 'Sosyal Bilgiler ve Türkçe Öğretmeni Adaylarının Okuma Durumları: (İlgi, Alışkanlık ve Okuma Stratejilerini Kullanım Düzeyleri)'
Sosyal Bilgiler ve Türkçe Öğretmeni Adaylarının Okuma Durumları: (İlgi, Alışkanlık ve Okuma Stratejilerini Kullanım Düzeyleri)	Başlık İNG: 'Reading Statures of Social Studies and Turkish Language Teacher Candidates (Their Attention, Habits and Levels of Using Reading Strategies)'
MAKALE DERGİ BAŞLIK TR BAŞLIK ING ABS.TR ABS.ING ANAHTAR SÖZ. KEYWORDS YAZARLAR KURUM	Yazar(lar): 'Zafer Kuş, Mustafa Türkyılmaz'
Reading Statures of Social Studies and Turkish Language Teacher Candidates (Their Attention, Habits and Levels of Using Reading Strategies)	Abstract TR: 'Okuma, yaşam boyu öğrenmenin gerçekleştirilmesi için gerekli olan bir edindir. Bireyler okuma becerilerini kullanabildikleri ve
Kaynakça	

Şekil 7. Örnek üst veri girişi

Diğer taraftan karakter sorunu nedeni ile okunamayan ya da sorunlu dosyaların veri girişini kolaylaştırmak için veri girişi yapılan makalenin tam metinlerine (pdf ve doc) bağlantı verilmiştir. Bu sayede sistemin öğeleri tanımlayamadığı durumlarda veri girişi yapılırken elle müdahale imkânı sunulmuştur. Üst veri giriş arayüzü ile temel öğelerin etiketlenmesi 30 saniye ile 1 dakika arasında zaman almaktadır. Üst veri bilgilerinin işaretlenmesi sonrasında derlemde toplam 423 Türkçe makaleye ait bilgi işaretlenmiştir.

4.4.2. Kaynakça Bilgilerinin Toplanması

Çalışma kapsamında her bir makalenin yaptığı atıfların farklı bir veri dosyasında hatasız olarak tutulması büyük önem taşımaktadır. Ancak daha önce birkaç çalışmada da belirtildiği gibi her iki dergide de kaynakçaların standardizasyonu açısından çeşitli sorunlar bulunmaktadır (Doğan, 2014; Kurbanoglu, 1996). Tüm kaynakçalar için standart bir yapının bulunmadığından hareketle arayüzün ikinci kısmında kaynakçalarda yer alan tüm kayıtların etiketlenmesi, tanımlanması ve standardizasyonu hedeflenmiştir.

Araştırma kapsamındaki dergilerde yer alan makalelerde kaynakçalar için standart bir başlık olmaması kaynakçaların nerede başlayıp nerede bittiğini saptamak açısından problem yaratmıştır. Bu nedenle olası ve kullanılan tüm başlıklar sisteme tanımlanmış (Kaynaklar, Kaynakça, Bibliyografya vb.) ve bu başlıktan sonra yer alan tüm künyeler

ikinci aşamada veri giriş ekranında sunulmuştur. Veri giriş işleminde ikinci aşamaya geçildiğinde karşılaşılan ekranda bu künyeler düz metin formatında gösterilmiştir. Bu aşamada veri girişini yapan kişiden beklenen her bir atfı bir belirteçle birbirinden ayırmasıdır. Veri tabanında ayırım klavyedeki yakınlığı ve ayırt edici olması sebebi ile seçilen öç. karakterleri tanımlanarak yapılmıştır. İşaretlemeyi yapan kişinin her bir atfın arasına öç. işaretlemesini yapması ile her bir atfı tek tek tanımlanabilmiş ve sistemde makale numarası ile kimliklendirilmiştir. Künyelerin belirlenmesi işleminin ardından ekranda oluşan görüntü Şekil 8'de gösterilmektedir.

Buckland, M. (1997). Redesigning library services: A Manifesto. (çevrimiçi). http:// sunsite.berkeley.edu/
Çakın, İ. (1998) Üniversitelerimizin bilgiye erişim ortamları: Genel değerlendirme. Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi, (Cumhuriyetimizin 75. Yılı Özel Sayısı), 37-67.
Lancaster, F.W. (1982). Libraries and librarians in an age of electronics. Arlington, Va.: Information Resources Press.
Lenzner, FT and Johnson, BB. (1997). Beeing things as they really are. Forbes, 59(5): 122-128.
Lucker, J.K. (1993). The library as place: Research libraries: today and tomorrow(Ed. byW.J.Welsh). London: Greenwood Press, 3-17.
Marcum, DB.(1997). Digital libraries: for whom? For what?. Journal of Academic Librarianship, 23(2): 81-84.
Meyer, R.W. (1997). Surviving the change: the economic paradigm of higher education in trasformation. Journal of Academic librarianship, 23(4): 291-301.
Neal, J. G. (1996). Academic Libraries: 2000 and beyond. Library Journal, 121 (12). 74-76.
Ruschoff, C. (1997). What are the most important issues confronting academic Librarianship as we approach the 21st century?. Journal of Academic librarianship, 23(4): 321-322.
Stevens, N.D. (1993). Research libraries: past, present and future. Advances in Librarianship, i7: 79-100

Şekil 8. Veri girişi sonrası künyelerin gösterimi

Her bir makalede yer alan künyelerin tek tek kimliklendirilmesinin ardından elle veri girişinden kaynaklanan hataların anında fark edilip düzeltilebilmesi için her bir referans düzenlenebilir formata getirilmiştir. Yani her bir kutucukta yer alan künye üzerine tıkladığında yeniden değiştirilebilir hale gelmektedir.

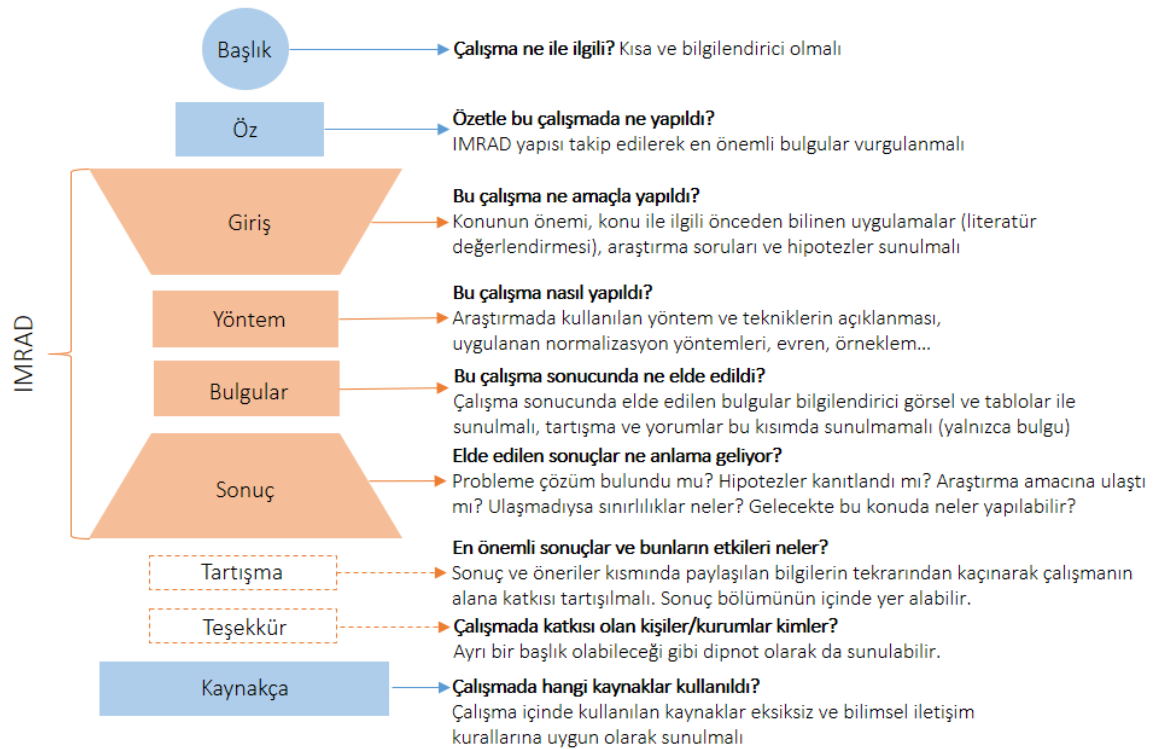
Bu kısımda bir makale için künyelerin düzenlenmesi yapılan atfı sayısı ile ilgili olarak 3 ile 7 dakika arasında değişen sürelerde gerçekleştirilmektedir. İşaretleme sonrasında 423 Türkçe hakemli makaleye ait toplam 12.881 kaynakça bilgisi veri tabanına kaydedilmiştir.

4.4.3. Bölüm Bilgilerinin ve Yapısal Tam Metinlerin Toplanması

Bu çalışmanın temel amacı makalelerde yapılan atfıların semantik ve sentaktik olarak sınıflandırılmasıdır. Bu bağlamda en önemli noktalardan biri makale tam metinlerinin veri tabanında eksiksiz olarak depolanabilmesidir. Bu sebeple atfıların hangi bölümde ve hangi amaçla yapıldığının saptanması bu çalışma için önem arz ettiğinden tüm metinlerin anlamlı bir şekilde çıkarımını yapmak amacı ile arayüzün üçüncü kısmı tam metinlerin

anlamli olarak işaretlenmesi işine ayrılmıştır. Burada her bir bölümün başlığının ve o bölümde yer alan metnin etiketlenmesi ve işaretlenmesi hedeflenmiştir.

Bölmelerin etiketlenmesi aşamasında uluslararası geçerliğe sahip IMRAD (giriş [introduction], yöntem [methods], araştırma [research] ile tartışma ve sonuç [discussion]) yapısının kullanılması tercih edilmiştir. IMRAD yapısına göre makaleler bir bilimsel araştırmanın ana çerçevesi sayılan bu dört başlık altında yapılandırılmış olmalıdır. Araştırmaların bu temel bölümlerinin yanında başlık, öz, teşekkür, dipnot ve ekler şeklinde diğer bölümleri de vardır (Swales ve Feak, 2000, s. 220). IMRAD'ın temel yapısı ve her bölümde bulunması beklenen öğeler Şekil 9'da gösterilmektedir (Wu, 2011, s. 1346).



Şekil 9. Bir makalenin temel bilimsel bölümleri ve IMRAD yapısı

Temel bir standart yapı sunması sebebi ile bu çalışma kapsamında her bir makale için yazarların kullanmayı tercih ettikleri başlıkların yanında Şekil 9'da da detayları aktarılan IMRAD yapısına göre metnin yer aldığı bölüm de işaretlenmiştir. Bu sayede metinler temel bilimsel bölümlerine ayrılabilmiş, bu ayırım ile birlikte atıfların hangi amaçlarla yapıldığının saptanması aşamasında bölümlere göre bir farklılık olup olmadığı kolaylıkla analiz edilmiştir. Metin ön işleme işlemlerinin son aşaması olan bölümlere yönelik veri giriş arayüzü Şekil 10'da gösterilmektedir.

TK201031 GİRİŞ

IMRAD : SEÇ

Alt Başlık :

Metin :

Kaydet

Kapat

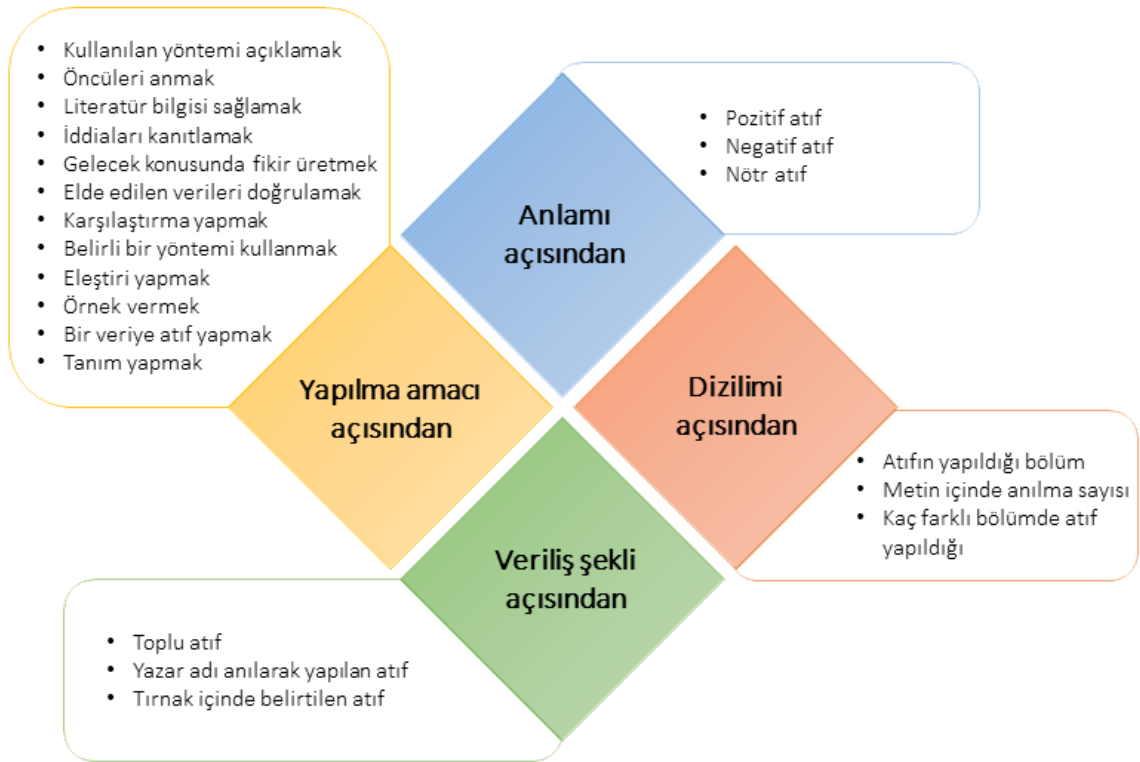
Şekil 10. Bölümlere göre tam metin verisi giriř ekranı

Veri toplanması aşamasında süre açısından üzerinde en fazla zaman harcanan yer tam metin verilerinin toplandığı bu kısımdır. Harcanan süre makalenin uzunluğuna, başlık ve paragraf sayısına göre deęişiklik gösterebilmektedir.

Veri setinde yer alan tüm makaleler için üç aşamalı bu veri girişinin tamamlanmasının ardından tüm veriler metin kategorizasyonu uygulaması için etiketleme sürecine hazır hale getirilmiştir. Tam metinlerin işaretlenmesi sonrasında Türkçe kütüphanecilik ve bilgilim literatüründe yer bulan toplam 101.019 cümle veri tabanında depolanmıştır.

4.5. ATIF SINIFLARININ BELİRLENMESİ

Çalışmanın ikinci bölümünde literatür çalışmalarında belirlenen sınıflamalar sunulmuş, bu sınıflamaların dayandığı temeller aktarılmıştır. Bu bağlamda literatürde yer alan sınıflamalar dikkate alındığında Şekil 11’de gösterildiği gibi bir taksonomik sınıflama yapılmış, verilerin etiketlenmesi sürecinde bu sınıflar üzerinden uygulama gerçekleştirilmiştir.



Şekil 11. Taksonomik atıf sınıfları

4.5.1. Anlamı Açısından Atıflar

Anlamı açısından atıfların değerlendirilmesindeki ana amaç bir yazarın olumlu ya da olumsuz hangi duygu temelinde atıf yaptığının kestirilebilmesidir. Yazar atıf yaptığı makalede belirtilen araştırma, yöntem ya da bulgular konusunda olumlu görüşe sahip ise pozitif atıf; yazar bu makalede belirtilen görüşü desteklemiyor, hatalı buluyor ya da hem fikir değilse negatif atıf olarak sınıflanmalıdır. Nötr atıflar ise yazarın herhangi bir duygu belirtmeksizin yaptığı atıfları temsil etmektedir. Aşağıda verilen üç örnek üzerinden anlatmak gerekirse;

Örnek 1: "Stueart ve Sullivan (2010) sekiz bölümden oluşan 'Developing Library Leaders' adlı kitabın yedinci bölümünde mentorluk konusu üzerinde durmuştur. Mentorluğun tanımı, türleri, fonksiyonları, başarılı mentorluk ilişkileri, mentorun rolleri ve sorumluluklarının açık bir şekilde ele alındığı kitap, bu çalışmanın teorik kısmının oluşturulmasında yol gösterici olmuştur."

Örnek 2: "Bu araştırmada ele aldığımız konu ile ilgili bir diğer yasa olan Araştırma ve Geliştirme Faaliyetlerinin Desteklenmesi Hakkında Kanun (2008)'un yazılı amacında araştırma geliştirme (ar-ge) etkinliklerinin çerçevesi ayrıntılı olarak çizilmekte, ancak kütüphane kurumuna hiçbir biçimde atıfta bulunulmamaktadır. Bir başka deyişle, kütüphane kurumunun doğrudan yer aldığı ve çok güçlü bir biçimde ilişkili olduğu ulusal ar-ge etkinliklerinin çerçevesini düzenleyen bir yasa metninde kütüphane kurumu anılmamaktadır."

Örnek 3: "Bilgi politikası; belirli bir ülke ya da bölgede, bilgi sağlama ve bilgiye erişim için yapılan ulusal plandır (Keenan ve Johnston, 2000, s. 135)."

İlk örnekte yazar atıf yaptığı kitabın konuyu çok açık bir şekilde incelemesi sebebi ile kendi çalışması için yol gösterici olduğunu belirterek kitaba olumlu bakış açısını açık bir dille ifade etmektedir. Örnek 2'de ise söz konusu yasanın kapsamı gereken önemli bir konuyu anlamıyor olması yazar tarafından olumsuz açıdan değerlendirilmektedir. Üçüncü örnekte ise bir tanımlama yapılmış, atıf yapılan yayına karşı herhangi bir duygusal belirteç kullanılmamıştır. Bu açıdan bakıldığında atıflar anlamlarına göre sınıflandırıldığında atıf yapan yazarın duygu durumunun ve duygusal olarak makaleye bakış açısının değerlendirilmesi gerekmektedir.

4.5.2. Yapılma Amacı Açısından Atıflar

Yapılma amacı açısından atıfların sınıflandırılmasındaki temel amaç yazarın duygusal olarak atıf yaptığı makaleye bakış açısından ziyade ne amaçla atıf yaptığının kestirilmesidir. Çalışmanın ikinci bölümünde de değinildiği gibi her atıf aynı amacı taşımamaktadır. Bu sebeple atıfların yapılma amaçları da değerlendirilmelidir. Çalışma kapsamında atıflar yapılma amaçları açısından sınıflandığında 12 temel sınıf belirlenmiştir. Bu sınıflar ve her sınıfın açıklandığı birer örnek Tablo 4'te sunulmaktadır.

Tablo 4. Yapılma amacına göre atıf türleri, tanımlamalar ve örnekler

Sınıf	Açıklama	Örnek
Yöntemi açıklamak	Yazarın kendi çalışmasında kullandığı yöntemi aktarmak üzere yaptığı tüm atıflar (yöntemin tanımı, kullanım alanları, araştırma tasarımı vb.)	"Yüzde 95 güven düzeyi ve ± 0.04 hoşgörü miktarında 2000 birimlik evren için uygun olan örneklem birimi sayısı 108'dir (Çingı, 1990, s. 262)."
Öncüleri anmak	Bir konuyu ilk tartışan, ilk kez tanımlayan, ilk kez geliştiren vb. kişi, kuruluş veya çalışmaların anıldığı atıflar	"Dünyada 1835 yılında Havas Haber Ajansı'yla başlayan çalışmaların öncüleri arasında Associated Press (AP) ve Reuters bulunmaktadır (Tokgöz, 2006, s. 105-111)."
Literatür bilgisi sağlamak	Konu ile ilgili temel arkaplan bilgisi vermek amaçlı yapılan atıflar	"Üniversitelerin oldukça eski bir geçmişi olmasına karşın, çağdaş anlamda üniversite kavramı, yani eğitim ve öğretimin yanı sıra araştırmaya da önem verilmesi XIX. yüzyılın başında Berlin Üniversitesi'nin kurulması ile başladı (Çakın 1998: 37)."
İddiaları kanıtlamak	Yazarın öne sürdüğü bir konuda kanıt bilgisi sunmak amaçlı yaptığı atıflar	"Bilgi, organizasyon için girdi olmasının yanı sıra, faaliyetleri yönlendirme ve faaliyetlerin sonucunda ortaya çıkan bir ürün özelliği de taşımaktadır. Buradan da anlaşılacağı gibi bir sistem bilgisiz devamlılığını sürdürmez (Paksoy, 1996, s. 56)."
Gelecek konusunda fikir üretmek	Yazarın gelecekte olması beklenen gelişmeler ile ilgili bilgi verirken yaptığı atıflar	"Tenopir ve King (1997, s. 165) önümüzdeki 5-10 yıl içerisinde çoğu dergilerin hem elektronik medya üzerinde, hem de kâğıt kopya olarak yayımlanacağına işaret etmektedirler."
Elde edilen verileri doğrulamak	Yazarın çalışmasının sonunda elde ettiği bulguların literatürdeki bulgularla doğrulamasını yaptığı atıflar	"Dergi kendine atfı ile ilgili elde ettiğimiz oran, daha önce Thomson Reuters (2004) tarafından gerçekleştirilen çalışmadaki dergi kendine atfı ile ilgili bulgulara paralellik göstermektedir."
Karşılaştırma yapmak	Yazarın önceki çalışmalar ya da çeşitli olgular arasında karşılaştırma yapmak üzere değindiği atıflar	"Çevrimiçi olarak verilen danışma hizmeti yüz yüze verilen hizmetten çok daha karmaşıktır (Ellis ve Francoeur, 2001)."
Belirli bir yöntemi kullanmak	Yazarın çalışmasında bir başkası tarafından geliştirilen yöntemi kullandığını aktarırken yaptığı atıflar	"Bu çalışmada en fazla 10 epokta öğrenen, hesaplamalar çıkartabilen ve hesaplamalarda birçok hususu çözümleyebilen Levenberg-Marquardt (LM) öğrenme algoritması kullanılmıştır (Levenberg, 1944; Marquardt, 1963)."
Eleştiri yapmak	Yazarın belirli konularda işleyiş ya da uygulamalara eleştiri getirmek amacı ile yaptığı atıflar	"Bilgi politikasının saptanması, planlaması ve uygulanması sürecinde kütüphaneciler geçmişteki başat rollerini kaybederek, bu süreç içerisinde yer alan aktörlerden birisi konumuna düşmüşlerdir (Toplu, 2006)."
Örnek vermek	Yazarın ortaya koyduğu görüşlerle ilgili destekleyici örnek vermek üzere yaptığı atıflar	"Doi numarasının nasıl geliştiği Şekil 2'de bir başka örnekle grafik olarak gösterilmektedir (Rosenblatt 1997)."
Veriye atıf yapmak	Yazarın bir başkasının verisini kullandığına ya da başka çalışmaların elde ettiği istatistikî verileri sunduğuna dair yaptığı atıflar	"Son verilere göre ABD'deki büyük araştırma kütüphanelerinin toplam bütçelerinin %6 ile %10'luk bir bölümünün elektronik bilgi kaynağı sağlamak için harcadığı görülmektedir (Blixrud ve Jewell 1998)."
Tanım yapmak	Yazarın bir başkası tarafından yapılmış tanımı anarken yaptığı atıflar	"Veri (data); islenmemiş, düzenlenmemiş ham gerçeklerdir (Davis 1991: 181)."

4.5.3. Veriliş Şekli Açısından Atıflar

Veriliş şekli açısından atıfların sınıflandırılmasında temel amaç atıfların şekilsel olarak sınıflandırılmasıdır. Bu bağlamda atıf yapılırken genellikle tercih edilen atıf verme şekilleri de

ortaya çıkarılabilecektir. Atıflar şekilsel olarak sınıflandırıldığında temel üç sınıf elde edilmiştir. Bu sınıflar toplu atıflar, yazar adı anılarak yapılan atıflar ve tırnak içinde belirtilen atıflardır. Toplu atıflarda temel mantık bir atıf parantezi içinde verilen birden fazla kaynağın varlığıdır. Yazar adı anılarak yapılan atıflarda yazarların “Taşkın’ın da belirttiği gibi..” şeklinde yazarın adını özel olarak andığı atıfların saptanması hedeflenmiştir. Son sınıf olan tırnak içinde belirtilen atıflar ise tırnak işareti ile ya da metin içinde direkt alıntı şeklinde (fazla girinti, italik yazım vb.) yapılmış atıflardır. Şekline göre atıfları diğer türlerden ayıran temel özellik bir atıfın aynı anda birden fazla sınıfta yer alabilmesidir. Örneğin, aşağıdaki atıf cümlesinde tırnak içinde belirtilen ve yazar adı anılan bir toplu atıf görülmektedir.

“Mesleki bağlılık konusunda önemli bir yere sahip olan Aranya ve arkadaşları (1986) “profesyonel bağlılık” ifadesini kullanmaktadır. Yazarlar profesyonel bağlılık kavramını “kimliği ön plana çıkarmak, mensubu olunan meslek için çaba sarf etmek, mesleki hedef, değer, norm ve etik ilkelere bağlılık” olarak tanımlamaktadırlar (Aranya ve Ferris, 1984, s. 3; Lanchman ve Aranya, 1986, s. 229).”

4.5.4. Dizilimi Açısından Atıflar

Dizilimi açısından atıfların sınıflandırılmasındaki ana amaç atıfların metin içinde geçtikleri bölüm ve atıf sayıları dikkate alınarak değerlendirme yapılmasıdır. Bu bağlamda bu taksonomik türde atıfların yapıldığı IMRAD bölümleri, metin içinde anılma sayıları ve kaç farklı bölümde atıf yapıldığı gibi temel tanımlayıcı istatistiksel bilgiler üzerinden değerlendirme yapılabilmektedir.

4.6. ETİKETLEME SÜRECİ

Türkçe kütüphanecilik ve bilginbilim literatüründe yayımlanmış hakemli makalelerin üst veri, künye ve kaynakça bilgilerinin veri tabanına kaydedilmesi ve atıflar için taksonomik türlerin belirlenmesinin ardından atıfların etiketlenmesi işlemine geçilmiştir. Bu aşamada yürütülen etiketleme sürecinde bir etiketleme arayüzü tasarlanmış, etiketleme arayüzüne ilişkin etiketleme eğitimleri verilmiş ve uzlaşmalı etiketleme süreci işletilmiştir.

4.6.1 Etiketleme Arayüzünün Tasarımı

Çalışma kapsamında atıfların sınıflarına göre etiketlenmesi işleminin yapılabilmesi için bir arayüz geliştirilmiştir. Sözü edilen bu arayüz kullanılarak 115K440 Projesi kapsamında çalışan bursiyerlerin hızlı ve pratik şekilde etiketleme işlemini yapabilmeleri hedeflenmiştir. Arayüzün genel görünümü Şekil 12’de gösterildiği gibidir.

Etiketleme arayüzünde her bir makale cümlelerine ayrılmış ve veri tabanında depolanmıştır. Cümlelere ayırma işlemi nokta işareti belirleyici olurken cümle bitişini işaret etmeyen nokta işaretleri (yüzyıl kısaltmaları, sayılardan sonra gelen noktalar, unvan kısaltmaları vb.) veri tabanına tanımlanmış ve en doğru cümle ayrımlarının yapılmasına çalışılmıştır. Etiketleme işleminin temel sistematığı şöyledir:

- Etiketçi kendi kullanıcı adı ve şifresini kullanarak sisteme giriş yapar. Bu denetimin temel sebepleri etiketlenmenin hangi etiketçi tarafından yapıldığının işaretlenmesi ve etiketçilerin birbirlerinin yaptığı etiketleri görmelerinin engellenmesidir.
- Sisteme kullanıcı adı ve şifresi ile giriş yapan etiketçi etiketleme yapacağı makaleyi üst bölmede yer alan açılır kutu aracılığı ile seçer.
- Seçilen makalede geçen tüm cümleler arayüzün sağ tarafında görüntülenir. Her cümlelerin sağında seçim için bir kutu bulunmaktadır. Etiketleme yapılacak cümle ya da cümleler bu kutular kullanılarak seçilir. Bir atıf bir cümleden oluşabildiği gibi birden fazla cümleden de oluşabilmektedir. Bu sebeple bu çalışma kapsamında kullanılan "atıf cümlesi" terimi tek cümleyi temsil etmemekte atıfın yapıldığı cümle ya da cümle gruplarını temsil etmektedir.
- Seçimi yapılan atıf cümlesinin kaynakçadaki karşılığı sol menüde yer alan "kaynakça" açılır kutusundan seçilir. Veri toplama aşamasında kaydedilen tüm kaynakça bilgileri makalelerle eşleştirilerek bu kutu aracılığı ile sunulur. Bu kutudan tek bir kaynak seçimi yapılabildiği gibi birden fazla kaynak seçimi de yapılabilmektedir. Bu sayede toplu atıflar da kolaylıkla işaretlenebilmektedir.
- Atıf cümlesini belirleyen etiketçi sırasıyla amacına, anlamına ve şekline göre atıf sınıflarını etiketler. Ardından bu sınıfların ayırımına hangi sözcükler sayesinde vardığını da her bir sınıfın altında yer alan kutucuğa kaydeder. Örneğin, pozitif atıfı ayırt etmesini sağlayan sözcük grubu "önemli bir çalışmadır" ise bu sözcüğü kutuya yazar. Birden fazla sözcük var ise sözcükleri birbirinden noktalı virgül ile ayırır. Şekline göre atıflarda bir atıf birden fazla sınıfa dâhil olabileceğinden bu amaçla tasarlanan açılır kutudan birden fazla seçenek işaretlenebilmektedir.
- Etiketleme işlemini tamamlayan etiketçi "Kaydet" butonunu kullanarak yaptığı atıf cümlesi etiketlemesi işlemini veri tabanına kaydeder. Mevcut girişte değişiklik yapıyorsa "Üzerine Yaz" butonunu kullanır. Daha önce söz konusu makale için yaptığı diğer atıf etiketlerini görüntülemek için de "Atıf Listele" butonunu kullanır.

Araştırmamız kapsamında atıf etiketleme işlemine başlanmadan önce tüm etiketçilere etiketleme eğitimi verilerek standart bir uygulama gerçekleştirilmesi sağlanmıştır.

PDF Aç

Kaynakça:

None selected ▼

Cümleler:

Amacına Göre Atfın Türü:

SEÇ ▼

Anlamına Göre Atfın Türü:

SEÇ ▼

Şekline Göre Atfın Türü:

None selected ▼

YENİ KAYDET

ÜZERİNE YAZ

ATIF LİSTELE

Cümle	Cümle No	Seç	İMRAD
giriş			
İnsanoğlu tarihin her döneminde "gelecekte ne olacak?" sorusunun yanıtı ile yakından ilgilenmiş, bilimsel ya da bilim-dışı yöntemlerle bu konuya eğilmiştir.	1	<input type="checkbox"/>	GIRIS
Tarih boyunca kahinler ve müneccimler, günümüzde de falcılar .fütüristler, planlamacılar ve yöneticiler geleceğe ilişkin beklentilerini çeşitli şekillerde ortaya koymaktadırlar.	2	<input type="checkbox"/>	GIRIS
Ancak, "gelecek" ile ilgili öngörülerde bulunmak, çeşitli güçlükleri birliğinde taşır.	3	<input type="checkbox"/>	GIRIS
Çünkü gelecek bir bakıma belirsizliktir.	4	<input type="checkbox"/>	GIRIS
Özellikle toplumsal ve teknolojik değişimlerin hızlı olduğu dönemlerde, geleceğe yönelik plan yapmak ve öngöründe bulunmak zordur.	5	<input type="checkbox"/>	GIRIS
Bu yüzdendir ki, örneğin, daha önceleri bir şirket için uzun vadeli bir planlama yaklaşık 20 yıllık bir dönemi kapsarken, günümüzde bu rakam 1,5-2 yıla inmiştir.	6	<input type="checkbox"/>	GIRIS
Geleceği kestirmenin ve plan yapmanın güçlüğüne karşın, her düzeydeki yönetici, kurumlarına ilişkin öngörülerde bulunmaktan kaçınmazlar, çünkü gelecekleri büyük ölçüde buna bağlıdır.	7	<input type="checkbox"/>	GIRIS
Denilebilir ki, geçmişi ve günümüzü iyi tahlil edebilirsek ileriye daha doğru görme olanağımız vardır.	8	<input type="checkbox"/>	GIRIS
Bu yazının konusu olan üniversite kütüphanelerinin de gelecekte alacakları biçim üzerinde düşünmek ve öngörülerde bulunmak, hem bu kurumlarda çalışanlar ve çalışacaklar hem de	9	<input type="checkbox"/>	GIRIS

Şekil 12. Etiketleme arayüzünün genel görünümü

4.6.2. Etiketleme Eğitimi

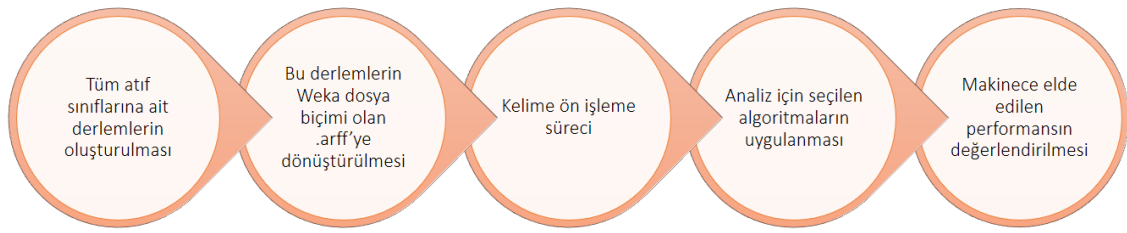
Etiketleme eğitimi öncesi tüm etiketçiler için atıf türlerini tanıtıcı bir rehber infografik hazırlanmıştır. Bu rehber http://115k440.hacettepe.edu.tr/etiketleme_egitimi.jpg adresinden erişilebilir durumdadır. Eğitim sürecinde üç farklı örnek makale üzerinden etiketleme işlemi yapılarak etiketçilerin sınıfları doğru bir şekilde ayırt edebilmelerinin sağlanmasının yanında, her sınıf detaylı olarak tartışılmış ve sınıfların ana hatları bu eğitimler sürecinde çizilmiştir.

4.6.3. Uzlaşmalı Etiketleme Süreci

Doğal dil işleme çalışmalarında en önemli koşullardan biri etiketçiler arasında uzlaşmanın sağlanmasıdır. Literatürde yapılan çalışmalar bu uzlaşma olmaksızın yapılan etiketleme süreçlerinde doğruluk ve objektiflik ile ilgili çeşitli sorunlar bulunabileceği konusunda çeşitli bulgular sunmaktadır (Artstein ve Poesio, 2008, s. 591; Landis ve Koch, 1977, s. 159). Bu sebeple çalışma kapsamında her bir atıf cümlesi en az iki etiketçi tarafından etiketlenmiş; bu etiketler arasından birbirine en çok benzeyen atıflar doğal dil işleme aşamasında analize tabi tutulmuştur. Çalışma kapsamında dördü doktora ikisi lisans öğrencisi toplam altı etiketçi görev yapmıştır. Etiketleme süreci sonunda birinci grup etiketçiler 14.259 atıf cümlesi saptarken, ikinci grup etiketçiler 14.840 atıf cümlesi elde etmişlerdir. Konu ile ilgili detaylı istatistiki bilgiler beşinci bölümde sunulmaktadır. Uzlaşmalı etiketleme süreci sonucunda anlamına göre atıf derleminde 13.866, yapıma amacına göre atıf derleminde 10.437 ve şekline göre atıf derleminde de 13.527 atıf cümlesi yer almıştır. Bu atıf cümleleri her iki etiketçi tarafından ortak olarak aynı sınıf etiketi ile işaretlenmiş cümle ya da cümle gruplarından oluşturulmuştur.

4.7. ATIFLARIN SINIFLANDIRILMASINDA DOĞAL DİL İŞLEME YÖNTEMLERİNİN UYGULANMASI

Etiketçiler tarafından tüm makalelerde yer alan atıf cümlelerinin işaretlenmesinin ardından bu etiketler karşılaştırılmış ve birbirine en çok benzeyen atıf cümleleri üzerinden doğal dil işleme çalışmaları yapılmıştır. Analizlerin yapılması amacı ile Weka veri madenciliği aracı kullanılmıştır. Weka aracı kullanılarak yapılan analizlerin detayları Şekil 13'te aktarılmaktadır.



Şekil 13. Türkçe atıflar için makine öğrenmesi süreci

4.7.1. Tüm Atıf Sınıflarına Ait Derlemlerin ve arff Dosyalarının Oluşturulması

Etiketçiler arasındaki benzerliklerin saptanmasının ardından her temel atıf sınıfı için birer derlem oluşturulmuştur. Bu derlemlerde atıf cümleleri ve bu cümlelerin ait oldukları sınıflar yer almıştır. Derlemlerin oluşturulmasının ardından ise bu derlemler Weka aracı tarafından kullanılan ve csv yapısına çok benzeyen arff (attribute-relation file format) dosya formatına dönüştürülmüştür. Bu dosya formatı iki temel yapının birleşmesi ile oluşmuştur. İlk kısımda başlık (header) ikinci kısımda ise veri (data) sunulmaktadır. Başlık kısmında her bir veri sınıfı (yani sütun) @ATTRIBUTE işareti ile sisteme tanıtılmakta (Örn: @attribute class {Positive, Negative, Neutral} gibi), ardından da veri kısmında veriler sunulmaktadır (Attribute-Relation File Format (ARFF), 2008). Çalışma kapsamında anlamı, amacı ve şekli açısından atıf sınıfları için ayrı ayrı arff derlemlerinin yaratılmasının ardından kelime ön işleme süreci için derlemler Weka yazılımı ile açılmıştır.

4.7.2. Kelime Ön İşleme Süreci

Türkçe atıflar için kelime ön işleme yöntemi olarak metinlerin sözcük vektörlerine dönüştürülmesi, bu dönüştürmede ise 1-2 gram (bigram) ve 1-3 gram (trigram) olarak ağırlıklandırılmış n-gram algoritması tercih edilmiştir. n-gram algoritmasının temel amacı verilen bir dizilimdeki tekrar oranının saptanması, yani ardışık n sayıda karakter dizisi yaratmaktır (Damashek, 1995, s. 843). n-gram'ın temel yapısını "Bu çalışmanın sonuçları olumludur" cümlesi üzerinden örneklendirmek gerekirse; "Bu-çalışmanın-sonuçları-olumludur" unigram, "Bu çalışmanın-çalışmanın sonuçları-sonuçları olumludur" bigram, "bu çalışmanın sonuçları-çalışmanın sonuçları olumludur" trigram kelime işleme olarak kabul edilebilir. Bu bağlamda çalışma kapsamında anlamı ve amacı açısından sınıflanan atıflar için kelimelerin birlikte geçme sıklıklarının dikkate alınarak değerlendirme yapan kelime n-gram işleme; şekli açısından atıflar için ise karakterlerin geçiş sıklığını değerlendiren karakter n-gram işleme kullanılmıştır. Şekli açısından atıflarda değerlendirme yönteminin

değiştirilmesinin temel sebebi şekline göre atıflarda kelimelerden çok kullanılan karakterlerin (kesme (') işareti, parantezler gibi) önemli olmasıdır.

Kelime ön işlemenin ikinci aşamasından beklenen dur sözcüklerinin (stop words) ayıklanmasıdır. Ancak, Türkçe için literatürde belirlenmiş dur sözcüklerini içeren derlemeler (Aksoy ve Öztürk, 2016; Alexander, 2016; Ranks.nl, 2017) kontrol edildiğinde atıfların ayırt edilmesini sağlayan pek çok sözcüğün aslında dur sözcüğü olduğu ortaya çıkmıştır.¹⁰ Dolayısıyla bu çalışma kapsamında yapılan analizlerde Türkçe atıfların ayırt edilmesini *ancak*, *dolayısıyla*, *dolayı*, *göre*, *hiçbir* gibi dur sözcükleri sağladığından bu sözcükler derlemden çıkarılmaksızın analiz gerçekleştirilmiştir.

4.7.3. Algoritmaların Uygulanması

Araştırma kapsamında Türkçe atıfların makinece ayırt edilmesi sürecinde istatistiğe dayalı bir algoritma olan Naive Bayes Multinomial ve temelde bir karar ağacı algoritması olan Random Forests algoritmaları en başarılı sonuçlar verdiği için çalışma kapsamında yalnızca bu iki algorithmadan elde edilen başarımlar raporlandırılmıştır. Her iki algoritma için de tanımlayıcı bilgi vermek gerekirse;

- *Naive Bayes Multinomial (Çok terimli Naive Bayes)*: Naive Bayes algoritması Bayesyen teoremini dikkate alarak sınıflama yapan basit olasılıkçı bir sınıflama türüdür (Duda ve Hart, 1973). Naive Bayes rastgele değişkenlerin koşullu olasılığını ölçer ve gözlenen verileri başka rastgele değişkenler üzerine uygular (Aggarwal, 2015, s. 306). Naive Bayes Multinomial algoritmasında ise belgedeki kelimelerin dağılımı çok terimli olarak modellenir ve belgeler kelime dizileri olarak ele alınır. Bu aşamada her kelimenin pozisyonunun diğerinden bağımsız olarak oluşturulduğu varsayılır (McCallum ve Nigam, 1998, s. 44). Yani multinomial algoritmasında Naive Bayes modeline göre terimlerin var olup olmadığından çok ne sıklıkla tekrar ettiği dikkate alınır.
- *Random Forests (Rastgele orman)*: Random Forests algoritması üçüncü bölümde metin kategorizasyonunda kullanılan tekniklerden karar ağaçlarının altında değerlendirilebilecek bir algoritmadır. Random Forests algoritmasında çok sayıda karar ağacından oluşan orman sınıflandırılmaktadır. Breiman (2001,

¹⁰ Pozitif ve negatif atıf cümlelerinde ayırt ediciliği sağlayan sözcüklerin listesi için http://115k440.hacettepe.edu.tr/atif_sozcukleri/atif_sozcukler.html adresi ziyaret edilebilir.

s. 25) tarafından geliştirilen bu algoritmanın temel amacı rastgele karar ağaçlarından öğrenilmiş öznitelikler kullanılarak bir orman yaratılmasıdır. Bu sayede olasılıklar üzerinden tahmin yürütme başarısı da sağlanabilmektedir.

Araştırmanın beşinci bölümünde sözü edilen iki algoritmanın uygulandığı derlemlerdeki sonuçlar sunulmaktadır.

4.7.4. Algoritma Performanslarının Değerlendirilmesi

4.7.4.1. Metodolojik Değerlendirme

Algoritmaların metodolojik açıdan isabetli sonuçlar verip vermediklerini doğrulamak amacı ile çalışma kapsamında K katlı çapraz doğrulama (K-fold cross validation) yapılması tercih edilmiştir. Bunun temel sebebi K katlı çapraz doğrulamanın en uzun süren doğrulama yöntemi olması zorluğunun yanında en doğru sonuçlara ulaştırmasıdır. Çünkü analiz aynı veri seti üzerinde defalarca test edilmektedir. Araştırma kapsamında kat miktarı olarak literatürde de sıklıkla tercih edilen 10 katlı doğrulama seçilmiştir. Çapraz doğrulama yönteminde veri seti rastgele 10 eşit parçaya ayrılır. Ayrılan bu veri parçalarından her defasında bir tanesi test için, geri kalan dokuz tanesi ise eğitim amaçlı kullanılır. Bu işlem 10 kez tekrarlanır ve bu tekrarların ortalaması alınarak sonucun doğruluğu hesaplanır (Kohavi, 1995, s. 1138).

4.7.4.2. Sayısal Değerlendirme

Çalışmada yapılan sayısal değerlendirmelerin raporlanmasında algoritma başarı oranı (doğru sınıflanan atıflar/veri setindeki tüm atıflar) ve üçüncü bölümde de detayları anlatılan *f*-ölçütü değerleri kullanılmıştır. Bunun yanında her bir uygulamanın başarımının ölçümünde elde edilen karışıklık matrisleri de raporlanmıştır. Karışıklık matrislerinde gerçek kategoriler ile tahmin edilen kategoriler karşılaştırılarak doğru pozitif, doğru negatif, yanlış pozitif ve yanlış negatiflerin ortaya çıkarılması hedeflenmektedir. Bu matrislerin genel yapısı üçüncü bölümde yer alan Tablo 3'te gösterildiği gibidir (bkz. sayfa 34).

5. BÖLÜM: BULGULAR VE DEĞERLENDİRME

5.1. GİRİŞ

Bu bölümün ilk kısmında araştırma sonucunda elde edilen bulgular sunulmakta, ikinci kısmında ise elde edilen bu bulgulardan hareketle Türkçe kütüphanecilik ve bilginbilim literatürü için bir değerlendirme modeli önerilmektedir.

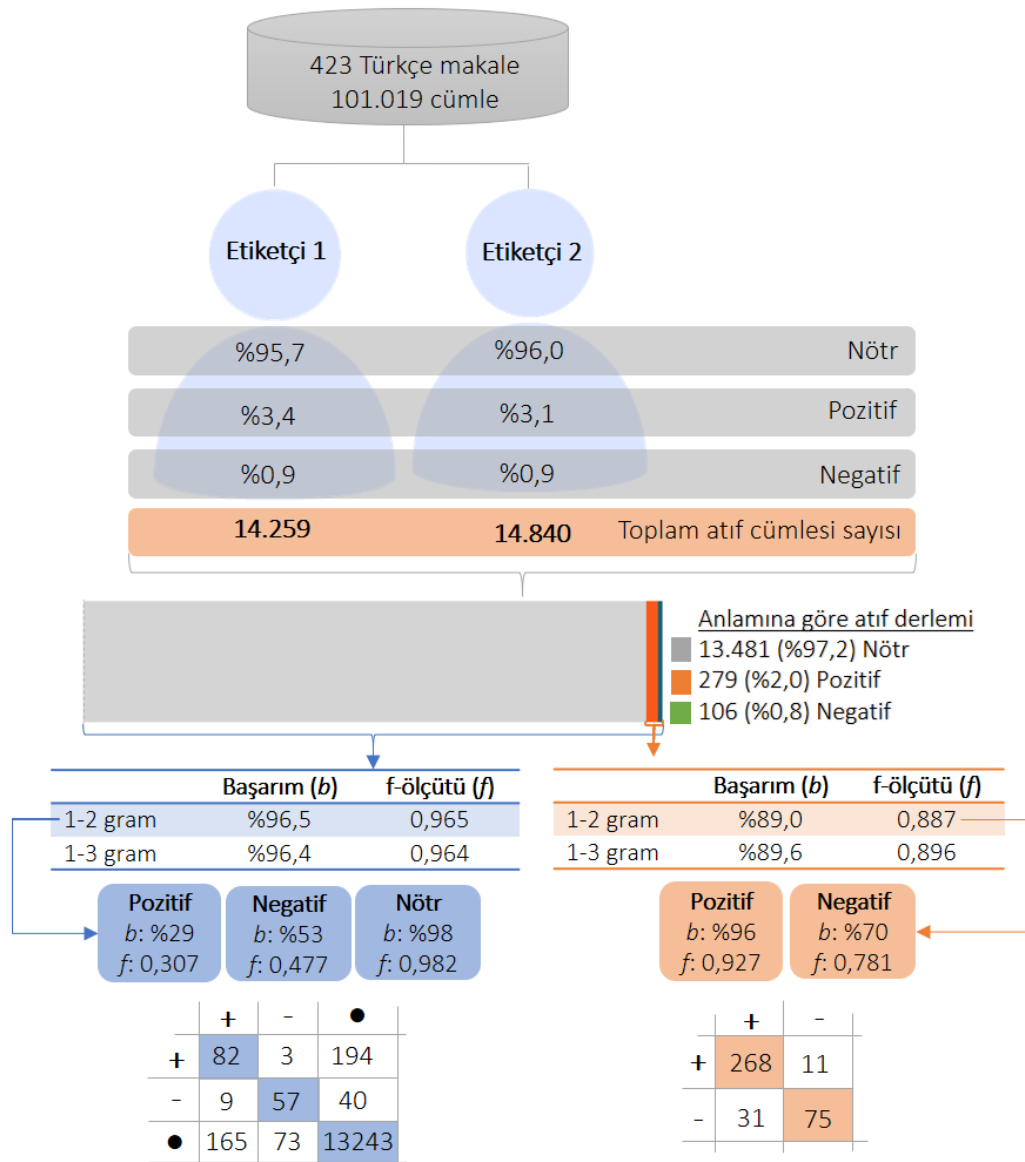
5.2. BULGULAR

Araştırma kapsamında “Yöntem” bölümünde aktarılan ve literatüre dayalı olarak oluşturulmuş taksonomik atıf sınıfları kullanılarak her bir sınıf için makine öğrenmenin başarımı test edilmiştir. Her bir bölümde ilgili bölüm için yapılan analiz bir infografik yardımı ile görselleştirilerek yürütülen işlemlerin anlaşılması kolaylaştırılmaya çalışılmıştır.

5.2.1. Anlamı Açısından (Pozitif, Negatif ve Nötr) Atıfların Sınıflandırılması

Etiketleme süreci sonucunda tüm derlem iki uzman etiketçi tarafından pozitif, negatif ve nötr olarak sınıflarına ayrılmıştır. Anlamı açısından atıflar ile ilgili işlem basamakları ve sınıflandırma işleminin başarım sonuçları Şekil 14’te gösterildiği gibidir.

Şekil 14’te de açıkça görüldüğü üzere 423 makalede yer alan 101.019 cümlelerin atıflarına ayrılması sürecinde etiketçiler birbirine yakın oranlarda pozitif, negatif ve nötr atıf cümlesi saptamışlardır. Araştırma kapsamında ilk etiketçi toplam 14.259 atıf cümlesi işaretlerken; ikinci etiketçi ise 14.840 cümle belirlemiştir. Bu atıfların anlamına göre sınıflara dağılımı birbirine benzer niteliktedir. Her iki etiketçinin yaptığı etiketleme sonuçları arasından birbirine benzeyen atıf cümleleri saptanmış ve bu sayede çalışmanın temel anlamına göre atıf derlemi oluşturulmuştur. Bu derlemde yer alan atıf cümlelerinin %97,2’si nötr, %2’si pozitif ve %0,8’i negatiftir.



Şekil 14. Anlamına göre atıfların sınıflandırılması ve makine tarafından saptanması süreci

Literatürde yapılmış pek çok çalışma metin içindeki anlam sınıfları dağılımının genellikle bu çalışma ile benzer sonuçlar verdiğini kanıtlamaktadır. Örneğin, literatürdeki bir çalışmada incelenen toplam 2.309 atıfın %2,4'ünün pozitif, %0,4'ünün ise negatif olduğu saptanmıştır (Spiegel-Rösing, 1977, s. 105). Pozitif, negatif ve nötr atıfları bilimsel makaleler üzerinden saptamaya çalışan bir araştırmada da 310 makale incelenmiş toplam atıfların %3'ünün negatif, %10'unun pozitif ve geri kalanının nötr atıflardan oluştuğu ortaya çıkarılmıştır (Athar, 2011, s. 82). Athar'ın hazırladığı diğer çalışmada da (2014, s. 36) atıfların %3,2'si negatif, %9,5'i pozitif ve %87,3'ü nötr olarak belirlenmiştir. Cano'da (1989, s. 286) yaptığı çalışmada tüm atıf sınıfları içinde en az karşılaşılan atıf türünün %2 oranı ile negatif atıflar olduğunu saptamıştır. Yüksek mahkemelerde

otoritelere yapılan atıfların içeriksel açıdan incelendiği bir makalede ise pozitif atıfların oranı %33, negatif atıfların oranı ise %8 olarak hesaplanmıştır (Johnson, 1985, s. 513). Klinik denemelerde yapılan atıfların anlamsal analizini içeren bir çalışmada da atıflar anlamına göre sınıflandırıldığında %17'sinin pozitif; %7'sinin negatif ve %76'sının nötr olduğu görülmüştür (Xu, Zhang, Wu, Wang, Dong ve Xu, 2015, s. 1338). Araştırma sonucunda elde edilen veriler ilgili literatür ile karşılaştırıldığında derlemlerde yer alan anlamsal sınıfların genellikle benzer şekilde dağıldığı görülmektedir. Diğer taraftan bu sonuçlar aynı zamanda negatif atıfların toplam atıflar içindeki yerinin oldukça az olduğunu da ortaya çıkarmıştır. Bunun temel sebebi yazarların negatif atıf yapmaktan kaçınması olarak yorumlanabilir. Bilimin eleştirisi ile geliştiği aşikârdır ancak bu eleştiriler tepki toplamamak amacı ile genellikle üstü kapalı bir dille ifade edilmektedir. Bu da sonuç olarak olumsuz yorumlardan kaçınılması şekline dönüşebilmektedir.

Pozitif ve negatif atıf cümlelerine daha yakından bakıldığında pozitif atıflar için genellikle “dikkat çekmektedir”, “güzel bir ömektir”, “incelemeye değerdir”, “son derece önemlidir” “açısından değerlidir” ve “açısından önemini ortaya koymuştur” gibi ifadeler kullanılırken negatif atıflarda ise “ancak”, “ama”, “-diği halde” ve “düşündürücüdür” şeklinde ifadelere yer verilmiştir.¹¹ Bu durum negatif atıfların genellikle üstü kapalı olarak yapıldığı savını güçlendirmektedir. Athar'ın (2011, s. 82) da ifade ettiği gibi yazarlar negatif atıf yapacakları zaman genellikle ya önce olumlu bir şeyden bahsedip ardından olumsuzluk anlamına gelen bir bağlaç yardımı ile eleştiriyi pozitif görüş ile birlikte sunarak törpülemekte ya da “yanlıştır” gibi sert sözcükler yerine daha yumuşak bir söylem olan “doğru değildir” gibi ifadeleri kullanmaktadırlar. Bir başka çalışmada da benzer bir konuya değinilmekte ve pozitif ve negatif atıf yapma motivasyonlarının genellikle aynı olduğu ve bu iki atıf türü arasında bir korelasyon bulunduğu iddia edilmektedir (Brooks, 1986, s. 34). Yani bir yazarın bir çalışmaya atıf yaparken tepki toplamamak için önce yazının olumlu taraflarını vurguladığını, ardından ise asıl eleştirinin yapıldığını belirtmektedir. Bu çalışmada da negatif ve pozitif atıflardaki bu söylem farklılıklarında benzer sonuçlar ortaya çıkmıştır.

Anlamına göre atıf cümleleri için bir derlem yaratılmasının ardından Weka yazılımında çeşitli analizler yapılabilmesi için derlem Weka'nın kullandığı dosya formatı olan arff formatına uygun hale getirilmiştir. “Yöntem” bölümünde de detaylı olarak bahsedildiği gibi analiz uygulanmadan önce kelime ön işleme safhası için söz dizileri (string) n-gram

¹¹ Negatif ve pozitif atıf yapılırken tercih edilen sözcüklerin listesi için http://115k440.hacettepe.edu.tr/atif_sozcukleri/atif_sozcukler.html adresi ziyaret edilebilir.

parçalayıcısı (n-gram tokenizer) kullanılarak sözcük vektörlerine dönüştürülmüştür. Bu işlem için 1-2 gram ve 1-3 gram seçenekleri test edilmiş, her iki sınıflama da birbirine benzer sonuçlar vermiştir. 1-2 gram ve 1-3 gram sonuçları Şekil 14'te yer almaktadır.

Kelime ön işleme aşamasının ardından öncelikle pozitif ve negatif atıf cümleleri arasında kullanılan dil açısından anlamlı bir fark olup olmadığını saptamak amacı ile yalnızca pozitif ve negatif atıf cümlelerinin yer aldığı veri seti analiz edilmiştir. Uygulamada en iyi sonuç benzer özelliklere sahip metinlerin sınıflandırılması için kullanılan Naive Bayes Multinomial ve Random Forests algoritmalarından alınmıştır. Random Forests algoritması pozitif atıf cümlelerinin tamamını, negatif cümlelerin ise %98'ini saptayarak oldukça isabetli bir başarıya ulaşsa da ($f=0,982$), nötr cümleler analize dâhil edildiğinde başarı oranı ciddi bir düşüş gösterdiğinden çalışma kapsamında yalnızca Naive Bayes Multinomial algoritmasından elde edilen sonuçlar raporlanmıştır. 10 kat çapraz doğrulama ile başarıyı test edilen analiz sonucunda pozitif ve negatif atıf cümlelerinin makinece sınıflandırılmasında %89'luk bir başarı elde edilmiştir. Pozitif atıf cümlelerinin %96'sı, negatiflerin ise %70'i saptanabilmiş; pozitif iken negatif olarak sınıflanan atıf cümlesi sayısı 11, negatif iken pozitif olarak sınıflananların sayısı ise 31 olarak belirlenmiştir. Analiz sonuçlarına ilişkin karşılıklı matrisi (confusion matrix) de Şekil 14'te gösterilen infografikte detaylandırılmaktadır. Bu bulguya oldukça benzer bir bulgu da literatürde yer alan bir çalışmada elde edilmiştir (Jha, Jbara, Qazvinian ve Radev, 2016, s. 103). Bu çalışmada da pozitif ve negatif atıflar doğal dil işleme teknikleri ile sınıflandırılmış ve pozitif atıflar için %93, negatif atıflar için de %78 oranında başarıya ulaşılmıştır. Tüm bu sonuçlar pozitif atıflar ile negatif atıflar arasında kullanılan dil açısından ayırt edilebilirliği sağlayan önemli bir farklılık bulunduğunu ortaya koymakta; pozitif atıfların negatif atıflara göre daha belirleyici bir dile sahip olduğu görüşünü desteklemektedir. Diğer taraftan negatif atıf cümlelerinde başarımın pozitiflere göre daha düşük olmasının bir sebebinin de veri setinde yer alan atıf sayılarının farklılığı olduğu düşünülmektedir. Bunun doğruluğunu test etmek amacı ile pozitif ve negatif atıf cümleleri rastgele pozitif atıflar çıkarılarak eşitlenmiş ve analiz tekrar edilmiştir. Bu analizde negatif atıfların başarımı önemli oranda artmış ($f=0,871$) ve pozitif atıf saptama başarımına ($f=0,874$) yaklaşmıştır. Bu durumda derlemde yer alan negatif atıf cümlesi sayısının fazla olması halinde başarı oranlarının artabileceği düşünülmektedir. Ancak negatif atıflara doğası gereği çok fazla rastlanmadığından bunu sağlamak oldukça zordur.

Pozitif ve negatif atıfların dil kullanımı açısından birbirinden otomatik olarak ayrılma başarımının test edilmesinin ardından aynı analiz tüm derlem üzerinden (pozitif, negatif ve nötr atıflar) gerçekleştirilmiştir. Bu analiz sonucunda sınıflama algoritmasının genel

başarımı %96 olarak hesaplanmıştır ($f=0,965$). Bu denli yüksek başarı oranının en temel sebebi derlemde yer alan nötr atıf cümlesi sayısının ve nötr atıf saptama başarısının genel başarıyı etkilemesinden kaynaklanmaktadır. Zira nötr atıflar %98 oranında doğru bir şekilde sınıflandırılabilir (= $0,982$). Bunun anlamı da çok az sayıda nötr atıfın pozitif ve negatif atıf sınıfları ile karıştığıdır. Şekil 14'te verilen infografikte yer alan karışıklık matrisi incelendiğinde de bu durum açıkça görülmektedir. Aynı yorumu pozitif ve negatif atıflar için yapmak ise zordur. Pozitif atıf saptama başarımı %29 oranında kalırken ($f=0,307$); negatif atıflarda ise bu başarı %53 olarak ölçülmüştür ($f=0,477$). Bu bulgulara göre tüm derlem üzerinden analiz yapıldığında Türkçe için negatif atıf cümlesi sınıflama başarımı sayıları daha az olmasına rağmen pozitif atıf cümlelerine oranla daha yüksek düzeydedir. İçerik tabanlı atıf analizi literatürü ile karşılaştırıldığında bu bulguyu destekleyen ve desteklemeyen bazı çalışmalar bulmak mümkündür. Bu bulguya benzer sonuç üreten ve analiz aracı olarak da benzer algoritma ve teknikleri kullanan bir çalışmada nötr atıflar için %86 ($f=0,883$), negatif atıflar için %68 ($f=0,614$) ve pozitif atıflar için %61 ($f=0,563$) oranında başarı elde edilmiştir (Athar, 2014, s. 79). Söz konusu çalışmada da negatif atıf başarımı pozitif atıf başarımından daha yüksek düzeyde çıkmıştır. Bu çalışmanın bulguları ile benzeşmeyen bir çalışmada (Xu, Zhang, Wu, Wang, Dong ve Xu, 2015, s. 1339) 1-2 gram kelime işleme ile elde edilen negatif atıf başarımında f -ölçütü 0,498, pozitif atıflarda $f=0,719$ ve nötr atıflarda $f=0,924$ olarak saptanmıştır. Sözü edilen çalışmalar ile karşılaştırıldığında bu çalışmada elde edilen nötr atıf cümlesi saptama başarımının diğer çalışmalara göre daha yüksek olduğu görülürken; bu çalışmadaki pozitif ve negatif atıf cümlesi saptama başarımının artırılması gerekliliği de ortaya çıkmıştır. Atıf yapılan iki çalışmada n-gram kelime işlemesine ek olarak çeşitli dilbilimsel eklemeler (duygu sözlüğü kullanılması, gövdeleme gibi) yapılarak makine başarımı artırılabilmiştir. Örneğin, Xu ve diğerlerinin yaptığı çalışmada (2015, s. 1339) 1-2 gram kelime işleme yanında duygu sözlüğü ve yapısal gövdeleme kullanılarak negatif atıf başarımı $f=0,551$ 'e; pozitif atıf başarımı ise $f=0,723$ 'e yükseltilebilmiştir. Athar'ın çalışmasında da 0,138 olan negatif atıf başarımının f -ölçütü çeşitli duygu analizi eklemeleri ile $f=0,614$ 'e yükseldiği görülmektedir.

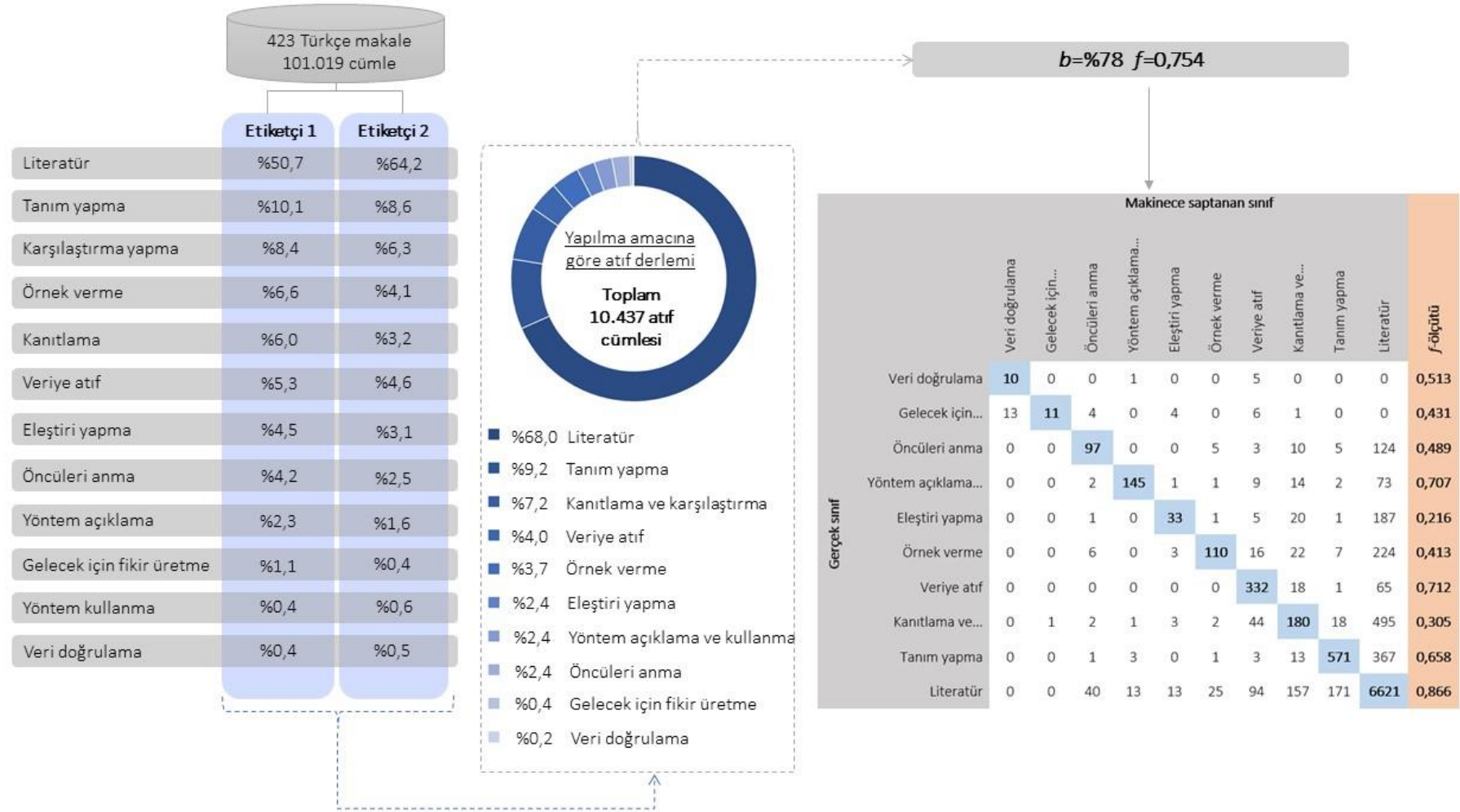
5.2.2. Yapılma Amacı Açısından Atıfların Sınıflandırılması

Yöntem kısmında da belirtildiği gibi yapılma amacı açısından atıflar 12 farklı sınıfa ayrılmıştır. Bu sınıflar literatür, tanım yapma, karşılaştırma yapma, örnek verme, kanıtlama, veriye atıf, eleştiri yapma, öncülerini anma, yöntem açıklama, gelecek için fikir

üretme, yöntem kullanma ve veri doğrulama olarak belirlenmiştir. Etiketçi performansları karşılaştırıldığında yöntem açıklama ve yöntem kullanma atıfları ile kanıtlama ve karşılaştırma atıflarının birbirlerine çok benzemeleri sebebi ile karıştırıldığı saptanmıştır. Sonuç olarak bu iki atıf türü birleştirilmiş ve analizler bu birleşimler üzerinden yapılmıştır. Belirlenen sınıflara ilişkin detaylı bilgiler, derlem özellikleri ve makine başarımları Şekil 15'te yer alan infografikte sunulmaktadır.

İnfografik incelendiğinde yapıma amacı açısından en çok atıf yapılan türün literatür atıfları olduğu görülmektedir. Her iki etiketçi de atıf cümlelerinin yarısından fazlasını literatür atıfı olarak işaretlemiştir. Etiketçi performansları karşılaştırılıp birbirine benzeyen atıf cümleleri ortaya çıkarıldığında analiz edilecek derlemde yer alan literatür atıflarının oranı %68 olmuştur. Literatür atıflarının ardından en çok rastlanan atıf türü ise tanım yapma atıflarıdır. Türü ne olursa olsun pek çok çalışma temel kavram ve terimlerin tanımlanması ile başladığından böyle bir bulgu ortaya çıkması beklenebilir. Derlemde yer alan tanım yapma atıflarının oranı %9,2 olarak belirlenmiştir. Kanıtlama ve karşılaştırma atıfları ise etiketçiler tarafından genellikle birbiri ile karıştırılmıştır. Türk kütüphanecilik literatüründeki yazarların genellikle kanıtlama amacı ile karşılaştırma yaptıkları görülmüştür. Bu sebeple bu iki sınıf birleştirilmiş ve analiz derleminde yer alan kanıtlama-karşılaştırma atıflarının oranı %7,2 olmuştur. Bu türü %4 ile veriye atıf, %3,7 ile örnek verme takip etmiştir. Üç atıf türü (yöntem açıklama ve kullanma, eleştiri yapma ve öncülerini anma) derlemde eşit oranda temsil edilmektedir (%2,4). Derlemde en az temsil edilen türler ise gelecek için fikir üretme atıfları ile veri doğrulama atıfları olmuştur. Tüm bu atıf türlerinin birleştirilmesi ile birlikte analiz derleminde toplam 10.437 atıf cümlesi yer almıştır.

Yapılma amacına göre atıflar 1-2 gram n-gram parçalayıcısı ile kelime ön işleme yapılmasının ardından Naive Bayes Multinomial algoritması kullanılarak makinece sınıflandırıldığında makinenin genel başarımları %78 olarak hesaplanmıştır. Bu oranın en önemli tespit başarısı literatür atıflarından kaynaklanmaktadır. Bu durum bir önceki kısımdaki nötr atıflarda saptanan başarımla benzerlik göstermektedir. Literatür atıflarının %92'si makine tarafından saptanabilmiştir ($f=0,866$). Literatür atıflarının en çok karıştığı sınıflar ise tanım yapma ve kanıtlama-karşılaştırma atıfları olmuştur. Literatür atıflarının ardından en yüksek başarımlar ise veriye atıf ($b=%80$, $f=0,712$) ve veri doğrulama ($b=%63$, $f=0,513$) atıfları için saptanmıştır. Veri doğrulama atıflarının derlem içindeki oranı oldukça düşük iken bu denli yüksek bir başarımla bu tür atıflarda kullanılan dilin oldukça ayırt edici olduğunu göstermektedir. Yöntem açıklama-kullanma atıflarının başarımları %59 olsa da f -ölçütü veri doğrulama atıflarından daha anlamlı sonuç vermiştir ($f=0,707$).

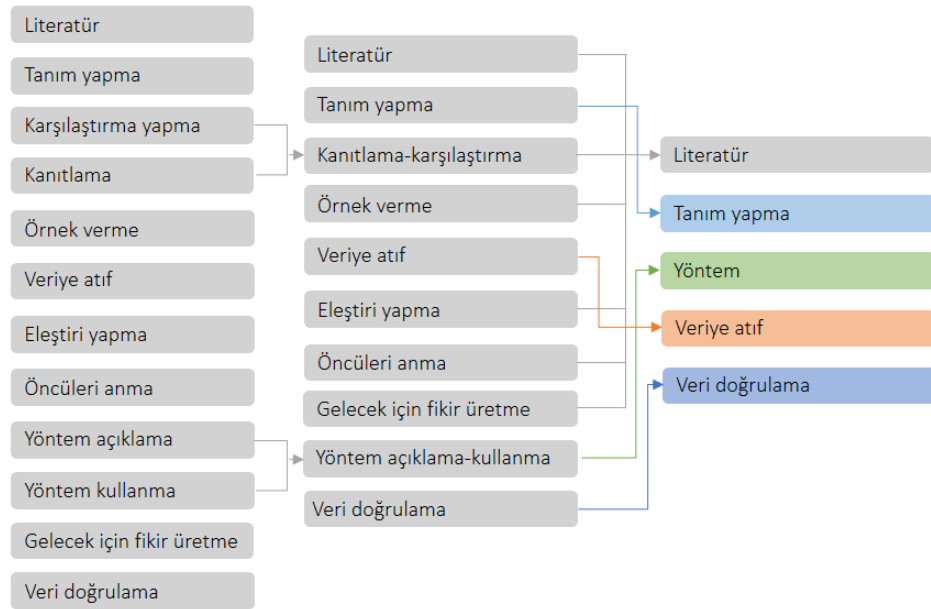


Şekil 15. Yapılma amacına göre atıfların sınıflandırılması ve makine tarafından saptanması süreci

Literatürde atıfların taksonomik olarak sınıflandırılması ile ilgili 1960'lı yıllardan günümüze çeşitli çalışmalar yapılmıştır ve bu çalışmalar yapılan tez çalışmasının “yapılma amacı açısından atıflar” bölümünün temel dayanağını oluşturmaktadır. Örneğin, 1965 yılı ile 1989 yılları arasında geliştirilen atıf sınıflama şemalarının değerlendirildiği bir çalışmada (Ding ve diğerleri, 2014, s. 1825) içerik tabanlı atıf analizi alanında öncü sekiz makale¹² incelenmiş ve ortak yönleri kullanılarak bir gruplama yapılmıştır. Bu gruplamaya göre yöntem atıfları ayrı bir grupta değerlendirilirken kavramsal çerçeve, arkaplan bilgi ve geçmiş deneyimlerin anılması diğer üç grubu oluşturmuştur. Karşılaştırma, kanıtlama, doğrulama, veriye atıf gibi alt gruplar ise geçmiş deneyimlerin anılması ana grubunun altında yer almıştır. Daha güncel çalışmalarda ise (Angrosh, Cranefield ve Stanger, 2010; Dong ve Schäfer, 2011; Teufel, Siddharthan ve Tidhar, 2006) sınıflama şemalarının genellikle çoğunda pozitif, negatif ve nötr sınıflar ana sınıflar olarak yer almış; bunun yanında literatür çalışmaları, alternatif yaklaşımlar, karşılaştırma, yöntem ve teknik ile ana fikir gibi temel sınıflar kullanılmıştır. Bu çalışmaya çok benzer bir sınıflama ortaya koyan Dong ve Schäfer (2011, s. 624), Naive Bayes ve Sequential Minimal Optimization (SMO) gibi algoritmalar kullanarak sınıfladığı atıfları dört temel gruba ayırmıştır. Bu gruplar arkaplan bilgi verme, çalışmanın ana fikrinin kaynağını anmak, teknik altyapı ve karşılaştırma olarak belirlenmiştir. Çalışmadaki makine sınıflamasının başarımları ise çeşitli farklı algoritmalar ile $f=0,510$ ile $f=0,670$ arasında hesaplanmıştır. Bir diğer çalışmada ise literatür (başka çalışmalardan kısaca bahseden) atıfları %60 oranında saptanabilirken diğer türden atıfların saptanmasında başarımlar %18'i geçememiştir (Sendhilkumar, Elakkiya ve Mahalakshmi, 2013, s. 417). Sayılan bu literatür çalışmalarından hareketle Türkçe kütüphanecilik ve bilgin bilim literatüründe 10 farklı sınıf üzerinden analiz gerçekleştirmektense bu sınıfların daha temel ana gruplara ayrılmasının daha anlamlı sonuçlar vereceği görülmüştür. Zira derlemede yer alan bazı sınıflar literatür atıfları ile benzer özellikler taşımaktadırlar. Örneğin, örnek verme, öncülerini anma, karşılaştırma-kanıtlama ve eleştiri yapma atıfları literatür ana başlığı altında değerlendirilebilecek atıflardandır. Örnek verilirken ya da literatürün öncülerini anılırken literatür arkaplan bilgisinden yararlanılmaktadır. Eğer bu atıflarda pozitif ya da negatif bir yaklaşım var ise bu zaten anlamı açısından yapılan sınıflamada %94 başarımlar oranı ile ayırt edilebilmiştir. Bu bakımdan bu kısımda temel amaç atıfların

¹² Söz konusu makaleler şunlardır: Chubin ve Moitra, 1975; Frost, 1979; Lipetz, 1965; McCain ve Turner, 1989; Moravcsik ve Murugesan, 1975; Oppenheim ve Renn, 1978; Peritz, 1983; Spiegel-Rösing, 1977.

yapılma amaçlarının sınıflandırılmasıdır. Atıflar yapılma amaçları açısından temel gruplara ayrıldığında beş ana sınıf belirlenmiştir. Bu beş ana sınıf Şekil 16'da gösterildiği gibidir.



Şekil 16. Yapılma amacı açısından temel atıf sınıfları

Atıflar temel sınıflara ayrıldığında oluşan sınıflar literatür, tanım yapma, yöntem, veriye atıf ve veri doğrulama olarak belirlenmiştir. Veri doğrulama atfı cümlelerinin sayısı derlemin çok küçük bir kısmını oluştursa da anlam itibariyle değerli atıflar olduğundan ayrı bir atıf türü olarak değerlendirme kapsamına alınmıştır. Çünkü veri doğrulama atıfları ile yazarlar kendi çalışmalarında buldukları sonuçları doğrulamak amacı ile geçmişteki çalışmaları anmaktadırlar. Bu anma da bu tür atıflara diğer literatür atıflarından daha fazla önem yüklediği anlamına gelmektedir. Araştırma yöntemini açıklamak ya da kullanılan yöntemi anmak üzere yapılan atıflar ise araştırmanın temelini dayandırıldığı kaynaklar olması açısından ayrı bir sınıf olarak değerlendirme kapsamına alınmıştır. Türkçe kütüphanecilik ve bilgilim literatüründe tanımların önemli bir yere sahip olduğu ortaya çıkarıldığından bu tür atıflar da ana sınıflardan birini oluşturmuştur. Bu çalışmada ortaya koyulan temel taksonomik sınıflandırma bir önceki paragrafta belirtilen literatür sınıflamaları ile benzerlik göstermesi açısından önemli görülmektedir.

Yapılma amacı açısından atıflar için gerçekleştirilen son analiz ise temel taksonomik sınıfların makinece saptanmasındaki başarımın hesaplanmasıdır. Yine kelime ön işleme aşamasında 1-2 gram sözcük parçalayıcısından yararlanıldığı uygulamada Naive Bayes

Multinomial algoritması kullanılarak analiz gerçekleştirilmiştir. Yapılan bu analizin genel başarımları oranı %90,4 ve *f*-ölçütü 0,905 olarak hesaplanmıştır. Analiz sonuçları ve karışıklık matrisi Tablo 5'te gösterildiği gibidir.

Tablo 5. Yapılma amacı açısından atıflar için analiz sonuçları ve karışıklık matrisi

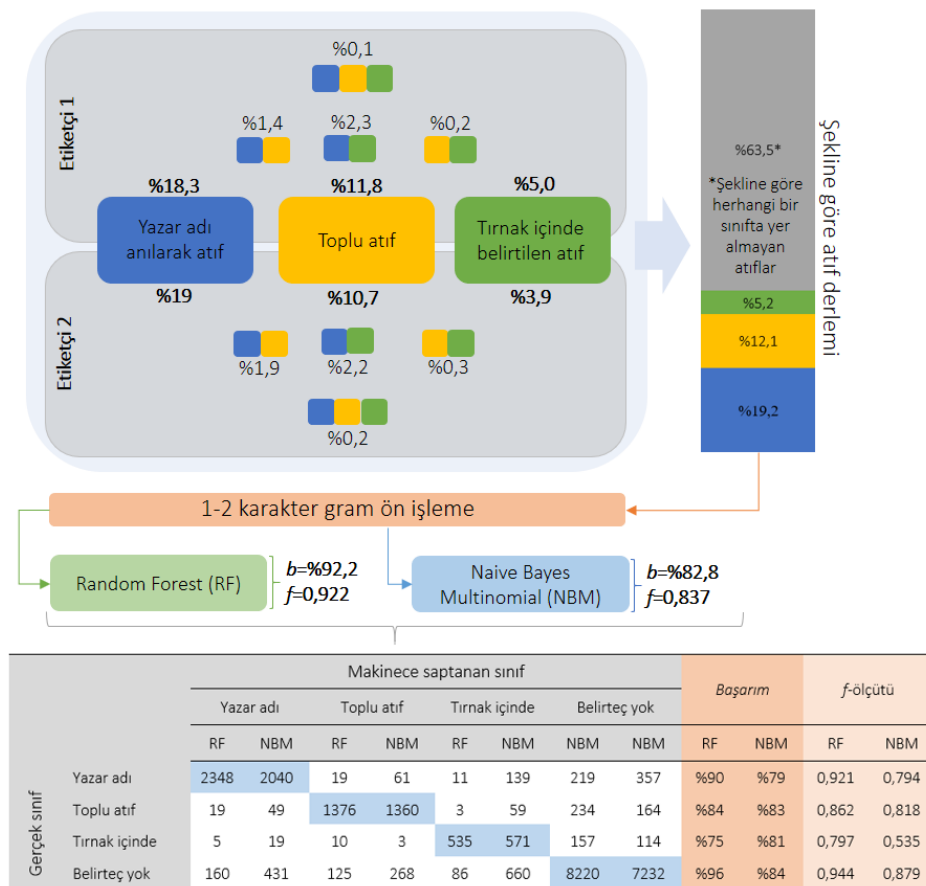
		Makinece saptanan sınıf					Başarım	<i>f</i> -ölçütü
		Literatür	Tanım yapma	Yöntem	Veriye atıf	Veri doğrulama		
Gerçek sınıf	Literatür	8286	270	30	207	6	%94	0,944
	Tanım yapma	328	621	4	6	0	%65	0,670
	Yöntem	73	3	161	10	0	%65	0,729
	Veriye atıf	61	0	0	355	0	%85	0,714
	Veri doğrulama	0	0	0	0	16	%100	0,842

Tablo 5'ten de görüldüğü üzere tüm başarımlar %65 ve üzerindedir. Veri doğrulama atıf cümlelerinin tamamı doğru sınıflanırken, literatür atıf cümlelerinin de %94'ü doğru bir şekilde sınıflandırılabilmiştir. Veri doğrulama atıfları ile ilgili bir diğer önemli bulgu da diğer sınıflardan yalnızca 6 atıf cümlesinin veri doğrulama olarak işaretlenmesidir. Bu durum veri doğrulama atıflarının oldukça karakteristik bir özelliği olduğunu ortaya koymaktadır. Bu sonuçların yanında diğer sınıflar ile karşılaştırıldığında en düşük başarımlar tanım yapma ve yöntem atıflarında görülmüştür. Bir diğer ilgi çekici bulgu ise yöntem atıflarının en çok tanım yapma atıfları ile karışıyor olmasıdır. Yani başarımların düşük olmasının sebebi bu iki türün genellikle birbiri ile karışmasıdır. Bu karışıklığın büyük çoğunluğu bir yöntemi açıklarken yapılan tanımlardan ileri gelmektedir. Genel çerçeveden bakıldığında ana sınıflara ayrılmış taksonomik atıf sınıfları üzerinden yapılan uygulama tüm sınıflar üzerinden uygulanan analize göre önemli oranda yüksek başarımlar vermiştir.

5.1.3. Yapılma Şekli Açısından Atıfların Sınıflandırılması

Türkçe kütüphanecilik ve bilginbilim literatüründe yapılmış atıf cümleleri yapımla şekillerine göre sınıflandırıldığında bu cümlelerin %63'ünde herhangi bir şekilsel belirteç bulunmadığı görülmüştür. En fazla karşılaşılan tür ise yazar adı anılarak yapılan atıflardır. Atıflar yapımla şekillerine göre, yani somut işaretlere dayanılarak sınıflandırıldığı için etiketçiler arasında yüksek oranda benzerlik görülmüştür. Şekline göre atıf derleminde yer alan atıf

cümlelerinin %19'unu yazar adı anılarak yapılan atıflar; %12,1'ini toplu atıflar ve %5,2'sini tırnak içinde yapılan atıflar oluşturmuştur (bkz. Şekil 17). Literatürdeki çalışmalar şekline göre en değerli atıf türlerinin yazar adı anılarak ve tırnak içinde özel olarak alıntılanarak yapılan atıflar olduğunu belirtmektedir (Bonzi, 1982, s. 211; Zhu, Turney, Lemire ve Vellino, 2015, s. 413). Uluslararası kütüphanecilik literatürünün değerlendirildiği bir çalışmada şekline göre en çok karşılaşılan atıf sınıfının yazar adı anılarak yapılan atıflar olduğunu doğrulamaktadır (Bonzi, 1982, s. 212). Şekline göre atıf derleminde yer alan toplam 13.527 atıf cümlesi için genel istatistikler ve sınıflama algoritmalarından elde edilen sonuçlar üzerine detaylı bilgiler Şekil 17'de sunulmaktadır.



Şekil 17. Yapılma şekli açısından temel atıf sınıfları

Şekil 17'de de gösterildiği gibi şekline göre atıfların sınıflandırılması amacıyla yapılan ön işleme sürecinde diğer iki türün aksine karakter n-gram ön işleme yapılmıştır. Bunun temel sebebi şekline göre atıflarda sözcüklerden ziyade harfler ve işaretlerin önemli olmasıdır (tırnak işareti, kesme işareti gibi - öm.: Taşkın'ın belirttiği gibi). Analiz sonuçları da bu öngörüğü doğrulamaktadır. 1-2 gram kelime ön işleme ile yapılan analizin başarımları %69'da kalırken (Naive Bayes Multinomial), 1-2 gram karakter ön işleme yapıldığında aynı algoritma için

başarım oranı %83'e çıkmıştır. Bu durum da Türkçe atflarda şekline göre sınıflama yapılırken karakter gram işlemenin daha isabetli sonuçlar verdiğini göstermektedir.

Daha önce pozitif ve negatif atfların sınıflandırılmasında başarı sağlayan ancak nötr atflar dâhil edildiğinde bu başarıyı koruyamayan Random Forests algoritması şekline göre atf sınıflarına uygulandığında ise yüksek oranda başarı sağlanabilmiştir (%92, $f=0,922$). Bu algoritma sayesinde özellikle yazar adı anılarak yapılan atflar ile herhangi bir belirteç taşımayan atfların sınıflandırılmasındaki başarı oranı önemli oranda artmıştır. Bunun yanında 0,797 ile 0,944 arasında değişen f -ölçütü değerleri de algoritmanın başarımını doğrulamaktadır. Bu algoritmanın daha yüksek başarı vermesinin en temel sebebi algoritmanın belirli kurallara dayanarak bir karar ağacı ormanı oluşturmasıdır. Şekline göre atflarda kurallar diğer sınıflara göre daha net olduğundan bu algoritmanın daha yüksek başarı verdiğini düşünülmemektedir. Algoritma başarımında en düşük başarı elde edilen sınıf ise tırnak içinde belirtilen atflar için saptanmıştır. Bunun temel nedeni "Yöntem" bölümünde de belirtildiği üzere metin içinde birebir alıntılanan bazı atfların tırnak işareti kullanılarak değil metnin biçimi değiştirilerek yapılmasıdır (paragrafın içeriden başlaması, daha küçük punto, italik yazılar gibi). Bunun yanında tırnak işareti diğer tür atflarda da özel olarak vurgulanmak istenen sözlerin veya cümle içindeki eserlerin ve yazıların adlarının gösterimi gibi amaçlarla da kullanılabilir (Noktalama işaretleri, 2016). Bu gibi durumlardan dolayı tırnak işareti içeren atfların saptanmasında diğer sınıflara oranla daha düşük başarı ve f -ölçütü değerlerinin elde edilmesi beklenen bir sonuçtur.

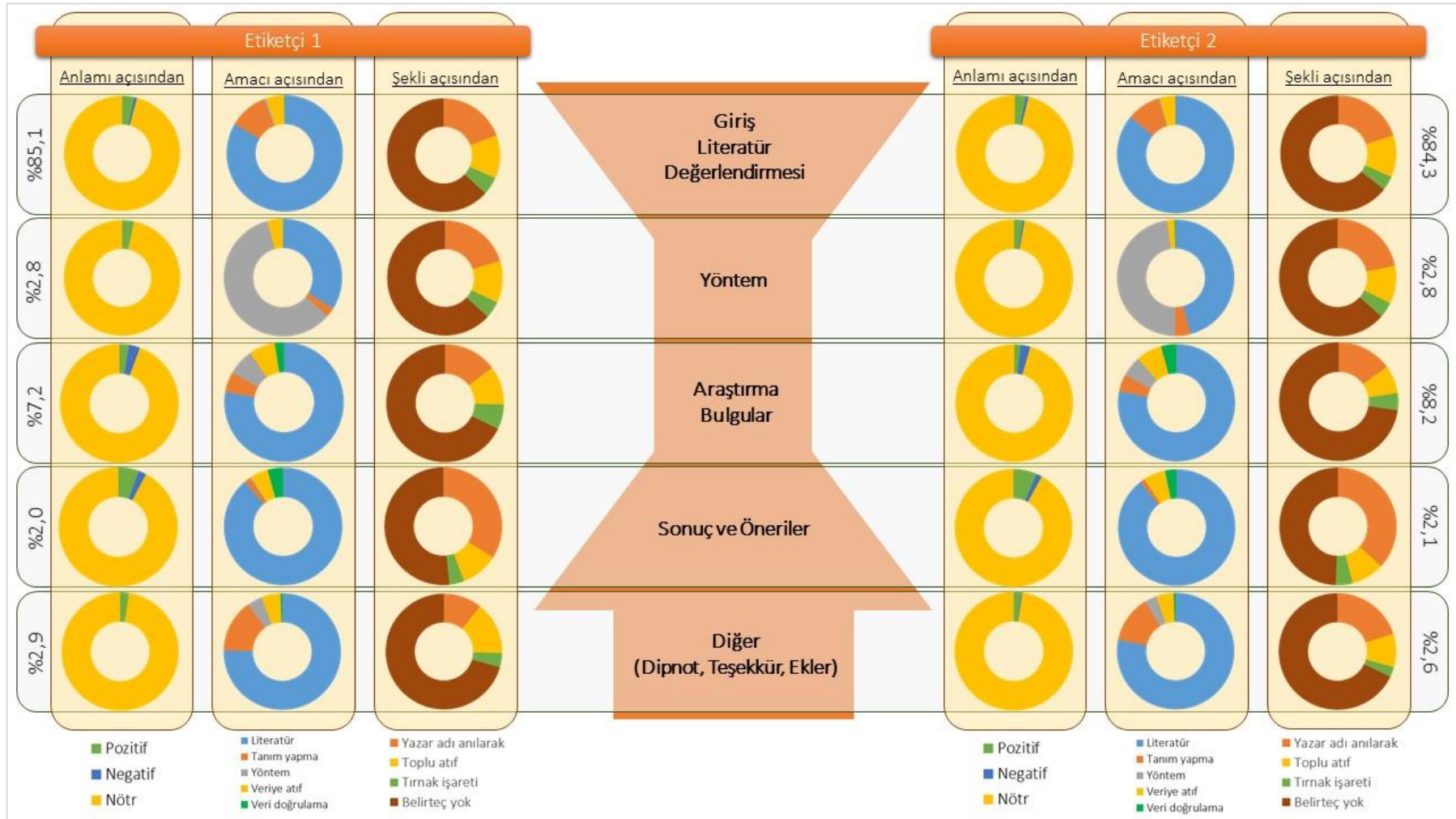
Şekline göre atf cümlelerinin makinece sınıflandırılması ile ilgili olarak literatürde fazla çalışma ile karşılaşılmamıştır. Önemli çalışmalardan birinde Athar (2014, s. 84-86) Naive Bayes ve Support Vector Machines (SVM) algoritmaları kullanarak hazırladığı algoritma aracılığı ile yazar adı içeren atfları saptama başarımında f -ölçütünü 0,463 olarak hesaplamıştır. Bu oran çalışmamızda elde edilen başarı oranı ile karşılaştırıldığında oldukça düşüktür.

5.1.4. Dizilimi Açısından Atfların Sınıflandırılması

Dizilimi açısından atfların sınıflandırılmasında atf yapılan bölüm ve metin içinde atf yapılma sayıları olarak iki yönlü bir analiz gerçekleştirilmiştir.

5.1.4.1. Atf Yapılan Bölümlerin Değerlendirilmesi

Türkçe kütüphanecilik ve bilginim literatüründeki atfların yapıldıkları bölümler ile ilgili analiz yapıldığında elde edilen sonuçlar Şekil 18'de görselleştirilmiştir.



Şekil 18. Dizilimi açısından atıfların bölümlere dağılımını gösterir infografik

Türkçe atıfların IMRAD yapısı içindeki dağılımı incelendiğinde en çok atıf yapılan bölümün literatür değerlendirmelerinin de yer aldığı giriş bölümü olduğu anlaşılmıştır. Her iki etiketçi de %84-%85 civarı atıf cümlesini bu bölümde işaretlemişlerdir. En çok atıf yapılan ikinci bölüm ise araştırma/bulgular olmuştur. Diğer bölümlerde yapılan atıfların oranları ise yüksek oranda benzer bir dağılım göstermiştir.

Şekil 18'de de görüldüğü gibi her iki etiketçi tarafından gerçekleştirilen etiketlemede de önemli benzerlikler saptanmıştır. Bu açıdan bakıldığında atıflar için bir genelleme yapılabilmesi açısından yakın sonuçlar elde edilmiş olması önemli bir bulguya işaret etmektedir. Her bir grubu kendi içinde detaylı incelemek gerekirse:

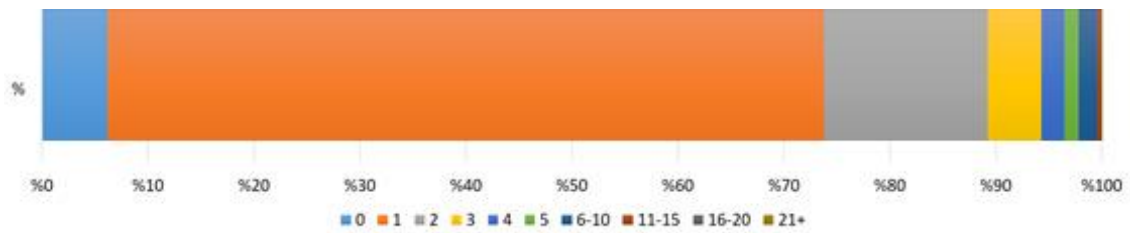
- *Anlamı açısından atıfların bölümlere dağılımı:* Anlamı açısından atıflar incelendiğinde negatif atıfların çoğunlukla araştırma/bulgular bölümü ile sonuç/öneriler bölümünde yapıldığı saptanmıştır. Birinci etiketçi yöntem ve diğer (dipnot, teşekkür ve ekler) bölümlerinde hiç negatif atıf etiketlememiştir. Bu durum negatif atıfların nerelerde aranabileceği konusunda fikir vermektedir. Öte yandan pozitif atıflara en yoğun olarak sonuç ve öneriler bölümünde yer verilmekle birlikte, makale içinde tüm bölümlerde rastlanmaktadır.
- *Amacı açısından atıfların bölümlere dağılımı:* Amacı açısından atıfların bölümlere göre dağılımında en belirgin sınıflar yöntem, tanım yapma ve veri doğrulama atıfları olmuştur. Bölümler arasında herhangi bir kullanım farklılığı olmayan atıflar literatür atıfları ile veriye atıflardır. Bu iki türe çalışmaların neredeyse tüm bölümlerinde rastlanabilmektedir. Yöntem açıklayıcı atıflar beklenildiği gibi sıklıkla yöntem bölümünde yapılırken, tanım yapma atıfları ise giriş ve dipnotlarda tercih edilmektedir. Veri doğrulama atıfları ise araştırma/bulgular bölümü ile sonuç/öneriler bölümünde yapılmaktadır. Bunun temel gerekçesi ise yapılan çalışmaların geçerliliğini ispatlamak için geçmiş çalışmaların doğrulamasının yapıldığı atıfların genellikle bu iki bölümde yer almasıdır.
- *Şekline göre atıfların bölümlere dağılımı:* Şekline göre atıfların bölümlere dağılımında belirgin bir farklılık görülmemekle beraber sonuç ve öneriler bölümlerinde yazar adlarının daha sık anıldığı saptanmıştır. Bu bulgunun dışında şekline göre atıf sınıfları ile araştırma bölümleri arasında önemli bir bağlantı görülmemiştir.

Literatürde makalelerin IMRAD bölümlendirmeleri ve atıflar üzerine yapılmış çeşitli çalışmalar bulunmaktadır. Bilgibilim alanında yapılmış atıfları bölümlerine göre değerlendiren bir çalışmada pozitif atıfların genellikle giriş ve literatür değerlendirmesi

bölümlerinde yapıldığı saptanmıştır (Ding, Liu, Guo ve Cronin, 2013, s. 583). Her atıfa eşit düzeyde değer atfedilmemesi gerektiğini vurgulayan ilk çalışmalardan biri olan makalede de (Voos ve Dagaev, 1976) en çok atıf yapılan bölümün giriş bölümü olduğu vurgulanmaktadır. Maričić, Spaventi, Pavičić ve Pifat-Mrzljak (1998, s. 539) yapmış oldukları çalışmada yöntem, bulgular ve sonuç bölümlerinde yapılan atıfların giriş ve literatür bölümünde yapılan atıflara oranla daha anlamlı atıflar olduğunu belirtmektedirler. Suppe (1998, s. 403)'nin çalışması da bilime yeni katkıların araştırmaların yöntem ve bulgular kısımlarında paylaşılması nedeniyle en çok katkı sağlayan bölümlerin yöntem ve bulgular olduğunu iddia etmektedir. Bu çalışma kapsamında incelenen amacına ve anlamına göre atıfların bölümlere dağılımı göz önüne alındığında benzer yorumu yapmak mümkündür. Pozitif atıflara makalelerin neredeyse tüm bölümlerinde rastlanırken, negatif atıflar çoğunlukla çalışmanın bulgular ve sonuç bölümünde yer almıştır.

5.1.4.2. Metin İçinde Atıf Yapılma Sayılarının ve Atıf Yapılan Bölümlerin Değerlendirilmesi

Kaynakçalarda yer alan referansların metin içinde geçiş sıklığı yalnızca etiketçi 1 tarafından yapılan etiketleme doğrultusunda değerlendirildiğinde hayli ilginç sonuçlara ulaşılmıştır. Atıf yapılan yayınların %67,5'i metin içinde sadece bir kez anılırken, kaynakçada yer alan referansların %6,1'ine metin içinde gönderme yapılmamıştır. Öte yandan makalelerde metin içinde geçen göndermelerin %1,1'inin kaynakçada yer almadığı da görülmüştür. Atıf yapılan kaynakçaların metin içinde anılma oranları görselleştirildiğinde Şekil 19'daki gibi bir görüntü ortaya çıkmaktadır. Şekle göre bir referansın metin içinde birden fazla kez kullanılma ihtimali yaklaşık %30'dur. Tüm makaleler içinde tek makalede 20'den fazla kez gönderme yapılan referans sayısı ise 4'tür.



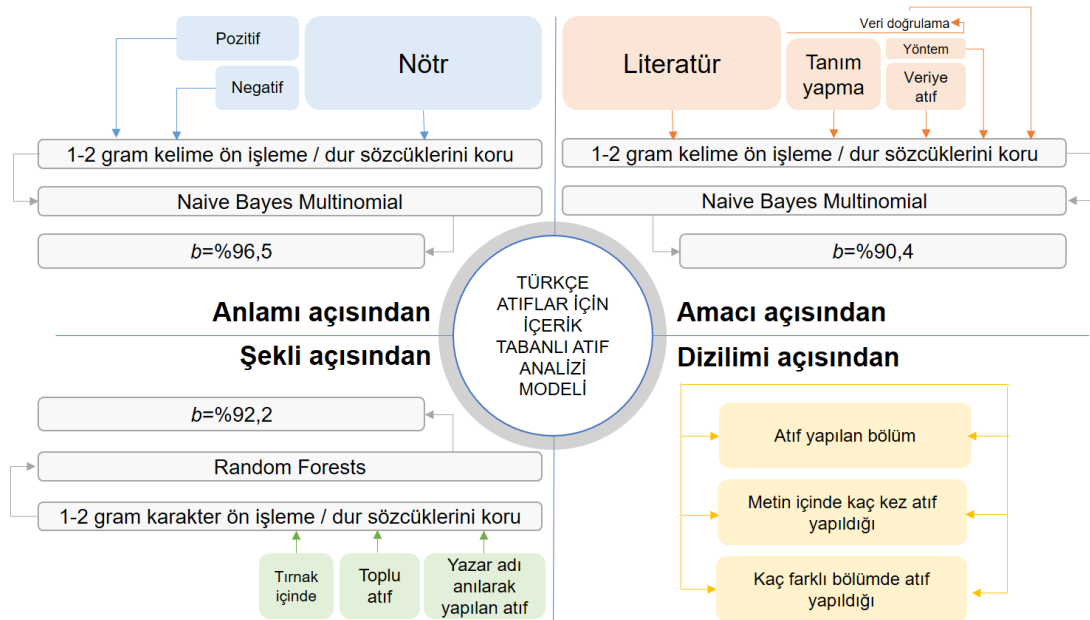
Şekil 19. Kaynakçalarda yer alan referansların metin içinde geçme sıklıkları

Bir referansa kaç farklı bölümde atıf yapıldığı değerlendirildiğinde ortaya çıkan sonuçlara göre yapılan göndermelerin %96'sının tek bir bölümde, %3'ünün iki bölümde ve kalan %1'inin 3 veya 4 bölümde yer aldığı sonucuna ulaşılmıştır. Dört bölümün tamamında atıf yapılan referans sayısı altı olarak bulunmuştur.

Araştırma sonucunda elde edilen bulgular literatürde yer alan çeşitli uygulamalar ile birlikte değerlendirilerek bir içerik tabanlı atıf analizi modeli geliştirilmiş ve bu model “Değerlendirme ve İçerik Tabanlı Atıf Analizi Modeli” bölümünde sunulmuştur.

5.2. DEĞERLENDİRME VE İÇERİK TABANLI ATIF ANALİZİ MODELİ

Bu çalışma kapsamında ilk aşamada literatür tabanlı olarak geliştirilen taksonomik sınıflama, ardından bu sınıfların Türkçe kütüphanecilik ve bilgilim literatürüne uyarlanması ve son aşamada ise bunun metin kategorizasyonu yöntemi ile otomatik olarak saptanabilmesi sonucunda Şekil 20’de gösterilen içerik tabanlı atıf analizi modeli ortaya çıkarılabilmektedir.



Şekil 20. Türkçe atıflar için içerik tabanlı atıf analizi modeli

Şekil 20’de yer alan model incelendiğinde Türkçe atıflar için dört temel taksonomik atıf sınıfı belirlenmiş, bu atıf sınıflarının alt grupları oluşturulmuştur. Bu bağlamda en önemli sınıfın anlamı açısından atıflar olduğu düşünülmektedir. Çünkü yazarların atıf yaptıkları kaynağa yönelik duygularının açıkça dile getirildiği atıf türü anlamı açısından atıflardır. Elbette bir yazarın bir kaynağa negatif eleştirilerde bulunması eleştirinin öznel olabileceği, alan temelli farklılıklar vb. gibi çeşitli sebeplerle atıf yapılan makaleyi kötü yapmaz. Ancak, atıf yapılan makalenin yazarının kötü yönde eleştirildiği bir atıftan akademik yarar sağlamanın önüne geçilmiş olur. Benzer şekilde pozitif atıflar ile nötr atıfların arasındaki farkın saptanması da önemlidir. Zira nötr atıflarda atıf yapılan yayınlara yönelik herhangi bir fikir beyan edilmezken; pozitif atıflarda konunun ne denli önemli olduğu, hangi açılardan

doğru noktalara değinildiği ya da ne denli emek yoğun çalışıldığı gibi çeşitli olumlu değerlendirmeler söz konusudur. Dolayısıyla bu üç anlamına göre atıf grubunun birbirinden ayrı ve içerikler dikkate alınarak değerlendirilmesi önem taşımaktadır.

Anlamına göre atıflar üzerine geliştirilen sınıflama algoritmasının başarıımı %96 gibi yüksek bir rakam olsa da başarımanın önemli bir kısmının verilerin %97'sini oluşturan nötr atıflardan sağlanması anlamı açısından atıflar ile ilgili başarıımı geliştirmeye yönelik ek çalışmalar yapma gereğini de beraberinde getirmektedir. Bunun için literatürde çeşitli yol ve teknikler (gövdeleme, duygu sözlüğü vb.) önerilmektedir.

İçerik tabanlı atıf analizinde önemli bir diğer grup yapıma amacı açısından atıfların belirlenmesidir. Amacına göre atıflar çalışma kapsamında beş temel alt gruba ayrılmıştır. Bu gruplarda yapılan atıfları kendi içlerinde değerlendirmek gerekirse:

- Yöntem atıfları ile veri doğrulama atıfları bu sınıfın en önemli atıf türleridir. Çünkü yöntem atıfları ile araştırmaya temel dayanak sağlayan altyapı kurulmaktadır. Veri doğrulama atıfları ise literatürde yapılmış çalışmaların tekrar test edilip benzer sonuçlara ulaşıldığını göstermesi açısından önemlidir. Veri doğrulama atıfları literatürde çok az rastlansa da karakteristik özellikleri diğer atıflardan farklı olduğundan hiçbir sınıf ile karışmamaktadır. Makinenin yapıyı ezberlediği için bu denli yüksek başarı vermiş olabileceği kaygısına cevap verebilmek için literatürdeki veri doğrulama atıflarının sayısının artması gerekmektedir.
- Bir yazar ya da kurum tarafından geçmişte elde edilmiş araştırma verilerinin farklı araştırmacılar tarafından tekrar kullanılması, ya da bir araştırmacının sonucunda elde edilmiş verilerin farklı çalışmalarda anılması bahse konu verinin önemini göstermesi açısından dikkate değerdir. Bu sebeple veriye yönelik atıflar da farklı açılardan değerlendirilerek yorumlanmalıdır.
- Bir yazar bir terim ya da kavram için literatüre yeni bir tanım getiriyorsa alanda kavramsal çalışmalar yaptığı anlamına gelmektedir. Eğer bu kavramsal tanımlar farklı çalışmalarda kullanılıyorsa, bu tanımın literatürde yer bulduğunu kanıtlamaktadır. Bu sebeple tanım yapma atıfları da alanda kavramsal çalışmalar yapan araştırmacıların ortaya çıkarılması ve kavramsal altyapının kimler tarafından sağlandığının gösterilmesi açısından önemlidir.
- Amacı açısından atıflar değerlendirilirken en az ağırlık literatür atıflarına verilmelidir. Bu durum literatür atıflarının gereksiz olduğu anlamına gelmemektedir. Bunun temel sebebi diğer atıf türlerine göre literatür atıflarının daha az etkili olduğu

gerçeğidir. Literatür atıfları değerlendirilirken söz konusu atıfların diğer sınıflarda hangi bölümlere dâhil olduğunun anlaşılması değerlendirme işlemi kolaylaştırıcaktır. Çünkü bir literatür atfının anlamına göre sınıfı pozitif ise değerlendirmede farklı bir yaklaşım sergilenmesi gerekecektir.

Amacı açısından atıfların saptanmasında tüm sınıflarda birbirine yakın sonuçlara ulaşılmış olması metin kategorizasyonu ile amacı açısından atıfların doğru bir şekilde sınıflandırılabilmesini kanıtlamaktadır. Bu da yönetici ve karar vericilerin işlerini oldukça kolaylaştırıcaktır.

Anlamı ve amacı açısından yapılan atıfların metinsel açıdan birbirlerine benzemeleri nedeniyle metin kategorizasyonu ile atıfların sınıflarının saptanması sürecinde benzer uygulamalar benzer sonuçlar vermiştir. Her iki taksonomik kategoride de en başarılı kelime ön işleme tekniği 2-gram işleme olurken en başarılı sonuç veren algoritma ise Naive Bayes Multinomial olmuştur.

Hazırlanan içerik tabanlı atıf analizi modelinde örnek olarak etiketçi 1'in etiketleme performansı ele alındığında saptanan pozitif atıfların %48'inde yazar adının anıldığı görülmüştür. Toplu atıflarda bu oran %8 iken tırnak içinde belirtilen atıflarda %4'tür. Bu durum pozitif atıfların en azından yarıya yakınının saptanmasını kolaylaştırıcı bir yöntemin de şekline göre atıfların doğru bir şekilde sınıflandırılması olduğunu kanıtlamaktadır. Anlamı ve şekli açısından sınıflandırılan atıflar karşılaştırılarak değerlendirildiğinde pozitif atıfların daha kolay saptanabileceği öngörülmektedir. Öte yandan aynı dağılımı negatif atıflar için söylemek mümkün değildir. Negatif atıfların %24'ünde yazar adı anılırken %65'inde herhangi bir belirteç kullanılmamaktadır. Bu sonuçlar negatif atıfların üstü kapalı bir şekilde yapıldığı gerçeğini destekler niteliktedir.

Türkçe literatürde şekli açısından atıfların saptanması sürecinde kelime ön işlemede 1-2 karakter gram ağırlıklandırması ve Random Forests algoritması tarafından yapılan sınıflandırma oldukça yüksek başarımla verilmekte, her 10 atıf cümlesinden 9,2'si doğru bir şekilde sınıflandırılabilir.

Geliştirilen modelde dizilimi açısından atıflar konusunda herhangi bir makine öğrenmesi metodu uygulanmamıştır. Bunun yerine referansların metin içinde kaç kez anıldığı hesaplanması ve bu anılmaların hangi bölümlerde kaçar defa yapıldığının saptanması dizilimine göre atıfların değerlendirilmesi için yeterlidir. Matematiksel bu işlemler sayesinde

metin içinde daha çok önemsenen (olumlu ya da olumsuz) referansların ortaya çıkarılması sağlanabilmekte, atfın yapıldığı bölümden hareketle atfın hangi amaçla yapıldığı tahmin edilebilmektedir. Eğer dizilimi açısından atıf bulguları amacı açısından atıf bulguları ile birlikte değerlendirilirse daha anlamlı sonuçlara ulaşılması mümkün olabilecektir.

İçerik tabanlı atıf analizi modeli geliştirilmesi sürecinde atıflara sayısal değer vermek (negatif atıflar eksi puan, pozitif atıflar artı puan gibi) anlamlı değildir. Bunun en önemli sebebi atıfların kendilerine has karakteristiklerinin olabileceğidir. Ayrıca önceki bölümlerde detaylandırılan atıflarla ilgili güncel etik problemlerden dolayı bu gibi bir ağırlıklandırma sisteminin kolaylıkla kandırılacağı gerçeği de göz önünde bulundurulmalıdır. Bu gibi bir ağırlıklandırma sistemi yazarların atıf verme motivasyonlarını etkileyecek, değiştirecek ve en çok puan getiren türden atıf almaya yönelik faaliyetlerin artışı ile yüz yüze kalınabilecektir. Bu durumda önerilen içerik tabanlı atıf analizi üzerinden atıf değerlendirmelerini yapma görevi yönetici ve karar vericilere düşmektedir. Yönetici ve karar verici pozisyonunda olan kişilerin (Akademik Teşvik Ödeneği Yönetmeliği Değerlendirme Kurulları, atama ve yükseltme jürileri, üniversite teşviklerinin şartlarını belirleyen kurulların üyeleri gibi) her atfın farklı değerinde olduğunun ayırdına varması ile atıflar gerçek anlamlarında kullanılacaklardır. Sonuç itibarıyla bu uygulamalar etik dışı motivasyonların önüne geçilmesini olanaklı kılacaktır.

6. BÖLÜM: SONUÇ VE ÖNERİLER

6.1. SONUÇ

Başlangıçta yalnızca ilişkili yayınlar arasında bağlantı kurulması amacını taşıyan atıfların kullanım amaçları yıllar içinde değişiklik göstermiş ve araştırma/cı değerlendirme ölçütü olarak kullanılmaya başlanmıştır. Akademik yükselmeler, teşvikler ya da ödüllendirme sistemlerinde en önemli puanlama kalemlerinden biri yazarların aldıkları atıflar olmuştur. Bu durum pek çok etik dışı uygulamayı da beraberinde getirmiş ve son yıllarda yayın sayısının artışı ile de paralel olarak atıf sayıları devasa boyutlara ulaşmıştır. Bu bağlamda bu tez çalışması kapsamında tüm atıfların eşit olmadığı hipotezinden hareketle bir içerik tabanlı atıf analizi modeli geliştirilmesi hedeflenmiş, atıfların çeşitli anlamlar yüklenerek değerlendirilmesinin daha doğru bir yaklaşım olduğu savunulmuştur.

Çalışmanın “Türkçe atıflar için otomatik metin kategorizasyonu tekniği kullanılarak semantik ve sentaktik temelli bir içerik tabanlı atıf analizi modeli yaratılması mümkün müdür?” şeklindeki temel araştırma sorusuna olumlu yanıt alınmış ve metin kategorizasyonu temelli algoritmalarla kayda değer makine öğrenmesi başarımları elde edilebilmiştir. Bunun yanında elde edilen bu sınıfsal yapılar kullanılarak da Türkçe kütüphanecilik ve bilgin bilim literatürü için içerik tabanlı bir atıf analizi modelinin tasarımı mümkün olabilmektedir.

Çalışma kapsamında yanıt aranan araştırma sorulardan biri olan “Türkçe atıfların taksonomik kategorizasyonu nasıldır ve bu atıflar için bir sınıflama şeması yaratılması mümkün müdür?” şeklindeki soruya cevaben bir taksonomik tür sınıflaması yapılabilmektedir. Çalışma kapsamında geliştirilen bu taksonomik sınıflama uluslararası literatürde önerilen sınıflamalara ve bu sınıflarda yer alan atıfların dağılımına oldukça benzemektedir; bu da Türkiye’de atıf motivasyonlarının uluslararası literatür ile benzerlik gösterdiğini kanıtlamaktadır. Bu benzerliğin ortaya koyulması sayesinde gelecekte yapılacak uluslararası çalışmaların Türkçe literatüre uyarlanması kolaylaşacağı düşünülmektedir.

“Türkçe yayınlarda pozitif, negatif ve nötr atıfların yeri nedir ve bu atıflarda ayırt edici bir dil kullanımı mevcut mudur?” şeklindeki soruya da özellikle pozitif ve negatif atıflar arasında ayırt edici bir dil yapısının bulunduğu yönünde yanıt elde edilmiştir. Ancak, literatürdeki pek çok çalışmada da dile getirildiği ve bu çalışmada da onaylandığı gibi pozitif atıflar daha açık bir dille yapılırken, negatif atıflarda bu denli açık bir dil kullanılmamaktadır. Bu sebeple pozitif atıfların makinece ayırt edilmesi negatif atıflara göre oldukça kolaydır.

Bu çalışmada elde edilen veriler ve sonuçlar kullanılarak tam metinler içinden atıf cümlelerinin saptanabilmesi mümkün görünmektedir. Bunun yanında temel hipotez olan “tüm atıflar eşit değildir” tezi de kanıtlanmış, atıfların birbirinden farklı kategorilerde çeşitli anlamlara sahip olabileceği ortaya koyulmuştur. Konu ile ilgili çıkarım ve yorumlar “Gelecek Çalışmalar” başlığı altında değerlendirilmektedir.

6.2. ÖNERİLER

Bu çalışma ile Türkçe kütüphanecilik ve bilgibilim literatüründe daha önce üzerinde fazla tartışılmamış içerik tabanlı atıf analizi çalışmaları değerlendirilmiş ve bir uygulama gerçekleştirilerek Türkçe içerik tabanlı atıf analizleri için bir model önerilmiştir. Bu model sayesinde araştırmacılar, editörler ve karar vericilerin makale yazım, inceleme ve değerlendirme süreçlerinde işlerinin kolaylaşacağı tahmin edilmektedir.

Önerilen içerik tabanlı atıf analizi modelinin bilimsel iletişim süreçlerinde yer alan dört rolün atıflarla ilgili uygulamalarının iyileştireceği düşünülmektedir. Bu rollerden ikisi “araştırmacı” ana rolünün farklı misyonlarıdır. Şöyle ki, bir araştırmacı çalışmaya başlarken, analiz gerçekleştirirken veya bunların sonuçlarını kaleme alırken “iz süren araştırmacı” misyonuna sahiptir. Bu araştırmacı bulduğu izleri takip ederek ve bu izlerin yanına kendi izlerini de ekleyerek yeni bir yol yaratır. Bu yoldaki önemli unsurlardan biri de atıflar ve bu atıfların anlamlarıdır. Bunun yanında diğer araştırmacı misyonu ise “akademik yükselme ya da teşvik başvurusunda bulunacak araştırmacı” olarak belirlenmiştir. Kendi yolunu belirleyen ya da en azından bu yolun temel noktalarını çizen araştırmacılar için emeklerinin karşılığının alınması önemlidir. Emeklerin karşılığının alınması sürecinde de atıfların doğru bir şekilde değerlendirilmesi ve anlamlandırılması gerekmektedir. Bir diğer önemli rol ise “yönetici ve karar vericiler”dir. Çünkü mevcut uygulamaların hayata geçirilmesinde en anlamlı kararların verilmesi, etik ilkelerden şaşmadan uygun düzenlemelerin yapılması ve değerlendirmelerin sorunsuz gerçekleştirilebilmesi için yönetici ve karar vericilerce temel ilkelerinin bir model kapsamında uygulanması önemlidir. Son rol “editörler” olarak belirlenmiştir. Editörler bilimsel iletişim sürecinin en önemli aktörlerinden biridir. Dolayısıyla bilimsel iletişim süreçlerinin her aşamada iyileştirilerek devam ettirilebilmesi editörlerin işini iyi yapması ile doğru orantılıdır. Belirlenen bu roller bağlamında çalışma kapsamında bilimsel iletişim süreci içinde atıflar konusunda her role düşen görevler ve bu görevlerin en doğru şekilde nasıl yapılabileceği Şekil 21’de gösterilmektedir.



Şekil 21. Bilimsel iletişim sürecindeki dört farklı rol için atıf yapma, değerlendirme, uygulama önerileri

Şekil 21’de bir bilimsel yayının hazırlanma aşamasından karar vericilerin değerlendirmesine sunulma aşamasına kadar olan süreçte her bir rolün dikkate alınması gereken noktalar sunulmuş, bu ilkelere göre davranıldığında da atıfların daha anlamlı birer değerlendirme aracı olabilecekleri öngörülmüştür.

Her bir rolün dikkat edilmesi gereken hususlar aşağıda tek tek incelenmektedir.

Makale yazım sürecinde iz süren araştırmacı: Literatür taraması benzer çalışmaların nasıl sonuç verdiğinin öğrenilmesi, mevcut durumun tespiti, yeni araştırma konularının

bulunması, araştırma duplikasyonlarının önlenmesi gibi çeşitli sebeplerden dolayı araştırma tasarımında önemli aşamalardan birini oluşturmaktadır. Dolayısıyla doğru yapılmış bir literatür taraması hem en ilgili literatüre erişmeyi sağlayacak, hem de birikimli bir şekilde çoğalan bilimsel çıktılarının daha anlamlı hale gelmesini olanaklı kılacaktır. Doğru literatür taraması yapabilmek için alanın özelliklerinin iyi bilinmesi, temel bilgi erişim becerilerine sahip olunması ve doğru bilgi kaynaklarının seçimi önemlidir. Bunun yanında her çalışmanın kendine has özellikleri sebebi ile ayrı literatür taraması gerektirdiği de unutulmamalıdır. Ayrıca genel kanının aksine literatür değerlendirmeleri araştırma yapılıp bitirildikten sonra yazıya dökme aşamasında değil, araştırmaya daha başlanmadan yani tasarım aşamasında yapılmalıdır. Bu çalışma kapsamında geliştirilen model ile iz süren araştırmacılar için özelleştirilmiş literatür taraması yapılması mümkün olabilecektir. Şöyle ki, bir araştırma yöntemi ile ilgili bilgi arayan bir araştırmacı makinece saptanmış yöntem atıflarının ortaya koyulması ile araştırmalarını daha hızlı bir şekilde yapabilecek ve kaynak kaçırma şansı en aza indirilebilecektir.

Makale yazan araştırmacıların atıflar konusunda en çok dikkat etmeleri gereken husus daha az çaba ile literatür taramasını tamamlayabilmek amacı ile çeşitli etik dışı davranışlara yönelmemeleri gerektiğidir. Etik dışı uygulamalardan en önemlisi başkasının yaptığı atfı aynen kopyalama ve orijinali görülmeyen yayınlara atıf yapılmasıdır. Orijinali görülmeden yapılan atıflarda hatalı bilgi aktarımı yapılması ihtimalinin çok yüksek olduğu unutulmamalıdır. Öte yandan makalelerin yalnızca başlığı ya da özetinin okunması konu hakkında genel bir fikir verse de, çalışma ile bağlantılar kuvvetliyse mümkün olduğunca tamamının okunması önemlidir. Bu aşamada zaman problemi olan yazarların sistematik okuma yaparak literatür değerlendirmesi yapmaları da mümkündür. Bu çalışma ve devam niteliğinde önerilen “atıf çıkarım sistemi” ile editör veya hakemler için bir atıf cümlesi havuzu yaratılması mümkün olabilecek, bu sayede de başka yayınlardan kopyalanmış atıfların saptanması kolaylaştırılabilecektir.

Makale yazımında önemli konulardan bir diğeri çeviri yazıdır. Farklı dilde yazılmış bir kaynağa atıf yapılacağı zaman çevirinin doğruluğundan emin olunmalıdır. Başkalarının yaptıkları çeviriler konusunda da dikkatli davranmak gerekmektedir. Çalışmanın ikinci bölümünde anılan Madeira Adası hakkında yazılmış bilimsel kaynaklardaki hataların tekrar edilmemesi için yazarlar tarafından çevirilerin doğru bir şekilde yapılması ve editörler/hakemler tarafından dikkatle incelenmesi gerekmektedir. Bu işlemin otomatize edilerek kolaylaştırılması için de makine çevirisi yöntemlerinden yararlanılabilir.

Yazarların atıflar konusunda farkındalık düzeylerinin artırılması gerekmektedir. Kaynakçalar ya da göndermeler okuyucular için yalnızca “bunları da beğenebilirsiniz” türü ilgili yazıların listesi değildir. Bir çalışmayı eleştirmek, öncülerini anmak, karşılaştırma yapmak ya da elde edilen bulguları doğrulamak amacı ile de atıf yapılabilir. Bu sebeple makale yazan her araştırmacının literatüre bu gözle bakması sağlanmalıdır. Kaynakçaların suya sabuna dokunmayan literatür listeleri formatından çıkarılıp atıf yapılan kaynak ile makale arasında ilişki kurmaya yönelik bir yaklaşım içinde hazırlanması gerekmektedir. Bu yaklaşım gözetilirken dikkat edilmesi gereken bir diğer konu da bu yaklaşımların bilimsel temelli olması gerektiği (kişisel inanışlar, sosyal statü vb. dışında) ve alanlara göre farklı atıf yapma motivasyonları olabileceğinin farkında olunmasıdır. Hâlihazırda bu çalışmada da atıfların farklı özellikleri olduğu kanıtlanmış ve bu konuda farkındalık yaratılmaya çalışılmıştır.

Editörler: Bir bilimsel yayının yayımlanmadan önceki en önemli denetim mekanizması editör kontrolüdür. Bu sebeple editörlere de ciddi görevler düşmektedir. Bu aşamada editörlerden beklenen en temel husus etik konularla ilgili ilke ve kuralların önceden belirlenmiş olmasıdır. Örneğin, bir çalışmada bilinçli bir etik ihlâl saptandı ise bu yazıya düzeltilmesi için ikinci bir şans verilmemelidir. Yapılan araştırmada somut olarak saptanan önemli sonuçlardan biri metin içinde yer aldığı halde kaynakçada listelenmeyen atıflar ile kaynakçada listelenip metin içinde gönderme yapılmayan atıflardır. Yazar eğer bu atıfları kontrol etmediyse dergi yayın kurulları makale yayımlanmadan önce kontrol etmeli ve eksikliklerin giderilmesini talep etmelidir. Mevcut durumda kaynakçada yer alan ancak gerçek anlamda atıf almamış makalelerin yazarları performans değerlendirmelerinde artı puan sağlarken; metin içinde atıf yapıldığı halde kaynakçada anılmadıkları için bazı makalelerin yazarları bu artı puanları edinmemişlerdir. Bu sebeple bu konuda dikkat gösterilmesi önemlidir. Otomatize edilmiş yöntemler editörlerin kontrollerini kolaylaştıracağı için mutlaka değerlendirilmesi gereken yöntemlerdir.

Son yıllarda editörler ile ilgili en sık karşılaşılan problemlerden biri zoraki atıf eklemeleri ile dergiye ya da yayın kuruluna yapılan atıfların sayısını artırmaya yönelik etik dışı davranışlardır. Bu gibi yaklaşımlar hem dergi prestijine ağır darbe vurmakta, hem de atıfların anlamlarını kaybetmesine neden olmaktadır. Bu durumun güncel örneklerinden biri “Konunun Önemi” bölümünde detayları aktarılan *Energy Education Science and Technology* dergisidir. Yazarlar bu zoraki atıf eklemesi talepleri karşısında *yayınla ya da yok ol* düsturunu kabul edip bu atıfları ilgisiz olmalarına rağmen çalışmalarına eklemekte

ve bu sayede anlamsız bir atıf yığını her geçen gün daha da çoğalmaktadır. Dergi editörlerinin bu gibi uygulamalardan uzaklaşması gerekmektedir.

Dergi editörlerine düşen önemli bir görev de atıf yapılan kaynağın gerçekten ilgili olup olmadığına saptanmasıdır. Ancak iyi bir editoryal değerlendirme süreci ile atıf yapılan kaynakların bağlama gerçekten uygun olup olmadığı ortaya çıkarılabilmektedir. Bu konuda en ilginç örneklerden birine 2016 yılında yayımlanan bir makalede rastlanmıştır. Aşağıda bir atıf cümlesi örnek olarak sunulan barsak hastalıkları hemşireliği üzerine yazılan makalede (Uran ve Yıldırım, 2016, s. 27) giriş bölümünde yapılan bir toplu atıfta web tabanlı uzaktan eğitim kaynakları ile ilgili bir makaleye yer verilmiştir.

Örnek: Ödem, ülserasyon, kanama, derin sıvı ve elektrolit kayıplarına neden olarak gastrointestinal sistemin herhangi bir kısmını etkilemekte ve iyileşme ve ataklarla seyretmektedir (Akyüz ve Akyüz, 2011; Al ve Mardan, 2004; Arslantaş, Metintaş, Ünsal ve Kalyoncu, 2006; Avdal, Kızılcı ve Demirel, 2011; Avşar ve Kaşıkçı, 2009; Aydınтуğ, 1992; O'Connor, Bager, Duncan ve Gaarenstrom, 2013; Stansfield, Fraser, Thompson ve Povey, 2007; Stretton, Currie ve Chauhan, 2014). Atıf yapılan kaynak: "Al, U., Mardan, O.R. (2004). Web tabanlı uzaktan eğitim sistemleri: Sahip olması gereken özellikler ve standartlar. Bilgi Dünyası, 5(2):259-271." (Metinden direkt alıntıdır)

Yukarıdaki atıf cümlesi ve atıf incelendiğinde yapılan en küçük hatanın yazar soyadının hatalı yazılması olduğu görülmektedir. Zira atıf yapılan bu makalede ne ödemlerin sebep olduğu problemlerden bahsedilmekte, ne de hemşirelik uygulamaları değerlendirilmektedir. Yazar tarafından yapılmış bu hatanın editoryal süreçte tespit edilmesi literatürde bu gibi atıfların sayısının artmasının önüne geçilmesi açısından önemlidir. Aynı konu editörün gözünden kaçıp karar vericilerin onayına sunulan performans değerlendirme dosyaları için de önem taşımaktadır. Bu konuya yönetici ve karar vericiler de azami dikkat etmelidirler. Bu gibi örnekler doğal dil işleme yöntem ve teknikleri sayesinde kolaylıkla saptanabilme potansiyeline sahiptir. Bu sebeple bilimsel iletişim süreçlerinde doğal dil işleme yöntem ve tekniklerinin etkin kullanımı önemli bir yere sahiptir.

Editörler tarafından saptanması gereken bir diğer önemli konu da makalenin içinde kullanılan atıfların dengeli bir şekilde dağılmış olmasıdır. Bir bilimsel araştırma makalesinin alana katkı yapması beklenir. Bu durumda çoğunluğu atıf cümlesinden oluşan makaleler literatüre ek olarak fazla şey söylemeyen makaleler olacaktır. Dolayısıyla bu dengenin korunması editörlerin görevidir. Dizilimi açısından atıfların ortaya çıkarılması ile bu dengenin korunup korunmadığı kolaylıkla saptanabilecektir.

Yükselme ve teşvik başvurusunda bulunacak araştırmacı: Bu gibi başvurularda bulunacak araştırmacıların ilk aşamada atıf verisini elde edecekleri kaynakları doğru seçmeleri ve yükselme/teşviklerde tercih edilen kaynakları doğru şekilde kullanmaları gereklidir. Kişiler kendilerine yapılan atıfları toplama aşamasında gerçekten kendilerine ait yayınlara yapılan atıfları raporladıklarından emin olmalıdırlar. Zira atıf dizinlerindeki yazar adlarında yer alan standardizasyon sorunları, aynı adla farklı yazarların olması ya da yukarıdaki barsak hastalıkları atıfı örneğinde olduğu gibi bazı atıfların konu bağlamında uyuşmaması gibi sebeplerle bazen tüm atıflar doğru bir şekilde listelenememektedir. Bu gibi durumlarda da hatalı atıf sayıları sunulabilmektedir. Örneğin, *Web of Science*'da dizinlenmiş iki farklı "Taşkın, Z." olduğundan atıf taraması yapıldığında "Taşkın, Z."ye yapılmış 116 atıfa erişilmektedir. Ancak bu atıfların 96'sı Boğaziçi Üniversitesi personeli olan Z. Caner Taşkın'a 20'si ise Hacettepe Üniversitesi personeli Zehra Taşkın'a aittir. Hâl böyle olunca atıf bilgisine erişme konusunda tüm akademisyenlerin ortak bir okuryazarlık becerisine sahip olmaları önemlidir. Bu konuda çeşitli eğitim çalışmaları gerçekleştirilerek akademisyenlerin bilgi ve beceri düzeyleri artırılabilir. Bilgi ve beceri düzeyleri konusunda bir diğer önemli konu da temel atıf göstergelerinin anlamlarının araştırmacılar tarafından bilinmesidir. Örneğin, kendine atıfın ne anlama geldiğini bilmeyen bir akademisyenin hatalı raporlama yapma potansiyeli bulunmaktadır. Bu sebeple atıflara erişimin öğretilmesinin yanında temel göstergelerin de aktarılması şarttır. Çünkü hem akademik teşvikler hem de diğer tüm yükseltme ve ödüllendirmelerde yazarların beyanları üzerinden değerlendirmeler gerçekleştirilmekte, bu beyanlar karar vericiler tarafından incelenmekte ve sonuçlandırılmaktadır. Bu durumda akademisyenlerin bu süreçleri tam doğru şekilde öğrenmesi ile yönetici ve karar vericilerin mevcut iş yükleri azaltılabilecek, ortak bir dil geliştirilebilecektir.

Yazarlar ile ilgili etik endişelerden biri farklı unsurlar gözetilerek (ilgili olmasına rağmen rakibin çalışmasına atıf yapılmazken ilgisiz yakın çevre atıfları gibi) atıf yapılmasıdır. Bunun yanında temeli "sen bana atıf yap, ben de sana atıf yapayım" esasına dayanan atıf takası da önemli sorunlardan biridir. Bu gibi etik dışı uygulamaların karar vericilerce tespit edilip yaptırım ile sonuçlanması etik dışı uygulamaların sayısını azaltabilecektir.

Yazarlar başka yayınlara atıf yaparlarken bazen atıf yaptıkları araştırmayı eleştirip çeşitli sorular yöneltebilmektedirler. Bu sebeple yazarların aldıkları atıfları niceliksel olarak saymalarındansa her yeni atıfın ne amaçla yapıldığını öğrenmeleri, gerekli durumlarda eleştirilere bir okuyucu mektubu ile cevap vermeleri daha anlamlıdır. Bu sayede bilimin eleştiriler ile gelişmesi sağlanabilecektir.

Yöneticiler ve karar vericiler: Yöneticiler ve karar vericilerin ayırdına varmaları gereken en önemli konu atıfların sadece bir sayıdan ibaret olmadığı gerçeğidir. Araştırmacıların performanslarının değerlendirilmesi için kullanılan ölçütlerden biri atıflardır ancak tek başlarına kullanıldıklarında hem bir anlam ifade etmemekte, hem de etik dışı uygulamalara dönüşebilmektedir. Öte yandan yönetici ve karar verici pozisyonunda yer alan kişi veya kurulların her başvuru sahibine eşit davranmayı sağlayacak temel ilke ve politikaları belirlemeleri objektif değerlendirmelerin yapılabilmesi için birinci şarttır. Bu politika ve uygulamaların güncel bilimsel ve teknolojik gelişmelere uyarlanabilir olması da önemlidir ve politikaların uzun yıllar uygulanabilmesini sağlayacaktır.

Yöneticiler ve karar vericiler için önemli olan bir diğer nokta da raporlanan atıfların kontrolü ve doğrulamasının yapılmasıdır. Daha önce editör ve başvuru sahibi yazarlar için yapılan öneriler bu rol grubu için de geçerlidir. Bu grupta yer alan kişilerin atıfların gerçekten rapor sahibi kişiye yapılıp yapılmadığının saptanması ve atıf yapan ile yapılan arasında konusal bir bağlamın var olması gereklidir. Bu kişi ve kurullar ayrıca son yıllarda oldukça yaygınlaşan yağmacı (predatory) dergiler konusunda bilgi sahibi olmalı, bu dergilerden yapılan atıflara karşı politika belirlemelidir.

Önemli bir diğer konu da hazırlanan tüm politikalarda alan temelli bir değerlendirme yapılması gerekliliğidir. Örneğin, bir arkeoloğun atıflarını en iyi değerlendirecek olan yine bir arkeologdur. Bu sebeple kurullar oluşturulurken, atıflar için değerlendirme modelleri yaratılırken veya uygulamalar geliştirilirken alana özgü davranışların iyi bir şekilde analiz edilmesi ve öğrenilmesi anlamlı olacaktır. Kurullarda yer alan kişiler belirlenirken bu konuya azami dikkat edilmesi gereklidir.

Yönetici ve karar verici pozisyonundaki kişilerin yazarlar tarafından raporlanan atıfların kontrolünde dikkatli olmaları gerekmektedir. Örnek üzerinden konuşmak gerekirse Don Randall tarafından kaleme alınan bir makalede (2015) Türkiye’de İngiliz Dili üzerine yapılmış çalışmalar değerlendirilmiş ve bu çalışmalara ciddi olumsuz eleştiriler yapılmıştır. Sözü edilen makale *Web of Science*’da dizinlenen bir dergide yayımlandığından bu makalede atıf yapılmış tüm yazarlar teşvik ve ödüllendirmelerde bu atıfları raporlayabilmekte ve kullanabilmektedirler. Böyle örnekler söz konusu olduğunda yöneticiler tarafından uygulanması gereken en temel ilke atıfların içerik olarak da değerlendirilerek bu gibi atıfların dosyalardan çıkarılması olmalıdır. Bu şekilde yazarların aldıkları atıflar üzerinde otokontrol yapmalarını sağlamaları da mümkün olabilecektir.

Günümüzde doküman tiplerinin değişmesi ile birlikte atıf yapılan kaynak tiplerinde de farklılıklar görülmeye başlanmıştır. Artık bilimsel çalışmalarda sosyal ağ gönderileri, videolar, ses kayıtları ve çevrimiçi dersler gibi kaynaklara da atıf yapılabilmektedir. Bu bu gibi içeriklerin aldıkları atıflar da değerlendirmeler kapsamında incelenmelidir.

Yukarıda sıralanan tüm roller kendilerine önerilen şekilde atıf yapma, değerlendirme ve raporlama yaptıkları takdirde her geçen gün anlamını daha da yitiren atıfların yeniden bilimsel bilginin içinde dondurulmuş ayak izleri olmaları sağlanabilecektir. Bunun yanında hesaplamalı dilbilim çalışmalarının gelişmesi ile birlikte yukarıda sözü edilen pek çok konu otomatize yöntemlerle kolaylıkla çözümlenebilmektedir. Bu nedenle bilimsel iletişim sürecindeki tüm aktörlerin bu seçeneklerin farkında olması ve gerektiğinde bu alanda çalışan uzmanlar ile ortak çalışarak niteliksel analizleri sürdürmeleri büyük önem taşımaktadır.

6.3. GELECEK ÇALIŞMALAR

Bu çalışmadan elde edilen veri ve bulgular kullanılarak yakın gelecekte aşağıda sıralanan iki çalışmayı yapmak mümkün olabilecektir.

- Tam metinler içinden atıfların otomatik olarak saptanmasını sağlayacak bir makine çıkarım modeli yaratılması,
- Anlamı açısından atıfların makinece saptanması başarımının artırılması için gövdeleme, duygu sözlüğü ve diğer teknikler kullanılarak analizlerin tekrarlanması ve en doğru tekniğin ortaya çıkarılması.

Bu iki konuda çalışmalar yapıldığında literatürde içerik tabanlı atıf analizi çalışmalarında aşama kat edilecek ve atıf değerlendirme süreçlerinin büyük bölümünün otomatize edilmesi sağlanabilecektir.

7. BÖLÜM: KAYNAKÇA

- Adedayo, A.V. (2015). Citations in introduction and literature review sections should not count for quality. *Performance Measurement and Metrics*, 16(3), 303-306.
- Aggarwal, C.C. (2015). *Data mining: The textbook*. İsviçre: Springer International Publishing.
- Akademik Teşvik Ödeneği Yönetmeliği, T.C. Resmi Gazete. (12813099, 14.12.2015). Erişim adresi: <https://goo.gl/AKClck>
- Akademik Teşvik Ödeneği Yönetmeliği, T.C. Resmi Gazete. (13271644, 27.12.2016). Erişim adresi: <https://goo.gl/GTtGkZ>
- Akba, F. (2014). *Duygu analizinde öznitelik seçme metriklerinin değerlendirilmesi: Türkçe film eleştirileri* (Yüksek Lisans Tezi). Hacettepe Üniversitesi, Ankara.
- Akpınar, H. (2014). *Data: Veri madenciliği, veri analizi*. İstanbul: Papatya Yayıncılık Eğitim A.Ş.
- Aksnes, D.W. (2005). *Citations and their use as indicators in science policy: Studies of validity and applicability issues with a particular focus on highly cited papers* (Doktora Tezi). University of Twente.
- Aksoy, A. ve Öztürk, T. (2016). *Trstop*. Erişim adresi: <https://github.com/ahmetax/trstop>
- Al, U. (2008). *Türkiye'nin bilimsel yayın politikası: Atıf dizinlerine dayalı bibliyometrik bir yaklaşım* (Doktora Tezi). Hacettepe Üniversitesi.
- Al, U. ve Soydal, İ. (2012). Dergi kendine atfının etkisi: Energy Education Science and Technology örneği. *Türk Kütüphaneciliği*, 26(4), 699-714.
- Al, U. ve Soydal, İ. (2014). Akademik atıf dizinleri ile savaşı. *Hacettepe Üniversitesi Edebiyat Fakültesi Dergisi*, 31(1), 23-42.
- Al, U. ve Soydal, İ. (2015). Bilimsel iletişimin farklı bir yüzü: Geri çekilen makaleler. Umut Al ve Zehra Taşkın (Yay. Haz.), *Prof. Dr. İrfan Çakın'a Armağan* içinde (s. 22-37). Ankara: Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü.
- Al, U. ve Tonta, Y. (2004). Atıf analizi: Hacettepe Üniversitesi Kütüphanecilik Bölümü tezlerinde atıf yapılan kaynaklar. *Bilgi Dünyası*, 5(1), 19-47.
- Alexander, A. (2016). Lucene example source code file (stopwords.txt). Erişim adresi: <http://alvinalexander.com/java/jwarehouse/lucene/contrib/analyzers/common/src/resources/org/apache/lucene/analysis/tr/stopwords.txt.shtml>
- Aljaber, B., Stokes, N., Bailey, J. ve Pei, J. (2010). Document clustering of scientific texts using citation contexts. *Information Retrieval*, 13(2), 101-131.

- Angrosh, M.A., Cranefield, S., ve Stanger, N. (2010). Context identification of sentences in related work sections using a conditional random field: Towards intelligent digital libraries. *Proceedings of the ACM, JCDL'10* içinde (s. 293-302). Queensland: ACM.
- Arısoy, E., Saraçlar, M., Roark, B. ve Shafran, I. (2010). Syntactic and sub-lexical features for Turkish discriminative language models. *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010 içinde (s. 5538-5541). Dallas, TX: IEEE.
- Artstein, R. ve Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, 34(4), 555-596.
- Arunachalam, S. ve Manorama, K. (1988). Are citation-based quantitative techniques adequate for measuring science on the periphery?. *Scientometrics*, 15(5-6), 393-408.
- Athar, A. (2011). Sentiment analysis of citations using sentence structure-based features. *HLT-SS '11 Proceedings of the ACL 2011 Student Session* içinde (s. 81-87). Stroudsburg: Association for Computational Linguistics.
- Athar, A. (2014). *Sentiment analysis of scientific citations* (Teknik rapor, UCAM-CL-TR-856). Cambridge: University of Cambridge Computer Laboratory.
- Atılğan, D., Atakan, C. ve Bulut, B. (2008). Türkçe kütüphanecilik dergilerinin atıf analizi. *Türk Kütüphaneciliği*, 22(4), 392-413.
- Attribute-Relation File Format (ARFF)*. (2008). Erişim adresi: <http://www.cs.waikato.ac.nz/ml/weka/arff.html>
- Batini, C., Cappiello, C., Francalanci, C. ve Maurino, A. (2009). Methodologies and data quality assessment and improvement. *ACM Computing Surveys*, 41(3), 1-52.
- Bertin, M. (2008). Categorizations and annotations of citation in research evaluation. 13. *Natural Language Processing; 13.1 Discourse*. Erişim adresi: <http://aaai.org/Papers/FLAIRS/2008/FLAIRS08-108.pdf>
- Biricik, G. (2011). *Metin sınıflama için yeni bir özellik çıkarım yöntemi* (Doktora Tezi). Yıldız Teknik Üniversitesi.
- Blake, C. (2013). Text mining. *Annual Review of Information Science and Technology*, 45(1), 121-125.
- Bonzi, S. (1982). Characteristics of a literature as predictors of relatedness between cited and citing works. *Journal of the American Society for Information Science*, 33(4), 208-216.
- Bonzi, S. ve Snyder, H.W. (1991). Motivations for citation: A comparison of self citation and citation to others. *Scientometrics*, 21(2), 245-254.
- Bornmann, L. ve Daniel, H-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45-80.

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Brooks, A.T. (1986). Evidence of complex citer motivations. *Journal of the American Society for Information Science*, 37(1), 34-36.
- Cano, V. (1989). Citation behavior: Classification, utility, and location. *Journal of the American Society for Information Science*, 40(4), 284-290.
- Carter, G.M. (1974). Peer review, citations, and biomedical research policy: NIH grants to Medical School Faculty. *Rand Report, R-1583*. Santa Monica: Rand. Erişim adresi: <http://www.rand.org/content/dam/rand/pubs/reports/2008/R1583.pdf>
- Cavalcanti, D.C., Prudêncio, R.B.C., Pradhan, S.S., Shah, J.Y. ve Pietrobon, R.S. (2011). Good to be bad? Distinguishing between positive and negative citations in scientific impact. *23rd IEEE International Conference on Tools with Artificial Intelligence (ICTAI)* içinde (s. 156-162). Boca Raton: IEEE.
- Chowdhury, G.G. (2003). Natural language processing. *Annual Review of Information Science and Technology*, 37(1), 51-89.
- Chubin, D.E. (1980). Letter to editor: Is citation analysis a legitimate evaluation tool? *Scientometrics*, 2(1), 91-94.
- Chubin, D.E. ve Moitra, S.D. (1975). Content analysis of references: Adjunct or alternative to citation counting? *Social Studies of Science*, 5(4), 423-441.
- Clarivate Analytics. (2017). Web of Science: Trust the difference. Erişim adresi: <http://wokinfo.com/>
- Cole, J. ve Cole, S. (1971). Measuring the quality of sociological research: Problems in the use of the Science Citation Index. *The American Sociologist*, 6, 23-29.
- Cole, J.R. (2000). A short history of the use of citations as a measure of the impact of scientific and scholarly work. Cronin, B. ve Atkins, H.B. (Yay. Haz.), *The Web of Knowledge A Festschrift in Honor of Eugene Garfield* içinde (s. 281-300). New Jersey: Information Today.
- COPE. (2012). *Citation manipulation*. Erişim adresi: http://publicationethics.org/files/u7141/Forum%20discussion%20topic_final.pdf
- Cozzens, S.E. (1985). Comparing the sciences: Citation context analysis of paper from neuropharmacology and the sociology of science. *Social Studies of Science*, 15, 127-153.
- Cozzens, S.E. (1989). What do citations count?: The rhetoric-first model. *Scientometrics*, 15(5-6), 437-477.
- Cronin, B. (1981). The need for a theory of citing. *Journal of Documentation*, 37(1), 16-24.
- CWTS Leiden Ranking. (2015). *Data collection*. Erişim adresi: <http://www.leidenranking.com/methodology/datacollection>

- Çarkı, K., Geutner, P. ve Schultz, T. (2000). Turkish LVCSR: Towards better speech recognition for agglutinative languages. *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings içinde* (s. 1563-1566). İstanbul: IEEE.
- Damashek, M. (1995). Gauging similarity with n-grams language independent categorization of text. *Science*, 267(5199), 843-848.
- Davis, P. (2017). Citation cartel or editor gone rogue? [Blog yazısı]. *Scholarly Kitchen*. Erişim adresi: <https://scholarlykitchen.sspnet.org/2017/03/09/citation-cartel-or-editor-gone-rogue/>
- Ding, Y., Liu, X., Guo, C. ve Cronin, B. (2013). The distribution of references across texts: Some implications for citation analysis. *Journal of Informetrics*, 7, 583-592.
- Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X. ve Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the Association for Information Science and Technology*, 65(9), 1820-1833.
- Doğan, G. (2014). *Türk Kütüphaneciliği dergisinden dergilere yapılan atıflarda kaynak gösterme hataları: 2010-2012*. Uluslararası Kütüphane ve Bilgibilim Felsefesi Sempozyumu, 3-5 Eylül 2014, Kastamonu. Erişim adresi: <http://www.bby.hacettepe.edu.tr/akademik/guledaduzuyoldogan/file/ukbbf-gd-v3.pptx>
- Dong, C., ve Schäfer, U. (2011). Ensemble-style self-training on citation classification. *5th International Joint Conference on Natural Language Processing, IJCNLP 2011 içinde* (s. 623-631). Chiang Mai: AFNLP.
- Duda, R.O., ve Hart, P.E. (1973). *Pattern classification and scene analysis*. New Jersey: Wiley and Sons, Inc.
- Elkiss, E., Shen, S., Fader, A., Erkan, G., States, D. ve Radev, D. (2008). Blind men and elephants: What do citation summaries tell us about a research article? *Journal of the American Society for Information Science and Technology*, 59(1), 51-62.
- Elsevier. (2015). *About Scopus*. Erişim adresi: <https://www.elsevier.com/solutions/scopus>
- Erikson, M.G. ve Erlandson, P. (2014). A taxonomy of motives to cite. *Social Studies of Science*, 44(4), 625-637.
- Eryiğit, G. (2014). *ITU Turkish natural language processing pipeline*. Erişim adresi: <http://tools.nlp.itu.edu.tr/MorphAnalyzer>
- Extension. (2013). *What's the difference between a supervised and unsupervised image classification?*. Erişim adresi: <https://articles.extension.org/pages/40214/whats-the-difference-between-a-supervised-and-unsupervised-image-classification>
- Feldman, S. (1999). NLP meets the jabberwocky: Natural language processing in information retrieval. *Online*, 23, 62-72.

- Frost, C.O. (1979). The use of citations in literary research: A preliminary classification of citation functions. *The Library Quarterly: Information, Community, Policy*, 49(4), 399-414.
- Fu, L.D. ve Aliferis, C.F. (2010). Using content-based and bibliometric features for machine learning models to predict citation counts in the biomedical literature. *Scientometrics*, 85(1), 257-270.
- Garfield, E. (1970). Can citation indexing be automated? *Essays of an Information Scientist*, 1, 84-90.
- Garfield, E. (1973). Citation frequency as a measure of research activity and performance. *Essay of an Information Scientist*, 1, 406-408.
- Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, 1(4), 359-375.
- Garfield, E. (1986). Do Nobel Prize winners write citation classics?. *Essays of an Information Scientist*, 9, 182-187.
- Garfield, E. (1997). Concept of citation indexing: A unique and innovative tool for navigating the research literatüre. Far Eastern State University. Philadelphia: The Scientist. Erişim adresi: <http://www.garfield.library.upenn.edu/papers/vladivostok.html>
- Garfield, E. ve Malin, M.V. (1968). Can Nobel Prize winners be predicted?. *135th Annual Meeting, American Association for the Advancement of Science, Dallas, Texas - December 26-31, 1968.* Erişim adresi: <http://www.garfield.library.upenn.edu/papers/nobelpredicted.pdf>
- Gingras, Y. ve Wallace, M.L. (2010). Why it has become more difficult to predict Nobel Prize winners: a bibliometric analysis of nominees and winners of the chemistry and physics prizes (1901–2007). *Scientometrics*, 82(2), 401-412.
- Goudsmith, S.A. (1974). Citation analysis. *Science*, 183(4120), 28.
- Hacettepe Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi Uygulama Esasları ve Araştırmacı Bilgilendirme Kılavuzu.* (2015). Erişim adresi: <http://apsis.hacettepe.edu.tr/GetDoc.ashx?ID=19>
- Halevi, G. ve Bar-Ilan, J. (2016). Post retraction citations in context. *BIRNDL 2016 Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries* içinde (s. 23-29). Newark: CEUR.
- Herlach, G. (1978). Can retrieval of information from citation indexes be simplified?: Multiple mention of a reference as a characteristic of the link between cited and citing article. *Journal of the American Society for Information Science*, 29(6), 308-310.
- Jha, R., Jbara, A-A, Qazvinian, V. ve Radev, D.R. (2016). NLP-Driven citation analysis for scientometrics. *Natural Language Engineering*, 23(1), 93-130.

- Johnson, C.A. (1985). Citations to authority in supreme court opinions. *Law & Policy*, 7(4), 509-523.
- Kaplan, A. (2013). Üniversitelerde bilimsel yayın çalışmaları [Sunum]. *Bilimsel Dergilerimiz ve Uluslararası İndekslerdeki Yeri Çalıştayı*. Erişim adresi: http://webftp.gazi.edu.tr/bidb/dergilercalistayi/sunular/ULAKBIM_Dergi_sunuu.pptx
- Kaplan, P. (2014). Akademisyenlerin atıf çetesi. *HaberTürk*, 28 Temmuz 2014 Pazartesi. Erişim adresi: <http://www.haberturk.com/yazarlar/pervin-kaplan/974474-akademisyenlerin-atif-cetesi>
- Kim, Y. ve Weber, B. (2006). Automatic reference resolution in astronomy articles. *20th International CODATA Conference Scientific Data and Knowledge within the Information Society*. Erişim adresi: <http://www.codata.info/06conf/abstracts/G8wasB3/KimYunhyong.htm>
- King, C. (2015). The 2015 Thomson Reuters citation laureates. Erişim adresi: <http://stateofinnovation.thomsonreuters.com/citation-laureates-2015>
- Kochen, M. (1974). *Principles of information retrieval*. Los Angeles: Melville.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI'95 Proceedings of the 14th international joint conference on Artificial intelligence* içinde (s. 1137-1143). Montreal: ACM.
- Krampen, G., Becker, R., Wahner, U. ve Montada, L. (2007). On the validity of citation counting in science evaluation: Content analyses of references and citations in psychological publications. *Scientometrics*, 71(2), 191-202.
- Kurbanoğlu, S.S. (1996). Kütüphanecilik literatüründe atıf hataları. *Türk Kütüphaneciliği*, 10(2), 137-148.
- Landis, J.R. ve Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Lee, M.L., Lu, H., Ling, T.W. ve Ko, Y.T. (1999). Cleansing data for mining and warehousing. T.J.M. Bench-Capon, G. Soda ve A.M. Tjoa (Yay. Haz.), *Database and Expert Systems Applications* içinde (s. 751-760). Berlin: Springer.
- Lerner, J. ve Wulf, J. (2007). Innovation and incentives: Evidence from corporate R&D. *The Review of Economics and Statistics*, 89(4), 634-644.
- Lewis, D.D. ve Jones, K.S. (1996). Natural language processing for information retrieval. *Communications of the ACM*, 39(1), 92-101.
- Liddy, E.D. (2010). Natural language processing. *Encyclopedia of Library and Information Sciences, Third Edition* içinde (s. 3864-3873). New York: Taylor and Francis.
- Lipetz, B-A. (1965). Improvement of the selectivity of citation indexes to science literature through inclusion of citation relationship indicators. *American Documentation*, 16(2), 81-90.

- Liu, M. (1993). Progress in documentation the complexities of citation practice: A review of citation studies. *Journal of Documentation*, 49(4), 370-408.
- Liu, S., Chen, C., Ding, K., Wang, B., Xu, K. ve Lin, Y. (2014). Literature retrieval based on citation context. *Scientometrics*, 101(2), 1293-1307.
- Liu, Y., Yan, R. ve Yan, H. (2013). Guess what you will cite: Personalized citation recommendation based on users' preference. Rafael E. Banchs, Fabrizio Silvestri, Tie-Yan Liu, Min Zhang, Sheng Gao, Jun Lang (Ed.). *Information Retrieval Technology: 9th Asia Information Retrieval Societies Conference, AIRS 2013 Singapore, December 2013 Proceedings* içinde (s. 428-239). Heidelberg: Springer Verlag.
- Love, B.C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review*, 9(4), 829-835.
- MacRoberts, M.H. ve MacRoberts, B.R. (1996). Problems of citation analysis. *Scientometrics*, 36(3), 435-444.
- Manning, C.D. ve Schütze, H. (1999). *Foundations of statistical natural language processing*. London: MIT Pres.
- Maričić, S., Spaventi, J., Pavičić, L. ve Pifat-Mrzljak, G. (1998). Citation context versus the frequency counts of citation histories. *Journal of the American Society for Information Science*, 49(6), 530-540.
- Markey, K. ve Cochrane, P.A. (1981). *Online training and practice manual for ERIC data base searchers* (2. bs). New York: ERIC Clearing House on Information Sciences. Erişim adresi: <http://files.eric.ed.gov/fulltext/ED212296.pdf>
- Matthew 25:29*. (2004). Erişim adresi: <http://biblehub.com/tur/matthew/25.htm>
- McCain, K.W. ve Turner, K. (1989). Citation context analysis and aging patterns of journal articles in molecular genetics. *Scientometrics*, 17(1-2), 127-163.
- McCallum, A. ve Nigam, K. (1998). A comparison of event models for naive bayes text classification. *AAAI/ICML-98 Workshop on Learning for Text Categorization* içinde (s. 41-48). Wisconsin: AAAI Pres.
- Meadows, A.J. (1967). The citation characteristics of astronomical research literature. *Journal of Documentation*, 23(1), 28-33
- Merton, R.K. (1968). The Matthew Effect in science. *Science*, 159, 56-63.
- Miller, J.C., Coble, K.H. ve Lusk, J.L. (2013). Evaluating top faculty researchers and the incentives that motivate them. *Scientometrics*, 97(3), 519-533.
- Moed, H. F. (2005). *Citation analysis in research evaluation*. Netherlands: Springer.
- Moravcsik, M.J. ve Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science*, 5(1), 86-92.

- Nanba, H. ve Okumura, M. (1999). Towards multi-paper summarization using reference information. *IJCAI'99 Proceedings of the 16th International Joint Conference on Artificial Intelligence (Volume 2)* içinde (s. 926-931). San Francisco: Morgan Kauffman.
- Nasukawa, T. ve Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. *Proceeding K-CAP '03 Proceedings of the 2nd international conference on Knowledge capture* içinde (s. 70-77). New York, ACM.
- Nicolaisen, J. ve Frandsen, T.F. (2015). The focus factor: A dynamic measure of journal specialisation. *Information Research*, 20(4).
- Noktalamalar işaretleri.* (2016). Erişim adresi: http://tdk.gov.tr/index.php?option=com_content&view=article&id=187:Noktalamalar-Isaretleri-Aciklamalar&catid=50:yazm-kurallar&Itemid=132
- Oppenheim, C. (1996). Do citations count? Citation indexing and the research assesment exercise (RAE). *Serials*, 9(2), 155-161.
- Oppenheim, C. ve Renn, S.P. (1978). Highly cited old papers and the reasons why they continue be cited. *Journal of the American Society for Information Science*, 29(5), 225-231.
- Oransky, I. (2017). Citation-boosting episode leads to editors' resignations, university investigation [Blog yazısı]. *Retraction Watch*. Erişim adresi: <http://retractionwatch.com/2017/03/03/citation-boosting-episode-leads-editors-resignations-university-investigation/>
- Oxford Dictionaries. (2015). *Citation*. Erişim adresi: <http://www.oxforddictionaries.com/definition/english/citation>
- Öğretim Üyeliğine Yükseltme ve Atanma Yönetmeliği. (1982, 28 Ocak). *Resmi Gazete* (Sayı: 17588). Erişim adresi: http://www.yok.gov.tr/web/guest/icerik/-/journal_content/56_INSTANCE_rEHF8BlSfYRx/10279/17641
- Öztürk, K. (2012). *Şişme dergiler ve etik ihlalleri* [Blog yazısı]. Erişim adresi: <https://mkoz.wordpress.com/2012/06/23/sisme-dergiler-ve-etik-ihlalleri/>
- Öztürk, K. (2013). *Şişme dergiler, yeniden* [Blog yazısı]. Erişim adresi: <https://mkoz.wordpress.com/2013/01/25/sisme-dergiler-yeniden/>
- Peritz, B.C. (1983). A classification of citation roles for the social sciences and related fields. *Scientometrics*, 5(5), 303-312.
- Price, D.J.D.S. (1986). *Little science, big science... And beyond*. New York: Columbia University Press.
- QS. (2015). *Faculty area normalization - technical explanation*. Erişim adresi: http://content.qs.com/qsui/Faculty_Area_Normalization_-_Technical_Explanation.pdf

- Randall, D. (2015). English studies in Turkey: An Assessment. *ariel: A Review of International English Literature*, 46(1-2), 45-68.
- Ranks.nl. (2017). *Turkish stopwords*. Erişim adresi: <http://www.ranks.nl/stopwords/turkish>
- Ritchie, A. (2008). *Citation context analysis for information retrieval* (Doktora Tezi). University of Cambridge.
- Schneider, J.W. ve Borlund, P. (2005). A bibliometric-based semi-automatic approach to identification of candidate thesaurus terms: Parsing and filtering of noun phrases from citation contexts. F. Crestani ve I. Ruthven (Yay. Haz.), *Context: Nature, Impact and Role, 5th International Conference on Conceptions of Library and Information Sciences, CoLIS 2005 Glasgow, UK, June 2005, Proceedings* içinde (s. 226-237). Heidelberg: Springer Verlag.
- Schultz, T. ve Waibel, A. (2001). Language-independent and language-adaptive acoustic modeling for speech recognition. *Speech Communication*, 35(1-2), 31-51.
- ScienceWatch. (2015a). *About ScienceWatch*. Erişim adresi: <http://sciencewatch.com/about-us>
- ScienceWatch. (2015b). *Successful predictions*. Erişim adresi: <http://sciencewatch.com/nobel/successful-predictions>
- Sendhil Kumar, S., Elakkiya, E. ve Mahalakshmi, G.S. (2013). *Citation semantic based approaches to identify article quality*. Computer Science & Information Technology (CS & IT) içinde (s. 411-420). Delhi: ICCSEA.
- Shadish, W.R., Tolliver, D., Gray, M. ve Gupta, S.K.S. (1995). Author judgements about works they cite: Three studies from psychology journals. *Social Studies of Science*, 25(3), 477-498.
- Shen, H-P., Wu, C-H. ve Tsai, P-S. (2015). Model generation of accented speech using model transformation and verification for bilingual speech recognition. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 14(2).
- Sher, I.H. ve Garfield, E. (1983). *New tools for improving and evaluating the effectiveness of research*. Erişim adresi: <http://www.garfield.library.upenn.edu/papers/onpaper.html>
- Shum, S.B. (1998). Evolving the web for scientific knowledge: First step towards an "HCI knowledge web". *Interfaces, British HCI Group Magazine*, 39, 16-21.
- Siddharta, A. ve Teufel, S. (2007). Whose idea was this, and why does it matter? Attributing scientific work to citations. *Proceedings of NAACL/HLT-07*. Erişim adresi: <http://www.cl.cam.ac.uk/~sht25/papers/naacl07.pdf>
- Silahtaroglu, G. (2013). *Veri madenciliği: kavram ve algoritmaları*. İstanbul: Papatya Yayıncılık Eğitim A.Ş.

- Silva, J.A.T. ve Dobránszki, J. (2017). Highly cited retracted papers. *Scientometrics*, 110(3), 1653-1661.
- Simkin, M.V. ve Roychowdhury, V.P. (2003). Read before you cite!. *Complex Systems*, 14, 269-274.
- Simkin, M.V. ve Roychowdhury, V.P. (2006). Do you sincerely want to be cited? Or: read before you cite. *Significance*, 179-181.
- SJR. (2015). *About us*. Erişim adresi: <http://www.scimagojr.com/aboutus.php>
- Small, H.G. (1978). Cited documents as concept symbols. *Social Studies of Science*, 8(3), 327-240.
- Small, H.G. (2011). Interpreting maps of science using citation context sentiments: A preliminary investigation. *Scientometrics*, 87(2), 373-388.
- Smith, L.C. (1981). Citation analysis. *Library Trends*, 30, 83-106.
- Spiegel-Rösing, I. (1977). Science Studies: Bibliometric and content analysis. *Social Studies of Science*, 7(1), 97-113.
- Stigler, S.M. (1980). Stigler's law of eponymy. *Transactions New York Academy of Sciences*, 39(1), 147-157.
- Suppe, F. (1998). The structure of a scientific paper. *Philosophy of Science*, 65(3), 381-405.
- Swales, J. ve Feak, C. (2000). *English in today's research world: A writing guide*. Michigan: Ann Arbor.
- Şeker, S.E. (2016). *Weka ile veri madenciliği*. İstanbul: Bilgisayar Kavramları Yayınları.
- T.C. Maliye Bakanlığı. (2017). *En yüksek devlet memuru brüt aylığı*. Erişim adresi: <http://www.bumko.gov.tr/Eklenti/10429,2017ocakeydmxls.xls?0>
- Tandon, N. ve Jain, A. (2012). *Citation context sentiment analysis for structured summarization of research papers*. Erişim adresi: <https://people.mpi-inf.mpg.de/~ntandon/papers/ki2012-tandon.pdf>
- TDK. (2015). *Atıf*. Erişim adresi: http://tdk.gov.tr/index.php?option=com_bts&arama=kelime&guid=TDK.GTS.566a83cf1f4645.05297359
- Testa, J. (2008). *Regional content expansion update: Web of Science 5.0*. Erişim adresi: <http://wokinfo.com/media/pdf/RExEssay.pdf>
- Teufel, S. (1999). *Argumentative zoning: Information extraction from scientific text* (Doktora Tezi). University of Edinburg.
- Teufel, S., Siddharthan, A. ve Tidhar, D. (2006). *Automatic classification of citation function*. Erişim adresi: <http://www.cl.cam.ac.uk/~sht25/papers/emnlp06.pdf>

- Thomson Reuters. (2014). *Essential Science Indicators & Journal Citation Reports*. Erişim adresi: <http://thomsonreuters.com/content/dam/openweb/documents/pdf/scholarly-scientific-research/fact-sheet/esi-jcr-brochure.pdf>
- Thomson Reuters. (2015). *Highly cited researchers*. Erişim adresi: <http://highlycited.com/>
- Tonta, Y. (2014). *Akademik performans, öğretim üyeliğine yükseltme ve yayın destekleme ölçütleriyle ilgili bir değerlendirme*. Erişim adresi: <http://yunus.hacettepe.edu.tr/~tonta/yayinlar/tonta-yukseltme-kriterleri-hakkinda-degerlendirme-11-Temmuz-2014.pdf>
- Tunç, M. (2012). *Bilimsel ve teknolojik bilginin niteliksel performansının ölçümü* (Doktora Tezi). Marmara Üniversitesi.
- Uran, B.N.Ö. ve Yıldırım, Y. (2016). Yeni bir uzmanlık alanı: İnflamatuar barsak hastalıkları hemşireliği. *İzmir Kâtip Çelebi Üniversitesi Sağlık Bilimleri Fakültesi Dergisi*, 1(2), 27-33.
- URAP. (2015). *2015-2016 Türkiye genel sıralaması genel bilgi*. Erişim adresi: <http://tr.urapcenter.org/2015/2015.php>
- Van Raan, A.F.J. (1998). In matters of quantitative studies of science the fault of theorists is offering too little and asking too much. *Scientometrics*, 43(1), 129-139.
- Van Raan, A.F.J. (2004). Measuring science: Capita selecta of current main issues. Moed, H.F., Glänzel, W. ve Schmoch, U. (Yay. Haz.), *Handbook of Quantitative Science and Technology Research* içinde (s. 15-50). Dordrecht: Kluwer Academic.
- Van Raan, A.F.J. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics*, 62(1), 133-43.
- Van Rijsbergen, C.J. (1979). *Information retrieval*. London: Butterworths.
- Vinkler, P. (1987). A quasi-quantitative citation model. *Scientometrics*, 12(1-2), 47-72.
- Vinkler, P. (1994). Words and indicators: As scientometrics stands. *Scientometrics*, 30(2), 495-504.
- Voos, H. ve Dagaev, K.S. (1976). Are all citations equal? Or did we Op. Cit. your Idem? *The Journal of Academic Librarianship*, 1(6), 19-21.
- Weinstock, M. (1971). Citation indexes. *Essays of an Information Scientist*, 1, 188-195.
- Weka *Eğitim Serisi*. (2015). Erişim adresi: https://www.youtube.com/watch?v=5s8lgMfH698&list=PLh9ECzBB8tJP8vpIM91W1k_qoG110ikh1
- Weka. (2016). *Weka 3: Data mining software in Java*. Erişim adresi: <http://www.cs.waikato.ac.nz/ml/weka/index.html>

- Wetterer, J.K. (2006). Quotation error, citation copying, and ant extinctions in Madeira. *Scientometrics*, 67(3), 351-372.
- Witten, I.H., Frank, E. ve Hall, M.A. (2011). *Data mining: Practical machine learning tools and techniques* (3. Bs.). Amsterdam: Elsevier.
- Woolgar, S. (1991). Beyond the citation debate: Towards a sociology of measurement technologies and their use in science policy. *Science and Public Policy*, 18(5), 319-326.
- Wu, J. (2011). Improving the writing of research papers: IMRAD and beyond. *Landscape Ecology*, 26(10), 1345-1349.
- Xu, J., Zhang, Y., Wu, Y., Wang, J., Dong, X., ve Xu, H. (2015). Citation sentiment analysis in clinical trial papers. *AMIA Annual Symposium Proceedings*, 2015, 1334-1341.
- Yu, B. (2013). Automated citation sentiment analysis: What can we learn from biomedical researchers. *ASIS&T 2013 Annual Meeting Montréal, Québec, Canada, November 1-5, 2013*. Erişim adresi: <https://www.asis.org/asist2013/proceedings/submissions/papers/105paper.pdf>
- Zang, G., Ding, Y. ve Milojević, S. (2012). Citation content analysis (CCA): A framework for syntactic and semantic analysis of citation content. *Journal of the American Society for Information Science*, 64(7), 1490-1503.
- Zemberek NLP. (2015). Erişim adresi: <http://zembereknlp.blogspot.com.tr/>
- Zhu, X., Turney, P., Lemire, D. ve Vellino, A. (2015). Measuring academic influence: Not all citations are equal. *Journal of the Association for Information Science and Technology*, 66(2), 408-427.
- Ziman, J.M. (1968). *Public knowledge: An essay concerning the social dimension of science*. Cambridge: Cambridge University Pres.
- Zipf, G. (1949). *Human behavior and the principle of least effort*. Cambridge: Addison Wesley Pres.



**HACETTEPE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
TEZ ÇALIŞMASI ETİK KURUL İZİN MUAFİYETİ FORMU**

**HACETTEPE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
BİLGİ VE BELGE YÖNETİMİ ANABİLİM DALI BAŞKANLIĞI'NA**

Tarih: 5 / 2 /2015

Tez Başlığı / Konusu: İçerik tabanlı atf analizi modeli tasarımı: Türkçe atıflar için metin kategorizasyonuna dayalı bir uygulama

Yukarıda başlığı/konusu gösterilen tez çalışmam:

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır,
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne müdahale içermemektedir.
4. Gözlemsel ve betimsel araştırma (anket, ölçek/skala çalışmaları, dosya taramaları, veri kaynakları taraması) niteliğinde değildir.

Hacettepe Üniversitesi Etik Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Tarih ve İmza

Adı Soyadı: Zehra Taşkın

Öğrenci No: N11244821

Anabilim Dalı: Bilgi ve Belge Yönetimi

Programı: Bilgi ve Belge Yönetimi - Doktora

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

DANIŞMAN GÖRÜŞÜ VE ONAYI

Doç. Dr. Umut Al

Telefon: 0-312-2976860

Detaylı Bilgi: <http://www.sosyalbilimler.hacettepe.edu.tr>

Faks: 0-3122992147

E-posta: sosyalbilimler@hacettepe.edu.tr



HACETTEPE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÜKSEK LİSANS/DOKTORA TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
BİLGİ VE BELGE YÖNETİMİ ANABİLİM DALI BAŞKANLIĞI'NA

Tarih: 01/06/2017

Tez Başlığı / Konusu: İçerik Tabanlı Atıf Analizi Modeli Tasarımı: Türkçe Atıflar için Metin Kategorizasyonuna Dayalı Bir Uygulama

Yukarıda başlığı/konusu gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler ve d) Sonuç kısımlarından oluşan toplam 87 sayfalık kısmına ilişkin, 01/06/2017 tarihinde tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı %3 'tür.

Uygulanan filtrelemeler:

- 1- Kabul/Onay ve Bildirim sayfaları hariç,
- 2- Kaynakça hariç
- 3- Alıntılar hariç
- 4- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Tez Çalışması Orjinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Tarih ve İmza

Adı Soyadı: Zehra Taşkın

Öğrenci No: N11244821

Anabilim Dalı: Bilgi ve Belge Yönetimi

Programı: Bilgi ve Belge Yönetimi

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

Doç. Dr. Umut Al