

# **Türkçe Atıflar İçin İçerik Tabanlı Atıf Analizi Modeli Tasarımı**

**Program Kodu: 1001**

**Proje No: 115K440**

Proje Yürütücüsü:

**Prof. Dr. Umut AI**

Araştırmacı:

Doç. Dr. Umut Sezen

Bursiyerler:

Kardelen Aktaş  
Ayşe Esra Özkan Çelik  
Güleda Doğan  
Elçin Keleş  
İpek Şencan  
Zehra Taşkın

Ocak 2018  
ANKARA

Hacettepe Üniversitesi  
Bilgi ve Belge Yönetimi Bölümü  
06800 Beytepe, Ankara  
Tel: +90 (312) 2978200  
Faks: +90 (312) 2992014  
Web: <http://www.bby.hacettepe.edu.tr>  
E-posta: [bby@hacettepe.edu.tr](mailto:bby@hacettepe.edu.tr)

Umut Al  
E-posta: [umutal@hacettepe.edu.tr](mailto:umutal@hacettepe.edu.tr)  
Web: <http://yunus.hacettepe.edu.tr/~umutal>

Umut Sezen  
E-posta: [u.sezen@ee.hacettepe.edu.tr](mailto:u.sezen@ee.hacettepe.edu.tr)  
Web: <http://www.ee.hacettepe.edu.tr/?lang=t&link=400201&sublink=208>

## ÖNSÖZ

Günümüzde atıflar ve atıfların değerlendirilmesi oldukça popüler araştırma konuları arasında yer almaktadır. Özellikle konuya uzak kişilerin atıflar ile ilgili çeşitli çalışmalar yapması, bu çalışmalarda kamuoyunun yanlış yönlendirilme potansiyeli ve konu ile ilgili farkındalığın artırılması gerekliliği bu projenin ortaya çıkmasında önemli rol oynamıştır. Ayrıca proje bursiyerlerinden Zehra Taşkın'ın Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümünde yürütülen “İçerik Tabanlı Atıf Analizi Modeli Tasarımı: Türkçe Atıflar için Metin Kategorizasyonuna Dayalı Bir Uygulama” başlıklı doktora çalışması bu araştırmaya yön vermiştir. Araştırma ile ilgili tüm süreç bu tip çalışmalar için oldukça önemli olduğu düşünülen açıklık politikası gereği proje web sitesi aracılığıyla (<http://115k440.hacettepe.edu.tr/>) kamuoyunun erişimine sunulmuştur. Projenin son döneminde gerçekleştirilen çalışmaya da Youtube (<https://www.youtube.com/channel/UCIjdn3PsSSTknv6-sgNa7bg/videos>) üzerinden erişmek olanaklıdır.

Çok kısaca ifade etmek gerekirse; TÜBİTAK tarafından 115K440 proje numarası ile desteklenen bu proje sonucunda araştırma ve araştırmacı değerlendirmelerinde önemli birer gösterge olarak kullanılan atıflar ile ilgili içerik tabanlı bir değerlendirme sistemi önerilmektedir. Bu sistem ile tüm atıfların eşit şekilde değerlendirilmesinin sebep olduğu problemlerin önüne geçilmesi hedeflenmektedir.

Proje boyunca elde edilen deneyimler ve ortaya çıkan bulgular uluslararası kamuoyu ile çeşitli platformlarda paylaşılmıştır ve hali hazırda yürütülmekte olan bir doktora tezi bu projeye ilgili altyapı temelinde sonuçlandırılmaya çalışılmaktadır (bkz. Ek 1). Alanın önemli dergilerinin başında gelen *Scientometrics* adlı dergide yayımlanmış bir çalışma bulunmaktadır. Yayıncı tarafından sağlanan veriler Almanya, Amerika Birleşik Devletleri, Ekvator, Hollanda, İngiltere, İspanya, Peru gibi ülkelerdeki araştırmacıların kısa süre içerisinde makaleye ilgi gösterdiğini ortaya koymaktadır. 14-16 Eylül 2016 tarihleri arasında Valencia'da düzenlenen *21st International Conference on Science and Technology Indicators (STI 2016)* adlı uluslararası toplantıda “Sub-fields of Library and Information Science in Turkey: A Visualization Study”; 6-8 Eylül 2017 tarihleri arasında Paris'de düzenlenen *STI 2017* toplantısında “First Stage of an Automated Content-Based Citation Analysis Study: Detection of Citation Sentences” başlıklı çalışmalar sunulmuştur. Bu toplantılar konuyla ilgili son gelişmelerin tartışıldığı platformlardır ve alanın en önemli toplantıları olarak kabul edilmektedir.

Projenin bursiyer kadrosunda bulunan ve proje süresi içerisinde doktora derecelerini alarak akademiye gerçek anlamı ile giriş yapan Güleda Doğan ve Zehra Taşkın başta olmak üzere Kardelen Aktaş, Ayşe Esra Özkan Çelik, Elçin Keleş ve İpek Şencan'a özverili çalışmalarından dolayı içtenlikle teşekkür ederiz. Ayrıca üniversite içindeki bürokratik işlemlerin halledilmesi konusunda yardımlarını esirgemeyen Leyla Sevim'e de müteşekkirimiz.

# İÇİNDEKİLER

|   |      |
|---|------|
| ÖNSÖZ .....   | iii  |
| İÇİNDEKİLER.....  | iv   |
| ŞEKİLLER DİZİNİ.....  | vi   |
| TABLolar DİZİNİ .....   | vii  |
| ÖZ.....   | viii |
| ABSTRACT .....  | ix   |
| 1. BÖLÜM: GİRİŞ.....  | 1    |
| 1.1. Giriş .....  | 1    |
| 1.2. Literatür Değerlendirmesi.....                             | 2    |
| 1.2.1. Atıf Analizleri ve Güncel Sorunlar .....                 | 2    |
| 1.2.2. İçerik Tabanlı Atıf Analizleri .....                     | 5    |
| 1.2.3. Doğal Dil İşleme Yöntemleri ve Uygulamalar .....         | 5    |
| 1.2.4. Türkiye’de Atıf Analizi Çalışmaları .....                | 7    |
| 2. BÖLÜM: YÖNTEM VE TEKNİK.....                                 | 8    |
| 2.1. Veri Setinin Belirlenmesi ve Veri Tabanı Tasarımı .....    | 8    |
| 2.2. Arayüzlerin Tasarımı ve Veri Toplama .....                 | 11   |
| 2.3. Taksonomik Atıf Türlerinin Belirlenmesi .....              | 12   |
| 2.3.1. Yapılma Amacı Açısından Atıflar .....                    | 14   |
| 2.3.2. Anlamı Açısından Atıflar .....                           | 14   |
| 2.3.3. Dizilimi Açısından Atıflar .....                         | 14   |
| 2.3.4. Veriliş Şekli Açısından Atıflar .....                    | 14   |
| 2.4. Etiketleme Süreci .....                                    | 15   |
| 2.4.1. Etiketleme Arayüzünün Tasarımı.....                      | 15   |
| 2.4.2. Etiketleme Eğitimi .....                                 | 16   |
| 2.4.3. Uzlaşmalı Etiketleme Süreci .....                        | 16   |
| 2.4.4. Veri Ön İşleme Süreci ve Algoritmaların Uygulanması..... | 18   |
| 2.4.5. Performans Değerlendirme.....                            | 19   |
| 2.4.6. Atıf Çıkarım Sistemi.....                                | 20   |
| 3. BÖLÜM: BULGULAR VE YORUM.....                                | 21   |
| 3.1. Atıfların Otomatik Sınıflandırılması .....                 | 21   |
| 3.1.1. Anlamına Göre Atıflar .....                              | 21   |
| 3.1.2. Amacına Göre Atıflar .....                               | 23   |

|   |    |
|---|----|
| 3.1.3. Şekline Göre Atıflar .....   | 24 |
| 3.1.4. Dizilimine Göre Atıflar .....  | 24 |
| 3.2. Atıf Çıkarım Sisteminin Oluşturulması .....  | 26 |
| 3.2.1. Atıf Çıkarım Sistemi için Kural Listesi .....  | 26 |
| 3.2.2. Atıfların Çıkarımı.....  | 27 |
| 3.3. İçerik Tabanlı Atıf Analizi Süreçleri için Uygulama Önerileri .....  | 29 |
| 4. BÖLÜM: SONUÇ VE GELECEK ÇALIŞMALAR .....   | 32 |
| 4.1. Sonuç.....   | 32 |
| 4.2. Gelecek Çalışmalar .....   | 34 |
| KAYNAKÇA.....   | 35 |
| EK 1: Proje Kapsamında Yapılan Yayınlar .....   | 41 |
| Tamamlanmış çalışmalar .....  | 41 |
| Devam eden çalışma .....  | 41 |
| EK 2: Atıfların Sınıflanması ve Değerlendirilmesi Üzerine Yapılmış ve Proje Kapsamında İncelenen Çalışmalar ..... | 42 |

## ŞEKİLLER DİZİNİ

|   |    |
|---|----|
| Şekil 1. Çalışma süreçleri ve zaman planı .....   | 8  |
| Şekil 2. Oluşturulan ilişkiyel veri tabanının yapısı .....                                    | 10 |
| Şekil 3. Taksonomik atıf sınıfları .....  | 13 |
| Şekil 4. Etiketleme arayüzünün genel görünümü .....   | 15 |
| Şekil 5. Uzlaşmalı etiketleme süreci sonunda elde edilen atıf cümleleri ve sınıflara dağılımı | 18 |
| Şekil 6. Algoritmaların uygulanması sürecinin aşamaları .....                                 | 18 |
| Şekil 7. Temel atıf sınıfları için sınıflama başarımları .....                                | 22 |
| Şekil 8. Atıfların makale bölümlerine dağılımı .....  | 25 |
| Şekil 9. Göndermelerin saptanması için hazırlanan gramer grafiği .....                        | 28 |
| Şekil 10. Örnek gönderme çıkarımı .....   | 28 |
| Şekil 11. Atıf çıkarım sistemi akış şeması .....  | 29 |
| Şekil 12. İçerik tabanlı atıf analizi süreci .....  | 30 |
| Şekil 13. Bilimsel iletişim sürecindeki farklı roller için uygulama önerileri .....           | 33 |

## TABLÖLAR DİZİNİ

|  |   |
|--|---|
| Tablo 1. Uluslararası üniversite sıralama sistemlerinde kullanılan bilimsel etki ölçütlerine ilişkin ağırlıklar (Doğan, 2017, s. 30) ..... | 3 |
|--|---|

## ÖZ

Araştırma ve araştırmacı performanslarının ölçümünde önemli unsurlardan biri de atıflardır. Araştırmacılar aldıkları atıfların sayıları üzerinden ödüllendirilmekte, yükseltilmekte veya teşvik almaktadırlar. Ancak tamamen saymaya dayalı bu sistem etik dışı uygulamaları da beraberinde getirmiş ve atıf sayılarının manipüle edilmesine kadar uzanmıştır. Bu projenin temel amacı tüm atıfların eşit değerde olmadığını ortaya çıkarmak ve bu farklılıkları semantik ve sentaktik yaklaşımlarla çözümlenecek bir makine tasarlamaktır. Bu amaçla Türkçe kütüphanecilik ve bilgilendirme alanından 423 makale incelenmiş ve içerik tabanlı atıf analizi gerçekleştirilmiştir. Atıflar için anlam, amaç, şekil ve dizilimi dikkate alan dördü bir taksonomik sınıflama belirlenmiş ve Weka yazılımı kullanılarak otomatize sınıflama gerçekleştirilmiştir. Sınıflama sonucunda anlam, amaç ve şekil atıfları için Naive Bayes Multinomial ve Random Forest algoritmaları kullanılarak %90'ın üzerinde başarımla elde edilmiştir. Dizilimi açısından atıflar değerlendirildiğinde atıfların %85'inin giriş ve literatür değerlendirme 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. Elde edilen tüm veriler kullanılarak çalışma sonunda Türkçe atıflar için bir atıf çıkarım sistemi sunularak içerik tabanlı atıf analizlerinde daha az insan gücü ile başarıya ulaşılması hedeflenmiştir. Proje sonucunda sunulan bilgiler sayesinde bilimsel iletişim sürecindeki tüm rollerin atıflar konusundaki farkındalık düzeylerinin artacağı, atıfların değerlendirilmesi için yeni bir bakış açısı sunulacağı ve açık erişime sunulan proje verileri ile Türkçe içerik analizi çalışmalarına destek olunacağı düşünülmektedir.

**Anahtar Sözcükler:** İçerik tabanlı atıf analizi, doğal dil işleme, metin kategorizasyonu, Weka, atıf sınıflandırması, atıf çıkarım sistemi, Türkçe kütüphanecilik ve bilgilendirme literatürü



## ABSTRACT

One of the important elements in measuring research and researcher performance is the analysis of citations. Researchers are being rewarded, tenured or encouraged by the number of citations they receive. However, current system which is totally based on counting citations created unethical practices and manipulations in order to increase the number of citations. The main aim of this project is to reveal that all citations are not equal and to design a machine that may resolve these differences with semantic and syntactic approaches. For this purpose, 423 articles in the field of library and information science in Turkey are examined and content-based citation analysis is performed on these articles. A quadruple taxonomic classification considering the meaning, purpose, shape and array is determined and automated classification is performed using Weka software. As a result of the classification, over 90% accuracy is obtained by using Naïve Bayes Multinomial and Random Forest algorithms in the determination of meaning, purpose and shape classes. The results on citations by arrays show that literature citation sentences are generally placed in introduction and literature review sections (85%) Similarly, 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 section. It is determined that 67% of the references are cited only once in the text, and 6% are not cited in the text at all. By using all obtained data, a citation extraction system for Turkish citations is presented at the end of the study, aiming to reach success with less human power. Through the information presented in the project report, it is considered that the level of awareness of all roles in the scholarly communication process will increase, a new perspective to evaluate citations will be presented, and open data of this study will support future studies focused on text categorization for Turkish language.

**Keywords:** Content-based citation analysis, natural language processing, text categorization, Weka, citation classification, citation extraction system, Turkish librarianship and information science literature

# 1. BÖLÜM: GİRİŞ

## 1.1. Giriş

Bilim birikimli bir şekilde ilerler. Yani her yeni buluş veya olgu bir önceki dönemden izler taşır ya da önceki çalışmaların üzerine inşa edilir. Bilimsel yayınlar aracılığı ile sunulan araştırma bulguları bilimin sürdürülmesine önemli katkılar sağlar. Araştırma bulgularının sunulması esnasında ise önceki çalışmalarla bağlantıların kurulması önemlidir (Shum, 1998, s. 19). Araştırmalardaki sözü edilen bağlantılar atıflar aracılığı ile kurulur (Teufel, 1999, s. 33).

Atıf sözcüğünün kelime anlamı Türk Dil Kurumu tarafından “yönelme, çevirme”, “ilişkili bulma” ve “gönderme” olarak tanımlanmaktadır (TDK, 2017a). Oxford sözlüğü ise “bir kitap, makale ya da yazarın özellikle bilimsel bir yayın içinde bir kaynağı kullanması/anması” şeklinde tanımlamıştır (Oxford Living Dictionaries, 2017). Yazarlar kendi yayınları ile daha önce yapılmış benzer çalışmalar arasında anlamlı bağlantıları atıflar sayesinde kurarlar (Garfield, 1997; Smith, 1981, s. 84). Ancak, başlangıçta tek amacı araştırmalar arasında bağlantı kurmak olan atıfların kullanım amaçları atıf dizinlerinin ortaya çıkması ve yaygın şekilde kullanılmaya başlanması ile değişmiştir. Günümüzde atıflar araştırma ve araştırmacıların bilimsel etkinliğini ölçmek için bir araç olarak kullanılmakta, en çok atıf alan yazarlar yüksek performanslı olarak anılmakta ve bilimin seviyesi atıfların sayılması sonucu ölçülebilmektedir. Yalnızca sıklık saymaya dayanan bu yöntem beraberinde pek çok sorun ve sorunlu uygulamayı getirmektedir. Yayın sayılarının devasa boyutlara ulaştığı günümüzde, niceliksel değerlendirmeler bir kenara bırakılarak niteliksel değerlendirmeyi olanaklı kılmak araştırma ve araştırmacı değerlendirmelerinin daha doğru ve etkili şekilde yapılmasını sağlayacaktır.

Bu projenin temel amacı tüm atıfların eşit değerde olmadığını ortaya çıkarmak bu farklılıkları semantik ve sentaktik olarak çözümleyebilecek bir makine tasarlamak olarak belirlenmiştir. Bu bağlamda Türk kütüphanecilik ve bilginbilim literatürü derinlemesine incelenmiş, atıf sınıfları oluşturulmuş ve bu atıf sınıflarının makinece belirlenmesi için metin kategorizasyonu yöntemleri uygulanmıştır. Proje sonucunda sunulan bilgiler sayesinde bilimsel iletişim sürecindeki tüm rollerin atıflar konusundaki farkındalık düzeylerinin artacağı, atıfların değerlendirilmesi için yeni bir bakış açısı sunulacağı ve açık erişime sunulan proje verileri ile Türkçe içerik analizi çalışmalarına destek olunacağı düşünülmektedir.

Proje sonuç raporu dört ana bölümden oluşmaktadır. İlk bölümde projenin amacı ve literatürde yer alan konu ile ilgili çalışmalar sunulmaktadır. İkinci bölümde çalışmanın nasıl yapıldığı ile ilgili detaylı bilgiler destekleyici görseller aracılığı ile aktarılmakta, üçüncü bölümde de yapılan araştırma tüm detayları ile açıklanmaktadır. Son bölümde ise araştırma sonucunda elde edilen

bulguların geleneksel atıf deęerlendirmelerini hangi boyutlara taşıyabileceęi üzerinde durulmakta ve gelecekte yapılabilecek alıřmalar konusunda fikir verilmektedir.

## **1.2. Literatür Deęerlendirmesi**

### **1.2.1. Atıf Analizleri ve Güncel Sorunlar**

Atıflar bilimsel bilginin içinde dondurulmuş ayak izleri olarak tanımlanan ve bilimsel bilginin gelişimini sağlayan öğeler olduęu düşünölen unsurlardır (Cronin, 1981, s. 16). Atıflar sayesinde alıřmalar arasında iliřkiler kurulabilmekte, bu iliřkiler de bilimin birikimli yapısını desteklemektedir. Atıfların bilim evrelerince bu denli önemli kabul edilmesinin ardından 1873 yılında atıf dizinlemenin ilk örnekleri görölmeye başlanmıştır (Shepherd's Citations) (Al ve Tonta, 2004, s. 21). Bunun ardından 1950'lerin ortasında Eugene Garfield önderliğinde en bilinen atıf dizini olan Web of Science'ın temeli atılmış ve Bilimsel Bilgi Enstitüsü (Institute of Scientific Information) bünyesinde sırasıyla Science Citation Index (SCI), Social Sciences Citation Index (SSCI) ve Arts and Humanities Citation Index (A&HCI) oluşturulmuştur (White, 1985, s. 39).

Atıf yapma motivasyonlarının zamana, disipline, dergiye, erişilebilirliğe ve bundan başka pek çok etkene baęlı olarak deęişiklik gösterdiği iddia edilmesine rağmen (Bornmann ve Daniel 2008, s. 46) en bilinen atıf motivasyonu sınıflaması atıf dizinlerinin yaratıcısı Garfield tarafından yapılmıştır. Garfield'a göre (1970, s. 82) yazarlar konunun öncülerine saygı duymak, iliřkili yayınları açıklamak, yöntemi açıklamak veya kullanmak, literatür bilgisi vermek, önceki alıřmaları eleřtirmek, kanıtlamak, fikir üretmek, veri doğrulamak, gelecek konusunda uyarmak gibi sebeplerle atıf yapmaktadırlar. Ancak günümüzde atıf dizinlerinden elde edilen sayısal verilere gereęinden fazla önem verilmeye başlanması ile atıf motivasyonlarına daha fazla gelir elde etmek, daha fazla atıf kazanmak için daha fazla atıf yapmak gibi yeni motivasyonlar da eklenmiştir.

Atıf dizinlerinden elde edilen sayılar ile arařtırmacıların bilimsel etkinlikleri deęerlendirilmekte, akademik yükseltmeler gerçekleştirilmekte ve ödöller verilmektedir. Örneęin, Hacettepe Üniversitesi Bilimsel Arařtırma Projeleri Koordinasyon Birimi tarafından sağlanan "Yüksek Atıf Bilimsel Başarı Desteęi" ile her bir alan için belirlenmiş azami atıf sayısına ulaşmış akademisyenlerin yurtdışında katılacakları etkinlikler desteklenmektedir (Hacettepe

Üniversitesi Bilimsel, 2017, s. 18, 26).<sup>1</sup> Benzer şekilde devlet üniversitelerinde görev yapan akademisyenler yıllık bilimsel üretimleri dikkate alınarak atıfların da önemli bir puan unsuru olduğu “Akademik Teşvik Ödeneği” ile ödüllendirilmektedirler (Akademik Teşvik Ödeneği Yönetmeliği, 2016).<sup>2</sup> Yine akademik yükselmelerde atıflar önemli bir puan grubunu oluşturmaktadır (Öğretim Üyeliğine Yükseltme ve Atanma Yönetmeliği, 1982). Uluslararası arenada da atıflara verilen değerlerin yansımaları görülmektedir. Örneğin, üniversiteleri değerlendiren pek çok uluslararası sıralama sisteminde atıfların önemli bir yeri bulunmaktadır. Tablo 1’de üniversite sıralama sistemlerinin atıflara verdikleri ağırlıklar yüzdesel olarak gösterilmektedir.

Tablo 1. Uluslararası üniversite sıralama sistemlerinde kullanılan bilimsel etki ölçütlerine ilişkin ağırlıklar (Doğan, 2017, s. 30)

| Bilimsel etki ölçütleri  | THE | QS  | URAP | US  | NTU |
|--|-----|-----|------|-----|-----|
| Atıf sayısı  | %30 |     | %21  | %10 | %25 |
| Atıf etki toplamı  |     |     | %15  |     |     |
| Alan, yayın yılı ve yayın türüne göre normalleştirilmiş yayın başına atıf sayısı |     |     |      | %10 |     |
| Akademisyen başına atıf sayısı   |     | %20 |      |     |     |
| Yayın başına atıf sayısı   |     |     |      |     | %10 |
| Toplam   | %30 | %20 | %36  | %20 | %35 |

Tablo 1’den de anlaşılacağı gibi atıflar uluslararası sıralama sistemlerinin sonuçlarını önemli oranda etkileyebilme potansiyeline sahiptir. Bu bağlamda Türkiye’deki bir örnekten uluslararası camiada olumsuz anlamda söz edildiğini hatırlatmakta yarar vardır. Ege Üniversitesi’nin 2013 yılına ait bir sıralamada matematik ve bilgisayar bilimleri alanında ikinci sırada yer alması (bkz. <http://www.leidenranking.com/ranking/2013>) bilim insanlarının dikkatini çekmiş ve ayrıntılı inceleme yapıldığında bu durumun bir kişi tarafından üretilen 65 yayın ve bu 65 yayına yapılan 421 atıf ile gerçekleştiği, aksi takdirde Ege Üniversitesi’nin 300 civarında bir sırada yer alacağı anlaşılmıştır (Wouters, 2013).

<sup>1</sup> Hacettepe Üniversitesi’nin “atıf başarı desteği” altında verdiği toplantı katılım desteği incelendiğinde, tıp, eczacılık, fen ve mühendislik alanları için 50 olan atıf sayısının dışılık için 27, sosyal bilimler için ise 17 olduğu görülmektedir. Bu durum hiç kuşkusuz kimya, kimya mühendisliği, pediatri, matematik, istatistik, biyoloji gibi alt alanlarda sanki disiplinlerin birbirleri ile aynı atıf yönelimlerine sahip olduğu izlenimi edinilmesine neden olmaktadır. Benzer şekilde bilgi yönetimi, psikoloji, dilbilim ve antropolojinin de aynı potada ele alınıyor olması sorgulanması gereken bir durumdur.

<sup>2</sup> Akademik teşvik ödeneğinde alınabilecek ödeneğin %30’unu atıflardan elde etmek mümkündür.

Ticari bir kurum olan Clarivate Analytics firması tarafından sunulan ve her yılın en çok atıf alan yazarlarının yer aldığı “Highly Cited Researchers” (<http://hcr.stateofinnovation.com/>) listeleri de akademik çevrelerce oldukça yakından takip edilmektedir. Gerek ulusal gerekse uluslararası literatürde atıf sayma ağırlıklı bir yol izlendiğinden atıfların manipülasyonu ile ilgili haberlere gün geçtikte daha fazla rastlanmakta ve bunun sonucu olarak atıfların niteliği tartışılmaktadır.

Günümüzde atıflar ve atıfların değerlendirilmesi ile ilgili en belirgin sorunlar zoraki atıf eklemeleri, atıf manipülasyonları ve “atıf çeteleri”dir. Bilimsel dergiler editoryal süreçlerinde yazarlardan kendi dergilerine ya da editöre ait diğer dergilere atıf yapılmasını talep edebilmekte, bu talebin karşılanmaması durumunda makaleyi yayımlamamaktadırlar. Literatürde zoraki atıf (COPE, 2012) olarak adlandırılan bu durum, ortaya çıkan ürünlerin kalitesini düşürmekte ve ilgisiz pek çok atıfın yapılmasına sebep olmaktadır. Editörlerin yarattığı bu oyun, yazısının yayımlanmama riskini göze alamayan yazarlarca kabul edilmekte, böylece bu durum da artarak sürmektedir (Wilhite ve Fong, 2012, s. 542). Bu gibi uygulamaların örneklerine Türkiye’de (ör. Al ve Soydal, 2012) ve dünyada (ör. Davis, 2017) rastlamak mümkündür. Öte yandan akademisyenler stratejik davranarak rakiplerine atıf yapmama veya çalışma arkadaşlarına atıf yapma eğilimi içine girmişlerdir. Bazı üniversiteler çalışanlarından birbirlerine atıf yapmalarını talep etmektedirler (Retraction Watch, 2017). Her geçen gün daha fazla atıf çetesi haberi ortaya çıkmakta ve atıfların temel misyonu bu şekilde anlamını yitirmektedir (ör. Kaplan, 2014; Oransky ve Marcus, 2017).

Geleneksel atıf analizlerinin temel problemini sadece atıf manipülasyonları başlığı altında sınıflamak eksik bir yaklaşım olacaktır. Atıf sayarak araştırma ve araştırmacıların değerlendirilmesi uzun yıllardır tartışılan konulardan biridir. Atıf saymanın anlamsızlığı üzerine yapılan araştırmalarda ortaya çıkan sonuçlara göre, en az bir atıf alan yayının gelecekte atıf alma potansiyelinin hiç atıf almayan yayına göre daha yüksek olduğu düşünülmektedir (Merton, 1968). Matthew etkisi olarak adlandırılan bu teoride daha yaşlı olan yazarların genç olanlara, daha fazla atıf alan yazıların hiç almayanlara göre birikimli bir avantajı olduğu ve bu avantajın da akademide eşitsizliği beraberinde getirdiği savunulmaktadır (Allison, Long ve Krauze, 1982, s. 615). Zaten Eugene Garfield bile atıfların araştırmacı değerlendirmelerinde tek başına kullanılmasının anlamsızlığı üzerine görüş bildirmiş ve atıfların araştırmacı değerlendirmelerinden çok araştırmaların etkinliğini ve görünürlüğünü ortaya çıkaran göstergeler olduğunu belirtmiştir (Garfield, 1973, s. 407).

Sayılan tüm olumsuzluklara rağmen sıklık saymaya dayanan araştırma ve araştırmacı değerlendirmeleri yöneticiler ve karar vericiler tarafından kolay elde edilebilir olma özelliği

nedeniyle tercih edilmektedir. Bu noktada karar vericilerin bilimsel iletişim, atıflar ve atıfların özellikleri ile ilgili konularda bilgi sahibi olmalarının çok önemli olduğu düşünülmektedir.

Uluslararası literatürde saymanın yerini alabilecek çalışmalar 1950'lerden beri dillendirilmekte ve atıfların içerik olarak değerlendirilmesinin saymaktan daha anlamlı olacağı konusu vurgulanmaktadır. Bu bağlamda yazarların neden atıf yaptıklarını araştırmaya yönelik olarak geliştirilen yeni nesil atıf analizlerine literatürde "içerik tabanlı atıf analizi" adı verildiği görülmektedir (Ding, Zhang, Chambers, Song, Wang ve Zhai, 2014, s. 1820).

### **1.2.2. İçerik Tabanlı Atıf Analizleri**

Literatürde günümüze kadar yapılmış içerik tabanlı atıf analizi araştırmaları kapsamı doğruğultusunda incelendiğinde aşağıda listelenen konularda çalışmalara rastlanmıştır:

- Atfın metin içinde yapıldığı yeri dikkate alan çalışmalar (sentaktik çalışmalar)
- Atıf cümlesinin anlamını dikkate alan çalışmalar (semantik çalışmalar)
- Yazarların atıf yapma motivasyonlarını ortaya çıkarmaya yönelik çalışmalar

İçerik tabanlı atıf analizlerinde yanıtı aranan en temel soru yazarların hangi amaçlarla atıf yaptıklarıdır. Bu amaçlara göre yapılan atıflar arasındaki farklılıkların anlaşılması önemlidir. Bu analiz yöntemine göre yazarların atıf yapma motivasyonlarının ortaya çıkarılmasının ardından her bir atfın anlamına ya da metin içinde konumlandırılmasına göre değerlendirmesinin yapılması gerekmektedir. İçerik tabanlı atıf analizlerinin yeni yapılmaya başlandığı yıllarda örneklem büyüklükleri ve kullanılan teknikler nedeni ile genele uyarlanamayacağı belirtilse de, günümüzde içerik tabanlı atıf analizleri doğal dil işleme teknikleri kullanılarak daha kolay şekilde büyük veri üzerine uygulanabilmektedir. Uluslararası literatürde yer alan semantik ve sentaktik çalışmaların detayları ve bu çalışma ile karşılaştırmalı sonuçları raporun yöntem ve bulgular kısımlarında bulunmaktadır.

### **1.2.3. Doğal Dil İşleme Yöntemleri ve Uygulamalar**

Doğal dil işleme doğal metinlerin, konuşmaların ya da diğer kullanışlı ürünlerin bilgisayarlarca nasıl kullanılacağıının anlaşılması ve öğretilmesidir (Chowdhury, 2003, s. 51). Bu işlemenin temel amacı metne dayalı çalışmalarda "insan gibi" işleme başarısına ulaşmak ve bu sayede insan emeğini en aza indirerek bilgisayarlara yüksek başarılı işlemler yaptırabilmektir (Liddy, 2010, s. 3864).

Doğal dil işleme çalışmalarının üç temel aşaması vardır. Bu aşamalar veri toplama, veri ön işleme ve analitik işleme olarak tanımlanabilir. Bu işleme süreçlerinde problemin doğru bir

şekilde tanımlanması, verilerin hazırlanması, modelin kurulması ve kullanılacak yöntem ve algoritmaların doğru bir şekilde belirlenmesi önemlidir (Akpınar, 2014, s. 77-80).

Doğal dil işleme uygulamaları pek çok alanda kullanılmakta, bu sayede dillerin özellikleri belirlenebilmekte, bu özellikler üzerinden çeşitli uygulamalar gerçekleştirilebilmektedir. En bilinen doğal dil işleme yöntemleri şunlardır:

- **Bilgi erişim:** Bilgi miktarının hızla arttığı günümüzde en doğru bilgiye en doğru zamanda ulaşabilmek hayati öneme sahiptir. Bu sebeple doğal dil işleme uygulamaları ile bilgi erişimin etkinliği artırılmaya çalışılmaktadır. Burada temel amaç ihtiyaç duyulan bilginin büyük yoğunluklu bir metnin içerisinden erişiminin sağlanabilmesidir (Lewis ve Jones, 1996, s. 92).
- **Bilgi çıkarımı:** Bilgi çıkarımında temel amaç büyük yoğunluklu metinlerin içinden anlamlı temel anahtar unsurlara erişimin sağlanmasıdır (Liddy, 2010, s. 3871). Bu uygulama bazı durumlarda bir doğal dil işleme çalışmasının başlangıcını oluşturabildiği gibi bazen de başlı başına bir araç olarak kullanılabilir.
- **Makine çevirisi:** Metinlerin bir dilden diğer dile makinece çevrilmesini sağlayan uygulamalardır. Bu uygulamalar sayesinde farklı kültürlerin kaynaşması ve yok olmaya başlayan dillerin yaşatılması amaçlanmaktadır (Manning ve Schütze, 1999, s. 463).
- **Özetleme:** Uzun metinler içinden anlamlı özetler oluşturmayı amaçlayan uygulamalardır. Bu uygulamalar dilsel veya istatistiksel yöntemleri kullanarak anlamlı bir özet çıkarmaya çalışırlar (Chowdhury, 2003, s. 60).
- **Metin Kategorizasyonu:** Daha önceki örüntülerden hareketle gelecekte gerçekleşebilecek olayları tahminlemeye çalışan bir yöntemdir (Silahtaroglu, 2013, s. 67). Bu yöntem ile hava durumu tahminlerinden hastalık tanılamaya pek çok uygulama gerçekleştirilebilmektedir.

Pek çok alanda olduğu gibi kütüphanecilik ve bilginbilim alanında da doğal dil işleme yöntemleri son yıllarda sıklıkla kullanılmaya başlanmıştır. Uluslararası literatürde yeni sınıflama sistemleri yaratma, ontoloji geliştirme, duygu analizi gerçekleştirme, bilimsel iletişim süreçleri için veri sağlama, araştırma ve araştırmacı değerlendirme için bu yöntemlerin kullanıldığı görülmektedir. Ulusal literatürde ise bu gibi çalışmalar yeni yeni yaygınlaşmaya başlamıştır. Sosyal ağlarda yer alan bilgiyi doğrulama, niteliksel bibliyometrik çalışmaların sürdürülmesi ve otomatik özetleme gibi konularda yapılan çalışmalar ile kütüphanecilik ve bilginbilim alanında benzer konular çalışılmaya başlanmıştır.

#### **1.2.4. Türkiye’de Atıf Analizi Çalışmaları**

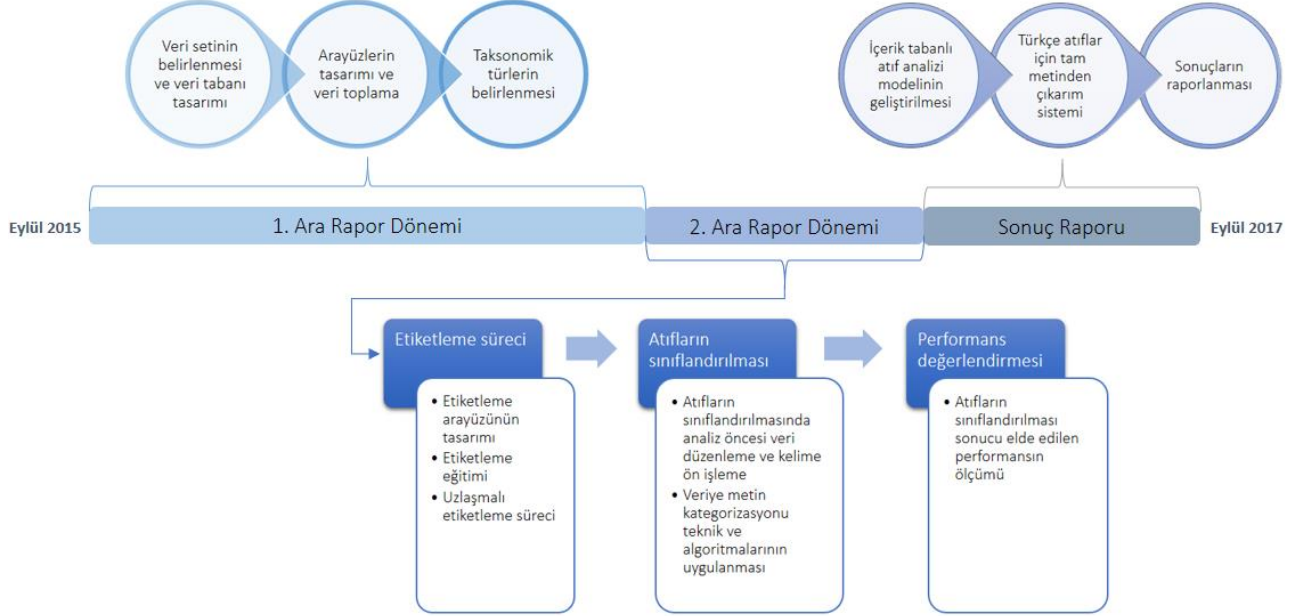
Türkiye kütüphanecilik ve bilginbilim literatüründe yer alan atıf analizi çalışmalarında atıflar sıklıkla niceliksel olarak değerlendirilmekte veya bibliyometrik haritalama yöntemi ile ortak atıf analizleri yapılmaktadır. Atıfların içeriklerinin değerlendirilmesine yönelik olarak ise az sayıda yayın üretilmiştir. Bu yayınların ilki Türkçe kütüphanecilik ve bilginbilim literatüründe yapılan temel atıf sınıflamasıdır (Atılgan, Atakan ve Bulut, 2008, s. 399). Bu çalışmada atıflar tek tek değerlendirilerek Türkçe atıflar için kanıt gösterme ve yönlendirme olarak iki temel sınıf oluşturulmuştur. Anılan çalışma Türkiye’de yapılmış ilk atıf motivasyonu sınıflaması olarak kabul edilebilir.

Niteliksel atıf analizi yapmayı hedefleyen tek çalışmada ise atıfların anlamsal bir şekilde sınıflandırılmasını dikkate alarak geliştirilecek bir h-endeks değerinin gerekliliğinden bahsedilmiştir (Tunç, 2012). Çalışmada ayrıca atıfların anlamsal sınıflamasının atıf yapan yazarlar tarafından belirtilmesine yönelik bir öneride bulunulmuştur. Bu öneriye göre yazarlar atıf yaptıkları zaman pozitif atıfların başına artı (+) negatif atıfların başına da eksi (-) işareti koyarak gönderme yaptıklarında bu gibi atıfların ayrıştırılması mümkün olabilecektir. Ancak, bu öneri yazarların duygu belirtmek istememeleri, geçmiş yayınların işaretlenemeyecek olması, atıf gelenekleri ve yazarlara ekstra bir külfet getireceğinden kullanışlı ve pratik bir yöntem değildir. 1970’li yılların sonunda uluslararası literatürde de bu konu tartışılmış ve pratik bir yöntem olmadığı o dönemde de vurgulanmıştır (Small, 1978, s. 329). Yukarıda sayılan iki yayın dışında Türkçe literatürde çalışma yapılmamıştır. Bu sebeple bu Proje ve bu Proje kapsamında gerçekleştirilen İçerik Tabanlı Atıf Analizi Modeli Tasarımı: Türkçe Atıflar için Metin Kategorizasyonuna Dayalı Bir Uygulama başlıklı doktora tezi (Taşkın, 2017) Türkiye’de atıf analizi ile doğal dil işleme yöntemlerini birlikte kullanarak yürütülen öncü içerik tabanlı değerlendirme örnekleridir.



## 2. BÖLÜM: YÖNTEM VE TEKNİK

Bu çalışmada yürütülen tüm süreçler ve zaman planı Şekil 1’de gösterildiği gibi planlanmış ve bu takvime uygun bir şekilde tamamlanmıştır. Bu bölümde her bir aşamada gerçekleştirilen işlemler hakkında kısa bilgiler verilmektedir.



Şekil 1. Çalışma süreçleri ve zaman planı

### 2.1. Veri Setinin Belirlenmesi ve Veri Tabanı Tasarımı

Türkçe atıfların semantik ve sentaktik özelliklerini ortaya çıkarmak amacı ile Türk kütüphanecilik ve bilgilim literatürü seçilmiş, bu literatürün iki temel kaynağı olan *Türk Kütüphaneciliği* ve *Bilgi Dünyası* dergilerinde yayımlanmış hakemli makaleler üzerinden çalışma yapılması kararlaştırılmıştır. Bu amaçla *Türk Kütüphaneciliği* dergisinde 1996-2015 yılları arasında yayımlanmış 251 makale ve *Bilgi Dünyası* dergisinde 2000-2015 yılları arasında yayımlanmış 172 Türkçe makale derinlemesine incelenmiştir. *Bilgi Dünyası* dergisinde yayımlanmış 23, *Türk Kütüphaneciliği*’nde yayımlanmış altı İngilizce makale çalışma kapsamının Türkçe literatür ile kısıtlı olmasından dolayı araştırmaya dâhil edilmemiştir.

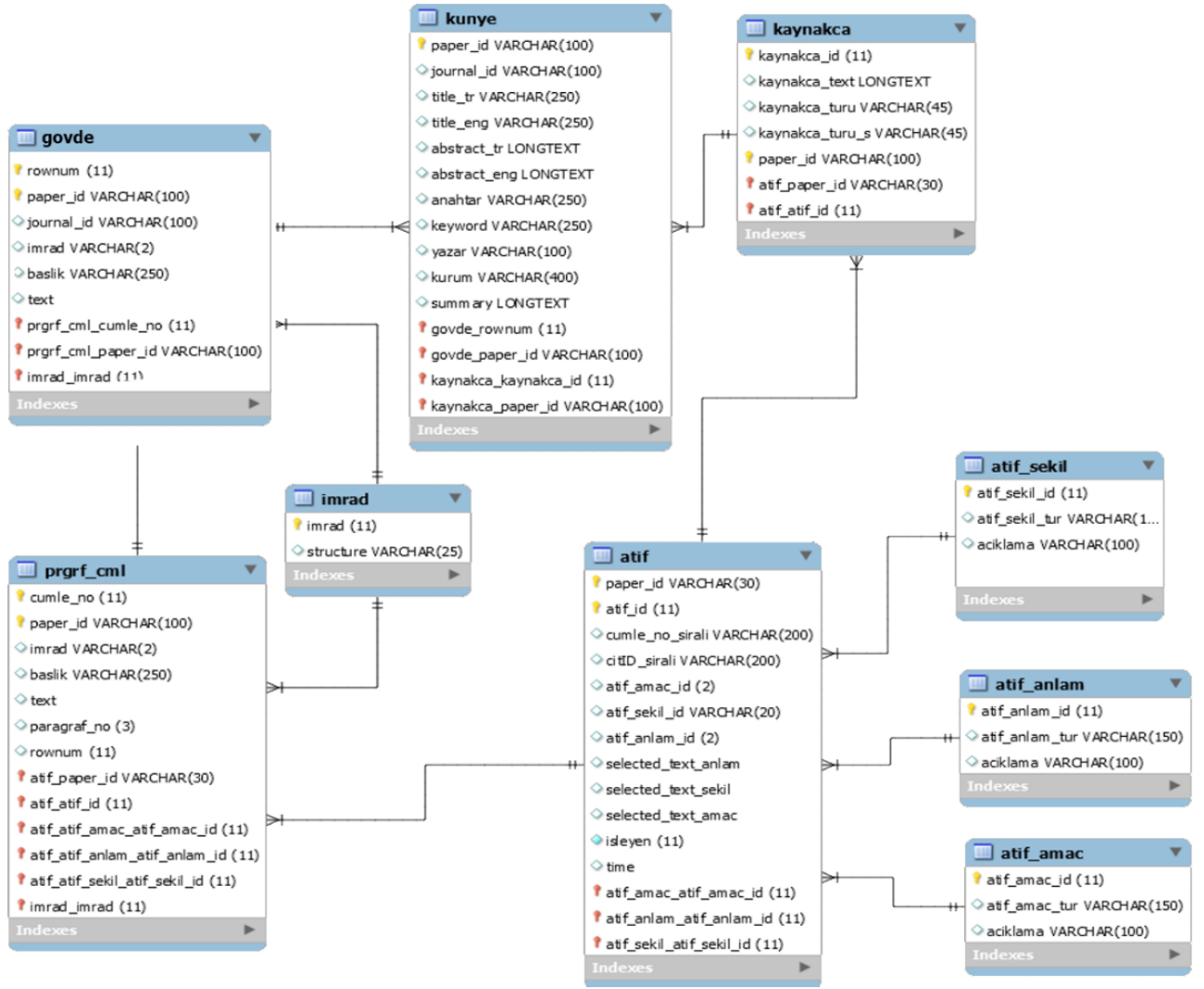
Çalışma kapsamına alınan her iki dergide yayımlanmış toplam 423 makale .pdf formatında kaydedilmiş ve OCR (Optical Character Recognition - Optik Karakter Tanımlama) işlemi yapılmamış dokümanlar için OCR işlemi yapılmıştır. Ardından indirilen tüm yayınlara akıllı birer numara verilerek dosyalar kimliklendirilmiştir. Bu kimlik numaralarında temel yapı aşağıdaki gibidir:

- *Dergi adı*+*yıl*+*sayı*+*no* (Ör: *TK201121*)

Bu numaralandırma sisteminde dergi adları kısaltmaları ile verilmiş ve *Bilgi Dünyası* dergisi için BD, *Türk Kütüphaneciliği* dergisi için TK kısaltması kullanılmıştır. Örnekte gösterilen numara *Türk Kütüphaneciliği* dergisinde 2011 yılının ikinci sayısında yayımlanmış ilk makaleyi temsil etmektedir.

Makalelerin pdflerinin optik tanımlamasının yapılması ve kimliklendirilmesinin ardından her bir makalenin .txt formatına dönüştürülmesi aşamasına geçilmiştir. Bu aşamada kodlama standardı olarak Türkçe karakter uyumsuzluklarını engellemek için UTF-8 seçilmiştir.

Dosyaların indirilmesi, kimliklendirilmesi ve dönüştürülmesi süreçlerinin ardından tüm makalelere ait künye, kaynakça ve tam metin bilgilerinin sistematik bir şekilde toplanabilmesi ve işlenebilmesi amacı ile MySQL tabanlı bir veri tabanı yaratılmış, yaratılan bu veri tabanı ile verilerin daha doğru ve etkili şekilde sağlanmasının yolu açılmıştır. Veri tabanı sunucusu olarak Hacettepe Üniversitesi Bilgi İşlem Dairesi Başkanlığı'nda bulunan sunuculardan 115K440 olarak isimlendirilen bölüm ayrılmıştır. Oluşturulan bu ilişkisel veri tabanının ana yapısı Şekil 2'de gösterildiği gibidir.



Şekil 2. Oluşturulan ilişkisel veri tabanının yapısı

Hazırlanan bu veri tabanında dokuz ana tablo bulunmaktadır. Bu tabloların amacını kısaca özetlemek gerekirse;

- Künye, gövde ve kaynakça tablolarında yayınlara ait temel bileşenler tutulmaktadır ve bu tablolar birbiri ile ilişki içindedir. Künye kısmında çalışmaların temel üst veri bilgileri tutulurken, çalışmalara ait tam metinler paragraflarına ayrılmış şekilde gövde tablosunda yer almaktadır. Kaynakça tablosunda ise yayınlarda yapılan atıflar tutulmaktadır.
- Tam metinlerde yer alan metinlerin çalışma içinde yer aldıkları bölümleri ayırt edebilmek amacı ile IMRAD tablosu yaratılmıştır. Paragraflar metin içinde yer aldıkları bölüme göre giriş, yöntem, bulgular, sonuç, teşekkür ve ekler olarak sınıflandırılmıştır.

- Atıf tablosu etiketlenen atıf cümleleri için tasarlanmıştır ve atıf\_şekil, atıf\_anlam ve atıf\_amaç tablolarından beslenmektedir. Atıf cümlelerinin sentaktik dizilimini ortaya çıkarmak üzere de prgrf\_cml tablosu kullanılmaktadır.

Veri tabanının oluşturulmasının ardından veri toplama aşamasına geçilebilmesi için arayüz tasarımları yapılmıştır.

## 2.2. Arayüzlerin Tasarımı ve Veri Toplama

Künye, kaynakça ve tam metin depolama çalışmaları için hazırlanan veri tabanına veri girişi yapılabilmesi için çeşitli arayüzler geliştirilmiştir. Bu arayüzlerde temel amaç makalelere ait tüm bilgilerin sistematik bir şekilde toplanabilmesidir. Bu arayüzler <http://115k440.hacettepe.edu.tr> adresinde erişilebilir durumdadır.

Veri giriş sayfası ilk açıldığında gelen ekranda makale seçimi yapılmaktadır. Makale seçimi için hiyerarşik yapı önce dergi seçimi, ardından yıl ve o yılda yayımlanan makalelerin seçimi şeklinde gerçekleştirilmektedir. Makale seçimi yapıldıktan sonra ise künye, kaynakça ve gövde bağlantılarından ilgili verilerin giriş işlemi yapılmaktadır.

Künye arayüzünde makalelerin Türkçe ve İngilizce başlıkları, özleri, yazar ve kurum bilgileri ile *Türk Kütüphaneciliği* dergisi makalelerinde bulunan özet (summary) bölümleri depolanmıştır. PHP programlama dili kullanılarak geliştirilen bu arayüzde veri tabanında depolanan .txt'ler tanımlanmış ve tüm metinler makinece okunabilir hale getirilmiştir. Bunu yapabilmek için her iki derginin şablonları ve kullandıkları diller incelenmiş ve standart bir yapı belirlenmiştir. Örneğin, “öz” sözcüğünden “anahtar kelimeler” sözcüğüne kadar olan kısım otomatik olarak gruplandırılmıştır. Benzer şekilde “anahtar kelimeler” ile “abstract” arasında kalan bölüm de arayüzde bir arada sunulmuştur. Ancak, gerek dergilerin standart bir yapıyı korumuyor olmaları, gerekse makinece yapılan hataları en aza indirmek açısından sınıflandırma işlemi elci sistemle kontrol edilmiştir. Makinece gruplanan bilgiler Proje bursiyerleri tarafından alanlarına ayrılmış ve veri tabanında depolanmıştır. Sonuçta toplam 423 makaleye ait tüm künye bilgileri veri tabanına eklenmiştir.

Kaynakça bilgilerinin toplanması için ise tam metinlerde “kaynakça”, “kaynaklar”, “referanslar” gibi sözcükler makinece başlangıç olarak kabul edilmiş, bu başlıklardan sonra gelen kısım açılan metin kutusunda görüntülenmiştir. Bu aşamada bursiyer ekibin yaptığı veri girişi çalışmasında her bir kaynakçanın düzenlenmesi ve birbirinden ayrılması sağlanmıştır. Bu sayede veri tabanında kullanılan kaynakların tamamı depolanabilmiştir.

Türk kütüphanecilik ve bilginbilim literatüründe önemli düzeyde kaynak gösterme sorunları bulunmaktadır (Kurbanoglu, 1996; Dogan, 2014). Bu nedenle kaynakça verilerinin makinece otomatik olarak toplanmasi verilerin dogruluđu ve tamliđi ađısından sorun yaratacaktır. Bu tür kaynak gösterme hatalari dikkate alınarak kaynakça verilerinin düzenlenmesi işi proje bursiyerlerince gerçekleştirilmiştir. Bursiyerlerin görevi kaynakçalarda yer alan her bir kaynađın tanımlanması, gereksiz alt-üst bilgilerin silinmesi ve kaynakçaların birbirinden bir belirteç ile ayrılmasıdır. Bu işlemin ardından veri tabanında toplam 12.881 kaynakça bilgisi depolanmıştır.

Tam metinlerin tamamının bölümlerine göre sistematik olarak depolanması proje ađısından büyük önem taşımaktadır. Tam metin girişı için kullanılan arayüzde veri girişı yapılacak makale bölümünün IMRAD yapısında hangi kısımda yer aldığı, yazarın tercih ettiđi başlıđı ve o bölümde yer alan metni etiketlemek hedeflenmiştir. Makalelerin kaçının bir araştırmaya dayandığı, kaçında yöntemin tanımlandığı, kaçında tartışma bölümünün yer aldığı gibi soruları yanıtlandırabilmek için IMRAD yapısı Proje ekibince etiketlenmiş ve araştırılmıştır.

Her ne kadar UTF-8 standardı ile optik tanıma yapılmış olsa da makale dosyalarında karakter hatalari saptanmıştır (ı,ı gibi). Bu hataların düzeltilmesi sürecinin tam metin girişlerini yavaşlatmasını engellemek amacı ile hatalı karakterler saptanmış ve veri tabanı üzerinden otomatik olarak düzeltilmiştir. Tam metin etiketleme işlemi sonrası veri tabanında Türk kütüphanecilik ve bilginbilim literatüründe yayımlanmış 101.019 cümle depolanmıştır.

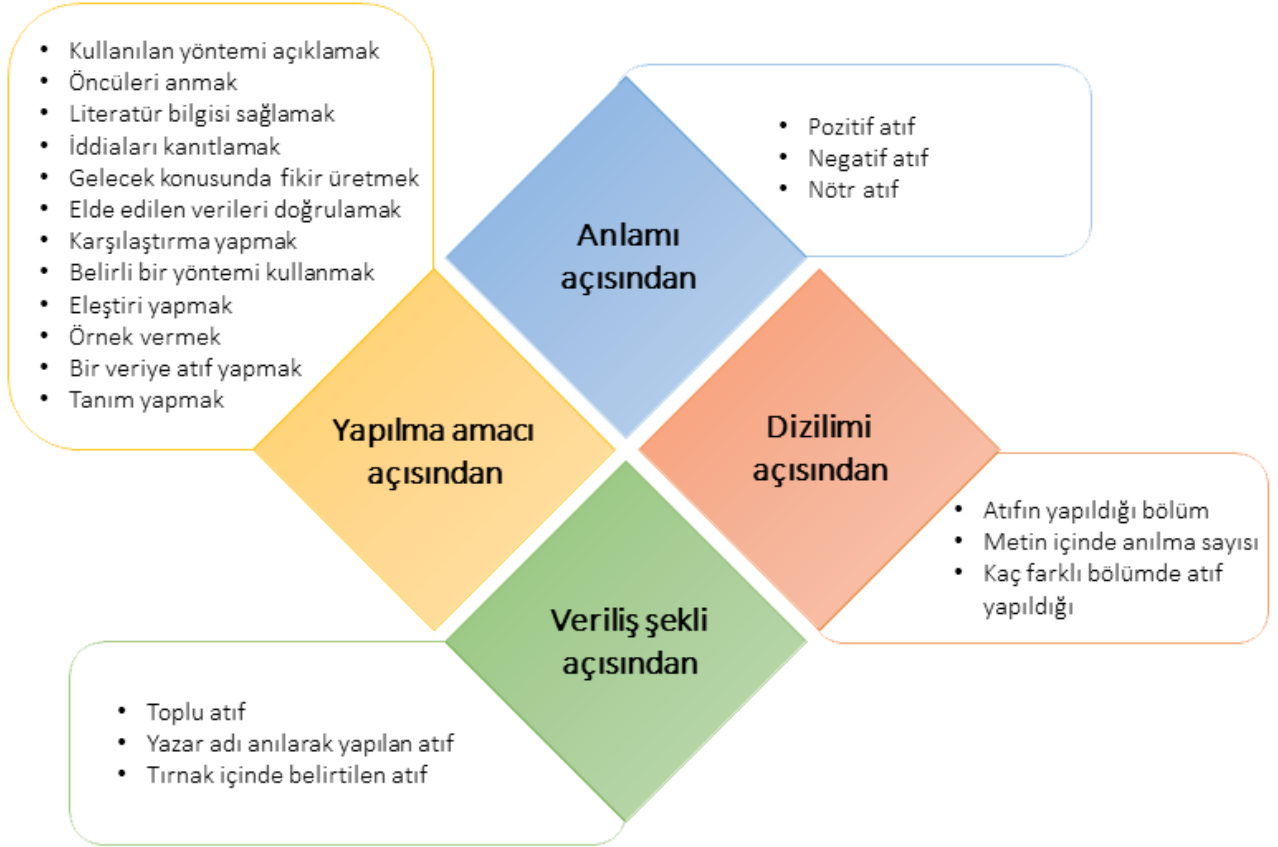
### **2.3. Taksonomik Atıf Türlerinin Belirlenmesi**

Makalelere ait tüm bilgilerin toplanmasının ardından içeriksel atıf etiketlemesi sürecinin başlatılabilmesi için atıfların taksonomik türlerinin belirlenmesi gerekmektedir.

Türk Dil Kurumu Güncel Türkçe Sözlük'te taksonomi sözcüğü "Sınıflandırılma ve bu sınıflandırmada kullanılan kurallar bütünü" şeklinde tanımlanmaktadır (TDK, 2017b). Taksonomi sözcüğü literatürde genellikle canlı ve hayvanların sınıflandırılması amacı ile kullanılırken, son yıllarda atıfların konusal olarak sınıflandırılması anlamında da taksonomi sözcüğünden yararlanılmaktadır (Erikson ve Erlandson, 2014). Taksonomi sözcüğüne benzer olarak atıfların sınıflandırılması için literatürde kullanılan diđer sözcükler ise atıf yapma motivasyonu (Bonzi ve Snyder, 1991), atıfların duygu analizi (Athar, 2011; Small, 2011), atıf kategorizasyonu (Bertin, 2008), atıf davranışı (Cano, 1989), atıf kapsam analizi (McCain ve Turner, 1989; Maričić, Spaventi, Pavičić, ve Pifat-Mrzljak, 1998) ve atıf yapma fonksiyonudur (Teufel, Siddharthan ve Tidhar, 2006). Bu gibi çalışmalarda en temel amaç araştırmacı değerlendirmelerinde önemli bir ölçüt olarak kullanılan atıf sayılarının aslında çok anlamlı

olmadığının gösterilmesi ve anlamlı olanın atıfların içeriksel bir değerlendirmeden geçirilmesi olduğunun ortaya çıkarılmasıdır. Çünkü atıf sayılarının yıllar içinde artması ile birlikte atıf yapma motivasyonlarında etik olmayan davranışlar da gözlenmeye başlanmıştır. Çalışmalara hiçbir katkısı olmayan (Öztürk, 2013) ya da çalışma arkadaşlarına kredi sağlamak adına yapılan atıfların sayısında artış gözlenmiştir (Goudsmith, 1974, s. 28). Bu durumda her atıfa eşit derecede anlam yüklenmesi çeşitli eleştirileri de beraberinde getirmiştir (Jha, Jbara, Qazvinian ve Radev, 2017; Voos ve Dagaev, 1976).

Literatürde atıfların sınıflandırılması üzerine pek çok çalışma yapılmış ve bu çalışmalar sonucunda çeşitli atıf sınıflama sistemleri geliştirilmiştir. Bu çalışmada literatürde atıf sınıflama ve değerlendirilmesi üzerine yapılmış çok sayıda araştırma incelenmiş (bu çalışmaların listesi için bkz. Ek 2) ve bu araştırmalar aracılığı ile ortaya koyulan taksonomik türler sentezlenmiştir. Bu türlerin temel sınıflandırması Şekil 3'te gösterildiği gibidir.



Şekil 3. Taksonomik atıf sınıfları

### **2.3.1. Yapılma Amacı Açısından Atıflar**

Atıfların yapılma amaçları birbirinden farklılıklar gösterebilmektedir. Burada önemli olan yazarların atıf yaparken hangi unsurları ön plana aldığıın ortaya çıkarılabımesidir. Atıfların yapılma amaçları dikkate alınarak yapılan sınıflandırmalar incelendiğinde 12 temel atıf yapma amacı saptanmıştır. Bu amaçlar yöntem açıklamak, öncülerini anmak, literatür bilgisi sağlamak, iddiaları kanıtlamak, gelecek için fikir üretmek, veri doğrulamak, karşılaştırma yapmak, yöntem kullanmak, eleştiri yapmak, örnek vermek, veriye atıf yapmak ve tanım yapmaktır (her bir sınıfa ait örnekler için bkz. Taşkın 2017, s. 47).

### **2.3.2. Anlamı Açısından Atıflar**

Literatürde üzerinde en çok tartışma yapılan konu pozitif, negatif ve etkisiz (nötr) atıflardır. Bazı görüşler negatif atıfların da bilimi geliştirdiğini savunurken (Cole ve Cole, 1971, s. 26; Carter, 1974; Garfield, 1979, s. 362), çoğunluğun görüşü negatif atıflar ile pozitif atıfların değerlerinin kesinlikle aynı olmadığı ve değerlendirme yapılırken atıfların anlamsal farklılıklarının ortaya çıkarılmasının önemli olduğudur (Voos ve Dagaev, 1976; Spiegel-Rösing, 1977; Chubin, 1980). Negatif atıfların yanında çalışmalara herhangi bir katma değer sağlamayan ve metin içinde kısaca anılan atıflar da literatürde eleştirilmiştir (Moravcsik ve Murugesan, 1975). Bu doğrultuda Proje kapsamında atıflar anlamı açısından pozitif, negatif ve nötr olmak üzere üç temel gruba ayrılmıştır.

### **2.3.3. Dizilimi Açısından Atıflar**

Bir yayına bir makale içinde birden fazla kere atıf yapılıyorsa (Herlach, 1978, s. 310) veya bulgular kısmında atıf yapılıyor ya da diğer makalede geliştirilen yöntem temel alınıyorsa (Maričić, Spaventi, Pavičić ve Pifat-Mrzljak, 1998, s. 530-540) bu yayına yapılan atıfların diğerlerine göre daha değerli olduğu düşünülmektedir. Bu bağlamda atıflar dizilimleri açısından da sınıflandırılmış ve metin içinde kaç kere atıf yapıldığı, hangi bölümlerde ve kaç farklı bölümde atıf yapıldığı bilgileri sınıf içindeki grupları oluşturmuştur.

### **2.3.4. Veriliş Şekli Açısından Atıflar**

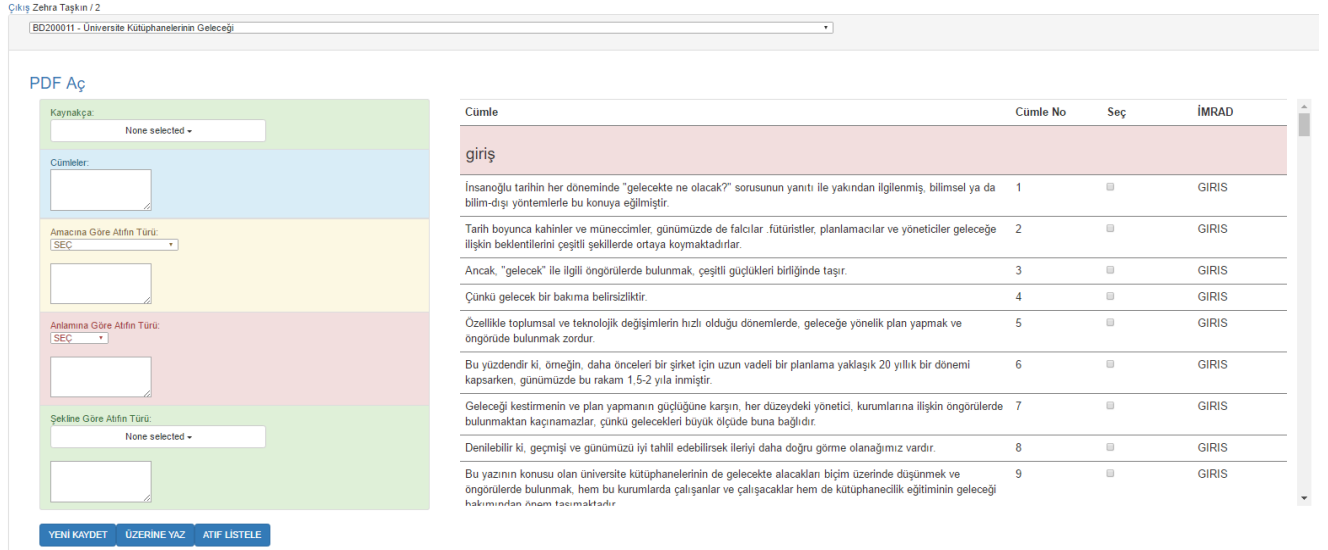
Bazı makaleler metin içinde yazar adı anılarak, tırnak işareti ile direkt alıntı yapılarak ya da sayfa numarası verilerek atıflanmaktadır (Bonzi, 1982, s. 211). Bu durumda atıf yapan yazarın daha fazla önem verdiği kaynaklar ortaya çıkmaktadır. Böylece Proje kapsamında hazırlanan son sınıfı veriliş şekli açısından atıflar oluşturmaktadır.

## 2.4. Etiketleme Süreci

Etiketleme süreci temel olarak etiketleme arayüzünün tasarımı, etiketleme eğitimi ve uzlaşmalı etiketleme olmak üzere üç alt bölümden oluşmaktadır.

### 2.4.1. Etiketleme Arayüzünün Tasarımı

Etiketleme arayüzünün tasarımında temel amaç Türkçe kütüphanecilik ve bilginilim literatüründe yapılmış tüm atıfların cümlelerinin tekil olarak toplanması, anlamlandırılması ve bu anlamlara göre sınıflandırılmasıdır. Bu amaçla hazırlanan arayüz <http://115k440.hacettepe.edu.tr/> adresinde yer alan "atıf işaretleme işlemleri" sekmesinin altından erişilebilir durumdadır.<sup>3</sup> Bu alana giriş için kullanıcı adı ve şifre ile sağlanan bir denetim mekanizması geliştirilmiştir. Sözü edilen mekanizma sayesinde hiçbir etiketçi bir diğerinin yaptığı etiket bilgisini görememekte, bu sayede etiketçilerin birbirinden etkilenme ihtimalleri en aza indirilmektedir. Etiketleme arayüzünün genel görünümü Şekil 4'te gösterildiği gibidir.



The screenshot shows a web interface for tagging. On the left, there are four sections for selecting tags based on different criteria: 'Kaynakça' (Source), 'Cümleler' (Sentences), 'Amacına Göre Atıfın Türü' (Type of Reference by Purpose), and 'Anlamına Göre Atıfın Türü' (Type of Reference by Meaning). Each section has a dropdown menu and a text input field. Below these sections are buttons for 'YENİ KAYDET' (New Save), 'ÜZERİNE YAZ' (Write on Top), and 'ATIF LİSTELE' (List References). On the right, there is a table with columns 'Cümle' (Sentence), 'Cümle No' (Sentence No), 'Seç' (Select), and 'İMRAD' (Status). The table contains several rows of text with corresponding sentence numbers and status indicators.

| Cümle  | Cümle No | Seç                      | İMRAD |
|--|----------|--------------------------|-------|
| giriş  |          |                          |       |
| İnsanoğlu tarihin her döneminde "gelecekte ne olacak?" sorusunun yanıtı ile yakından ilgilmiş, bilimsel ya da bilim-dışı yöntemlerle bu konuya eğilmiştir.   | 1        | <input type="checkbox"/> | GIRIS |
| Tarih boyunca kahinler ve müneccimler, günümüzde de falcular, 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 belirsizdir.  | 4        | <input type="checkbox"/> | GIRIS |
| Özellikle toplumsal ve teknolojik değişimlerin hızlı olduğu dönemlerde, geleceğe yönelik plan yapmak ve öngörülerde bulunmak zordur.   | 5        | <input type="checkbox"/> | GIRIS |
| Bu yüzdendir ki, örneğin, daha önceleri bir şirket için uzun vadeli bir planlama yaklaşık 20 yıllık bir dönem kapsarken, günümüzde bu rakam 1,5-2 yıla inmiştir.   | 6        | <input type="checkbox"/> | GIRIS |
| Geleceği kestirimin ve plan yapmanın güçlüğüne karşın, her düzeydeki yönetici, kurumlarına ilişkin öngörülerde bulunmaktan kaçınmazlar, çünkü gelecekteki büyük ölçüde buna bağlıdır.  | 7        | <input type="checkbox"/> | GIRIS |
| Denilebilir ki, geçmiş ve günümüze iyi tahili edebilirsek ileriyi 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 kütüphanecilik eğitiminin geleceği bakımından önem taşımaktadır. | 9        | <input type="checkbox"/> | GIRIS |

Şekil 4. Etiketleme arayüzünün genel görünümü

Etiketleme arayüzünde yapılması gereken ilk işlem sayfanın en üst kısmında yer alan açılır kutu içinden etiketlemenin yapılacağı makalenin seçilmesidir. Etiketlenecek tüm makaleler uzman bursiyerler arasında paylaşılmış ve her bir atıf cümlesinin en az iki etiketçinin etiketlemesi sağlanmıştır. Bu sayede görev tanımı önceden yapılmış etiketçiler açılır kutudan kendilerine tanımlanan makaleleri seçerek etiketleme işlemine başlayabilmislerdir.

<sup>3</sup> Sisteme giriş yapılabilmesi kullanıcı adı ve şifre ile mümkündür. Görüntülemek için burada belirtilen kullanıcı bilgileri kullanılabilir: Kullanıcı adı: 115440 ; Şifre: q1q1q1



Etiketleme arayüzünün temel öğeleri şu şekildedir:

- Atıf cümlesinin eşleştirileceği kaynakça bilgisi (bu alan bir cümle için birden fazla kaynak seçilebilecek şekilde yapılandırılmıştır).
- Atıfların taksonomik türlerine ait sınıfların seçimi ve bu sınıfları seçmeyi gerektiren sözcüklerin işaretlenmesi.
  - Anlamına ve amacına göre atıflarda etiketçi yalnızca tek bir sınıf seçebilir. Yani bir atıf yalnızca bir sınıfa ait olabilir.
  - Şekline göre atıflarda bir atıf cümlesi birden fazla sınıfa dâhil olabileceğinden bir cümle için birden fazla sınıf seçilebilir.
- Makalede geçen tüm cümleler ve atıf cümlelerinin işaretlenmesini sağlayacak işaret kutuları.

Dizilimi açısından atıflar taksonomik sınıfı bu etiketleme sürecine dâhil edilmemiştir. Bu tür atıflar metin içinde alıntılanma sayısı, göndermenin yapıldığı bölüm gibi temel istatistiki bilgilere dayandığından ve veri tabanından kolaylıkla çekilebilen bilgiler olduğundan herhangi bir etiketleme süreci gerektirmemektedir. Etiketleme arayüzünün tasarımının ardından bursiyerlere etiketleme eğitiminin verilmesi ikinci aşamayı oluşturmuştur.

#### **2.4.2. Etiketleme Eğitimi**

Görevli bursiyerlerin hepsinin atıflar konusunda benzer yaklaşımlarla etiketleme yapabilmeleri sağlamak amacı ile çeşitli eğitim toplantıları düzenlenmiştir. Tüm Proje ekibinin katıldığı söz konusu eğitim toplantılarında örnek makale etiketlemeleri gerçekleştirilmiş ve bu sayede temel ilkeler belirlenebilmiştir. Bunun yanında bireysel olarak etiketleme yapılması sırasında karşılaşılabilecek problemler için etiketleme eğitimi infografiği oluşturulmuş ve Proje web sayfasından yayımlanmıştır. Infografik [http://115k440.hacettepe.edu.tr/etiketleme\\_egitimi.jpg](http://115k440.hacettepe.edu.tr/etiketleme_egitimi.jpg) adresinden erişilebilir durumdadır.

#### **2.4.3. Uzlaşmalı Etiketleme Süreci**

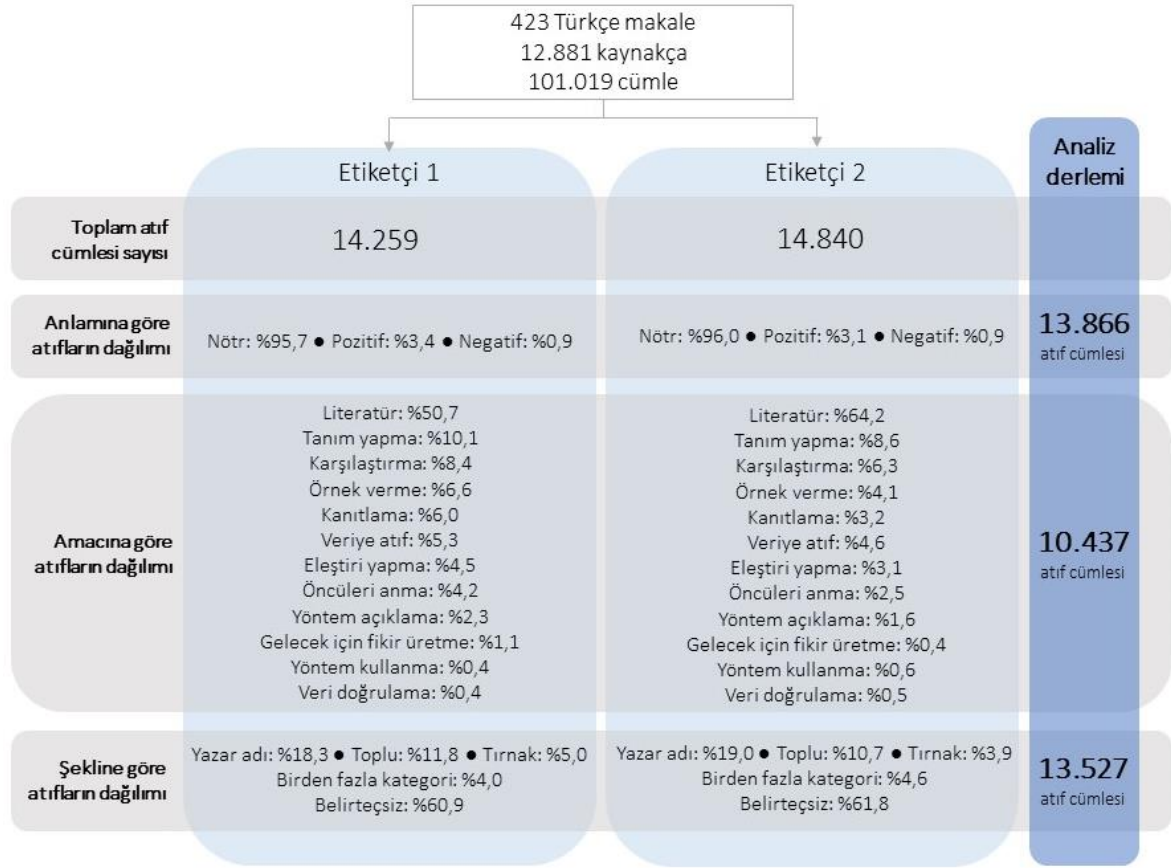
Yukarıda da bahsedildiği gibi çalışma kapsamında yapılacak etiketlemenin planı her bir atıf cümlesinin en az iki etiketçi tarafından etiketlenmesi şeklinde tasarlanmıştır. Bunun temel nedeni ise ortak karar doğrultusunda üzerinde fikir birliği sağlanmış atıfların analiz edilmesi ve bu sayede araştırma sonunda geliştirilecek modelde tutarlı sonuçlara ulaşılabilmesidir. Çünkü herhangi bir sabit veriye dayanmayan ve etiketçilerin duygu tahmin etmesi ile gerçekleştirilen sınıflama çalışmalarında sözü edilen tutarlılığın sağlanması çok önemlidir. Bu araştırma özelinde örnek vermek gerekirse, atıf cümlelerinin anlamsal veya amacına göre etiketlenmesinde tam doğru kabul edilebilecek kurallar listesi oluşturmak imkânsızdır. Zira bir

etiketçiye olumsuz görünen görüş diğer etiketçi için nötr olabilir. Kişiler tarafından farklı yorumlanabilecek durumlar olabileceğinden hareketle de bu çalışma için uzlaşmalı etiketleme (inter-annotator agreement) adı verilen ve iki ya da daha fazla etiketçinin bir etikette işaretlediği aynı kategoriler ile tutarlılığın sağlanması (Bhowmick, Mitra ve Basu, 2008, s. 58) anlamına gelen bu doğrulama yöntemi seçilmiştir.

Uzlaşmalı etiketleme sürecinde her bir etiketçinin yürüttüğü uygulamanın aşamaları şöyledir:

- Etiketçi kullanıcı adı ve şifresi ile giriş yapar,
- Etiketleme yapacağı makaleyi seçer,
- Açılan makaleden atıf yapılmış cümleyi seçer (bir atıf cümlesi bir ya da daha fazla cümleden oluşabilir),
- Seçtiği atıf cümlesinin göndermesinin bulunduğu kaynakça öge/lerini işaretler,
- İşaretlenen atıfın amacına göre (literatür, yöntem açıklama, tanım yapma vb.) hangi sınıfta yer aldığını işaretler ve bunu gösteren sözcükleri metin kutusuna kopyalar,
- İşaretlenen atıfın anlamına göre (pozitif, negatif veya nötr) hangi sınıfta yer aldığını işaretler ve bunu gösteren sözcükleri metin kutusuna kopyalar,
- İşaretlenen atıfın şekilsel özelliklerini değerlendirerek şekil sınıfını (yazar adı, tırnak işareti veya toplu atıf) belirler,
  - Bir atıf birden fazla şekline göre atıf kategorisinde yer alabilir,
  - Yalnızca yazar adı anılarak yapılan atıflar için atıfı tanımlamayı kolaylaştıracak sözcük seçilir,
- Bir atıf cümlesinin etiketlenmesi işlemi tüm bu işlemlerin yapılmasının ardından “Yeni Kaydet” butonunun tıklanması ile tamamlanmış olur.

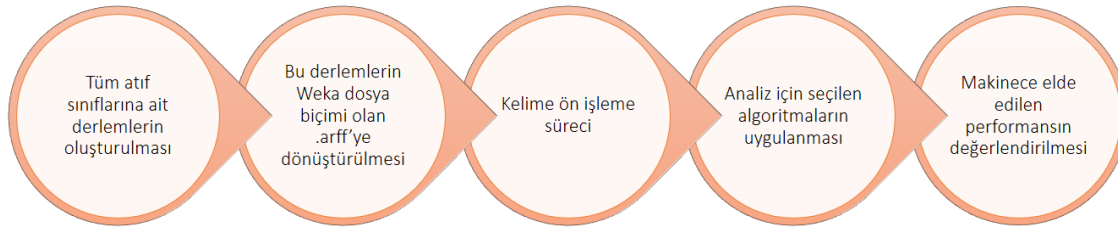
Uzlaşmalı etiketleme sürecinde bir atıf cümlesinin işaretlenmesi için harcanan süre cümlelerin uzunluğu, karmaşıklığı, çok fazla gönderme içermesi vb. değişkenler göz önüne alındığında 2 dakika ile 8 dakika arasında değişmektedir. Eylül 2016 tarihinde başlatılan etiketleme süreci Aralık 2016 tarihinde sona ermiştir. Her iki etiketçi grubunun saptadığı atıf cümlesi sayıları ve bunların sınıflara dağılımı Şekil 5'te gösterilmektedir.



Şekil 5. Uzlaşmalı etiketleme süreci sonunda elde edilen atıf cümleleri ve sınıflara dağılımı

#### 2.4.4. Veri Ön İşleme Süreci ve Algoritmaların Uygulanması

Atıf cümlelerinin tespit edilmesinin ardından metin kategorizasyonu algoritmalarının uygulanması sürecine geçilmiştir. Bu süreçte Waikato Üniversitesi tarafından geliştirilen Weka yazılımı kullanılmıştır. Algoritmaların uygulanması sürecinde geçirilen aşamalar Şekil 6'da gösterilmektedir.



Şekil 6. Algoritmaların uygulanması sürecinin aşamaları

Analiz öncesi yürütülen ilk işlem veri derlemlerinin analizde kullanılacak araç olan Weka (<http://www.cs.waikato.ac.nz/ml/weka/>)'nin temel dosya biçimi olan arff (attribute-relation file format)'ye dönüştürülmesidir. Bu dosya biçimi başlık ve veriden oluşan iki ayrı bölümden oluşur

(ARFF Book Version, 2017). Başlık kısmında her bir öznelik (attribute) tanımlanırken, veri kısmında ise veriler sınıf bilgileri ile birlikte tutulur. Örneğin, şekline göre atıf derleminde başlık bölümü @ATTRIBUTE class {yazar\_adi, toplu, tirnak} olarak belirlenirken veri bölümünde veriler @data cümle, sınıf şeklinde kodlanır. Bu sayede her bir cümlenin hangi sınıfta olduğu makinece kolayca anlaşılır hale getirilmektedir.

Dosyaların uygun biçime getirilmesinin ardından kelime ön işleme süreci başlatılmıştır. Bu araştırma kapsamında kelime ön işleme tekniği olarak n-gram sözcük parçalayıcısı seçilmiştir. Ardışık n sayıda karakter dizisi yaratmayı hedefleyen bu parçalayıcıda cümle içindeki dizilimde tekrar oranları saptanır (Damashek, 1995, s. 843). Çalışma kapsamında anlamı ve amacı açısından atıfların ön işleminde 1-2 gram (bigram) sözcük parçalayıcısından elde edilen sonuçlar raporlanmıştır. Şekline göre atıfların ön işleminde ise bu tür atıflarda kesme işareti, iyelik belirten tek karakterler ya da parantez gibi işaretler önemli olduğundan karakter 2-gram parçalayıcısı kullanılmıştır.

Doğal dil işleme çalışmalarında önemli olan bir ön işleme aşaması da dur sözcüklerinin (stop words) ayıklanmasıdır. Ancak, bu çalışma kapsamında dur sözcükleri ayıklanmadan analiz gerçekleştirilmiştir. Bunun temel sebebi atıfların ayırt edilmesini sağlayan sözcüklerin önemli bir kısmının “ancak”, “dolayısıyla” ve “göre” gibi dur sözcüklerinden oluşmasıdır. Çalışma kapsamında yürütülen analizler bu varsayımı doğrulamıştır. Dur sözcükleri çıkarılarak yapılan analizler çıkarılmadan yapılanlara göre oldukça başarısız sonuçlar vermiştir.

Kelime ön işleme uygulamalarının tamamlanmasının ardından veriler analize uygun hale gelmiş ve makine tarafından bu sınıfların otomatik olarak saptanması sürecine geçilmiştir.

Sınıflamanın başarımını ölçmek amacı ile çeşitli algoritmalar veri üzerinde test edilmiş ve bu algoritmalarından başarımı en yüksek olanlar raporlanmıştır. En yüksek başarımı veren algoritma anlamına ve amacına göre atıflar için Naive Bayes Multinomial; şekline göre atıflar için ise Random Forests olmuştur. Naive Bayes Multinomial algoritması metin içinde geçen terim sayısını dikkate alan bir Bayesyen sınıflamayı temel alırken (McCallum ve Nigam, 1998, s. 44); Random Forests birden fazla karar ağacının birleştirilmesi sonucu oluşturulmuş ormanı temsil eden algoritmadır (Breiman, 2001, s. 25).

#### **2.4.5. Performans Değerlendirme**

Sınıflama başarımının performansının değerlendirilmesi için iki temel değerlendirme yöntemi esas alınmış, analiz sonuçlarının elde edilmesi ve sunulmasında bu iki yöntem kullanılmıştır. Bunlar metodolojik ve sayısal değerlendirmedir.

- **Metodolojik değerlendirme:** Makine başarımlarının hesaplanmasında kullanılan çeşitli metodolojik değerlendirme yöntemleri vardır (eğitim ve test verisi yaratma, veriyi %66 ve %33'lük iki parçaya bölme ve k kat çapraz doğrulama gibi). Bu çalışma kapsamında analiz sonucu elde edilecek bilgilerin doğrulanması ve test edilmesi amacı ile 10 kat çapraz doğrulama seçeneği tercih edilmiştir. Bu doğrulamada veri seti rastgele 10 eşit parçaya bölünür ve ayrılan bu veri parçalarından her defasında bir tanesi test için, diğerleri ise makineyi eğitmek üzere kullanılır. Bu işlemin 10 kez tekrar edilmesi ile analiz sonuçları doğrulanmış olur (Kohavi, 1995, s. 1138).
- **Sayısal değerlendirme:** Çalışma kapsamında sayısal performans değerlendirme yapmak üzere seçilen yöntem başarımlar oranının (doğru sınıflanan atıflar/sınıfta yer alan tüm atıfların sayısı) ve f-ölçütü değerlerinin hesaplanmasıdır. f-ölçütü değeri anma ve duyarlılık sonuçlarının analiz performansını yeteri kadar ölçmediği fikri sonucu doğmuş ve bu iki analizin birleştirilmesi ile oluşturulmuştur (Coşkun ve Baykal, 2011, s. 4). Anma tüm sınıflar içinden erişilen doğru sınıf sayısını temsil ederken, duyarlılık ise doğru olarak sınıflanmış pozitif örnek sayısının tüm pozitif örnek sayısına oranıdır. Bu durumda f-ölçütünün temel misyonu anma ve duyarlılık formüllerini birleştirip her ikisini de değerlendirme kapsamına alarak tutarlılığı sağlamaktır (Hripcsak ve Rothschild, 2005, s. 297). Buna göre f-ölçütü formülü aşağıdaki şekilde gösterilebilir.

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

#### 2.4.6. Atıf Çıkarım Sistemi

Proje kapsamında içerik tabanlı atıf analizlerinde ihtiyaç duyulan insan gücünü en aza indirebilmek için tam metinlerden atıf cümlelerini otomatik saptayacak bir çıkarım sistemi geliştirilmiştir. Bu çıkarım sistemi için öncelikle kurallar listesi hazırlanmış, ardından sonlu durum algoritması ve diğer yöntemlerle metin içinden çıkarım yapabilecek bir sistem önerilmiştir. Detaylı bilgiler 3.2.'deki Atıf Çıkarım Sisteminin Oluşturulması başlığı altında sunulmaktadır.

### 3. BÖLÜM: BULGULAR VE YORUM

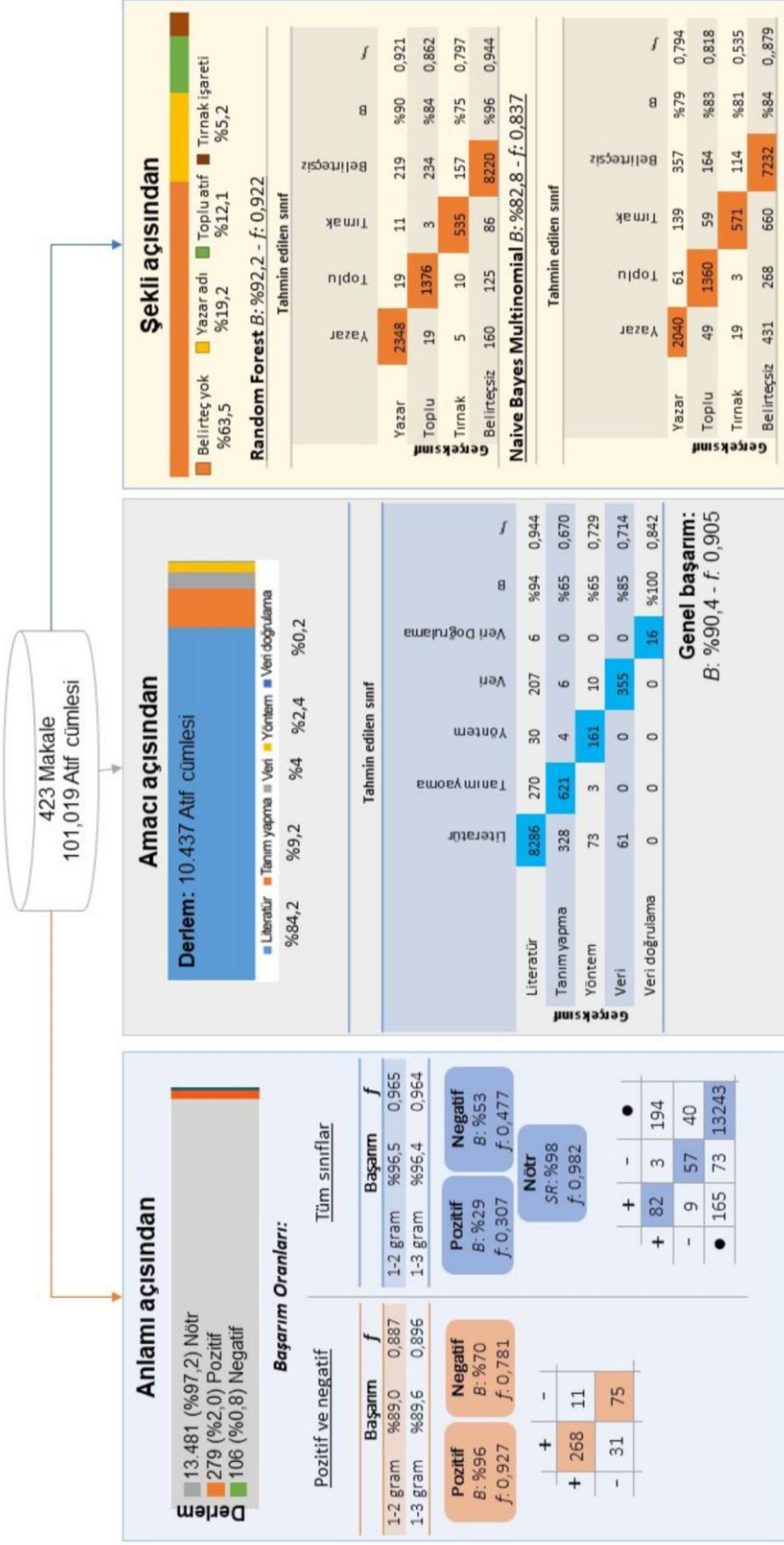
Araştırma sonucunda algoritmaların uygulandığı üç temel sınıf için elde edilen başarı oranları Şekil 7'de gösterildiği gibidir. Araştırmada kullanılan derlemlerin tamamı [http://115k440.hacettepe.edu.tr/115K440\\_data.7z](http://115k440.hacettepe.edu.tr/115K440_data.7z) adresinden açık erişimli olarak sunulmaktadır.

#### 3.1. Atıfların Otomatik Sınıflandırılması

##### 3.1.1. Anlamına Göre Atıflar

Atıflar anlamlarına göre sınıflandırıldığında ilk etiketçinin 14.259, ikinci etiketçinin ise 14.840 atıf cümlesi saptadığı anlaşılmaktadır. Etiketleme sonuçları karşılaştırıldığında derlemde yer alan çakışan atıf cümlelerinin %97,2'sini nötr,%2'sini pozitif ve %0,8'ini ise negatif atıflar oluşturmaktadır. Literatürdeki pek çok çalışmada da benzer dağılımlara rastlanmıştır. Örneğin, 2.309 atıfın incelendiği bir çalışmada atıfların %2,4'ü pozitif, %0,4'ü negatif olarak tanımlanmıştır (Spiegel-Rösing, 1977, s. 105). Bir diğer çalışmada ise atıfların %3'ü negatif olarak işaretlenirken %10'luk pozitif atıf oranına ulaşılmıştır (Athar, 2011, s. 82). Cano (1989, s. 286)'da en az görülen atıf tipinin negatif atıflar olduğunu doğrulamıştır. Literatürde yer alan bu örnekler değerlendirildiğinde, Türkçe kütüphanecilik ve bilginbilim literatüründeki anlamına göre atıfların nitelik açısından uluslararası örnekler ile benzerlik gösterdiğini söylemek yanlış olmayacaktır.

Literatürde negatif atıfların daha az yapılmasının temel nedeninin negatif atıfların daha kapalı bir dille yapılması olduğu düşünülmektedir. Athar (2011, s. 82), yazarların negatif atıf yapacakları zaman genellikle önce iyi düşüncelerini söyleyip ardından olumsuz görüşlerini yumuşatarak söylediklerini iddia etmektedir.



Şekil 7. Temel atıf sınıfları için sınıflama başarımları

Anlamına göre atıf derleminin yaratılmasının ardından algoritmalar uygulanmış ve başarımlar oranları Şekil 7’de sunulmuştur. Buna göre Anlamına göre atıfların makinece sınıflandırılması işleminde öncelikle pozitif ve negatif atıflar arasında dil açısından farklılık olup olmadığını saptamak amacı ile yalnızca bu iki sınıf üzerinden analiz gerçekleştirilmiştir. Analiz sonucunda pozitif ve negatif atıflar %89’luk başarımla ( $f=0,887$ ) ile sınıflandırılabilmiştir. Sınıflandırmada pozitif atıfların %96’sı doğru tahmin edilirken, negatif atıflarda bu oran %70’te kalmıştır. Bu durum pozitif atıfların negatif atıflara oranla daha açık ve net bir dille yapıldığını göstermektedir. Nötr atıflar da analize dâhil edildiğinde makine başarımları %96’ya ulaşmakta ancak nötr atıf sayısının fazla olması bu artışta etkili olmaktadır. Zira bu aşamada pozitif atıfları ayırma başarımları %29, negatif atıfları ayırma başarımları %53 ve nötr atıfları ayırma başarımları ise %98 olarak ölçülmüştür. Bu durumda pozitif ve negatif atıfların metin içinde ayırt edilmesini kolaylaştırmaya yönelik çeşitli farklı tekniklerin (gövdeleme, duygu sözlüğü yaratma gibi) eklenmesi konusu değerlendirilebilir. Literatürdeki çalışmalarda bu gibi düzeltme ve iyileştirmelerle başarımlar yarı yarıya artırılabilmiştir (Athar, 2014; Xu, Zhang, Wu, Wang, Dong ve Xu, 2015, s. 1339).

### 3.1.2. Amacına Göre Atıflar

Amacına göre atıflar sınıflandırılmadan önce kanıtlama ve karşılaştırma atıfları ile yöntem kullanma ve yöntem açıklama atıfları birleştirilmiştir. Çünkü bu iki türün etiketçiler tarafından genellikle birbirine karıştırıldığı saptanmış, bu sebeple de veri yapısı itibarıyla birbirine benzediği düşünülen bu iki grubun birlikte değerlendirilmesinin daha anlamlı sonuç vereceği anlaşılmıştır. Toplamda 10 atıf sınıfı üzerinde yapılan uygulamada makine sınıflamasının başarımları %78 ( $f=0,754$ ) olarak ölçülmüştür. Bu sınıflamada en yüksek başarımla %92’si doğru olarak sınıflanan literatür atıfları için hesaplanmıştır. Literatür atıflarının %79’u doğru sınıflanan veriye atıf sınıfı ile %59’u doğru sınıflanan yöntem açıklama kullanma atıfları olmuştur. Amacı açısından atıfların sınıflara dağılımı ve başarımlar oranları göz önüne alındığında bu tür atıflarda üst sınıfların belirlenmesinin anlamsal olarak atıfların temel sınıflarına ayrılması açısından önemli olduğu görülmüştür. Bu bağlamda birbiri ile benzer özellikler gösteren kanıtlama-karşılaştırma, örnek verme, eleştiri yapma, öncülerini anma ve gelecek için fikir üretme atıfları literatür temel başlığı altında toplanmıştır. Diğer sınıflar ise tanım yapma, yöntem, veriye atıf ve veri doğrulama olarak belirlenmiştir. Bu durumda toplamda beş temel amacına göre atıf sınıfı yaratılmıştır. Bu sınıflar üzerinden analiz tekrar edildiğinde makine genel başarımları %90,4 ( $f=0,905$ ) olarak ölçülmüştür. Gruplanmış yeni sınıflamada veri doğrulama atıflarının tamamı doğru olarak sınıflanabilirken, literatür atıflarının %94’ü, veriye atıfların %85’i ve tanım yapma ile yöntem atıflarının %65’i doğru olarak sınıflandırılabilmiştir (bkz. Şekil 7).



### 3.1.3. Şekline Göre Atıflar

Şekline göre atıfların sınıflandırılması için seçilen Random Forests algoritmasının sınıflama başarımı %92,2 (f=0,922) olarak hesaplanmıştır. Bu sınıflamada şekli açısından herhangi bir belirteci olmayan atıf cümleleri %96 oranında doğru sınıflanırken, onu %90 ile yazar adı anılarak yapılan atıflar, %84 ile toplu atıflar ve %75 ile tırnak içinde belirtilen atıflar takip etmektedir (bkz. Şekil 7).

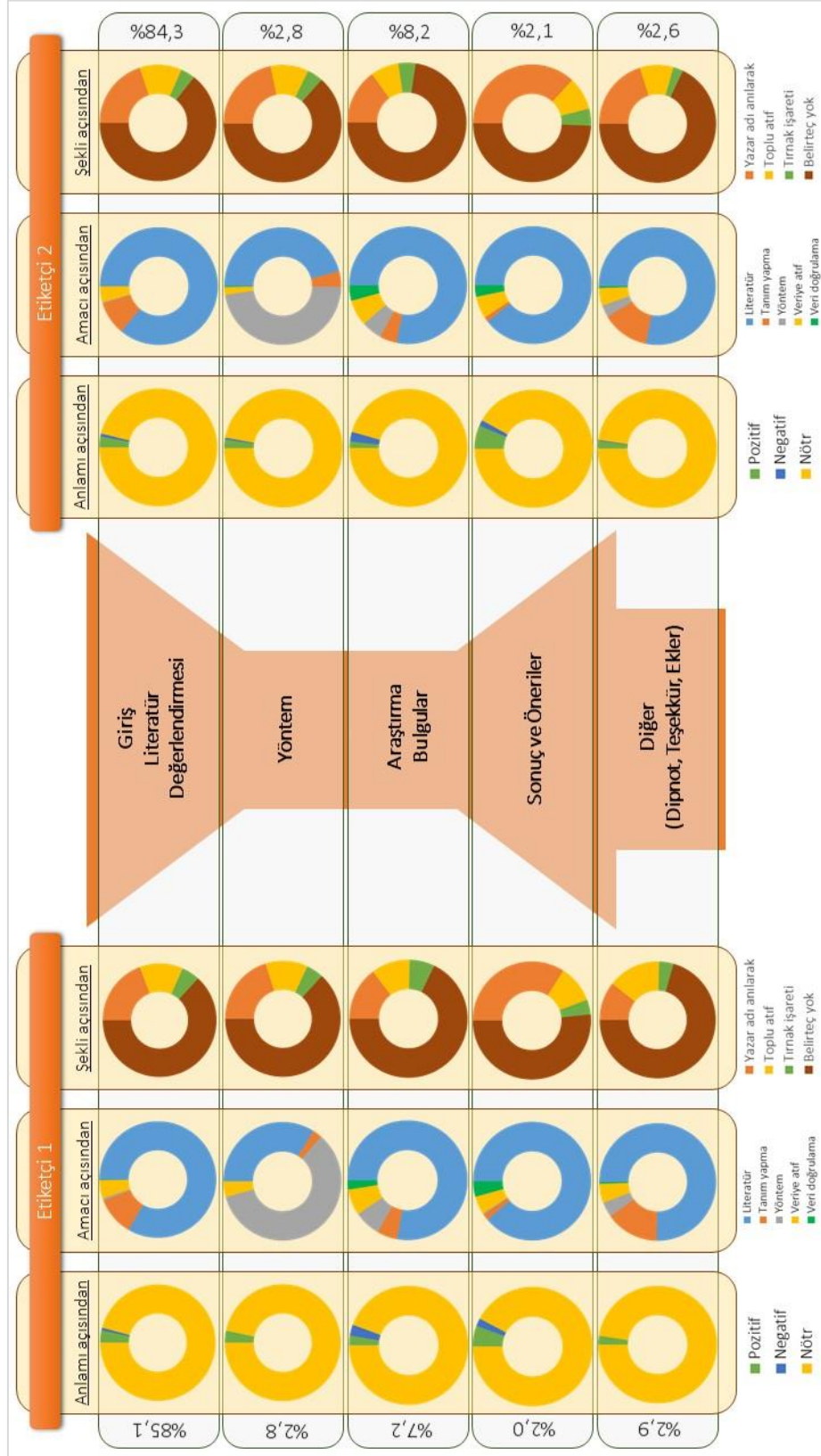
### 3.1.4. Dizilimine Göre Atıflar

Atıfların yapıldıkları bölümler incelendiğinde Şekil 8'de gösterilen sonuçlara ulaşılmıştır. Şekilden de görüldüğü üzere atıfların büyük çoğunluğu giriş ve literatür bölümlerinde yapılmaktadır. Onu araştırma ve bulgular bölümleri takip etmektedir.

Yazarların negatif atıf yapmak için tercih ettikleri bölümlerin araştırma ve bulgular ile sonuç ve öneriler bölümleri olduğu ortaya çıkmıştır. Bu bulgu analiz yapacak kişilere negatif atıfları nerelerde arayabileceğini göstermesi açısından önemlidir. Öte yandan pozitif atıflar çalışmaların tüm bölümlerinde yer alabilmektedir.

Amacı açısından atıfların bölümlere dağılımında ise en belirgin farklılık yöntem atıflarında görülmüştür. Çalışmaların yöntem kısımlarında çoğunlukla yöntem atıflarının yapıldığı sonucuna ulaşılmıştır. Tanım yapma atıfları ise en çok dipnotlarda kullanılmaktadır.

Şekline göre atıfların sınıflandırılmasında bölümler arasında önemli bir farklılık olmamakla birlikte sonuç ve öneriler bölümlerinde yazar adlarının daha sık anıldığı saptanmıştır.



Şekil 8. Atıfların makale bölümlerine dağılımı

Atıfların metin içinde yapıma sıklıkları değerlendirildiğinde ulaşılan sonuçlar şunlardır:

- Atıf yapılan yayınların %67,5'i metin içinde yalnızca bir kez kullanılmıştır.
- Kaynakçalarda yer alan yayınların %6,1'ine metin içinde gönderme yapılmadığı saptanmıştır.<sup>4</sup>
- Metin içindeki göndermelerin ise %1,1'i kaynakçada yer almamaktadır.

Bu bulgular bilimsel iletişim sürecinde yazar ve editörlerin gözünden kaçan unsurların boyutunu ispatlaması açısından önemlidir.

### 3.2. Atıf Çıkarım Sisteminin Oluşturulması

Bu çalışma kapsamında atıf cümlelerinin tam metin içinden saptanması işlemi elci sistemle etiketçiler tarafından yapılmıştır. Çünkü bir atıf cümlesi birden fazla cümleden oluşabilmektedir ve bu sebeple uzman görüşü alınarak atıf cümlelerinin saptanması önemlidir. Çalışmada elde edilen sonuçlar bu görüşü kanıtlamaktadır. Zira negatif atıf cümlelerinin %80'i, pozitif atıf cümlelerinin de %60'ına yakını en az iki cümleden oluşmaktadır. Atıf cümlesi örüntülerini bilmeden bir atıf çıkarım sistemi yaratmak başarılı bir sonuç vermeyecektir.

Proje için atıf çıkarım sistemi oluşturma sürecinde iki farklı yol önerilmiştir. Bu yollardan biri sonlu durum tekniği kullanılması, diğeri ise mevcut kural listelerinin geleneksel kodlama becerileri ile hayata geçirilmesidir. Her iki sistem de atıfları yüksek başarı ile sınıflayabilmiştir. Sistemler için kural listelerinin doğru şekilde oluşturulması, göndermenin metin içinden saptanması ve birden fazla cümle için çıkarım formüllerinin yaratılması zorunludur.

#### 3.2.1. Atıf Çıkarım Sistemi için Kural Listesi

Atıf çıkarım sistemi için atıf cümleleri yapısal olarak iki temel gruba ayrılmıştır. Her bir grup kendi içinde alt bölümlere ayrılmıştır. Bu gruplar ve grupların yapıları şu şekildedir:

- Tek cümlelik atıflar
  - Yalnızca yazar ve tarih bilgisi olan atıflar: (Taşkın, 2017), (Taşkın ve Al, 2017), (Taşkın, Al ve Sezen, 2017) gibi.
  - Sayfa numarası belirtilerek yapılan atıflar: (Taşkın, 2017, s. 4) gibi.
  - Toplu atıflar: (Taşkın, 2017; Taşkın, Al ve Sezen, 2017; Taşkın ve Al, 2017) gibi.
  - Yazar adı anılarak yapılan atıflar: Taşkın (2017, s. xx)'a göre gibi.

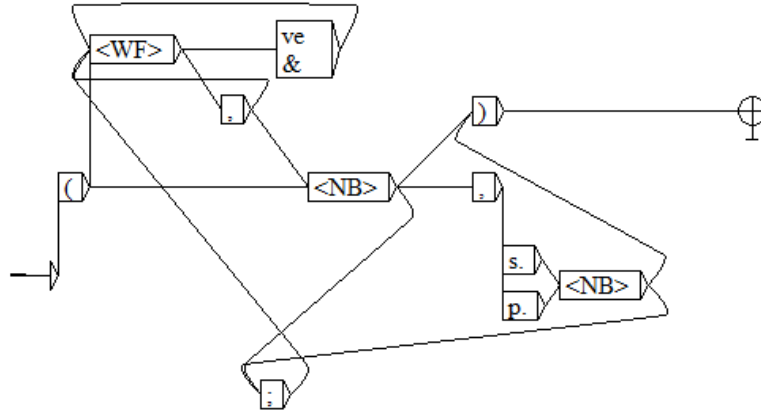
<sup>4</sup> Proje kapsamındaki iki derginin kaynakçalarındaki yayınların %6,1'i 786 çalışmaya karşılık gelmektedir. Bunların 331'i *Bilgi Dünyası*'nda, 455'i ise *Türk Kütüphaneciliği*'ndedir. *Türk Kütüphaneciliği*'ndeki yayınların kaynakçalarında yer alan toplam 7594 çalışmadan 455'ine (%6); *Bilgi Dünyası*'ndeki yayınların kaynakçalarında yer alan toplam 5287 çalışmadan 331'ine (%6,3) metin içinde gönderme yapılmadığı saptanmıştır.

- Birden fazla cümleden oluşan atıflar
  - Atıf cümlesi şu sözcüklerle başlıyorsa atıf cümlesinden bir önceki cümle de atıf cümlesinin parçasıdır: Bu nedenle, Bu bağlamda, Bunun yanı sıra, Buna göre, Bu konuda, Burada, Raporda, Zira, Zaten, Yine, Yine de, Yazıda, Yazara göre, Yazar, Yazarlar, Yapılan bu, Yani, Üstelik, Tüm bunlarla, Tüm bu, Tabi ki, Tabii ki, Sözü edilen, Söz konusu, Sonuçta, Sonuç olarak, Planda, Özellikle, Oysa, Oysa ki, Oysaki, Öyle ki, Öte yandan, Örneğin, Örnek olarak, Örnek vermek gerekirse, Ona göre, O halde, Nitekim, Ne var ki, Ne yazık ki, Makalede, Kuşkusuz, Konuyla ilgili, Kısaca, Keza, Kaldı ki, İşte, İncelenen, İnceleme, İlk, İlkinde, İlk olarak, İlgili, İkinci, Her iki, Her üç, Her bir, Hepsi de, Hatta, Görüldüğü üzere, Gerçi, Fakat, Eğer, Durum böyle iken, Dolayısıyla, Diğer taraftan, Diğer bir deyişle, Dahası, Daha sonra, Daha, Çünkü, Çalışmada, Çalışma sonucunda, Çalışma, Buradan, Burada, Bununla birlikte, Bunun, Bununla, Bunu, Bunların, Bunlardan, Bunlara, Bunlar, Bundan dolayı, Buna örnek olarak, Buna rağmen, Buna karşın, Buna (göre, örnek olarak vs.), Bulgulara göre, Bu, Böyle, Böylelikle, Böylece, Bir başka deyişle, Benzer, Belirtilen, Ayrıca, Aynı, Araştırmaya, Araştırmanın, Anılan, Ama, Aksi, Adı geçen.
  - Atıf cümlesinden sonraki cümle şu sözcüklerle başlıyorsa atıf cümlesinden bir sonraki cümle de atıf cümlesinin parçasıdır: Ancak, Oysa, Oysa ki, Oysaki, Öte yandan, Nitekim, Ne var ki, Ne yazık ki, Makalede, Gerçi, Fakat, Eğer, Durum böyle iken, Dolayısıyla, Diğer taraftan, Diğer bir deyişle, Dahası, Daha sonra, Daha, Çünkü, Çalışmada, Çalışma sonucunda, Çalışma, Burada, Bunu, Bununla, Bunların, Bunlardan, Bunlara, Bunlar, Bundan dolayı, Buna örnek olarak, Buna rağmen, Buna karşın, Buna, Bulgulara göre, Bu, Böyle, Böylelikle, Böylece, Bir başka deyişle, Benzer, Belirtilen, Ayrıca, Aynı, Araştırmaya, Araştırmanın, Anılan, Ama, Aksi, Adı geçen.

Yukarıda sözü edilen kurallar herhangi bir bilgi çıkarım sistemi ile veriye uygulandığında atıf cümlelerinin otomatik olarak çıkarımı sağlanabilecektir.

### 3.2.2. Atıfların Çıkarımı

Sonlu durum tekniği bir dize ya da cümlelerin düzenli bir dile sahip olup olmadığını belirlemek için kullanılmaktadır (Galvez ve Moya-Anegón, 2007). Çalışmamız için sonlu durum algoritmalarının uygulanması için Nooj yazılımı tercih edilmiştir. İlk aşamada atıf cümlelerinin belirlenmesi için Şekil 9'da gösterilen gramer grafiği kullanılmıştır.



Şekil 9. Göndermelerin saptanması için hazırlanan gramer grafiği

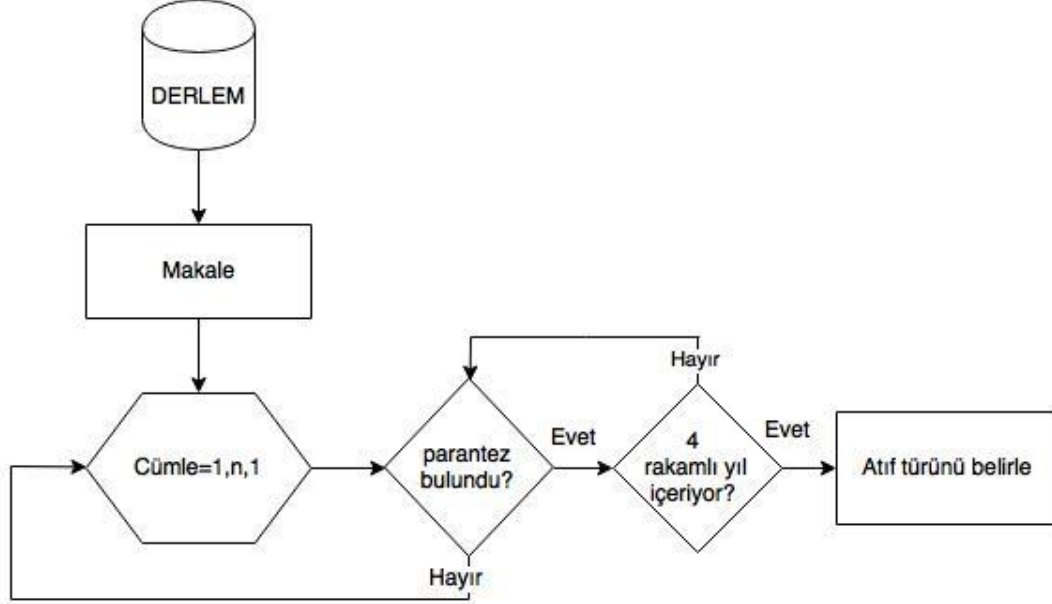
Şekilde gösterilen gramer grafiği ile atıf cümlelerinde geçen göndermeler saptanabilmektedir. Örnek çıkarım Şekil 10'da gösterilmektedir.

| Before   | Seq.   | After  |
|--|--|--|
| ve "açıklık" anlayışına bırakmasını sağlamıştır yönetiminin etkinliğinin artmasını da sağlayabilmektedir biri bilgiye özgürce erişebilmenin sağlanmasıdır ülkeler için temel haldardan biridir bir unsur olarak ortaya çıkmıştır edinme özgürlüğünün de temelini oluşturmaktadır anlayışın özümsemesine de katkı sağlamaktadır ve belgelere özgürce erişilmeye bağlıdır gereğince uygulanmasına da katkı sağlamaktadır aykırı davranışların da önüne geçmektedir (Özdemirci, 2004, s. 66). Gülle'ye göre dışında devleti şeffaflaştırdığını söylemek mümkündür giren bir başka kanun izlemiştir dünya genelinde yasalaşmasına zemin hazırlamıştır bir bölümü olarak yer almıştır ve özgürlükler arasında yer verilmiştir hakkına sahip olduğuna dikkat çeker yılında İngiltere'de uygulamaya koymuştur gibi çeşitli etkinlikler de düzenlemektedir düşünceye/ bilgiye özgür erişimi savunmaktadır -i Vakayı"nın çıkarılmasıyla tamamlanmıştır yine bu hak çerçevesinde gözetilmektedir bir bölümü olarak yer almıştır kalan düzenlemeleri içerdiğine dikkat çekilmektedir doğru bir şekilde gerçekleştirmekle sorumludurlar :ğerlendirme Kurulu tarafından değerlendirilmesini içermektedir yer aldığı 29 ile 33. maddeleri kapsamaktadır edinme hakkı kapsamındaki işlemlerin gerçekleştirilmeleri hatalı davranışları bu süreyi değiştirmez Edinme Değerlendirme Kurulu'na gönderirler incelenmesini kapsayan bir araştırma tekniğidir | (Köseoğlu, Selçuk ve Yurttaş, 2005, s.149; Çukurçayır, Özer ve Turgut, 2012, s. 9) (Uzun, 2005, s. 232) (Evren, 2010, s. 114) (İyimaya, 2003, s. 41; Canöz, 2008, s.144) (Uzun, 2005, s.231; Şavran, 2006; Yılmazoğlu, 2009, s. 1) (Tonta ve Çelik, 1996, s. 4) (Koçak, 2010, s.115) (İnaç ve Ünal, 2007, s. 16; Soykan, 2007, s. 65) (Uzun, 2005, s. 234) (Özdemirci, 2004, s. 66) (2006, s. 51) (Gülle, 2002, s. 89) (İyimaya, 2003, s. 42; Uzun, 2005, s. 237; Soykan, 2007, s.64; Canöz, 2008, s. 14... (Şavran, 2006, s. 47) (İyimaya, 2003, s. 42) (Cin, 2005, s.60) (İyimaya, 2003, s. 42) (Şavran, 2006, s. 48) (FAIFE, 2015) (Yılmaz, 2015, s. 139) (Demir, 2014, s. 59) (Cin, 2005, s. 42; Soykan, 2006, s. 27) (İyimaya, 2003, s. 42) (BEHK, 2003) (BEHK, 2003) (BEHK, 2003) (BEHK, 2003) (BEHK, 2003) (BEHK, 2003) (BEHK, 2003) (Yıldırım ve Şimşek, 1999) | . Demokratik olabilmeyen gereklerinden "açıklık" diğer İdari işleyişin halkın karar ve Bilgi edinme özgürlüğü, bireylerin idarenin Yürüncü yüzyılın ikinci yarısından itibaren Bilgi Edinme Hakkı Kanunu, katılımcı Bilgi Edinme Hakkı Kanunu ile Yönetimin denetlenmesi yürütülen işler hakkında Bu açıdan Bilgi Edinme Hakkı Bilgi edinme hakkını gözetilen yasalar Gülle'ye göre (2006, s. 51) kamu kamu kurumlarının ürettikleri resmi bilgilere Bilgi edinme ya da bilgi 1786 Virginia Bildirgesi, 1789 Fransa İnsan ve Bilgi edinme hakkı, özerk bir 1996 Güney Afrika Anayasası'nda bilgi Bu düzenlemelerin yanı sıra, bilgi Bilgi edinme hakkını başlı başına Bilgiye erişimi kısıtlayan etkenlere tepki IFLA, Amerikan Kütüphane Derneği (American Bilgi edinme özgürlüğüyle ilgili bir Takvim-i Vakayı halkın resmi e-Devlet çalışmaları ve Avrupa Türkiye'de 09.10.2003'te kabul edilen 4982 kamu İkinci bölüm, "Bilgi Edinme Hakkı Üçüncü bölüm ise "Bilgi Edinme "Bilgi Edinme Hakkının Sınırları"na Buna göre, kamunun uygulanmasından sorumlu Başvuru sahipleri gerçek ve tüzel Başvurunun reddedilmesi halinde ret kararının Türkiye'de Bilgi Edinme Hakkı Bu kapsamda 2004-2014 yılları arasında yayımlanan |

Şekil 10. Örnek gönderme çıkarımı

Çalışmaların içinde yer alan göndermelerin çıkarılmasının ardından iki veya daha fazla cümleden oluşan atıf cümlelerinin çıkarımı için de algoritma geliştirilerek tam atıf cümlelerine ulaşmak mümkündür. Ancak gramer grafiklerinde bu kuralları tanımlamak oldukça meşakkatli olduğundan alternatif yöntem aranmış ve geleneksel kodlama becerileri ile daha hızlı ve kolay bir şekilde atıf çıkarımı yapılabileceği sonucuna ulaşılmıştır.

Atıf cümlelerinin çıkarımı konusunda kodlama yeteneklerini kullanarak çıkarım yapmak hem daha pratik hem de isabetli sonuçlar vermektedir. Bu sebeple Şekil 11’de akış şeması verilen şekilde atıf çıkarımı yapıldığında da doğru sonuçlara ulaşılabilmektedir.



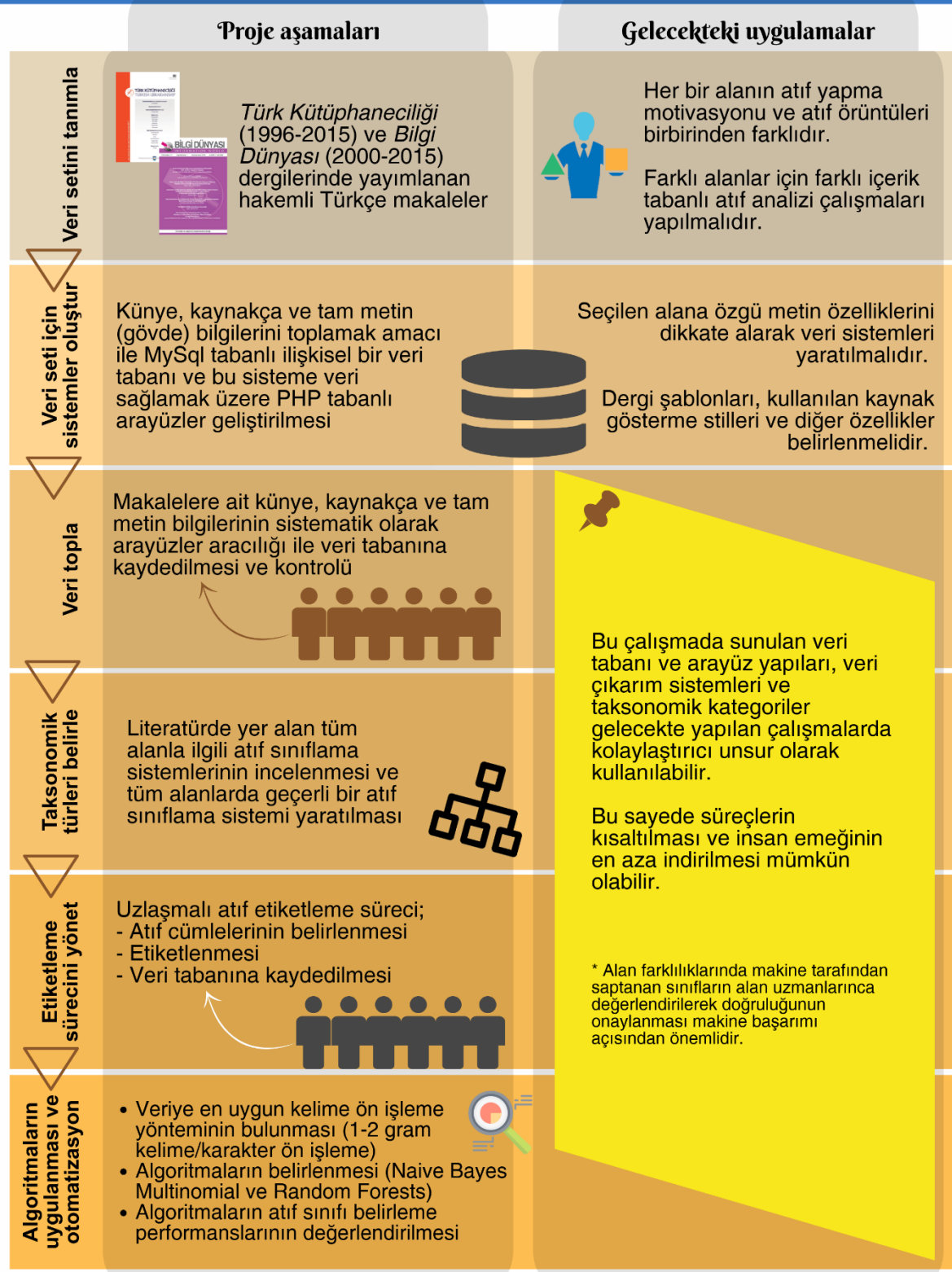
Şekil 11. Atıf çıkarım sistemi akış şeması

Atıf çıkarım sisteminin yaratılmasında önemli olan hangi algoritma ya da tekniğin kullanıldığından çok kural listelerinin doğru bir şekilde oluşturulmasıdır. Bu sebeple çalışmamızda sunulan kural listeleri kullanılarak pek çok farklı algorithmadan başarılı sonuçlar alınabilecektir.

### 3.3. İçerik Tabanlı Atıf Analizi Süreçleri için Uygulama Önerileri

İçerik tabanlı atıf analizinin otomatikleştirilmesi ile ilgili olarak Proje kapsamında gerçekleştirilen her aşama ve gelecekte neler yapılabileceği Şekil 12’de özetlenmektedir.

## İçerik Tabanlı Atıf Analizi Süreci



Şekil 12. İçerik tabanlı atıf analizi süreci

Proje çalışması kapsamında yürütülen içerik tabanlı atıf analizi süreci, Türk kütüphanecilik ve bilgilim literatürü temelli olmasına rağmen, pek çok alanda uygulanabilir niteliğe sahiptir. İçerik tabanlı atıf analizi sürecinde verilerin tanımlanmasından toplanmasına ve sınıflanmasına kadar her aşama dikkatle planlanmalı ve uygulanmalıdır. Herhangi bir alanda yürütülecek içerik tabanlı atıf analizi çalışmasında alanın belirlenmesi ve yayın özelliklerinin ortaya çıkarılmasının ardından Proje kapsamında sunulan veri tabanı ve arayüz yapılarının, veri çıkarım sistemlerinin ve taksonomik kategorilerin kullanımı ile yoğun uzman emeği gerektiren pek çok sürecin kolaylaştırılması ve hızlandırılması mümkündür. Bunun yanında otomatize sistemler tarafından sunulan sonuçların değerlendirilmesinde alan uzmanları tarafından yapılacak kontroller alan farklılıkları kaynaklı sorunları ortadan kaldıracaktır.



## 4. BÖLÜM: SONUÇ VE GELECEK ÇALIŞMALAR

### 4.1. Sonuç

Başlangıçta amaçları yalnızca araştırmalar arasında ilişki kurmak olan atıfların kullanım amaçlarının değişiklik göstermesi ve araştırmacı değerlendirmelerinde kullanılmaya başlanması ile birlikte atıfların kalitesi konusunda soru işaretleri doğmaya başlamıştır. Atıf takası, atıf manipülasyonları, zoraki atıflar gibi uygulamalar sebebi ile atıflar anlamlarını yitirmeye başlamışlardır. Bu bağlamda bu projenin temel amacı tüm atıfların eşit değerde olmadığını ispatlamak ve atıfların sınıflandırılması için bir model sunmaktır.

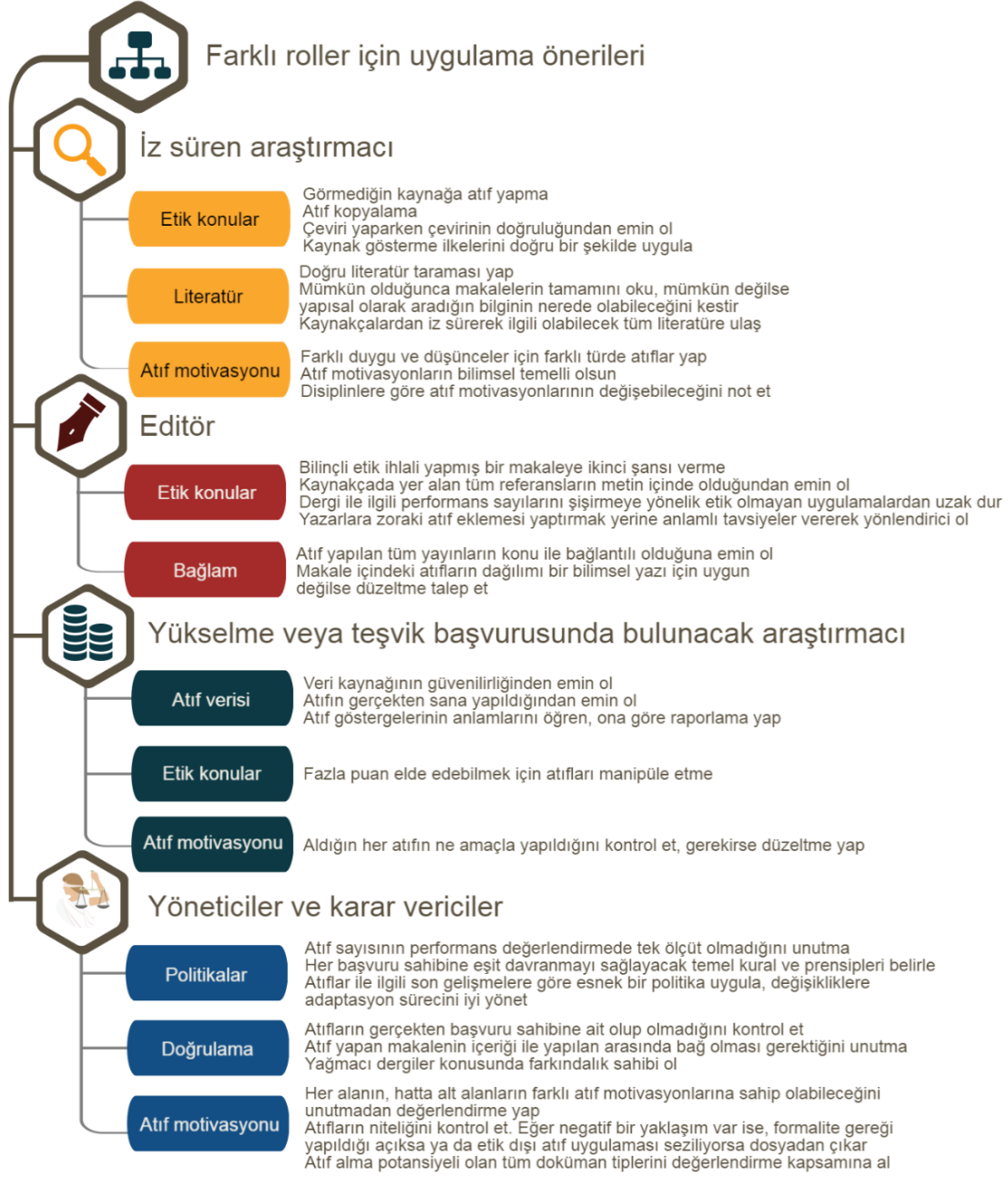
Çalışma kapsamında temel amaç olarak belirlenen semantik ve sentaktik yapıları çözümleyici atıf analizi modeli oluşturulmuş ve politika yapıcılara çeşitli öneriler sunulmuştur. Öte yandan Türkçe literatür için daha önce yapılmamış atıf sınıflaması da yine bu çalışma aracılığı ile gerçekleştirilmiştir.

Çalışmada ayrıca atıf çıkarım sistemleri için kural listeleri oluşturularak disiplin veya dilden bağımsız büyük yoğunluklu bilimsel veriden atıf cümlelerinin tespitini yapmak olanaklı hale gelmiştir. Bu sayede gelecekte yapılacak içerik tabanlı atıf analizi çalışmaları daha az insan gücü ile daha yüksek isabet oranları ile gerçekleştirilebilecektir.

Çalışmada kullanılan sınıflama verileri açık erişimli olarak Proje web sayfasından yayımlanmıştır. Bu veriler sayesinde analizlerin tekrar edilmesi, Türkçe için yapılacak duygu analizlerine veri sağlanması ve bu sayede de tüm doğal dil işleme çalışmalarına katkı sağlanması hedeflenmektedir.

Bu çalışma ile Türkçe literatürde daha önce detaylı olarak çalışılmamış içerik tabanlı atıf analizleri üzerinde durulmuş ve bu analizlerin gerekliliği ortaya koyulmuştur. Bu model sayesinde hem atıfların gerçek anlamı ile varlığını sürdürebilmesi hem de değerlendirmelerin daha kolay ve isabetli yapılmasının sağlanması hedeflenmektedir.

İçerik tabanlı atıf analizi dört temel araştırmacı/değerlendirici grubuna yardımcı olabilecek çıktılar sunabilecektir. Atıflar ile ilgili mevcut algı ve güncel uygulamalar dikkate alınarak bu dört temel gruba atıflar ile ilgili çeşitli tavsiyelerde bulunmaktadır. Bunlar Şekil 13'te sunulmaktadır.



Şekil 13. Bilimsel iletişim sürecindeki farklı roller için uygulama önerileri

Bilimsel iletişim sürecinde çok farklı rollere sahip kişiler yer almaktadır. Her bir rol için çalışmamızın kapsamına giren öneriler Şekil 13'te ayrıntıları ile verilmektedir. Yazarlar, yöneticiler, editörler ve araştırmacılar buradaki önerileri dikkate alıp uyguladıklarında atıf yapma ve atıf değerlendirme motivasyonlarının tutarlı ve etik sınırlar çerçevesinde kalması sağlanabilecektir.

En nihayetinde sıralanan tüm roller kendilerine önerilen şekilde atıf yaptıkları veya değerlendirdiklerinde atıflar kaybettikleri anlamlarını tekrar kazanabilecek ve bilimsel iletişimde önemli öğelerden biri haline gelebilecektir.

#### **4.2. Gelecek Çalışmalar**

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

- Anlamı açısından atıfların makinece saptanması başarımının artırılması için gövdeleme, duygu sözlüğü ve diğer teknikler kullanılarak analizlerin tekrarlanması ve en doğru tekniğin ortaya çıkarılması,
- İçerik tabanlı atıf analizi modelinin diğer alanlara uygulanması ve farklılıkların izlenmesi,
- İçerik tabanlı atıf analizlerinin atıf dizinlerine uygulanabilmesi için yöntemin belirlenmesi ve öneri sunulması.

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

## KAYNAKÇA

- Akademik Teşvik Ödeneği Yönetmeliği, T.C. Resmi Gazete. (13271644, 27.12.2016).  
<https://goo.gl/GTtGkZ>, Son erişim tarihi: 13 Eylül 2017.
- Akpınar, H. 2014. "Data: Veri madenciliği, veri analizi". İstanbul: Papatya Yayıncılık Eğitim A.Ş.
- Al, U. ve Soydal, İ. 2012. "Dergi kendine atfının etkisi: Energy Education Science and Technology örneği", Türk Kütüphaneciliği, 26, 699-714.
- Al, U. ve Tonta, Y. 2004. "Atıf analizi: Hacettepe Üniversitesi Kütüphanecilik Bölümü tezlerinde atıf yapılan kaynaklar", Bilgi Dünyası, 5, 19-47.
- Allison, P.D., Long, S.L. ve Krauze, T.K. 1982. "Cumulative advantage and inequality in science", American Sociological Review, 47, 615-625.
- ARFF Book Version. 2017.  
<https://weka.wikispaces.com/ARFF+%28book+version%29#Overview>, Son erişim tarihi: 14 Eylül 2017.
- 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, 392-413.
- Bertin, M. 2008. Categorizations and annotations of citation in research evaluation. 13. Natural Language Processing; 13.1 Discourse.  
<http://aaai.org/Papers/FLAIRS/2008/FLAIRS08-108.pdf>, Son erişim tarihi: 25 Mart 2016.
- Bhowmick, P.K., Mitra, P. ve Basu, A. 2008. "An agreement measure for determining inter-annotator reliability of human judgements on affective texts." Coling 2008: Proceedings of the workshop on Human Judgements in Computational Linguistics içinde (s. 58-65).  
Manchester: ACM.

- 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, 208-216.
- Bonzi, S. ve Snyder, H.W. 1991. "Motivations for citation: A comparison of self citation and citation to others". *Scientometrics*, 21, 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, 45-80.
- Breiman, L. 2001. "Random forests". *Machine Learning*, 45, 5-32.
- Cano, V. 1989. "Citation behavior: Classification, utility, and location". *Journal of the American Society for Information Science*, 40, 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.
- Chowdhury, G.G. 2003. "Natural language processing", *Annual Review of Information Science and Technology*, 37, 51-89.
- Chubin, D.E. 1980. "Letter to editor: Is citation analysis a legitimate evaluation tool?". *Scientometrics*, 2, 91-94.
- 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.
- COPE. 2012. "Citation manipulation", [http://publicationethics.org/files/u7141/Forum%20discussion%20topic\\_final.pdf](http://publicationethics.org/files/u7141/Forum%20discussion%20topic_final.pdf), Son erişim tarihi: 14 Eylül 2017.
- Coşkun, C. ve Baykal, A. 2011. "Veri madenciliğinde sınıflandırma algoritmalarının bir örnek üzerinde karşılaştırılması". AB'11'de sunulan bildiri. <http://ab.org.tr/ab11/bildiri/67.pdf>, Son erişim tarihi: 12 Eylül 2017.
- Cronin, B. 1981. "The need for a theory of citing", *Journal of Documentation*, 37, 16-24.
- Damashek, M. 1995. "Gauging similarity with n-grams language independent categorization of text". *Science*, 267, 843-848.
- Davis, P. 2017. "Citation cartel or editor gone rogue?" [Blog yazısı], *Scholarly Kitchen*, <https://scholarlykitchen.sspnet.org/2017/03/09/citation-cartel-or-editor-gone-rogue/>, Son erişim tarihi: 12 Eylül 2017.

- 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, 1820-1833.
- Dođan, G. 2014. "Türk Kütüphaneciliđi dergisinden dergilere yapılan atıflarda kaynak gösterme hataları: 2010-2012". Uluslararası Kütüphane ve Bilgibilim Felsefesi Sempozyumu, 3-5 Eylül 2014, Kastamonu, <http://www.bby.hacettepe.edu.tr/akademik/guledaduzyoldogan/file/ukbbf-gd-v3.pptx>, Son erişim tarihi: 13 Eylül 2017.
- Dođan, G. 2017. "Akademik performans odaklı uluslararası üniversite sıralama sistemlerinin genel sıralamalarına ve ölçütlerine göre değerlendirilmesi" (Doktora tezi), Hacettepe Üniversitesi, Bilgi ve Belge Yönetimi Anabilim Dalı.
- Erikson, M.G. ve Erlandson, P. 2014. "A taxonomy of motives to cite". Social Studies of Science, 44, 625-637.
- Galvez, C. ve Moya-Anegón, F. 2007. "Standardizing formats of corporate source data". Scientometrics, 70, 3-26.
- Garfield, E. 1970. "Can citation indexing be automated?", Essays of an Information Scientists, 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, 359-375.
- Garfield, E. 1997. "Concept of citation indexing: A unique and innovative tool for navigating the research literature". Far Eastern State University. Philadelphia: The Scientist. <http://www.garfield.library.upenn.edu/papers/vladivostok.html>, Son erişim tarihi: 12 Eylül 2017.
- Goudsmith, S.A. 1974. "Citation analysis". Science, 183, 28.
- Hacettepe Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi. "BAP uygulama esasları". <http://research.hacettepe.edu.tr/dosyalar/dokumanlar/BAP%20Uygulama%20Esaslar%C4%B1%202017..pdf>, Son erişim tarihi: 12 Eylül 2017.

- 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, 308-310.
- Hripcsak, G. ve Rothschild, A.S. 2005. "Agreement, the F-Measure, and Reliability in Information Retrieval". *JAMIA*, 12, 296-298.
- Jha, R., Jbara, A.-A., Qazvinian, V. ve Radev, D.R. 2017. "NLP-driven citation analysis for scientometrics". *Natural Language Engineering*, 23, 93-130.
- Kaplan, P. 2014. "Akademisyenlerin atıf çetesi", *HaberTürk*, <http://www.haberturk.com/yazarlar/pervin-kaplan/974474-akademisyenlerin-atif-cetesi>, Son erişim tarihi: 12 Eylül 2017.
- 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.
- Kurbanoglu, S.S. 1996. "Kütüphanecilik literatüründe atıf hataları". *Türk Kütüphaneciliği*, 10, 137-148.
- Lewis, D.D. ve Jones, K.S. 1996. "Natural language processing for information retrieval". *Communications of the ACM*, 39, 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.
- 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, 530-540.
- McCain, K.W. ve Turner, K. 1989. "Citation context analysis and aging patterns of journal articles in molecular genetics". *Scientometrics*, 17, 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.
- Merton, R.K. 1968. "The Matthew Effect in science", *Science*, 159, 56-63.

- Moravcsik, M.J. ve Murugesan, P. 1975. "Some results on the function and quality of citations". *Social Studies of Science*, 5, 86-92.
- Oransky, I. ve Marcus, A. 2017. "Gaming the system, scientific 'cartels' band together to cite each others' work", *StatNews*, <https://www.statnews.com/2017/01/13/citation-cartels-science/>, Son erişim tarihi: 14 Eylül 2017.
- Oxford Living Dictionaries. 2017. "Citation", <https://en.oxforddictionaries.com/definition/citation>, Son erişim tarihi: 12 Eylül 2017.
- Öğretim Üyeliğine Yükseltme ve Atanma Yönetmeliği. T.C. Resmi Gazete (Sayı: 17588, 28 Ocak 1982). [http://www.yok.gov.tr/web/guest/icerik/-/journal\\_content/56\\_INSTANCE\\_rEHF8BIsFYRx/10279/17641](http://www.yok.gov.tr/web/guest/icerik/-/journal_content/56_INSTANCE_rEHF8BIsFYRx/10279/17641), Son erişim tarihi: 13 Eylül 2017.
- Öztürk, K. 2013. "Şişme dergiler, yeniden" [Blog yazısı]. <https://mkoz.wordpress.com/2013/01/25/sisme-dergiler-yeniden/>, Son erişim tarihi: 14 Eylül 2017.
- Retraction Watch. 2017. "One way to boost your uni's ranking: Ask faculty to cite each other", <http://retractionwatch.com/2017/08/22/one-way-boost-unis-ranking-ask-faculty-cite/>, Son erişim tarihi: 14 Eylül 2017.
- 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.
- Silahtaroglu, G. 2013. "Veri madenciliği: kavram ve algoritmaları". İstanbul: Papatya Yayıncılık Eğitim A.Ş.
- Small, H.G. 1978. "Cited documents as concept symbols", *Social Studies of Science*, 8, 327-240.
- Small, H.G. 2011. "Interpreting maps of science using citation context sentiments: A preliminary investigation". *Scientometrics*, 87, 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, 97-113.



- Taşkın, Z. 2017. "İçerik tabanlı atıf analizi modeli tasarımı: Türkçe atıflar için metin kategorizasyonuna dayalı bir uygulama" (Doktora tezi), Hacettepe Üniversitesi, Bilgi ve Belge Yönetimi Anabilim Dalı.
- 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". <http://www.cl.cam.ac.uk/~sht25/papers/emnlp06.pdf>, Son erişim tarihi 13 Eylül 2017.
- Tunç, M. 2012. "Bilimsel ve teknolojik bilginin niteliksel performansının ölçümü" (Doktora Tezi), Marmara Üniversitesi, Türkiyat Araştırmaları Enstitüsü.
- Türk Dil Kurumu (TDK). 2017a. "Atıf". [http://www.tdk.gov.tr/index.php?option=com\\_gts&arama=gts&guid=TDK.GTS.59b982e578b023.81613580](http://www.tdk.gov.tr/index.php?option=com_gts&arama=gts&guid=TDK.GTS.59b982e578b023.81613580), Son erişim tarihi: 12 Eylül 2017.
- Türk Dil Kurumu (TDK). 2017b. "Taksonomi". [http://www.tdk.gov.tr/index.php?option=com\\_gts&arama=gts&guid=TDK.GTS.59bc1ef3b57b67.48962180](http://www.tdk.gov.tr/index.php?option=com_gts&arama=gts&guid=TDK.GTS.59bc1ef3b57b67.48962180), Son erişim tarihi: 13 Eylül 2017.
- Voos, H. ve Dagaev, K.S. 1976. "Are all citations equal? Or did we Op. Cit. your Idem?" The Journal of Academic Librarianship, 1, 19-21.
- White, E.C. 1985. "Bibliometrics: From curiosity to convention", Special Libraries, 76, 35-42.
- Wilhite, A.W. ve Fong, E.A. 2012. "Coercive citation in academic publishing", Science, 335, 542-543.
- Wouters, P. (2013). "May university rankings help uncover problematic or fraudulent research?" [Blog yazısı] <https://citationculture.wordpress.com/2013/07/04/may-university-rankings-helpuncover-problematic-or-fraudulent-research/>, Son erişim tarihi: 14 Eylül 2017.
- 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.

## EK 1: Proje Kapsamında Yapılan Yayınlar

### Tamamlanmış çalışmalar

1. Taşkın, Z. ve Al, U. 2018. "A Content-Based Citation Analysis Study based on Text Categorization". *Scientometrics*, 114(1): 335-357. <https://doi.org/10.1007/s11192-017-2560-2>
2. Taşkın, Z., Al, U. ve Sezen, U. 2017. "First Stage of an Automated Content-Based Citation Analysis Study: Detection of Citation Sentences". STI2017, 6-8 Eylül 2017, ESIEE, Paris, Fransa.
3. Taşkın, Z. 2017. "İçerik Tabanlı Atıf Analizi Modeli Tasarımı: Türkçe Atıflar için Metin Kategorizasyonuna Dayalı Bir Uygulama" (Doktora Tezi). Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü. Tam metin: [http://www.bby.hacettepe.edu.tr/akademik/zehrataskin/file/ZT\\_PhD.pdf](http://www.bby.hacettepe.edu.tr/akademik/zehrataskin/file/ZT_PhD.pdf)
4. Taşkın, Z., Doğan, G. ve Al, U. 2016. Sayfa 1494. "Sub-fields of Library and Information Science in Turkey: A Visualization Study". Book of Proceedings, STI2016 (21<sup>st</sup> International Conference on Science and Technology Indicators), Valencia. 14-16 Eylül 2016, Universitat Politècnica de València. Editörler: Ismael Ràfols, Jordi Molas-Gallart, Elena Castro-Martínez, Richard Woolley. Valencia: Universitat Politècnica de València.

### Devam eden çalışma

1. Çelik, A.E.Ö. (Tez Aşamasında). "Türkçe Akademik Yayınlar için Otomatik Metin Özetleme".

## EK 2: Atıfların Sınıflanması ve Değerlendirilmesi Üzerine Yapılmış ve Proje Kapsamında İncelenen Çalışmalar

- Adedayo, A.V. (2015). Citations in introduction and literature review sections should not count for quality. *Performance Measurement and Metrics*, 16(3), 303-306.
- 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.
- Aljaber, B., Stokes, N., Bailey, J. ve Pei, J. (2010). Document clustering of scientific texts using citation contexts. *Information Retrieval*, 13(2), 101-131.
- 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.
- 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>
- 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.
- 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.

- 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.
- 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.
- 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.
- 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.
- 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.
- Dong, C., ve Schäfer, U. (2011). Ensemble-style self-training on citation classification. *5<sup>th</sup> International Joint Conference on Natural Language Processing, IJCNLP 2011* içinde (s. 623-631). Chiang Mai: AFNLP.
- Erikson, M.G. ve Erlandson, P. (2014). A taxonomy of motives to cite. *Social Studies of Science*, 44(4), 625-637.
- 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.
- Garfield, E. (1970). Can citation indexing be automated? *Essays of an Information Scientists*, 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.
- Goudsmith, S.A. (1974). Citation analysis. *Science*, 183(4120), 28.
- 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.
- Johnson, C.A. (1985). Citations to authority in supreme court opinions. *Law & Policy*, 7(4), 509-523.
- 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>
- 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.
- 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.
- MacRoberts, M.H. ve MacRoberts, B.R. (1996). Problems of citation analysis. *Scientometrics*, 36(3), 435-444.

- 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.
- 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.
- Meadows, A.J. (1967). The citation characteristics of astronomical research literature. *Journal of Documentation*, 23(1), 28-33
- 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.
- Oppenheim, C. (1996). Do citations count? Citation indexing and the research assessment 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.
- Peritz, B.C. (1983). A classification of citation roles for the social sciences and related fields. *Scientometrics*, 5(5), 303-312.
- 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.
- Sendhilkumar, 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.

- 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/onrpaper.html>
- Siddhartan, 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>
- 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.
- 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.
- 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>
- 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>
- 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.
- 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.
- Wetterer, J.K. (2006). Quotation error, citation copying, and ant extinctions in Madeira. *Scientometrics*, 67(3), 351-372.
- 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.
- 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.
- 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.