



**METİN ANALİTİĞİ VE AĞ ANALİZİ İLE İŞ YAPMA YÖNTEMİ PATENTLERİ
VERİLERİNİN ANALİZİ**

Furkan İNCEDAYI

**YÜKSEK LİSANS TEZİ
YÖNETİM BİLİŞİM SİSTEMLERİ ANA BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ**

OCAK 2023

Furkan İNCEDAYI tarafından hazırlanan “METİN ANALİTİĞİ VE AĞ ANALİZİ İLE İŞ YAPMA YÖNTEMİ PATENTLERİ VERİLERİNİN ANALİZİ” adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ ile Gazi Üniversitesi Yönetim Bilişim Sistemleri Ana Bilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Doç. Dr. Hakan ÇERÇİOĞLU

Endüstri Mühendisliği Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

Başkan: Prof. Dr. Cem İYİĞÜN

Endüstri Mühendisliği Ana Bilim Dalı, Orta Doğu Teknik Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

Üye: Prof. Dr. Nursal ARICI

Yönetim Bilişim Sistemleri Ana Bilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum.

Tez Savunma Tarihi: 18/01/2023

Jüri tarafından kabul edilen bu tezin Yüksek Lisans Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

.....
Prof. Dr. Aslıhan TÜFEKÇİ
Bilişim Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Bilişim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmada;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmada yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Furkan İNCEDAYI

18/01/2023

METİN ANALİTİĞİ VE AĞ ANALİZİ İLE İŞ YAPMA YÖNTEMİ PATENTLERİ VERİLERİNİN ANALİZİ

(Yüksek Lisans Tezi)

Furkan İNCEDAYI

GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ

Ocak 2023

ÖZET

İş yöntemi patentleri şirketler için rekabetçi bir avantaja sahip olma adına önemli seçeneklerden biri olmuştur. Bu çalışma, Türk Patent ve Marka Kurumu'na (TÜRKPATENT) Türkiye'de yerleşik özel veya tüzel başvuru sahipleri tarafından başvurusu yapılan en güncel iş yöntemi buluşlarını odağına alarak bilgisayar tabanlı iş yöntemlerindeki trendleri ve iş yöntemlerinin diğer teknoloji alanları ile bağlantısını tema modelleme ve Uluslararası Patent Sınıflandırma (IPC) ile İşbirlikçi Patent Sınıflandırma (CPC) kodları kullanılarak gerçekleştirilen ağ analizi yaklaşımları ile açıklamayı hedeflemektedir. 897 adet kayıttan oluşan nihai veri setinde, Türkçe patent özetleri üstünde Örtülü Dirichlet Yerleşimi (Latent Dirichlet Allocation - LDA) yöntemi ve sınıflandırma kodları üstünde Sosyal Ağ Analizi (SNA) yaklaşımı uygulanmıştır. Örtülü Dirichlet Yerleşimi modelinden elde edilen çıktılarla iş yöntemi patent başvurularındaki son dönem eğilimleri ortaya konmuş ve Sosyal Ağ Analizi ile iş yöntemlerinin teknoloji alanları ile ilişkisi sunulmuştur. Ayrıca, iki yöntemin sentezinden elde edilen bulguların değerlendirilmesi ile aktif başvuru sahiplerinin sürekli ve tutarlı biçimde kendi faaliyet alanları ile ilişkilendirilen teknik unsurları geliştirdikleri iş yöntemleri başvurularına aktardıkları saptanmıştır. Başvuru sahiplerinin ortaya koyduğu bu davranış, ulusal iş yöntemi patentlerindeki eğilimlerin dominant başvuru sahiplerinin etkinlik gösterdiği alana kaymasına neden olmaktadır.

Bilim Kodu : 90618

Anahtar Kelimeler : İş yöntemi patenti, tema modelleme, ağ analizi

Sayfa Adedi : 62

Danışman : Doç. Dr. Hakan ÇERÇİOĞLU

ANALYSIS OF BUSINESS METHOD PATENT DATA USING TOPIC MODELING
AND NETWORK ANALYSIS

(M. Sc. Thesis)

Furkan İNCEDAYI

GAZI UNIVERSITY
GRADUATE SCHOOL OF INFORMATICS

January 2023

ABSTRACT

Business Method Patents have been a considerable option for companies to get a competitive advantage in the market. This paper focuses on the most up-to-date computer-implemented business method patent applications filed by residents in Türkiye to the Turkish Patent and Trademark Office (TÜRKPATENT). Topic modeling and International Patent Classification (IPC) / Cooperative Patent Classification (CPC) codes network analysis approaches were used to reveal the recent trends and potential inter-relationships with technology fields. Latent Dirichlet Allocation (LDA), a well-known topic modeling technique, operating on patent abstracts in Turkish was utilized to report the recent trends of the business method patents in Türkiye while Social Network Analysis (SNA) built with classification codes demonstrates the inter-relationships between business methods and technology fields. The findings on the final dataset containing 897 business method patent applications also showed that active applicants persistently include technical characteristics of their industrial domain in their business methods. That behavior consequently causes the business method trends to shift towards the main fields of those applicants.

Science Code : 90618

Key Words : Business method patent, topic modeling, network analysis

Page Number : 62

Supervisor : Assoc. Prof. Dr. Hakan ÇERÇİOĞLU

TEŞEKKÜR

Tez çalışmam boyunca bilgi birikimini ve değerli önerilerini aktarmayı esirgemeyerek bana yardımcı olan danışmanım sayın Doç. Dr. Hakan ÇERÇİOĞLU'na; eğitim-öğretim hayatımın tamamında her zaman desteklerini hissettiğim annem, babam ve kardeşime; bu çalışma boyunca hiçbir desteği esirgemeyen ve çabaları olmasa bu çalışmayı bitirebileceğimden emin olmadığım çok sevgili R. Merve'ye teşekkür ve şükranlarımı sunarım.



İÇİNDEKİLER

Sayfa

ÖZET	iv
ABSTRACT.....	v
TEŞEKKÜR.....	vi
İÇİNDEKİLER	vii
ÇİZELGELERİN LİSTESİ.....	ix
ŞEKİLLERİN LİSTESİ.....	x
SİMGELER VE KISALTMALAR.....	xi
1. GİRİŞ.....	1
2. İŞ YÖNTEMİ PATENTLERİ: KISA BİR TARİHÇE.....	3
2.1. İş Yöntemi Patentlerinin Sınıflandırılması.....	6
2.2. İlgili Çalışmalar.....	7
3. TEMEL KURAMLAR VE KAVRAMLAR.....	11
3.1. Tema Modelleme ve Örtülü Dirichlet Yerleşimi (LDA)	11
3.2. Sosyal Ağ Analizi (SNA).....	14
Derece merkeziet (Degree centrality).....	16
Yakınlık merkeziet (Closeness centrality).....	16
Arasındalık merkeziet (Betweenness centrality)	17
Özvektör merkeziet (Eigenvector centrality)	17
Düğüm kuvveti (Node strength).....	18
4. METODOLOJİ.....	19
4.1. Veri Çekme, Ayırıştırma ve Düzenleme	20
4.2. Keşifsel Veri Analizi.....	21
4.3. Metin Ön İşleme.....	23
Yazım konsolidasyonu	24
Doküman koleksiyonu oluşturma.....	25
Bağlam içi dolgu kelimeleri ayıklama.....	25

Bölütleme (Tokenization).....	25
Yalınlaştırma (Lemmatization).....	26
4.4. Doküman-Frekans Matrisi (DFM) Çıktıları.....	27
4.5. Örtülü Dirichlet Yerleşimi (LDA) ile Tema Modelleme	28
4.6. IPC Ağ Analizi	33
5. BULGULAR ve DEĞERLENDİRME.....	41
6. SONUÇ.....	49
KAYNAKLAR	51
EKLER.....	60
Ek – 1: Çalışmada kullanılan veri seti	61
ÖZGEÇMİŞ	62

ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 2.1: IPC sisteminde bölümler ve tanımları [12].....	6
Çizelge 2.2: İş yöntemi patentlerinde IPC grupları ve tanımları [12].....	7
Çizelge 3.1: Örnek bir DTM.....	13
Çizelge 4.1: İş yöntemi patentleri için araştırma raporu içeriği.....	22
Çizelge 4.2: Patent özetlerinde terim yazımları ve konsolidasyon	24
Çizelge 4.3: Yalınlaştırma ve kökleştirme örneği.....	26
Çizelge 4.4: Algoritmaya göre IPC kodu belirleme.....	34
Çizelge 4.5: İş yöntemi IPC alt sınıfı ile teknoloji alanları ilişkisi.....	35
Çizelge 4.6: En yüksek derece merkeziyete sahip teknoloji alanları.....	39
Çizelge 4.7: En yüksek düğüm kuvvetine sahip teknoloji alanları.....	39
Çizelge 5.1: LDA ile oluşturulan tema etiketleri	43
Çizelge 5.2: Hedef düğümlere ait IPC kod tanımları [12]	45
Çizelge 5.3: Başvuru sayısına göre başvuru sahipleri	46

ŞEKİLLERİN LİSTESİ

Şekil	Sayfa
Şekil 2.1: USPTO’da iş yöntemi patentlerinin tescil oranı [5]	3
Şekil 2.2: “Tek Tıkla Sipariş” patentinin bibliyografik veri görüntüsü [7]	5
Şekil 3.1: Orijinal yayında yer alan LDA grafik gösterimi [24]	12
Şekil 3.2: Yönsüz bir SNA diyagramı örneği	16
Şekil 4.1: Çalışma için kurgulanan süreçler ve akış şeması	19
Şekil 4.2: Araştırma uzmanına göre rapor içeriği dağılımı	23
Şekil 4.3: Patent özetlerine uygulanan metin ön işleme adımları	24
Şekil 4.4: Patent özetlerinden elde edilen sözcük bulutu	27
Şekil 4.5: Terimlerin kullanıldığı doküman sayısı ve toplam frekansları	28
Şekil 4.6: Dirichlet parametresi (α), tema sayısı (K) ve karmaşıklık skoru ilişkisi	29
Şekil 4.7: Çeşitli parametreler ile ideal tema sayısı (K) ilişkisi	30
Şekil 4.8: Harmonik istatistiksel ortalama ve tema sayısı (K) ilişkisi	31
Şekil 4.9: LDA ile oluşturulan temalar ve üst terimler	32
Şekil 4.10: Veri setindeki dosyalarda IPC tayini için uygulanan algoritma	34
Şekil 4.11: IPC kodu veri setleri ile oluşan SNA diyagramı	37
Şekil 4.12: Serbest konumlandırılmış düğümlerle SNA	38
Şekil 5.1: LDavis ile örnek bir tema ve içerik gösterimi	42
Şekil 5.2: En yüksek olasılık değerine göre doküman sayısı – tema dağılımı	44

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler

Açıklamalar

α	Dirichlet hiperparametresi
θ	Doküman başına tema oranı
β	Tema
η	Tema hiperparametresi
z	Sözcük başına tema ataması
w	Gözlenen sözcük
N	Doküman başına sözcük sayısı
M	Doküman sayısı
K	Tema sayısı

Kısaltmalar

Açıklamalar

CII	Bilgisayar Tabanlı Buluş (Computer Implemented Invention)
DFM	Doküman-Frekans Matrisi (Document-Frequency Matrix)
DTM	Doküman-Terim Matrisi (Document-Term Matrix)
EPO	Avrupa Patent Ofisi (European Patent Office)
IPC	Uluslararası Patent Sınıflandırması (International Patent Classification)
LDA	Örtülü Dirichlet Yerleşimi (Latent Dirichlet Allocation)
LSA	Örtülü Semantik Analiz (Latent Semantic Analysis)

NLP	Natural Language Processing (Doğal Dil İşleme)
<i>p</i>LSA	Olasılıksal Örtülü Semantik Analiz (Probabilistic Latent Semantic Analysis)
SNA	Sosyal Ağ Analizi (Social Network Analysis)
SMK	Sınai Mülkiyet Kanunu
tf-idf	Terim Frekansı – Ters Doküman Frekansı (Term Frequency – Inverse Document Frequency)
TÜRKPATENT	Türk Patent ve Marka Kurumu
USPTO	Birleşik Devletler Patent ve Marka Kurumu (United States Patent and Trademark Office)
WIPO	Dünya Fikri Mülkiyet Teşkilatı (World Intellectual Property Organization)

1. GİRİŞ

Patent, bir probleme getirilen teknik bir çözümün bir ürün, sistem, yöntem veya bunların bir kombinasyonu ile korunarak üçüncü kişiler tarafından üretilmesini, ticaretinin yapılmasını engelleyen, başvuru sahibine sınırlı bir süre için verilmiş bir tekel hakkıdır. İş yöntemi patentleri ise genel patent kavramının altında belirli bir patent türünü ifade etmektedir. İş yöntemi patentleri hakkında detaylı bilgi sahibi olunması için öncelikle bir iş yönteminin ne ifade ettiğinin kavranması önemlidir. Araştırmacılar geçmişten günümüze bir iş yöntemini tarif etmek için farklı tanımlamalar geliştirmişlerdir. Örneğin, Hammer ve Champy iş yöntemlerini “alıcı için değerli bir çıktı üreten ve sıklıkla iş süreçlerini ifade eden bir yapılandırılmış faaliyetler dizisi” olarak tanımlamıştır [1]. Patent perspektifinden ise iş yöntemi patentleri Birleşik Devletler Patent ve Marka Ofisi (USPTO) tarafından “bir işin yapılma yöntemini koruyan uygulanabilir bir patent” olarak açıklanmıştır [2]. Burada geçen “uygulanabilir” kavramı patentlenebilirlik kriterlerinden biri olan “sanayiye uygulanabilirlik” olarak anlaşılmalıdır.

Bu çalışmanın amacı, Türk Patent ve Marka Kurumuna (TÜRKPATENT) Türkiye’de yerleşik başvuru sahipleri tarafından başvurusu yapılan iş yöntemi patentleri için, daha sonra detaylarıyla açıklanacağı, üzere patent özetlerinde gerçekleştirilen tema modelleme ve ilişkili IPC (Uluslararası Patent Sınıflandırması)/CPC (İşbirlikçi Patent Sınıflandırması) kodlarına uygulanan ağ analizi teknikleri kullanılarak ilgili buluşlardaki eğilimleri ortaya koymak ve iş yöntemi patentlerinin diğer teknoloji alanları ile potansiyel ilişkisini açıklamaktır (bundan sonra anlatımı kolaylaştırmak amacıyla yalnızca IPC kodları veya belli kısımlarda sınıflandırma kodları olarak anılacaktır).

Çalışmanın teori ve pratikler bakımından literatür katkıları aşağıda verilmiştir:

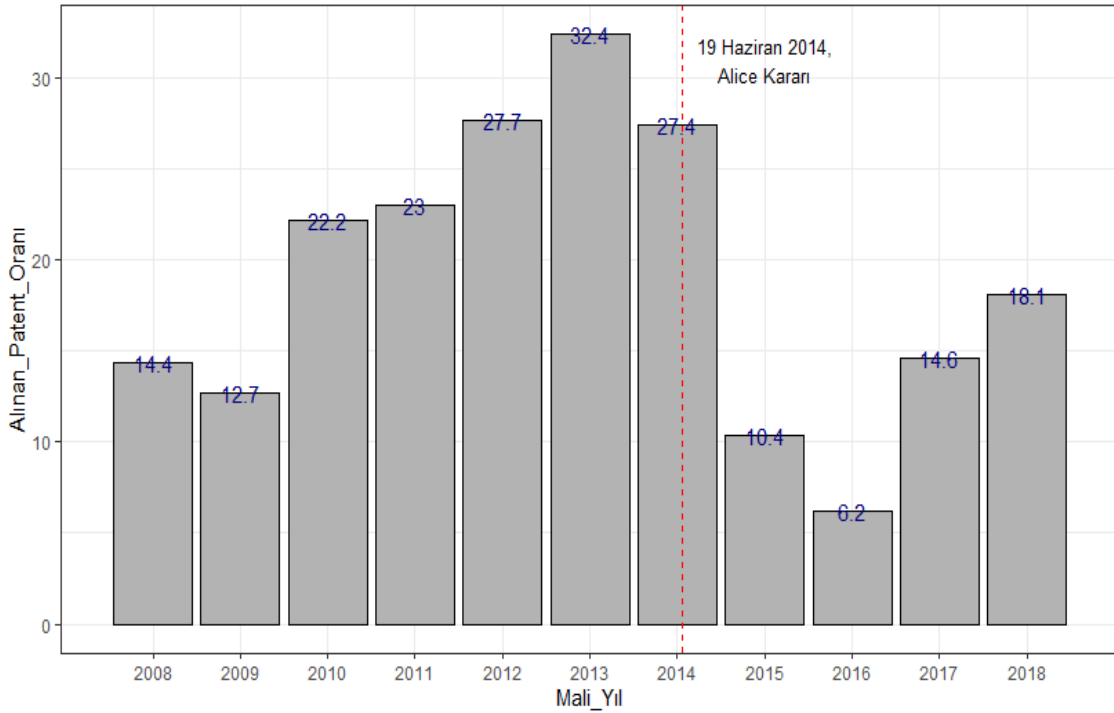
- Yapılandırılmamış patent verileri sınıfına giren patent özetlerinin eğilim gözlemi için kullanılabilirliği gösterilmiştir.
- Ulusal iş yöntemi patentleri ile teknoloji alanı ilişkisi ortaya konmuştur.
- Ön kriterlere dayalı algoritmik seçim ile veri güvenilirliğinin artırılması önerilmiştir.
- Daha büyük veri setleri ile yapılabilecek gelecekteki çalışmalar için bir uygulama alanı sunulmuştur.

- İş yöntemi – teknoloji alanı ilişkisini açıklamak için dosya tabanlı IPC ağ analizi ve tema modelleme tekniklerinin beraber kullanıldığı literatürdeki ilk çalışmadır.

Yukarıda açıklanan katkıların detaylı aktarımı için bu tez çalışması altı ana bölüm altında incelenmiştir. Bölüm 2’de iş yöntemi patentlerinin kısa bir tarihçesi sunulurken aynı zamanda bu tip patentlerin sınıflandırılması ve iş yöntemleri temel alınarak yapılmış önceki araştırmalardan kesitler açıklanmıştır. Bölüm 3’te Temel Kavramlar ve Kuramlar başlığı altında tez çalışmasının ana bileşenleri olan Örtülü Dirichlet Yerleşimi (Latent Dirichlet Allocation - LDA) ve sosyal ağ analizi (SNA) ile ilgili kapsamlı bilgilere yer verilmiştir. Bölüm 4’te çalışmanın dayandığı yöntemler üzerine odaklanılmış ve çalışmada kullanılan materyal açıklanarak uygulanan metodolojinin ayrıntıları ve işlem basamakları verilmiştir. Bölüm 5’te söz konusu metodoloji ile yürütülen çalışmadan elde edilen çıktılar verilmiş ve bulgular açıklanmıştır. Söz konusu bulgulardan yola çıkılarak bir değerlendirme yürütülmüş ve bulguları etkileme potansiyeline sahip nedenler, dayanakları ile beraber aktarılmaya çalışılmıştır. Son bölüm olan Bölüm 6’da çalışma ile ilgili sonuçlar ve gelecekteki çalışmalara yönelik öneriler ifade edilmiştir.

2. İŞ YÖNTEMİ PATENTLERİ: KISA BİR TARİHÇE

İş yöntemi patentleri genellikle bilgisayar tabanlı buluşlar (CII) kapsamına girmektedir. Bu tip buluşların patentlenebilirliği araştırmacılar, patent uzmanları, patent profesyonelleri ve hukukçular arasında tartışma konusu olmuştur [3]. Bunun genel nedeni olarak CII'lerin soyut veya bilinen bir fikrin salt bilgisayar gibi bir donanım aracını kullanmasının patentlenebilirliğin önemli ön şartlardan biri olan teknik karakteri sağlamada yeterli olmayacağı görüşüdür. Bu görüşün kaynağı, Birleşik Devletler Yüksek Mahkemesinin Alice Kararı olarak bilinen, Alice Corp ve CLS Bank International firmalarının taraf olduğu davada aldığı karardır [4]. Günümüzde Avrupa Patent Ofisi (EPO) Temyiz Kurulu'nun mevcut kararları, alınan bu önemli kararın Avrupa'da da uygulanmasını sağlamlaştırmıştır. Alice Kararından sonra USPTO özelinde iş yöntemi buluşlarının patent tescil oranlarında Şekil 2.1'de görülebileceği gibi fark edilebilir değişiklikler yaşanmıştır [5].



Şekil 2.1: USPTO'da iş yöntemi patentlerinin tescil oranı [5]

Bununla beraber, literatürde patentlenmek istenen buluşun teknik karakterinin CII'ler ve bilhassa iş yöntemi patentleri için nasıl sağlanabileceği ve bu tür patentler için EPO'da yapılan incelemeler ve değerlendirmeler hakkında çalışmalar da mevcuttur [6].

Alice Kararı'nı takiben patent ofisleri tarafından iş yöntemleri patentleri ve daha geniş kapsamda yazılım patentleri için iki aşamalı bir değerlendirme süreci yürütülmeye başlanmıştır. Bu aşamalar şu şekilde özetlenebilir:

- (i) Buluş konusu soyut bir kavramla mı ilişkili?
- (ii) Buluş konusunda yer alan unsurlar bu soyut kavramı patentlenebilir bir uygulamaya dönüştürüyor mu?

Eğer inceleme uzmanının ilk soruya cevabı hayır ise buluş konusunun teknik bir karaktere sahip olduğu anlaşılabilir. Bu nedenle, ikinci adımdaki soru uygulanmaz ve doğrudan patentlenebilirlik kriterleri olan yenilik ve buluş basamağı değerlendirmesi yapılır. Ancak, ilk soruya uzman tarafından verilen cevap evet ise, ikinci aşamada buluş konusunda yer alan unsurların soyut bir kavramı patentlenebilir hale getirip getiremediği sorgulanır ve verilen cevap evet olduğu takdirde yenilik/buluş basamağı değerlendirmesi yapılır. İkinci aşamadaki soruya verilen cevap hayır olduğunda ise başvurunun reddi beklenir. Yenilik ve buluş basamağı kriterleri yalnızca iş yöntemleri, CII'ler veya dijital/yazılım patentlerine özgü koşullar değildir. Adı geçen kriterler tüm patent başvuruları için sorgulanmaktadır ve kısa açıklamaları aşağıda verilmiştir:

Yenilik: Buluş konusunun başvuru tarihinden önce kamuoyuna açılmış herhangi bir dildeki tek bir doküman, açıklama, makale, video vb. karşısında yeni olup olmadığının değerlendirildiği safhadır.

Buluş Basamağı: Başvurunun yenilik kriterini sağladığı durumlarda başvuru tarihinden önce kamuoyuna açılmış herhangi bir dildeki bir veya birden fazla doküman, yazılı açıklama, makale, video vb.nin kombinasyonu karşısında teknikte uzman kişi için aşikâr olup olmadığının değerlendirildiği safhadır.

Yukarıda açıklanan bilgilerin daha iyi anlaşılabilmesi adına iş yöntemi patentlerine ilişkin bazı somut örneklerin sunulmasının yerinde olacağı öngörülmüştür. Örneğin, Amazon firması tarafından yapılan ve kamuoyunda “Tek Tıkla Sipariş” olarak bilinen yöntem 12 Eylül 1997 tarihinde Amazon tarafından USPTO'ya “Bir Haberleşme Ağı Üzerinden Bir Sipariş Oluşturma Yöntemi ve Sistemi” başlığıyla başvurusu yapılmış ve ofis tarafından 28 Eylül 1999 tarihinde US5960411A yayın numarası ile üçüncü kişilere ulaşılabilir kılınmıştır [7]. Şekil 2.2'de Amazon'un patent başvurusunun bibliyografi verilerini içeren sayfası gösterilmiştir.

Home > Results > US5960411A

1 >

☆ **US5960411A** Method and system for placing a purchase order via a communications network

Bibliographic data	Description	Claims	Drawings	Original document	Citations	Legal events	Patent family
Applicants	AMAZON COM INC [US] +						
Inventors	HARTMAN PERI [US]; BEZOS JEFFREY P [US]; KAPHAN SHEL [US]; SPIEGEL JOEL [US] +						
Classifications							
IPC	G06Q10/08; G06Q20/04; G06Q20/12; G06Q30/04; G06Q30/06; (IPC1-7): G06F17/60;						
CPC	G06Q10/087 (EP,US); G06Q20/04 (EP,US); G06Q20/12 (EP,US); G06Q30/04 (EP,US); G06Q30/06 (EP,US); G06Q30/0605 (EP,US); G06Q30/0633 (EP,US); G06Q30/0635 (EP,US); G06Q30/0641 (EP,US); G07F9/002 (EP,US); Y10S715/962 (EP,US);						
Priorities	US92895197A 1997-09-12						
Application	US92895197A 1997-09-12						
Publication	US5960411A 1999-09-28						
Published as	US2007106570A1; US2013085896A1; US5960411A; US8341036B2						

Şekil 2.2: “Tek Tıkla Sipariş” patentinin bibliyografik veri görüntüsü [7]

Bu sayede Amazon kullanıcılarının daha etkili ve hızlı sipariş verebilmelerine imkân tanınmıştır. Bu sistem ve metot, bilindiği üzere müşterinin hesap ve kullanıcı bilgilerinin önceden kaydedilmesi sayesinde ürünün sanal alışveriş sepetine atılmadan tek bir hamle ile siparişinin verilmesini sağlar ve günümüzde halen kullanımdadır. Bir başka örnek Priceline firması tarafından patentlenen ve kamuoyunda “Tersine Müzayede” olarak bilinen buluştur [8]. Geleneksel açık arttırmalarda tek bir satıcı ve birden fazla alıcı bulunurken Priceline firması tersine bir süreç öngörmüş ve birden fazla satıcının tek bir alıcı ile etkileşimde olduğu bir müzayede sistemi kurmuştur. Dolayısıyla, inovatif bir fikrin teknik bir çözüme dönüştürüldüğü sistem ve yöntemlerin patent ile koruma altına alınmasının başvuru sahibine rekabetçi bir avantaj kazandırması muhtemeldir.

Patent bilgilerinin kullanımı kurumsal düzeyde teknoloji yönetim stratejilerinin belirlenmesinde kullanılabilen veriye dayalı önemli bir bileşendir [9]. İş yöntemi patentlerinin diğer tip patentlere (örneğin, bir ürün patenti) oranla daha az değerli olduğuna ilişkin klasik görüşe karşı Allison ve Tiller patent veri setlerini kullanarak iş yöntemlerinin herhangi bir teknoloji alanına ait ortalama bir patentten daha kötü olmadığını; aksine potansiyel olarak daha iyi olduğu sonucuna ulaşmıştır [10]. Bununla birlikte, Martin ve Mykytyn bir iş yöntemi patentine sahip olmanın patent sahibine faaliyet alanında finansal avantaj getirip getirmediğine ilişkin araştırmayı literatüre kazandırmıştır [11].

2.1. İş Yöntemi Patentlerinin Sınıflandırılması

Patentler, ait oldukları teknolojik alanlara göre Dünya Fikri Mülkiye Teşkilatı (WIPO) tarafından yayımlanan ve IPC kısaltması ile bilinen uluslararası patent tasnif sistemine tabidir. Patentlerin sınıflandırılmasının önemli bir avantajı patent dosyalarının dilden bağımsız biçimde araştırılmasına olanak tanınmasıdır. Patent sınıflandırılmasına ilişkin notasyon Çizelge 2.1’de sunulduğu gibi en üst seviyede bölüm adı verilen ve A-H harfleri ile ifade edilen kısımlarla başlamaktadır [12].

Çizelge 2.1: IPC sisteminde bölümler ve tanımları [12]

Bölüm	Tanım
A	İnsan İhtiyaçları
B	İşlemlerin Uygulanması; Taşıma
C	Kimya; Metalürji
D	Tekstil; Kâğıt
E	Sabit Yapılar
F	Makine Mühendisliği; Aydınlatma; Isıtma; Silahlar
G	Fizik
H	Elektrik

IPC kodları patent sisteminde öncelikli olarak yukarıda belirtilen bölüm harflerinden biri ile başlar. Hiyerarşik olarak bölüm harfini iki adet rakam takip ederek IPC sınıfını oluşturur. IPC sınıfını takip eden bir harf ile IPC alt sınıfı meydana getirilir. Daha alt seviyede alt sınıfı bir veya iki rakam takip ederek IPC grubunu oluşturur. IPC grubundan sonra / karakteri ayırıcı olarak kullanılır ve bu karakterden sonra gelen rakamlar IPC alt gruplarını ifade eder. Örneğin; H04W8/20 koduna sahip bir patent dosyası tasnif şemasında “Kullanıcı veya abone verisi transferi” olarak tanımlanmıştır. Bu yüzden, belirtilen koda sahip bir patent dosyası hangi dilde kaleme alınmış olursa olsun; buluş konusu, IPC sistemi vasıtasıyla dilden bağımsız olarak kategorize edilebilir ve hangi teknoloji alanı ile ilgili olduğu saptanabilir. IPC sistemi tarihsel süreçte birden fazla kez güncellenmiştir. Mevcut sürümde bilgisayar tabanlı iş yöntemi patentleri çoğunlukla G06Q IPC alt sınıfıyla ilişkilendirilir. Günümüzde kullanılan IPC sürümü IPC’nin sekizinci versiyonu olarak isimlendirilmiştir [13]. G06Q alt sınıfı “bilhassa idari, ticari, finansal, yönetsel, denetsel veya tahminsel amaçlar için

uyarlanmış veri işleme sistemleri veya yöntemleri” olarak tanımlanmıştır [12]. Bu sınıfın altında konumlanan IPC gruplarına ilişkin tanımlar Çizelge 2.2’de yer almaktadır:

Çizelge 2.2: İş yöntemi patentlerinde IPC grupları ve tanımları [12]

IPC grup	Tanım
G06Q10	İdare; Yönetim
G06Q20	Ödeme mimarileri; şemaları ve protokolleri
G06Q30	Ticaret; örneğin alışveriş veya e-ticaret
G06Q40	Finans; Sigorta; Vergi Stratejileri; Kurumsal veya gelir vergisi işleme
G06Q50	Özellikle iş sektörlerine uyarlanan sistem ve yöntemler; örneğin kamu hizmetleri veya turizm
G06Q90	Özellikle yoğun veri işleme içermeyen idarî, ticari, finansal, yönetsel, denetimsel veya tahmin amaçlarına uyarlanan sistem ve yöntemler
G06Q99	Buluş konusu bu alt sınıfta yer alan diğer gruplarda tanımlanmamış (sistem ve yöntemler)

WIPO tarafından sunulan IPC sisteminin yedinci versiyonunda iş yöntemleri büyük ölçüde G06F17/60 altında konumlandırılmıştır [14]. Bu nedenle, bir önceki sürümde iş yöntemlerinin çoğunlukla tek bir IPC alt grubu ile ilişkilendirildiği anlaşılabılır. Ancak bu alandan gelen başvuru sayılarındaki önemli artış otoriteler tarafından dikkate alınmış ve Ocak 2006’da yapılan güncelleme ile iş yöntemleri Avrupa genelinde yeni oluşturulan G06Q alt sınıfı içinde değerlendirilmeye başlanmıştır. USPTO’da ise iş yöntemi patentleri Sınıf 705 adı verilen patent topluluğu içinde değerlendirilmektedir.

2.2. İlgili Çalışmalar

Literatürde gerek araştırmacılar gerekse akademisyenler tarafından yapılmış ve iş yöntemleri patentleri ile doğrudan veya dolaylı olarak bir bağı bulunan çok sayıda araştırma bulunmaktadır. Bu çalışma için gerçekleştirilen literatür taraması hem iş yöntemi patentlerine ilişkin araştırmaların hem de işbu çalışmada kullanılan patent verisi analitiği ile ilgili yapılmış çalışmaların kısa bir dökümünü içermektedir.

Giriş kısmında bahsedildiği üzere, çalışmanın amaçlarından biri ilgili patent verilerini kullanarak iş yöntemi patentleri için bir trend gözlemi ortaya koymaktır. Literatürde, patent verilerinin bir teknoloji alanındaki eğilimleri belirlemek için kullanımına ilişkin çalışmalar mevcuttur. Örneğin, Dereli ve Durmuşoğlu TÜRKPATENT veritabanından alınan tekstil patentleri verilerinde bulanık tabanlı kümeleme yöntemini kullanarak tekstil endüstrisindeki geleceği en parlak alt sektörleri sektörleri tespit etmeye çalışmıştır [15].

Patent verisinin incelenmesi için kullanılan veri analitiği yaklaşımlarında belli bir yazılım platformunun kullanılması sık rastlanan bir uygulamadır. Örneğin Jain, Tripathi, Agarwal ve Murthy gelişmiş bir kümeleme tekniğinin kullanılması ile patent dokümanlarındaki teknolojik kapsamın irdelenebileceğini göstermek için açık kaynak kodlu istatistiksel bir yazılım dili olan R dilinden faydalanmıştır [16].

Jiang, Goel ve Zhang, yazılım tabanlı iş yöntemi patentlerinden kaynaklanan bilgi akışını keşfetmek için patent atıflarını kullanmıştır. Çalışmada yazarlar USPTO'ya başvurusu yapılan iş yöntemlerinden istifade etmiş ve Sınıf 705'a ait patent kayıtlarını filtrelemek için belirli anahtar sözcükler kullanmışlardır. Yazarlar, patentlerin aldıkları atıf sayısına göre firmaları konumlandırmak için bir patent sosyal ağ analizi uygulamıştır [17]. Benzer şekilde No, An ve Park da iş yöntemleri kaynaklı bilgi akışı konusuna değinmiştir. İlgili çalışmada USPTO Sınıf 705'e ait patent verileri alınarak iş yöntemlerinden doğan akış örüntülerinin belirlenmesi için patent atıf analizi ve metin madenciliği teknikleri kullanılmıştır [18].

Moehrle, Wustmans ve Gerken, radyo frekans tanımlayıcı cihaz alanındaki patentleri elde etmek için USPTO veritabanında bir anahtar sözcük araması gerçekleştirmiştir. Karşılaştırmalı bir analiz ortaya koymak için kayıtlar arasından Sınıf 705'e ait iş yöntemi patentlerini ayırtıran araştırmacılar, semantik ve informetrik yaklaşımları kullanarak bir firmanın teknolojik inovasyonlarının iş yöntemleri tarafından eşlik edilip edilmediği bulgusuna ulaşmaya çalışmışlardır [19].

Lee ve Sohn, finans alanı ile ilişkili iş yöntemi patentlerindeki eğilimi göstermeyi amaçlamıştır. Yazarlar, patent özetleri üzerinde tema modelleme tekniği olarak LDA'ı seçmiş ve olasılık eşik değeri yaklaşımı ile trend olan temaları 'sıcak', popülerliğini yitiren temaları ise 'soğuk' olarak sınıflandırmışlardır. Çalışmada kullanılan veri seti USPTO Sınıf 705'in belirli alt sınıfları ile G06Q40 IPC alt sınıfı kullanılarak elde edilmiştir [20].

Chang, Lai ve Chang, iş yöntemi patentlerinde sınıflandırma ve teknoloji difüzyonu üzerine yoğunlaşmıştır. USPTO Sınıf 705 kayıtlarının elde edilmesini takiben, yazarlar teknoloji difüzyonu bağlantısını göstermek için bir patent atıf ağı oluşturmuş; temel patentleri sınıflandırmak içinse bir hiyerarşik küme analizi gerçekleştirmişlerdir [21].

Önceki çalışmalarda, bilginiz ölçüsünde, iş yöntemi patentlerindeki eğilimleri ve diğer teknoloji alanları ile ilişkilerini hem tema modelleme tekniğı hem de tekil dosya tabanlı bir IPC kodu ağı ile sunmaya çalışan bir çalışma mevcut değildir. İlerleyen kısımlarda görüleceğı üzere, bu tez çalışması bulguları ortaya koymak adına her iki yaklaşımın bir sentezini kullanmıştır.





3. TEMEL KURAMLAR VE KAVRAMLAR

Çalışmanın bu bölümünde metin analitiği kapsamında kullanılan tema modelleme tekniği ve patent sınıflandırma kodları kullanılarak gerçekleştirilen ağ analizine ilişkin teori ve uygulamalarla ilgili kapsamlı açıklamalar alt başlıklarla verilmektedir.

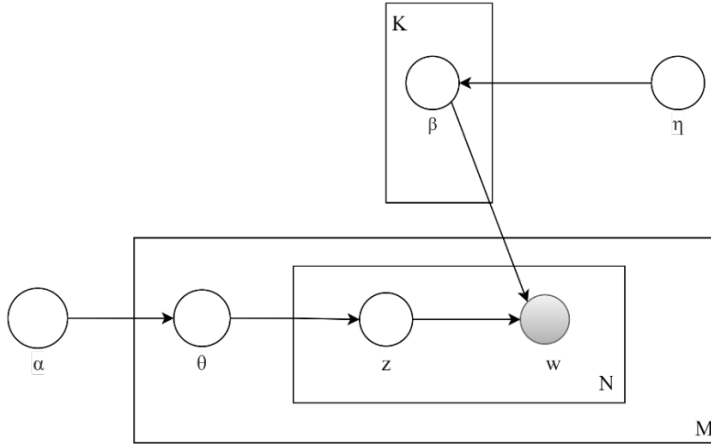
3.1. Tema Modelleme ve Örtülü Dirichlet Yerleşimi (LDA)

Tema modelleme tanımsal olarak, her bir dokümanın tema adı verilen örtülü değişkenlerin bir fonksiyonu olduğu varsayımıyla metin dokümanlarını kümeleme yaklaşımıdır [22]. Dokümanlar bir araya gelerek derlem (*corpus*) adı verilen bir doküman koleksiyonunu oluşturur. Birden çok derlemin bir araya getirilmesi ile derlemler (*corpora*) adı verilen çoklu doküman koleksiyonları oluşturulabilir. Tema modellemede bir doküman içinde bölüt (*token*) olarak tanımlanan birimler bulunmaktadır. Doğal Dil İşleme (NLP) uygulamalarında bölütler tek bir sözcük olabileceği gibi, sıralı iki sözcük, sıralı üç sözcük, bir cümle veya bir paragraf da olabilir. Doküman içinde tek bir sözcüğün bölüt olarak belirlendiği durumda bölütler tek ağırlıklı bölüt (*unigram token*) olarak ifade edilirken çoklu sıralı sözcük dizileri için iki ağırlıklı (*bigram*), üç ağırlıklı (*trigram*) vb. gibi tanımlar kullanılmaktadır. Tema modelleme tekniğinin uygulama alanlarında kullanılan doküman tipleri çeşitlidir. Bunlar arasında, gazete köşe yazıları, anket cevapları, müşteri geribildirimleri gibi metinler bulunmaktadır.

Blei, Ng ve Jordan tarafından tanıtılan ve tema modellemede yaygın olarak kullanılan Örtülü Dirichlet Yerleşimi (LDA) üç seviyeli hiyerarşik bir Bayesian modeldir [23]. LDA, aynı zamanda bir karışımli üyelik modelidir. Karışımli üyelik tanımının altında yatan temel fikirler LDA tekniğinin tarifini yapan Blei tarafından şu şekilde açıklanmıştır [24]:

- Veri gruplandırılır.
- Her bir grup bir karışım ile modellenir
- Karışım bileşenleri tüm gruplar tarafından paylaşılır.
- Karışım oranları gruptan gruba değişkenlik gösterir.

Şekil 3.1’de Blei ve diğerleri tarafından ortaya konan LDA’nın bir grafiksel gösterimi sunulmuştur. Kullanılan parametre sembollerinin açıklaması aşağıda verilmiştir:



Şekil 3.1: Orijinal yayında yer alan LDA grafik gösterimi [24]

LDA, tema modelleme için her bir dokümanın birden fazla tema ile ilişkilendirilebildiği bir yaklaşım sağlamaktadır. Bu nedenle, bir olasılık dağılımı da göz önünde bulundurularak bir doküman birden fazla tema ile bağdaştırılabilir. Olasılık dağılımında en yüksek değere sahip olan tema dokümanın en ilgili olduğu tema olarak anlaşılabilir. Ancak, metin dokümanlarının heterojen olması beklendiği için dokümanla ilişkili diğer temaların varlığı metnin hangi temaları içerdiğinin anlaşılmasına katkı sağlar ve dokümanı kavramaya yardımcı olur. Bu yönüyle, K-means gibi kümeleme tekniklerinden farklı olarak, LDA bir ‘yumuşak kümelendirme (soft clustering)’ mekanizması olarak da bilinmektedir.

Şekil 3.1’de yer alan dikdörtgen şekilleri yinelenen süreçleri ifade etmek için kullanılırken, oklar değişkenler arası tek yönlü bağlantıları göstermek için kullanılmıştır. Grafik gösterimde “w” harfinin gölgelendirildiği görülmektedir. Bu gösterim gözlenen değişkenin yalnızca sözcükler olduğunu anlatmak için kullanılmıştır. Diğer değişkenler hiperparametreler ve LDA tekniğinde gözlenemeyen örtülü değişkenlerdir.

İteratif bir sürecin yürütüldüğü LDA yönteminde doküman başına genellikle bir sözcük yapısında olan her bir bölüt (w) için bir tema ataması (z) yapılır. Seçilen bir dokümanın tema dağılımı ise Dirichlet parametresi (α) ile bağlantılı olduğu görülen doküman başına tema dağılım parametresi (θ) ile gerçekleştirilir. Bir tema (β) üzerindeki sözcük dağılımı tema hiperparametresi (η) tarafından belirlenir. Süreç, derlem içinde her bir dokümanda yer alan tüm sözcükler (N), dolayısıyla tüm dokümanlar (M) modellenene kadar devam eder.

Hiperparametlerden α ve η ile oluşturulacak tema sayısını bildiren K değişkeni modellemeyen önce kullanıcı tarafından belirlenebilmektedir. Bu tez çalışmasında, adı

geçen hiperparametreler ve tema sayısının nasıl belirlendiğine dair yöntem, metodoloji kısmında detaylı olarak açıklanmıştır.

LDA tekniği metin dokümanlarında “sözcük torbası” yaklaşımını kullanılır. Sözcük torbası bir doküman içindeki sözcüklerin yeri ve semantiği ile ilgilenmemekte; yalnızca sözcüğün doküman içindeki frekansına odaklanmaktadır. Diğer bir deyişle, sözcüklerin sırası göz ardı edilmektedir [21]. Sözcük torbası yaklaşımının nasıl uygulandığına ilişkin aşağıda bir örnek verilmektedir.

“Eğer bir ördek gibi yürüyor, bir ördek gibi vaklıyor ve bir ördek gibi yüzüyorsa muhtemelen o bir ördek.”

Yukarıdaki cümlede sözcük torbası yaklaşımı uygulanmak istendiğinde sözcük dizilimi göz ardı edilerek sözcük terimlerinin frekans tablosu oluşturulur. Dolayısıyla verilen örneğin bir doküman olduğu varsayıldığında Çizelge 3.1’deki gibi bir doküman-terim matrisi (DTM) oluşur.

Çizelge 3.1: Örnek bir DTM

Terim	eğer	bir	ördek	gibi	yürüyor
Frekans	1	4	4	3	1
Terim	vaklıyor	ve	yüzüyorsa	muhtemelen	o
Frekans	1	1	1	1	1

Çizelge 3.1 örnek bir doküman üzerinden oluşturulmuş bir terim-frekans tablosudur ve NLP uygulamalarında sıklıkla karşılaşılan bir doküman-terim matrisi (DTM) örneğidir. DTM’ler LDA tekniğinin de önemli bir bileşenidir. Zira DTM’ler üzerinden yapılan analizde iki sözcüğün bir doküman içinde beraber bulunuyor olması bu iki sözcüğün aynı temaya ait olma olasılığını artırır [22]. LDA, dokümanlar üzerinde bu yaklaşımı uygulayarak derlemi oluşturan örtülü temaları bulmayı amaçlar.

Gözetimsiz bir makine öğrenmesi tekniği olan LDA’de sözcük sırasının göz ardı edilebilmesi işlemsel yoğunluk ve işlem zamanı açısından avantajlıdır. Bunun yanında LDA, karışımli modelleme kabiliyeti sayesinde dokümanlara modelleme süreci sonucunda ortaya çıkan temaların olasılıksal bir karışımını atayabilir. Bir başka deyişle, LDA yöntemi her bir

dokümanı bir olasılık skoru ile beraber birden fazla tema ile ilişkilendirerek derlemi modeller. Ayrıca LDA, tema modellemede kullanılan diğer yaklaşımlar olan Örtülü Semantik Analiz (Latent Semantic Analysis) LSA, Olasılıksal Örtülü Semantik Analiz (Probabilistic Semantic Analysis – *p*LSA) ve terim frekansı – ters doküman frekansı (tf-idf) gibi yöntemlerin sahip olmadığı genelleme özelliğine sahiptir [25].

LDA’ın bahsedilen avantajlarına rağmen veriler ile ortaya konmuş kısıtları mevcuttur. Bunlardan biri LDA tekniğinin kısa metinler ve küçük doküman koleksiyonlarında kötü modelleme performansı göstermiş olmasıdır [26]. Örneğin, Pietsch ve Lesmann anket sorularına verilen açık uçlu cevaplar üzerinde yaptığı çalışmada LDA yerine daha uygun olabilecek farklı yaklaşımları önermiştir [27]. Yazarlar, açık uçlu cevapların birden fazla tema içerdiği ancak LDA’ın kısa dokümanlardaki yetersiz performansı dolayısıyla başka tekniklerin kullanımına ve bu tekniklerin LDA’dan farklarını açıklamaya odaklanmıştır. Tek bir temanın söz konusu olabileceği tweetler gibi kısa metin dokümanları içinse LDA’ın modelleme performansı benzer şekilde yetersiz bulunmuş ve tek temalı LDA yaklaşımı önerilmiştir [28]. Bir başka kısıt ise LDA’ın geniş kapsamlı temaları spesifik olanlara göre daha iyi saptayabildiği, dolayısıyla dar kapsamlı temaları tespit etmede yetersiz olduğuna ilişkin çalışmalar olarak görülmektedir [29].

Patent özetleri, tweetler veya anket sorularına verilen açık uçlu cevapların aksine, genellikle yarım sayfayı aşmayacak şekilde düzenlenen ve LDA modelleme için yeterli sözcük sayısına sahip olduğu düşünülen metinlerdir. Bu nedenle, LDA’ın kısa metinlere uygulanırken beraberinde getirdiği dezavantajların patent özetleri için deneyimlenmeyeceği göz önüne alınarak bu çalışmada tema modelleme yöntemi olarak LDA kullanılmıştır.

3.2. Sosyal Ağ Analizi (SNA)

Sosyal Ağ Analizi (SNA) bir grubun, bireyin veya bir varlığın başka bir grup, birey veya varlık ile ilişkisini inceleyen bir araştırma yöntemidir. SNA için verilen bu üst seviye tanım farklı araştırmacılar tarafından farklı biçimde tarif edilmiştir. Örneğin; Chang, Hall, Xu, Doan ve Wang, SNA’i “bireyler veya organizasyonlar, topluluklar, şirketler, ülkeler veya diğer kolektif gruplardan oluşan bir varlık grubunun birbirleriyle ilişkilerini analiz eden bir araştırma yöntemi” olarak tanımlamıştır [30]. SNA yaklaşımı, arkadaşlık ilişkileri ve dil becerisi analizi [31], psikoloji alanındaki çalışmalar [32], bir hayvan grubunun içine sonradan dâhil edilen bir üyeyi konu alan araştırmalar [33] gibi çok çeşitli ve farklı alanlarda kullanılmıştır. Patent dokümanları da bu konuda bir istisna değildir. Kim, Jun, Jang ve Park

yapay zekânın sürdürülebilir teknoloji analizini gerçekleştirmek üzere sosyal ağ modelleri kullanarak yapay zekâ ile ilgili en baskın 20 IPC alt sınıfı kodunu sunmuştur [34].

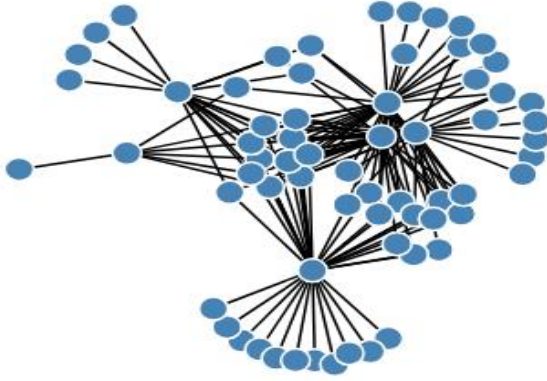
SNA diyagramlarında aktörler arası ilişki bağlantı olarak tanımlanan doğru parçaları veya oklarla sağlanır. Sosyal ağlar, düğüm adı verilen bileşenlerinin bağlantı biçimlerine göre çift yönlü bağlantılar, tek yönlü bağlantılar ve yönsüz bağlantılarla oluşturulabilir. Çift yönlü bağlantılarda düğümler arası ilişkili karşılıklıyken, tek yönlü bağlantılarda düğümler arası ilişki tek taraflıdır. Çift yönlü bağlantı içeren sosyal ağlar için popüler bir örnek Facebook arkadaşlığıdır. Facebook'ta bir X kullanıcısı bir Y kullanıcısı ile arkadaşsa Y kullanıcısının da X kullanıcısıyla arkadaş olduğu bilinir. Dolayısıyla Facebook arkadaşlıkları çift yönlü ve karşılıklıdır. Tek yönlü bağlantıya sahip ağlar için ise Instagram örneği sıklıkla verilmektedir. Facebook arkadaşlıklarının aksine Instagram uygulamasında bir X kullanıcısı Y kullanıcısının takipçisi iken Y kullanıcısının da X kullanıcısını takip ettiği doğrudan söylenemez. Y kullanıcısının X kullanıcısını takip etmediği durumda kullanıcılar arası ilişki tek yönlü bağlantı olarak tarif edilir. Yönsüz bağlantılar ise Şekil 3.2'de bir örneği görülebileceği gibi düğümlerin birbiri ile ilişkili olduğu bilgisini verirken bu ilişkinin doğrultusu ile ilgilenmez.

SNA ile oluşturulan diyagramlar temel olarak aşağıda kısa açıklamaları verilen terimleri içerir ve diyagramlar bu terimlere atanan değerler ve etiketler gözetilerek oluşturulur:

Kaynak: İki kullanıcı, birey, grup, varlık vb. arasındaki ilişkide bağlantının başlatıldığı düğüm olarak tanımlanır.

Hedef: Kaynaktan gelen bağlantıyı alan düğüm olarak tanımlanır.

Bağlantı: Kaynak-Hedef arasındaki ilişkiyi göstermek için kullanılan terimdir.



Şekil 3.2: Yönsüz bir SNA diyagramı örneği

SNA'in konu alındığı çalışmalarda düğümlerin merkeziyetleri ve kuvvetleri sosyal ağ diyagramlarının anlamlandırılmasında önemli rol oynamıştır. Dolayısıyla, özellikle temel kavramların okuyucu tarafından anlaşılması, diyagramların isabetli bir değerlendirilmesini yapmak için gereklidir. SNA diyagramlarında yer alan düğümler için farklı merkeziyet ve kuvvet ölçütleri bulunmaktadır. Adı geçen ölçütler matematiksel formülleri ve ifadeleri ile beraber takip eden alt başlıklar altında açıklanmıştır.

Derece merkeziyet (Degree centrality)

Derece merkeziyet SNA diyagramlarındaki en temel ölçüt olarak değerlendirilmiştir.

$G = \langle V, E \rangle$ bağıntısında G bir sosyal ağ, V düğüm dizisi, E bağlantı dizisi olmak üzere bir düğümün derece merkeziyeti şu şekilde hesaplanabilir [35]:

$$CD(n_i) = \Phi(n_i) \quad (3.1)$$

Denklem 3.1'de Φ katsayısı n_i düğümünün diğer düğümlerle yaptığı bağlantı sayısını ifade eder. Örneğin bir n_i düğümü n_x, n_y, n_z, n_t düğümleri ile bağlantılı ve ağdaki diğer düğümler ile bir bağlantısı yoksa Φ ve dolayısıyla derece merkeziyet değeri 4 olarak alınır. Bir ağda derece merkeziyeti yüksek düğümler daha merkezi düğümler olarak tanımlanır. Düşük derece merkeziyet skoruna sahip düğümler ise bu ölçüt özelinde SNA diyagramında perifer bölgelerde yer bulurlar.

Yakınlık merkeziyeti (Closeness centrality)

$G = \langle V, E \rangle$ bir SNA diyagramı olmak üzere, yakınlık merkeziyeti bir düğümün sosyal ağ içinde yer alan diğer düğümlere yakınlığını belirten bir merkeziyet ölçütüdür. Yakınlık

merkeziyet ölçümü, bir düğümün diğer tüm düğümlere olan en kısa mesafesinin ortalaması alınarak hesaplanır [36]:

$$CC(i) = \frac{N-1}{\sum_{j=1 \rightarrow N} (D_{ij})} \quad (3.2)$$

Denklem 3.2’de D_{ij} , i ve j düğümleri arasındaki en kısa mesafenin gösterimi için kullanılır. Eğer i ve j düğümleri birbirleri ile bağlantılı değilse D_{ij} sonsuza yakınsar. Dolayısıyla yakınlık merkeziyeti skoru görece yüksek bir düğüm, SNA diyagramında bulunan ve bir çoğunluğu oluşturan düğümlerle doğrudan bağlantılı veya onlara yalnızca birkaç düğüm uzaklıktadır [36]. Bu durum, sosyal ağda yüksek yakınlık merkeziyeti değerine sahip bir düğüm üstünden iletilen bilginin diğer düğümlere oldukça hızlı aktarılabilceği anlamına gelir [35].

Arasındalık merkeziyeti (Betweenness centrality)

$G = \langle V, E \rangle$ diyagramında birbiriyle doğrudan bağlantısı bulunmayan iki düğüm arasındaki etkileşim bu iki düğüm arasında yer alan diğer düğümlerle ilişkilidir [37]. Arasındalık merkeziyeti skoru, bir düğümden geçen en kısa mesafelerin sayısı ile hesaplanır [36].

$$BC(i) = \sum_{u,v \neq i} \frac{\sigma_{u,v}(i)}{\sigma_{u,v}} \quad (3.3)$$

Denklem 3.3’te $\sigma_{u,v}(i)$ terimi bir i düğümünden geçen en kısa yol sayısı için kullanılırken; $\sigma_{u,v}$ diyagramdaki herhangi bir düğüm çifti arasındaki en kısa yolların toplamını ifade etmek için kullanılmıştır [36]. Arasındalık merkeziyet değeri yüksek olan düğümler veya aktörler düğümler arasındaki etkileşimi kontrol ettikleri için SNA diyagramlarında önemli rol oynarlar [38].

Özvektör merkeziyeti (Eigenvector centrality)

Özvektör merkeziyet değeri, SNA diyagramındaki bir düğümün kendi merkeziyet değerinden daha yüksek skora sahip bir düğümle bağlantılı olmasının söz konusu düğümün özvektör merkeziyet değerini artırması ile sonuçlanacağını ifade eden bir ölçüttür [35]. Bu tanımdan anlaşılması gereken sosyal ağda bir düğümün farklı düğümlerle daha çok bağlantı sayısına sahip bir düğümle ilişkide olmasının kendi özvektör merkeziyet skorunu arttıracığıdır.

Düğüm kuvveti (Node strength)

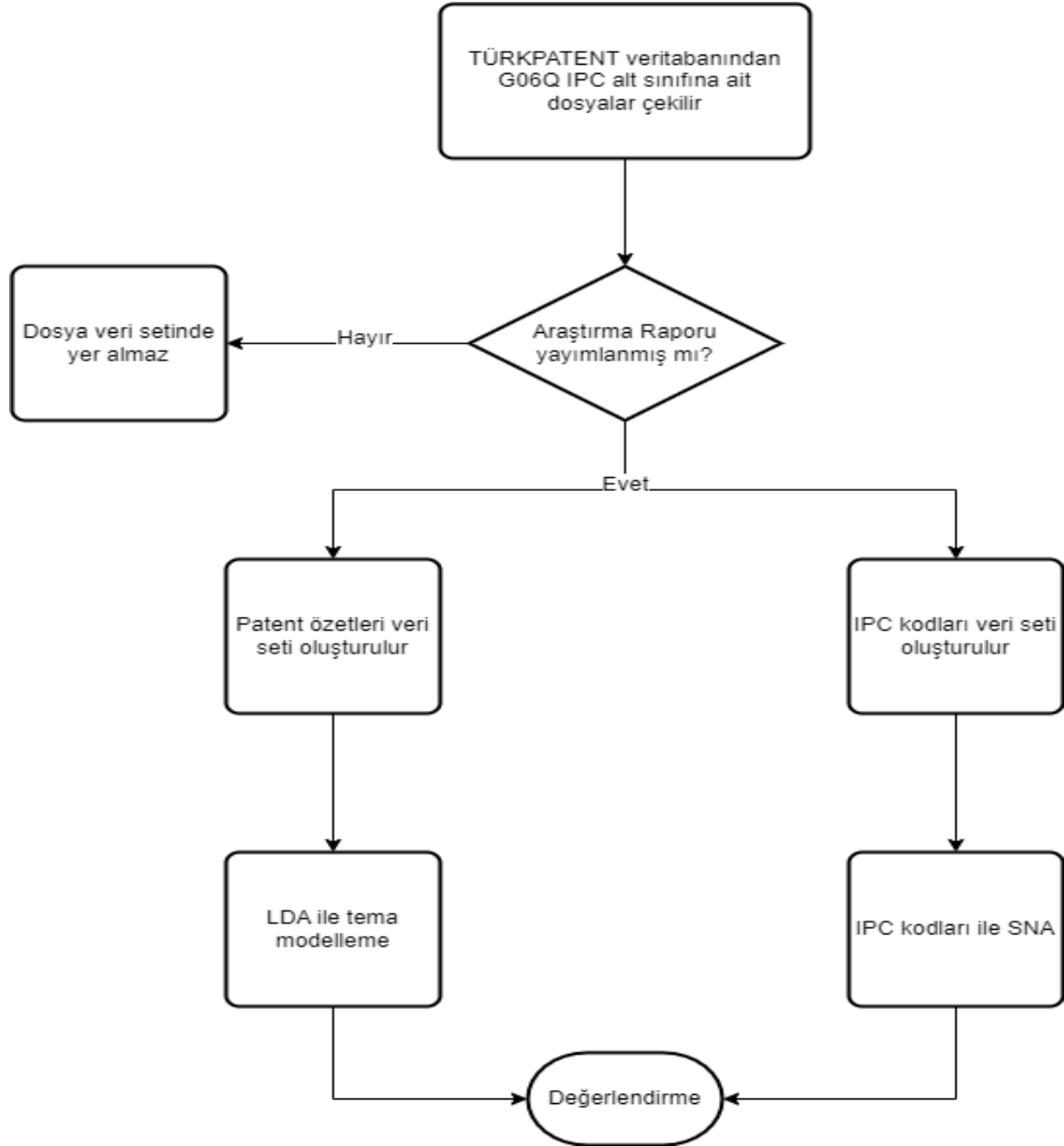
$G = \langle V, E \rangle$ diyagramında bir n_j düğümünün kuvveti şu şekilde hesaplanır:

$$S(n_j) = \lambda(n_j) \quad (3.4)$$

Denklem 3.4'te eşitliğin sağında yer alan λ terimi düğümün diğer düğümlerle olan toplam bağlantı sayısını ifade eder. Bağlantılara verilen ağırlıklar, n_j düğümünün ilişkili olduğu diğer herhangi bir düğümle olan bağlantı kuvvetini ifade etmek için kullanılır. Kuvvet, bağlantılara verilen katsayılarla veya SNA diyagram görsellerinde bağlantı genişliği ile ifade edilebilmektedir. Düğüm kuvveti veya diğer merkezîyet skorları gibi ilave parametrelerin kullanıldığı diyagramlar 'ağırlıklı diyagram' olarak adlandırılır.

4. METODOLOJİ

Çalışmada yer bulan temel süreçler aşağıdaki akış şemasında sunulmuştur:



Şekil 4.1: Çalışma için kurgulanan süreçler ve akış şeması

Şekil 4.1’de gösterilen temel süreçler için daha açık bir anlatım sunmak amacıyla çalışma yöntemi Veri Çekme, Ayırıştırma ve Düzenleme; Keşifsel Veri Analizi; Metin Ön İşleme; LDA ile Tema Modelleme ve IPC Ağ Analizi alt başlıklarında incelenmiştir. Tüm alt başlıklarda yapılan incelemelerde istatistiksel programlama dili R kullanılmıştır [39].

4.1. Veri Çekme, Ayırıştırma ve Düzenleme

Çalışmaya TÜRKPATENT veritabanında iş yöntemi patentleri için ayrılmış IPC alt sınıfı olan G06Q IPC kodlu dosyaların çekilmesi ile başlanmıştır. Elde edilen patent dosyaları aşağıdaki öncüller gözetilerek ayırıştırılmıştır:

- (i) Patent başvuru tarihi 2017 yılı ve sonrası olmalıdır.
- (ii) Patent araştırma raporu 2022 yılı öncesinde bültende yayımlanmış olmalıdır.

Öncül (i) hem iş yöntemi patentlerindeki son dönem eğilimleri bulmak ve artık popülerliğini yitirmiş alanları filtrelemek hem de dosyaların Ocak 2017 itibarıyla yürürlüğe giren 6769 sayılı SMK'ye tabi olmasını sağlamak için belirlenmiştir. Bu yolla, her başvurunun günümüzde de yürürlükte olan SMK maddelerine göre değerlendirilmesi durumu gözetilmiştir. Dolayısıyla, çekilen her dosya için TÜRKPATENT tarafından işletilen süreçlerin aynı olması ve başvurular arasında patent süreçleri bakımından hukuksal bir farkın olmaması sağlanmıştır.

Öncül (ii) daha katı bir IPC kod kontrolü yapılmasına olanak tanımıştır. İş yöntemi patentlerinin ve genel olarak CII'lerin teknik karakteri sağlamada potansiyel olarak diğer patent türlerine kıyasla daha fazla problem yaşadıkları bilinmektedir. Bunun önemli nedenlerinden biri, EPO'nun inceleme kılavuzunda yer alan ve 'iki engel' olarak bilinen yaklaşımın uygulanmasıdır [40]. Araştırma raporunun yayımlanmış olması ilgili patent başvurusundaki buluş konusunun teknik karaktere sahip olması ön koşulunu sağladığı anlamına gelmektedir. Bunun yanında, her bir dosya ile ilişkilendirilen IPC kodu patent süreçlerinde değişime uğrayabilmektedir. Öyle ki, bir başvuru için dosyayla ilişkilendirilen IPC kodu patent uzmanına göre değişebileceği gibi, bir patent ailesine sahip olan başvurular için farklı ülkelerin fikri mülkiyet ofislerinin farklı IPC kodlarını tercih ettiği bulgusu sunulmuştur [41]. Ancak araştırma raporunun yayımlanmış olması genellikle ilgili uzmanın dosya ile ilgili IPC atamasını tamamladığı anlamına gelir. Bu nedenle, bu aşamayı geçmiş olan bir dosya için IPC kodunun değiştirilmesi henüz araştırma raporu tamamlanmamış, dolayısıyla patent sürecinde yalnızca şeklen kontrolü gerçekleştirilmiş bir başvuruya kıyasla daha az olasıdır. Bu yaklaşım ile nihai veri setinde yer alan kayıtlarda veri güvenilirliğinin arttırılması hedeflenmiştir.

Yukarıda açıklanan öncüllere dayalı yapılan ayırıştırma sonucu çalışmada kullanılan veri seti oluşturulmuştur. Bu veri seti ile 897 adet tekil iş yöntemi patenti; başvuru numarası, özet, IPC kodları, araştırma raporu içeriği ve araştırma uzmanı değişkenleri ile beraber

listelenmiştir. Değişkenlerin veri setinde yer alma nedeni ve kullanım amacı kısa açıklamalarla aşağıda verilmiştir:

Başvuru Numarası: Her patent başvurusu sistem tarafından oluşturulan benzersiz bir başvuru numarasına sahiptir. Başvuru numarasının yer alması ile aynı başvurunun veri seti içinde birden fazla kez kayıt oluşturması kolaylıkla engellenebilir.

Özet: Patent özetleri; tarifname, istemler ve isteğe bağlı olarak sunulabilen teknik resimlerle beraber patent başvurularının temel unsurlarından biridir. Bu çalışmada LDA ile tema modelleme tekniği işlemsel yük ve metin uzunluğu standardı gözetilerek özetlere uygulanmıştır.

IPC kodları: Patent başvuruları buluşun ait olduğu teknik alanı dilden bağımsız olarak belirten en az bir IPC kodu ile ilişkilendirilir. Çalışmada yer bulan her bir dosya ile ilişkilendirilmiş tüm IPC kodları veri setinde yer almıştır. Dosya tabanlı IPC kodları üzerinden bir SNA diyagramı oluşturulmuştur.

Araştırma Raporu İçeriği: Araştırma raporları başvuru sahibine ulusal patent ofisleri tarafından bildiri yapılar ve buluş konusunun önceki tekniğe ait dokümanlarla bir mukayesesini içeren raporlardır. Bu çalışmada veri setinde yer alan dosyalar için TÜRKPATENT uzmanları tarafından gönderilen araştırma raporları incelenmiş ve içerikleri ‘olumlu’ veya ‘olumsuz’ olarak sınıflandırılmıştır. Bu sınıflandırmadaki amaç keşifsel veri analizine katkı sunmaktır. Olumlu içeriğe sahip raporlar buluşun tescillenebilirlik olasılığını yükseltirken olumsuz raporlara karşı başvuru sahibi itiraz ve karşı görüşler sunarak uzmanın buluşa bakış açısını olumlu yöne çevirmeye çalışır.

Araştırma Uzmanı: Her bir araştırma raporu görevlendirilen bir uzman tarafından hazırlanır. Bu değişken keşifsel veri analizi ile elde edilen bulgulara katkı sunması için veri setine dâhil edilmiştir.

4.2. Keşifsel Veri Analizi

İş yöntemleri patentleri diğer patentlere göre düşük tescil oranı nedeniyle kötü şöhrete sahip patent türleri olarak değerlendirilmiştir [42]. Bu durumun nedenlerinden biri olarak Alice Kararı’nı takip eden incelemelerde patent uzmanlarının bilgisayar tabanlı iş yöntemleri için buluş basamağı çitasını diğer patent türlerine oranla daha yukarıda konumlandırmaları gösterilebilir. Her ne kadar EPO buluş basamağı ile ilgili değerlendirmeleri nesnel hale

getirmeye çalışsa da [43] buluş basamağı kriterinin değerlendirilmesindeki belirsizlikler patent alanında yapılan çalışmalarda araştırma konusu olmuştur [44].

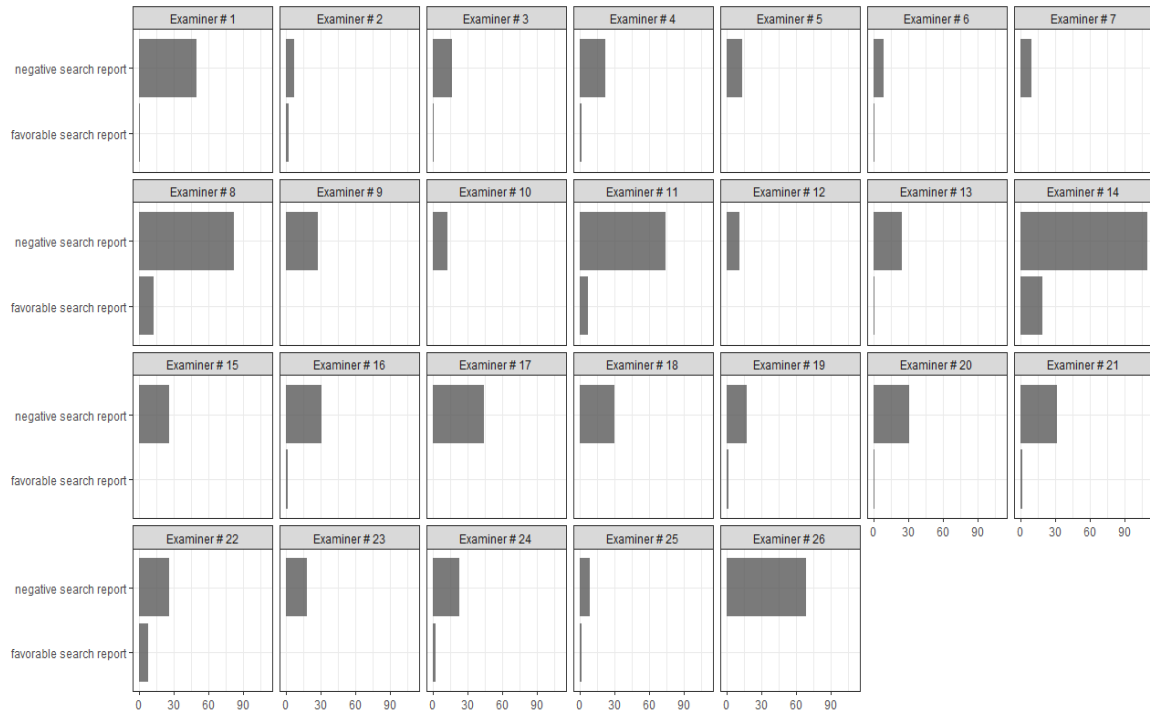
Çalışmada yer alan patent dosyalarının araştırma raporu içeriğine ilişkin sonuçlar Çizelge 4.1’de yer almıştır:

Çizelge 4.1: İş yöntemi patentleri için araştırma raporu içeriği

Araştırma Raporu İçeriği	Gözlem Sayısı	Yüzde
Olumlu	68	7,6
Olumsuz	829	92,4

Anlaşılabacağı üzere veri setinde yer alan iş yöntemi patentlerinde başvuru sahipleri büyük oranda olumsuz bir araştırma raporu bildirimi almıştır. Bu durum TÜRKPATENT nezdinde yapılan bilgisayar tabanlı iş yöntemi buluşlarının patentlenebilirliği hakkında önemli bir ön bilgi vermektedir.

Araştırma raporlarında bir diğer etken uzman karakteristiğidir. Uzman karakteristiklerinde patent uzmanı olarak çalışma süresi, incelenmiş patent sayısı, diğer patentler tarafından daha sonra atıf yapılmış patent incelemesi sayısı gibi çok farklı değişkenler söz konusudur [45]. Bu çalışmada patent uzmanlarının araştırma raporunu hazırladıkları iş yöntemi patentlerinin rapor içeriğine göre bir dağılımı hesaplanmış ve bulgular Şekil 4.2’de sunulmuştur. Söz konusu dağılım uzman karakteristiklerine dair herhangi bir çıkarım önermemekte veya uzmanların yaklaşımı hakkında bir saptama yapmamaktadır. Zira bu tip saptamalar eldeki verinin çok detaylı bir analizi sonucunda dahi okuyucuyu yanlış yönlendirme olasılığı bulunan tespitlerdir. Bu dağılımın verilmesindeki amaç keşifsel veri analizi sürecinde yalnızca ilave bir çıktı sağlamaktır.



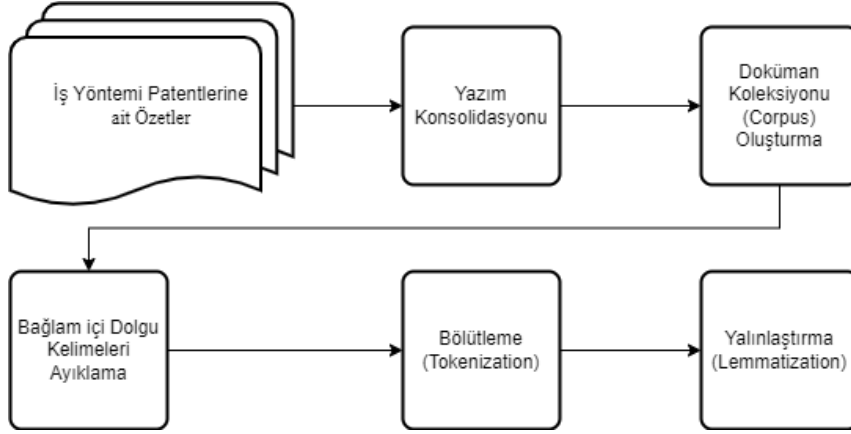
Şekil 4.2: Araştırma uzmanına göre rapor içeriği dağılımı

4.3. Metin Ön İşleme

Metin analitiği, NLP uygulamalarında metin içeriklerinin analizine yardımcı olan veri madenciliği tekniğinin bir uzantısı olarak tanımlanmıştır [46]. Metinler gerçek dünyada genelde yapılandırılmamış dokümanlar olarak kendilerini gösterir. Yapılandırılmamış veriler herhangi bir veri modelinde kullanılabilecek kadar işlem görmeyen verilerdir.

Patent özetleri de yapılandırılmamış metinlere bir örnektir. Tema modelleme yaklaşımını patent özetlerinde işlevsel bir şekilde kullanılabilmek için metin ön işleme adımlarının uygulanması önemli bir gerekliliktir. Bu çalışmanın metin analitiği ve ön işleme içeren kısmında R ortamında kullanılmak üzere ‘quanteda’ kütüphanesinden faydalanılmıştır [47].

Tanımsal olarak orijinal dokümana uygulanan her bir dönüştürme veya değişiklik işlemi metin ön işleme anlamına gelir. Metin ön işleme isabetli bir metin sınıflandırması yapmak üzere yapılandırılan bir modelin performansında önemli etkilere sahip olabilir [48]. Bu nedenle, iş yöntemi patentleri ile oluşturulan veri setinde yer alan özetler üzerinde modelleme aşamasına geçmeden önce Şekil 4.3’te verilen akıştaki basamaklar uygulanmış ve devam eden alt başlıklarda bu işlemlere ilişkin açıklamalara yer verilmiştir.



Şekil 4.3: Patent özetlerine uygulanan metin ön işleme adımları

Yazım konsolidasyonu

Patent dokümanlarında başvuru sahipleri veya vekil firmalar olabildiğince dikkatli bir yazım kontrolü gerçekleştirirler de teknik dokümanlarda teknoloji alanına göre oldukça sık kullanılan bazı kelimelerle ilgili bir fikir birliğinin oluşması güçtür. Bu nedenle, bu çalışmada kullanılan veri setinde yer alan özetlerde geçen bazı terimlerin farklı başvuru sahipleri tarafından farklı biçimde yazılmasına istinaden terim yazımlarında bir konsolidasyona gerek duyulmuştur. Çünkü ‘sözcük torbası’ yaklaşımı kullanılan modellerde terimlerin semantik anlamı aynı olsa dahi yazımlarındaki farklılık nedeniyle ulaşılan nihai modelde birbirinden farklı terimler olarak değerlendirilmektedir. Bu durumun önüne geçmek amacıyla belli terimler için yazım kontrolü yapılmış ve farklı yazım tercihleri ile özetlerde yer alan kelimeler tek bir yazım ile değiştirilmiştir. Böylece tema modelleme aşamasında aynı anlama sahip birden fazla terimin, oluşturulan temalar içinde yer almasının engellenmesi amaçlanmıştır.

Yazım konsolidasyonu uygulamasına ilişkin bir örnek Çizelge 4.2’de verilmiştir.

Çizelge 4.2: Patent özetlerinde terim yazımları ve konsolidasyon

Yazım	Konsolidasyon
online, on-line, çevrimiçi, çevrim içi	çevrimiçi
sunucu, server	sunucu
database, veri tabanı, veritabanı	veritabanı

Doküman koleksiyonu oluşturma

Çalışmada yazım konsolidasyonunun ardından modellemeye hazırlık aşamalarından biri olan doküman koleksiyonu oluşturma adımı gerçekleştirilmiştir. Doküman koleksiyonu (corpus) veri setinde yer alan ve metin alanı olarak sınıflandırılmış bölümün programlama diline tanıtılması ile oluşturulur. Belirtildiği üzere, çalışmadaki veri setinde metin alanı iş yöntemi patentlerine ait özetlerdir.

Bağlam içi dolgu kelimeleri ayıklama

Dolgu kelimeler (stop-words) bir doküman içinde yer alan ancak sözcük torbası yaklaşımı özelinde dokümandan çıkartıldığında anlam kaybının oluşmadığının değerlendirildiği sözcüklerdir. Dolgu kelimeler her bir dil için farklıdır ve bu tip kelimeler için çeşitli listeler bulunmaktadır. Örneğin, Türkçede ‘bir, o, dolayısıyla, kendi’ gibi sözcükler dolgu kelimeler kapsamında değerlendirilebilmektedir.

Ancak doküman koleksiyonu gözetilerek oluşturulan bağlam içi dolgu kelimelerin filtrelenmesi de oluşturulan modelde ortaya çıkabilecek gürültüyü engelleyebilir [49]. Bu yüzden, patent alanı özelinde özetlerin yazımında sık kullanılan ancak bilgi verici olmayan kısa bir bağlam içi dolgu kelimeler sözlüğü oluşturulmuştur. Bu kelimelerin ayıklanması ile model performansında iyileşme ve oluşturulan temaların ayırt edilebilirliğinin artırılması hedeflenmiştir. Patent özetler için doküman koleksiyonuna özel olarak hazırlanan bu sözlük, sıkça kullanılan ‘buluş, sistem, yöntem’ gibi hâlihazırda patentin doğası gereği zaten önceden bilinen terimleri ve ‘gerçekleştirmek, amaçlamaktadır’ gibi jenerik eylemleri içermektedir.

Bölütleme (Tokenization)

Bölütleme (*tokenization*), metni modellemede kullanılacak anlamlı birimlere ayırma işlemine verilen addır. Bilgisayar bilimlerinde bu ön işleme adımı sıklıkla bölütleme (*tokenization*), oluşturulan birimler de bölüt (*token*) olarak tarif edilir. Çalışmada token olarak sözcükler belirlenmiştir. Her bir sözcük tek başına bir bölüt ifade ettiği için özetlerdeki tüm sözcükler tek ağırlıklı bölüt (*unigram token*) olarak tanımlanır. Bir önceki alt başlıkta detayları verilen dolgu sözcük ayrıştırması bölütleme adımında Türkçe dili ile ilişkilendirilen dolgu sözcükler için yapılabilir. Bu adımda ayrıca, diğer alt adımlar olan noktalama işaretlerinin kaldırılması, sayı ve sembollerin kaldırılması, bölütlerin tamamen küçük harfe çevrilmesi işlemleri de uygulanmıştır.

Yalınlaştırma (Lemmatization)

Metinlerde tokenization sonrası gerçekleştirilen önemli ön işleme adımlarından bir tanesi de tokenlerin hangi biçimde DFM’de tutulacağıdır. Tokenler olduğu haliyle herhangi bir işlem yapılmadan bırakılabileceği gibi kökleştirme (stemming) veya yalınlaştırma (lemmatization) işlemleri kullanılarak normalize edilebilir. Kökleştirme işlemi keşifsel bir algoritma ile tokenin bittiği karakterden başlayarak sağdan sola bir budama işlemi yapar ve bu işlem sonucunda ortaya çıkan kök bir sözlük girdisi olabilecek anlamlı bir sözcük olabileceği gibi, kendi başına anlamsız bir terim de olabilir. Ancak sonuç üzerinden hareketle kökün aslında hangi sözcük olduğu yorumlanabilir. Yalınlaştırma ise dildeki sözcük morfolojisini temel alarak tokenin bir sözlük girdisi olacak biçimde en sade hâlini sunmayı amaçlayan bir ön işleme yöntemidir. Yalınlaştırma yaklaşımının İngilizce dışındaki bazı dillerde kökleştirmeye kıyasla daha iyi bir sözcük normalizasyonu sağlayabileceği ortaya konmuştur [50]. Ayrıca bazı dillerde, yalınlaştırma işlemi kökleştirme işlemiyle hariç bırakılan aynı semantik ve sentaktik özelliklere sahip tüm sözcük biçimlerine işaret etmektedir [51].

Yalınlaştırma ve kökleştirme operasyonlarının karşılaştırılması için patent dokümanlarından alınmış iki ifade aşağıda verilmiştir:

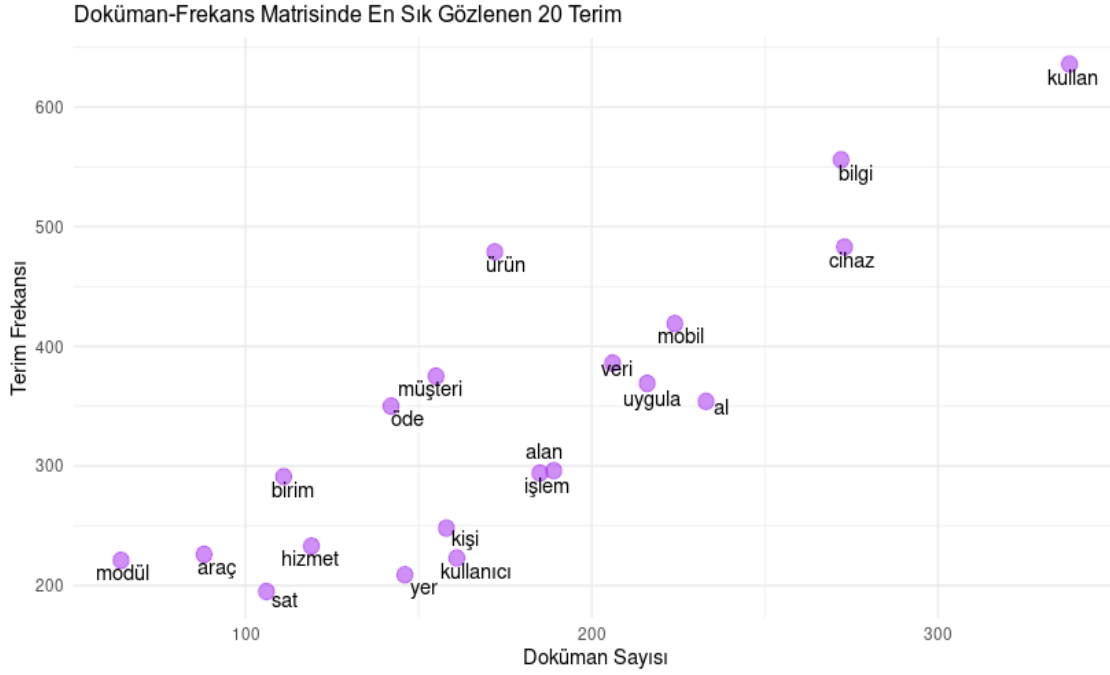
İfade 1: “...Kişiye göre özelleştirilmiş kampanyaları müşterilere sunan bir yöntem olup...”

İfade 2: “...Toplanan verilerin en az bir dış sunucuya aktarılması işlem adımını içerir...”

Çizelge 4.3: Yalınlaştırma ve kökleştirme örneği

Sözcük	Kökleştirme	Yalınlaştırma
sunan	sun	sunan
sunucuya	sun	sunucu

Çizelge 4.3’te iki sözcük semantik olarak farklı anlamlara sahip olmasına rağmen aynı köke sahiptir. Dolayısıyla kökleştirme işlemi sonucu elde edilen sonuca bakarak sözcüğün orijinini tahmin etmek güçtür. Alanda uzman bir kişi aynı köke ait farklı orijinleri tahmin edebilir ve model sonucu temalar altında gösterilen kök terim içinde hem sunma eyleminin hem de bir donanım bileşeni olan sunucunun olabileceğini bu örnek özelinde öngörebilir. Ancak bu durumda bile sözcük orijinlerinin frekansına dair bir tahmin yürütülemez.



Şekil 4.5: Terimlerin kullanıldığı doküman sayısı ve toplam frekansları

4.5. Örtülü Dirichlet Yerleşimi (LDA) ile Tema Modelleme

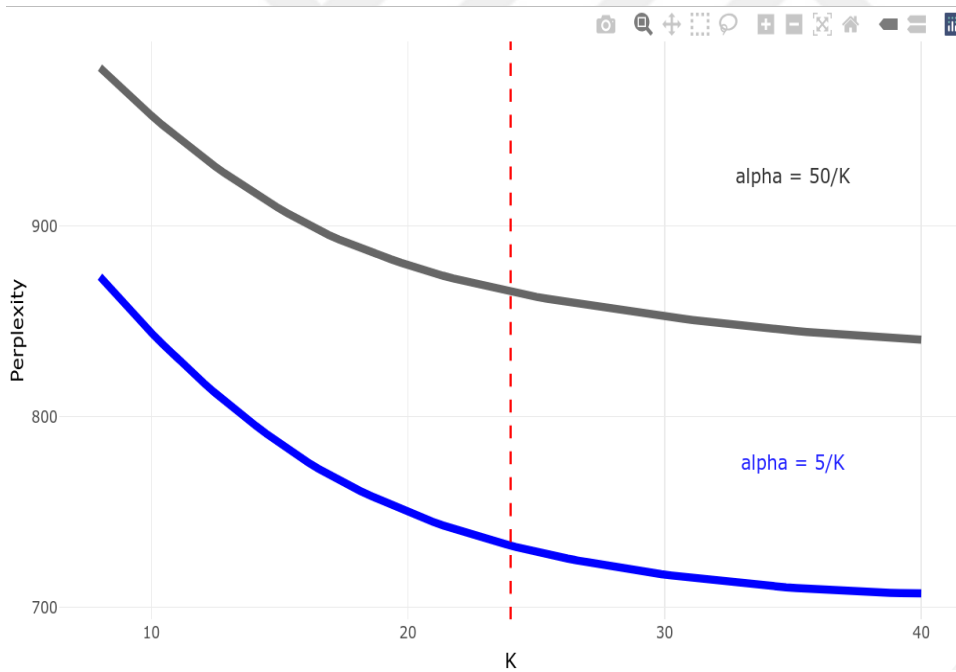
LDA yaklaşımının çoğu makine öğrenmesi tekniğinde de gözlenen belli dezavantajlara rağmen patent dokümanları için kullanışlı bir yöntem olduğu gösterilmiştir [54]. Çalışmada, oluşan DFM'in modellemeye uygun hâle getirilmesi sonrası iş yöntemi patentlerindeki örtülü temaları belirlemek için LDA yaklaşımı uygulanmıştır. Bu çalışmanın tema modelleme bölümünde R dili ile kullanılan topicmodels kütüphanesinden faydalanılmıştır [55].

LDA tekniğin uygulanırken kullanıcı örneklem yöntemini de açık etmelidir. Örneklem yöntemi kullanıcı tarafından karar verilen bir parametredir. LDA ile tema modellemede en popüler örneklem uygulamalarından biri Gibbs örneklemidir [56]. Gibbs örnekleme, “Dirichlet öncülü altında parametre kestirimine olanak tanıyan bir yöntemdir” [57]. Açıklanan özelliklerinden dolayı çalışmada bu örneklem yöntemi tercih edilmiştir.

Bölüm 3.1’de belirtilen parametrelerden tema sayısı (K) ve dirichlet parametresi (α) modelleme öncesi belirlenmiştir. Gözetimsiz makine öğrenme tekniklerinin çoğunda olduğu gibi, LDA’de de en iyi sonucu veren tema sayısını belirlemek ve oluşan temalara bir tanım atamak için analitik bir formül veya yöntem bulunmamaktadır. Bu nedenle hem tema sayısı hem de oluşan temalarla ilişkilendirilen etiketler kullanıcı tarafından belirlenmek durumundadır.

Tema sayısının (K) belirlenmesine yardımcı olmak için kullanılabilecek metriklerden biri karmaşıklık skorudur (perplexity). Karmaşıklık skorunun yüksek olması oluşturulan temalar arası ayırt ediciliğin düşük olduğunu ifade ederken, düşük karmaşıklık skoru model tarafından oluşturulan temalarda daha iyi bir ayırt ediciliğe işaret eder. Jacobi, Van Atteveldt ve Welbers gazete makaleleri gibi tematik metinler için daha düşük bir Dirichlet parametresi (α) belirlendiğinde daha düşük bir karmaşıklık skoru elde edilebileceğini belirtmiştir [58]. Patent özetlerinin uzun olmayan yapısı ve doğası gereği tematik metinler olduğu düşünüldüğünde, benzer bir fikrin çalışmaya da uyarlanabileceği fikri ortaya çıkmıştır.

Dirichlet parametresi (α), tema sayısı (K) ve karmaşıklık skoru arasındaki ilişkiyi ortaya koymak için 5-katlamalı çapraz geçerlilik (5-fold cross validation) yöntemi kullanılmıştır. Tema sayısı (K) için 4'ten 40'a ardışık aralığının ve Dirichlet parametresi (α) için $5/K$ ve $50/K$ değerlerinin atandığı parametreler ile yürütülen analiz sonucu bir grafik oluşturulmuştur.



Şekil 4.6: Dirichlet parametresi (α), tema sayısı (K) ve karmaşıklık skoru ilişkisi

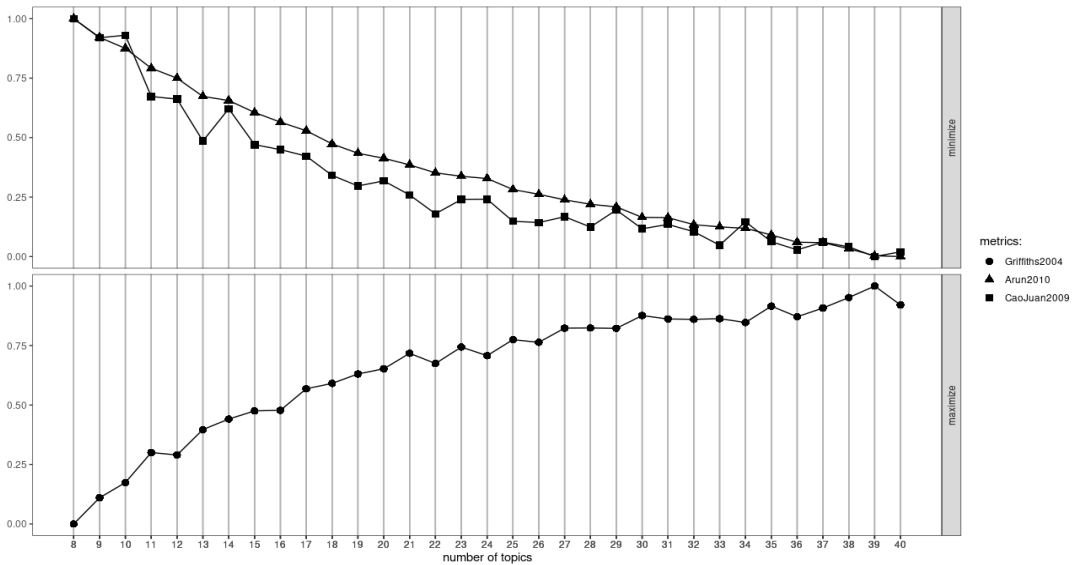
Şekil 4.6 ile $5/K$ 'lık Dirichlet parametresi değerinin (α) beklendiği gibi $50/K$ 'lık ön tanımlı değere göre daha düşük bir karmaşıklık skorunu sağladığı görülmüştür. Dolayısıyla $\alpha = 5/K$ ile gerçekleştirilen tema modellemenin daha ayırt edilebilir tema setleri vermesi beklenebilir. Fakat uç nokta olarak değerlendirilebilecek kadar küçük bir Dirichlet parametresi (α), LDA'nın karışımly üyelik yapısını bozarak yöntemi K-means gibi katı kümeleme yaklaşımına ait bir modelleme tekniğine benzemesine neden olabilir. Eğitim seti için ince ayar yapılmış

bir Dirichlet parametresi (α) ise görülmeyen veri kestiriminde model performansının çok düşük olmasına sebebiyet verebilir. Bu nedenle, çalışmada 5/K ve ön tanımlı değer olan 50/K haricinde farklı değerler için bir mukayese veya analiz yürütülmemiştir.

Bir sonraki adımda kullanıcı modelleme için tema sayısına (K) karar vermelidir. Karmaşıklık skoru grafiğinden iyi bir tema sayısı elde edebilmek için önerilen yöntemlerden biri grafik üzerinde bir ‘dirsek noktası’ tespit etmektir [59]. Bu ifade karmaşıklık skorunun gözle görülür biçimde kırıldığı ve eğrinin yataya dönmeye başladığı tema sayısını ifade eder. Şekil 4.6’da verilen grafikte böyle bir noktanın saptanmasının zor olduğu açıktır.

Karmaşıklık skoru tema sayısını (K) belirlemeye potansiyel olarak yardım eden bir ölçüt olmasına karşın bu ölçütün iyi çözümler üreteceğine dair bir güvence yoktur. Özellikle karmaşıklık skorunun aynı veri seti için bile seçilen ‘random seed’ değerine göre değişken sonuçlar üreten istikrarsız yapısı düşünüldüğünde [60], farklı yöntemlerin de değerlendirilmesi gerekliliği öne çıkmıştır.

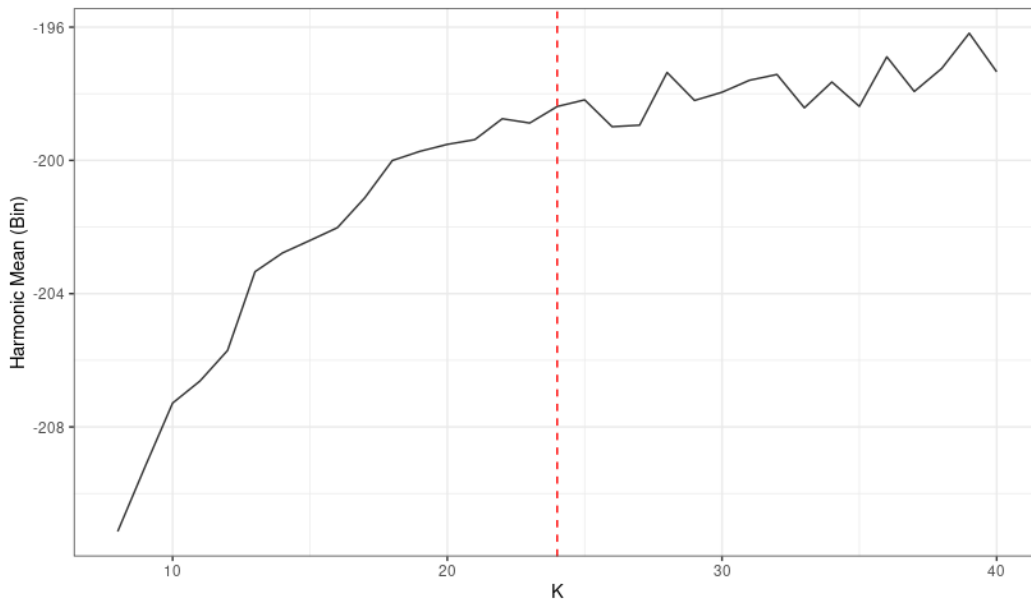
Tema sayısını (K) belirlemek için yararlanılacak bir başka yöntem, R dili için geliştirilmiş ve farklı metriklerle göre [56], ideal tema sayısını (K) bulmaya yardımcı olan ldatuning kütüphanesidir [61]. Bu yaklaşım uygulanarak adı geçen metriklerle tema sayısı (K) ilişkisini gösteren bir grafik çıktısı Şekil 4.7’de verilmiştir.



Şekil 4.7: Çeşitli parametreler ile ideal tema sayısı (K) ilişkisi

Arkasındaki literatüre rağmen, NLP uygulamaları içeren çoğu gözetimsiz makine öğrenmesi modelinde olduğu gibi, bu yöntem de kesin bir sonuç vermek için değil; kullanıcının tema sayısını (K) belirlemesine yardımcı olması için geliştirilmiştir.

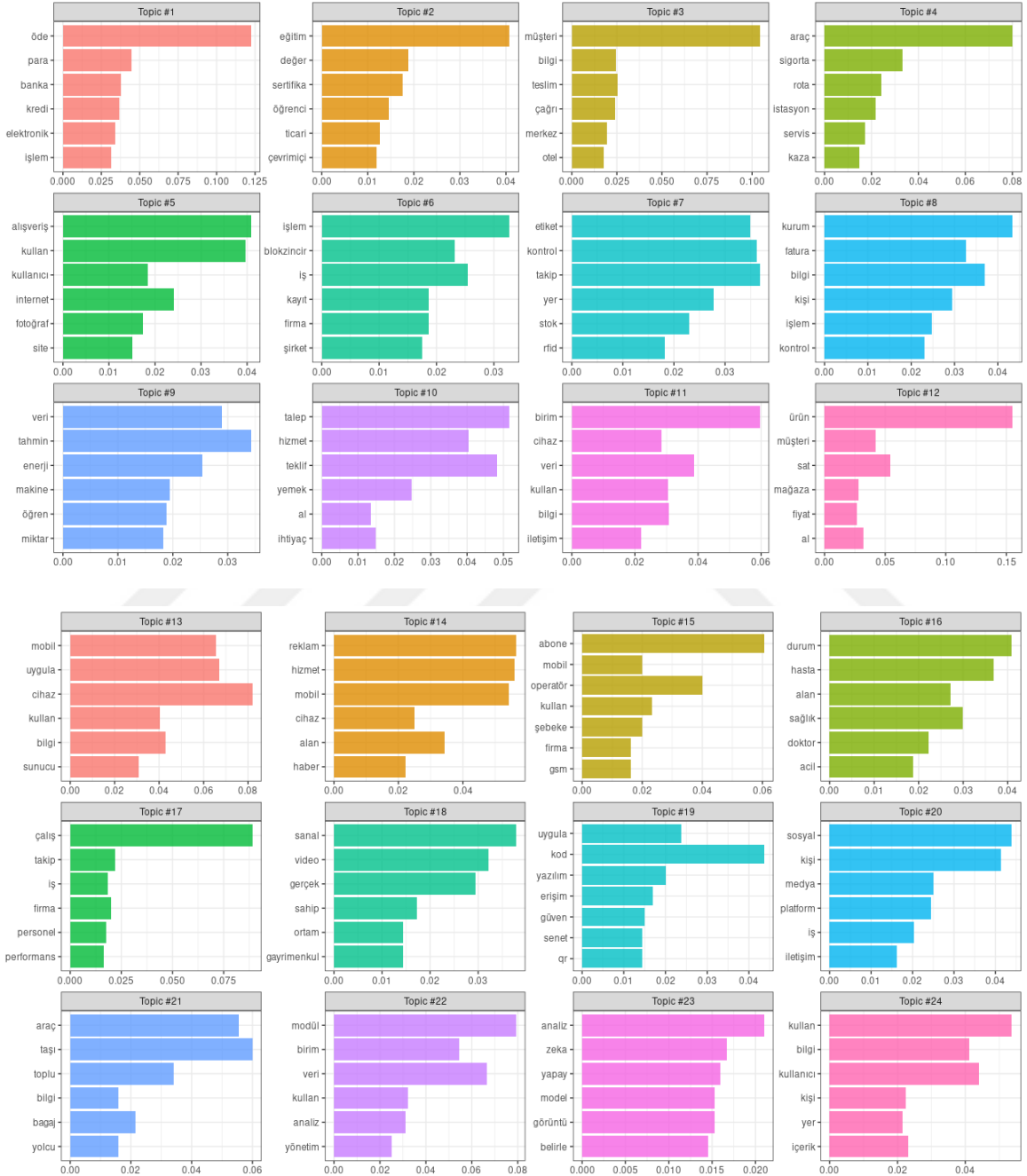
İdeal tema sayısını (K) belirlemek için kullanılabilecek bir başka yöntem ise harmonik istatistiksel ortalamayı temel alan yöntemdir [62]. Harmonik istatistiksel ortalama “veri bir doküman koleksiyonu içindeki sözcükler (w) ve model tema sayısı (K) olmak üzere $p(w|K)$ olasılığının hesaplanması” olarak tarif edilmiştir [56]. Bu yöntem, harmonik istatistiksel ortalama değerinin yüksek olduğu yerel maksimum değerlere karşılık gelen tema sayısının (K) potansiyel diğer tema sayılarına (K) göre daha iyi bir sonuç verdiği yaklaşımı üzerine kurgulanmıştır. Çalışmada bu yaklaşım da uygulanmış ve grafik Şekil 4.8’de verilmiştir:



Şekil 4.8: Harmonik istatistiksel ortalama ve tema sayısı (K) ilişkisi

Yukarıda açıklanan ve ideal tema sayısını (K) belirlemek için kullanılan yöntemler kullanıcıya belli seviyede fikir vermektedir. Ancak bu yöntemlerin hiçbiri ölçüt aldığı metrikler üzerinden ideal tema sayısının (K) bulunmasını garanti etmez. Bunun nedeni, LDA yaklaşımının gözetimsiz bir makine öğrenmesi yöntemi olması ve gözetimsiz makine öğrenmesi yöntemlerinde insan yorumuna danışılmasıdır. İnsan yorumu makine değerlendirmesinden farklı biçimde işler. Çünkü makineler veri içinde nüanslar yakalayıp insanlar için semantik olarak bir anlam ifade etmeyen tema setleri ile bir model oluşturabilir [63]. Bu çalışmada da insan yorumuna danışılarak tema setlerinin ayırt edilebilirliği ve yüzey geçerliliği mercek altına alınmıştır. Bu doğrultuda yukarıda açıklanan yöntemler de gözetilerek farklı tema sayıları (K) ile oluşturulan modellerdeki tema setlerinin birbirinden

ayrıt edilebilirliği değerlendirilmiştir. Bu değerlendirme sonucunda Şekil 4.6 ve 4.8’de dikey kesikli çizgi ile gösterilen kısma karşılık gelen tema sayısı ve daha düşük Dirichlet parametresi (α) kullanıldığında modelin daha ayrıt edilebilir bir tema seti oluşturduğu sonucuna ulaşılmıştır. ($K = 24$ ve $\alpha = 5/24$)



Şekil 4.9: LDA ile oluşturulan temalar ve üst terimler

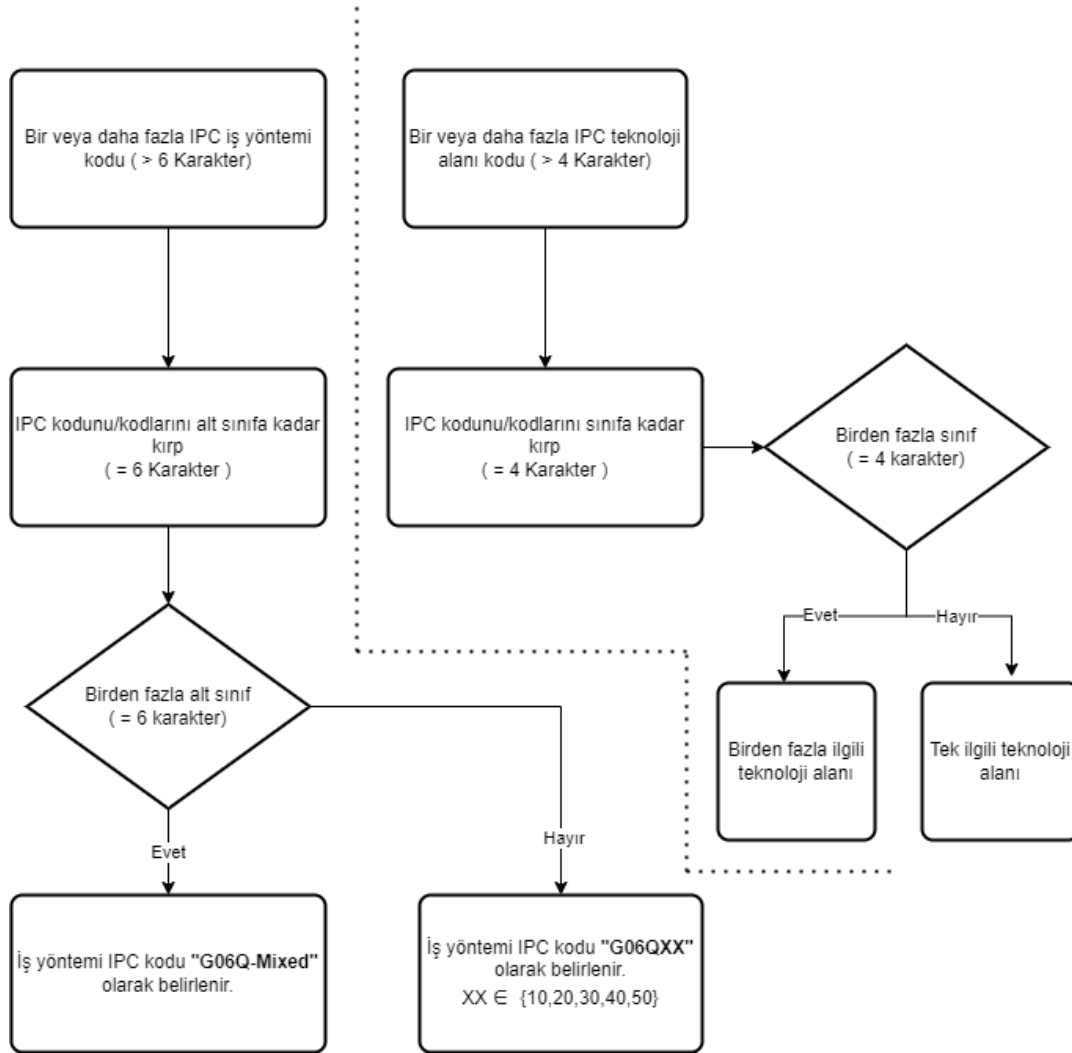
Şekil 4.9’da LDA ile üretilen 24 tema ve bu temalara ait en yüksek olasılık değerine sahip üst terimler gösterilmiştir. Temaların değerlendirilerek üretilen her bir tema için bir tanımlayıcı etiket atanmasına ilişkin açıklamalara Bulgular bölümünde yer verilmiştir.

4.6. IPC Ağ Analizi

IPC kodları, daha önce açıklandığı üzere patent uzmanları, buluşçular, araştırmacılar ve patent profesyonellerinin patent dosyalarını dilden bağımsız biçimde aramasını sağlamaktadır. Her bir patent dosyası en az bir IPC kodu ile ilişkilendirilir. Dosyaya atanan IPC kodu en üst seviyede bir IPC grubuna ait olmalıdır. Dolayısıyla, herhangi bir patent dokümanı ile ilişkilendirilmiş bir IPC kodu en azından bir IPC grubu içerir. Daha üst kümede yer alan bir alt sınıf veya sınıf ile IPC kodu ataması genel pratikler içinde uygulanmamaktadır. Bu nedenle, örneğin bir iş yöntemi patenti G06Q 50/02 koduna sahip olabilirken bir üst basamak olan G06Q50 ile ifade edilmez. Böylece dosyanın ait olduğu alan daha hassas ve spesifik biçimde tasnif edilebilir.

Patentler için otomatize edilmiş bir tasnif yöntemi öneren çeşitli çalışmalar gerçekleştirilmiştir [64,65]. Fakat bütün patent ofislerinde patentleri sınıflandırmada patent uzmanlarının görüşü öncelikli olmaya devam etmektedir. Çünkü patent dosyası tasnifini başarılı bir biçimde yapmak tarifname takımının derinlemesine bir incelemesi ile gerekli detayların yakalanması ve yorumlanması ile mümkündür.

Bu çalışmada IPC gruplarının çok çeşitli ve sayıca fazla olması gözetilerek dosyalar için incelenen kodlarda belli dönüşümler uygulanmıştır. Önerilen dönüşümün amacı patentleri daha üst bir sınıfta kümelemek ve temel alanları tespit ederek kompakt bir SNA diyagramı oluşturmak adına gerçekleştirilmiştir. Bu yaklaşımla iş yöntemi patentlerinin ait olduğu temel alanlar ve ana teknoloji alanlarının potansiyel ilişkisini ortaya koymak amaçlanmıştır. Aynı zamanda ağ analizi sonucu ortaya çıkan iş yöntemlerindeki ilgili teknoloji alanlarının yorumlanması LDA ile üretilen temalardan yapılan çıkarımlara da yardımcı olmuştur. Sınıflandırma kodları ile ilgili önerilen dönüşüm algoritması Şekil 4.10'da verilmiştir.



Şekil 4.10: Veri集中的 dosyalarda IPC tayini için uygulanan algoritma

İşletilen algoritmaya göre IPC kodunun nasıl belirlendiğine dair bir örnek Çizelge 4.4'te gösterilmiştir:

Çizelge 4.4: Algoritmaya göre IPC kodu belirleme

Kod Tanımı	Orijinal	Dönüşüm	Sonuç
IPC – 1 (iş yöntemi)	G06Q 10/10	G06Q10	G06Q-Mixed
IPC – 2 (iş yöntemi)	G06Q 30/06	G06Q30	
IPC – 3 (teknoloji alanı)	H04W 4/02	H04W	H04W

Veri setinde yer alan dosyalara uygulanmış IPC dönüşümleri ile elde edilen IPC veri seti 897 adet iş yöntemi patentinden 433 adedinin diğer teknoloji alanları ile de ilişkilendirildiğini göstermiştir. Bu dosyaların teknoloji alanları ile toplamda 590 bağlantı oluşturduğu saptanmıştır. Bu bağlantıların iş yöntemi IPC kodlarına göre dağılımı Çizelge 4.5’te sunulmuştur.

Çizelge 4.5: İş yöntemi IPC alt sınıfı ile teknoloji alanları ilişkisi

IPC	Teknoloji Alanları ile Toplam Bağlantı
G06Q-Mixed	88
G06Q10	175
G06Q20	70
G06Q30	140
G06Q40	8
G06Q50	109

Çizelge 4.5 ve Şekil 4.10’dan da çıkarımı yapılabileceği üzere uzmanlar herhangi bir iş yöntemi patentini tek IPC kodu olarak G06Q90 veya G06Q99 alt sınıfları ile bağdaştırmamıştır.

SNA uygulamalarında oluşturulacak diyagramın hangi biçimde çizildiğinin diyagramı yorumlayan insanlar tarafından yapılan çıkarımlarda önemli bir etkiye sahip olduğu tespit edilmiştir. Ayrıca, sosyal ağlarda en iyi yerleşim planının olmadığı; bu nedenle “araştırmacıların yerleşim düzenini tasarlarken ağın en önemli özelliklerini dikkate alması gerektiği” belirtilmiştir [66].

SNA diyagramları elle çizilebileceği gibi, birden fazla düğümün ve çok sayıda bağlantının izlendiği ağlarda bilgisayar programları tercih edilmektedir. SNA diyagram çizimlerinde tek düzeyli düğüm bağlantısı yaklaşımı yaygın bir uygulama olduğu ve bu çizim yönteminin en çok tercih edilen uygulamalarından birinin kuvvet güdümlü (force-directed) çizimler olduğuna literatürde yer verilmiştir [67]. Kuvvet güdümlü çizimlerin tipik kullanımına ilişkin “büyük diyagramlar için organik görünümlü, düz bağlantılara sahip bir yerleşim düzeni oluşturmak iken küçük ve orta boyutlu diyagramlar için genelde katımsal tekniklerin tercih edilir” açıklaması ve “kuvvet güdümlü tekniklerin amacı düğümler

arasında gerçekte oluşan ayırım ile ideal durumda oluşan ayırımı en aza indirmektedir” açıklamaları yapılmıştır [68].

Kuvvet güdümlü SNA diyagramı çizimine önemli bir örnek Fruchterman-Reingold algoritmasıdır. Bu çizim yönteminde temel olarak düğümlerin simetrik yerleşimi ve bağlantıların olabildiğince birbiri ile kesişmemesi ve bağlantı uzunluklarının eş oranlı olması amaçlanmıştır [69].

Bu çalışmada başka bir kuvvet güdümlü çizim tekniği olan ve görsel olarak da cazip bir yerleşim planı sağlayan Yifan-Hu algoritması kullanılmıştır [70]. Bu algoritma, SNA diyagramında düğüm yerleşimini Fruchterman-Reingold’un aksine simetrik olarak göstermemekle birlikte daha izole düğümlerin dışarıda konumlandırıldığı ve düğümlerin üst üste çakışmasının olabildiğince önlendiği bir yapı sağlamaktadır. Algoritmanın işletilebilmesi için IPC veri seti “düğüm seti” ve “bağlantı seti” olarak ikiye ayrılmış ve bu setler Gephi uygulamasına [71] tanıtılmıştır. Ayrıca, düğüm setine Kısım 3.2 altında açıklanan “düğüm kuvveti” parametresi ilave edilerek “ağırlıklı diyagram” oluşturulması sağlanmış ve nihai diyagram Şekil 4.11’de verilmiştir. Elde edilen SNA diyagramında düğüm yarıçapı ve bağlantı genişliği yüksek aktörler daha kuvvetli düğümler olarak değerlendirilmiştir.

Çizelge 4.6: En yüksek derece merkeziyete sahip teknoloji alanları

İlgili Teknoloji Alanı (Hedef) IPC Kodu	Derece Merkeziyet Skoru
G06F	6
G06N	6
H04L	6
H04M	6
H04W	6
G07F	5

Çizelge 4.7: En yüksek düğüm kuvvetine sahip teknoloji alanları

İlgili Teknoloji Alanı (Hedef) IPC Kodu	Düğüm Kuvveti Değeri
G06F	113
H04W	69
H04L	66
G06K	56
G06N	43
H04M	35

Çizelge 4.6’da SNA’de yer alan en yüksek derece merkeziyete sahip teknoloji alanları görüntülenmektedir. Yüksek derece merkeziyet skoru, söz konusu çizelgede yer alan teknoloji alanlarına ait IPC kodlarının daha fazla birbirinden farklı iş yöntemi IPC kodu ile bağlantı yaptığı anlamına gelmektedir. Çizelge 4.7’de ise en yüksek düğüm kuvvetine sahip teknoloji alanları büyükten küçüğe verilmiştir. Bu çizelgedeki IPC kodlarının nicel olarak ifade edilen kuvvetleri SNA diyagramında sahip oldukları bağlantı genişliği ve düğüm yarıçapı ile doğru orantılıdır.

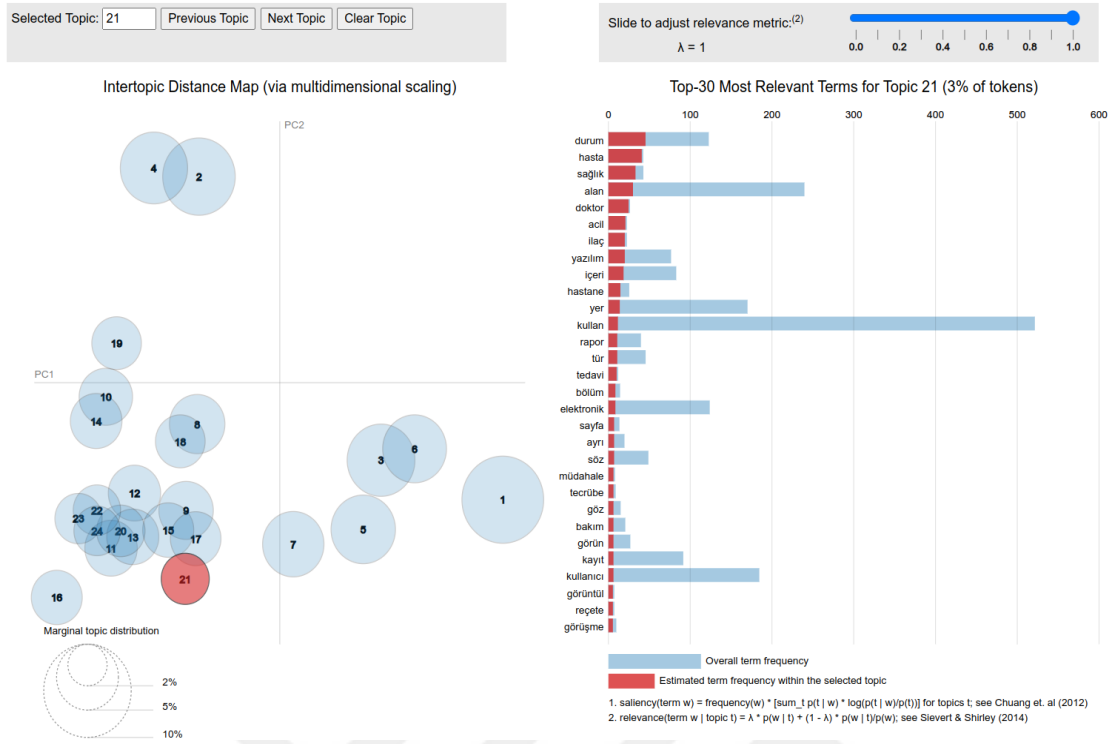


5. BULGULAR ve DEĞERLENDİRME

Çalışmanın bu bölümünde, iş yöntemi patentlerinin hem bibliyografik verilerinden biri olan IPC kodları üzerinde yapılan SNA hem de yapılandırılmamış veri sınıfına giren özetleri üzerinde gerçekleştirilen LDA ile tema modelleme çalışmalarından elde edilen bulgular sunulmuştur. Her iki çalışmanın kümülatif bir yorumlamasına Değerlendirme bölümünde yer verilmiştir.

LDA yaklaşımı kullanılarak oluşturulan modelin ürettiği 24 adet temanın üst terimleri Şekil 4.9'da incelenmişti. Karmaşıklık skoru, ilgili R kütüphanesi ve harmonik istatistiksel ortalama ölçütleri ile yapılan analizler tema sayısının artması ile daha ayırt edilebilir bir tema seti üretilebileceğini göstermiştir. Ancak LDA, yapısı gereği aynı zamanda bir boyut indirgeme yaklaşımıdır. Bu nedenle, insan yorumlamasının güçleşeceği sayıda tema seçimine gitmek özellikle çalışmaya konu olan küçük boyutlu veri setleri için tercih edilmemektedir.

Tema modelleme sonucu oluşan temaların hangi konuya ilişkin olduğuna araştırmacı karar vermektedir. LDA yaklaşımını ortaya koyan araştırmacılar üst terimlerin tema için bir etiket üretmede genellikle yeterli olduğu bilgisine yer vermiştir [21]. Ancak bu kesin bir çıkarım değildir ve uygulayıcılar tema etiketlerini oluşturabilmek için gerektiğinde üst terimler içinde yer almayan ve perifer terimler olarak adlandırılan öğelerden de yararlanabilirler. Bu doğrultuda, LDA ile üretilen temaların öncelikle üst terimleri incelenmiş ve bu terimlere göre her tema için bir etiket ataması yapılmıştır. En üst terimlerden bir etiket üretmenin güç olduğu kısımlarda ise LDavis [72] ile oluşturulan ve Şekil 5.1'de de bir çıktısı verilen tema-terim bağıntılarını sunan interaktif görsellerden yararlanılmıştır. Uygulama, model tarafından üretilen temaları bir temel bileşenler (principal components) kartezyen düzlemi üzerine yerleştirerek temalara ait en ilgili 30 terimin toplam terim frekansını gösterir. Ayrıca seçilen bir tema içindeki terimlerin tüm temalarda gözlenen toplam frekansının tema içi frekansına göre bir karşılaştırması da uygulama tarafından sağlanmaktadır.



Şekil 5.1: LDAvis ile örnek bir tema ve içerik gösterimi

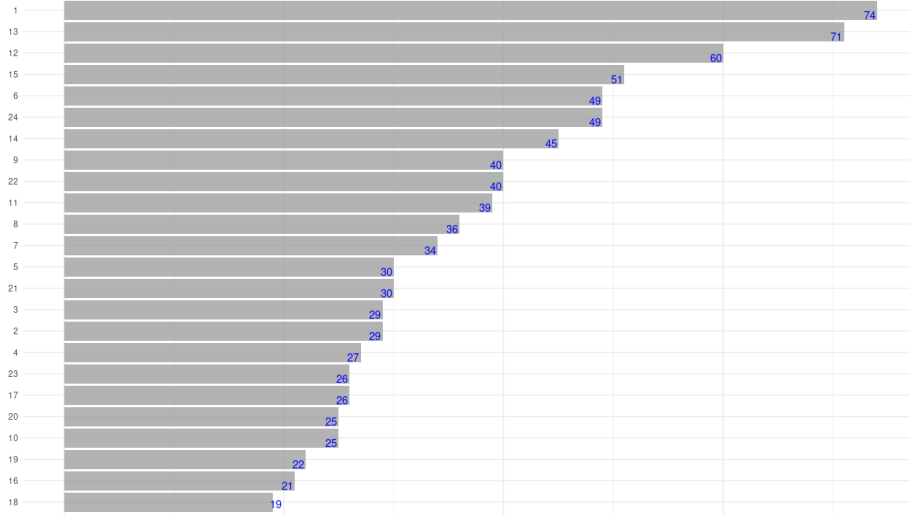
Tema etiketlerinde yorumlayıcı kişinin bir alan bilgisine sahip olması terimlerin daha derinlikli yorumlanabilmesine ve daha uygun bir tanım üretilmesine katkı sağlayabilir. Ancak, alanda uzmanlığı olsun veya olmasın, temalardaki terimler gözetilerek yapılan etiket atamaları öznelidir. Dolayısıyla aynı alan uzmanlığına sahip iki kişinin aynı temalar için farklı etiketler uygun görmeleri olasıdır. Ancak bu olasılığın, teknik dokümanlarda herhangi bir uzmanlığı bulunmayan iki farklı yorumlayıcıya göre daha düşük olması beklenmektedir. Bu nedenle, oluşturulan tema etiketlerinin makine tarafından üretilmediği ve açıklanan yöntemler kullanılarak oluşturulan öznel görüşler sonucu ortaya çıktığı yeniden vurgulanmıştır.

LDA yaklaşımı ile üretilen temalara ait üst terimler ile beraber LDAvis ile oluşturulan tema dağılımlarının bir değerlendirmesi yapılmıştır. Yapılan bu inceleme sonucunda her bir tema için uygun görülen tema etiketleri Çizelge 5.1’de verilmiştir.

Çizelge 5.1: LDA ile oluşturulan tema etiketleri

Tema No	Tema Etiketi	Tema No	Tema Etiketi
1	Banka İşlemleri	13	Mobil Uygulamalar
2	Eğitim Uygulamaları	14	Mobil Reklam
3	Müşteri Hizmetleri	15	Mobil Servis
4	Araç Takip	16	Sağlık Hizmetleri
5	E-alışveriş	17	Çalışan Verimliliği
6	Blokszincir Teknolojisi	18	Mekân ve Sanal Gerçeklik
7	Stok Yönetimi	19	Belge Doğrulama
8	Kurumsal Ödemeler	20	Sosyal Medya
9	Enerji Kullanımı	21	Toplu Taşıma
10	Öneri Sistemleri	22	Bilgi Gösterimi
11	Veri İşleme	23	Yapay Zekâ
12	Alışveriş	24	Kullanıcı Verisi

LDA'nın olasılığa dayanan bir yaklaşım olması ile çalışmadaki veri setinde yer alan 897 dokümanın her biri için dokümanın en yüksek olasılıkla ait olduğu tema belirlenerek bir doküman-tema dağılımı veya toplam doküman sayısı-tema dağılımı gösterilebilir [59]. Ancak karışıklı üyelik modellerde bu dağılımlar dokümanların tek bir temayla ilişkilendirilmiş gibi gösterilmesine neden olabileceği için bu dağılım üzerine yapılan yorumlarda dikkatli olunmalıdır.



Şekil 5.2: En yüksek olasılık değerine göre doküman sayısı – tema dağılımı

Şekil 5.2’de, örneğin, en fazla doküman ile ilişkilendirilen 1 numaralı temanın 74 ayrı doküman içerdiği görülmektedir. Ancak grafik verisi yorumlanırken bu dokümanların tamamının farklı temalara ait olma olasılığının da hesaplandığı ve dokümanların birden fazla temayı içerebileceği göz ardı edilmemelidir.

Çalışmanın bir diğer bileşeni olan SNA uygulamasında, IPC kodları ile yapılandırılan veri seti ile elde edilen ve düğüm bağlantısı yaklaşımıyla oluşturan ağ diyagramına ilişkin çıktılar Metodoloji bölümünde verilmiştir. Bununla birlikte, düğüm kuvveti değişkeninin ağırlık olarak eklendiği verilerde en yüksek düğüm kuvvet değerine sahip ilgili teknoloji alanlarına ait bir çizelge de Metodoloji bölümünde yer almıştır. En yüksek kuvvete sahip ilgili teknoloji alanlarının IPC kodlarına ilişkin tanımlar WIPO’nun belirlediği IPC kodu tanımlarına uygun biçimde ve düğüm kuvveti sırasına göre Çizelge 5.2’de sunulmuştur.

Çizelge 5.2: Hedef düğümlere ait IPC kod tanımları [12]

IPC Kodu	Tanım
G06F	Sayısal Veri İşleme
H04W	Kablosuz Haberleşme Ağları
H04L	Sayısal Bilginin İletimi, örn: Telgraf Haberleşmesi
G06K	Grafik Veri Okuma
G06N	Belirli İşlemsel Modeller Tabanlı Hesaplama Ayarlamaları
H04M	Telefon Haberleşmesi

Tema modelleme ve SNA ile elde edilen bulgularda iş yöntemlerinin potansiyel olarak diğer teknoloji alanları ile bir ilişkisi olduğu gösterilmiştir. SNA ile gözlenen en yüksek kuvvete sahip hedef düğümün G06F olmasının beklenen bir durum olduğu düşünülmüştür. Özellikle IPC'nin yedinci versiyonunda iş yöntemi patentlerinin G06F alt sınıfında yer alan bir grup olarak değerlendirildiği hatırlanırsa iş yöntemlerinin eski ana kümesi ile bağlantısını korumasının sürpriz olmadığı kolayca değerlendirilebilir.

G06F alt sınıfının ardından düğüm kuvveti sıralamasında H04 ana sınıfına ait H04W ve H04L alt sınıflarının geldiği görülmüştür. Kablosuz haberleşme teknolojisi ve sayısal bilginin iletimi gibi teknoloji alanları “haberleşme teknolojisi” içinde değerlendirilir.

Bu alt sınıflardan sonra gelen G06K ve G06N, tıpkı iş yöntemleri gibi “hesaplama, işleme ve sayma” sınıfı altında konumlanmaktadır. Grafik veri okuma alanını kapsayan G06K ve genelde makine öğrenmesi yöntemlerini işaret eden G06N sınıflarının da iş yöntemleri ile bağlantılı olmasının sürpriz olarak değerlendirilemeyeceği öngörülmüştür. Zira iş yöntemi patentlerinin “bilhassa idari, ticari, finansal, yönetsel, denetsel veya tahminsel amaçlar için uyarlanmış veri işleme sistemleri veya yöntemleri” olarak tanımlandığı düşünüldüğünde iş yöntemlerinin bu yakın alt sınıflara ait teknolojilerle teması da şaşırtıcı değildir. Ayrıca, makine öğrenmesi tekniklerini içeren G06N alt sınıfının günümüzde popüler bir teknoloji haline geldiği ve iş yöntemlerini de içine alan CII'ler kapsamında patent dokümanlarında sıkça yer almaya başlamış bir teknoloji olduğu düşünüldüğünde bu tekniklerin iş yöntemlerine uyarlanmasının oldukça makul ve beklenir olduğu değerlendirilmiştir.

Listedeki son alt sınıf olan H04M ise telefon haberleşmesi olarak tanımlanan ve yine haberleşme teknikleri altında konumlanan bir alt sınıf olarak kendisini göstermiştir. Nihai olarak G06 sınıfı altında konumlanan teknoloji alanlarının iş yöntemleri ile bağlantısı için geçerli ön nedenler olduğu görüşüne karşın H04 sınıfı altında konumlanan teknoloji alanlarının (H04W, H04L, H04M) iş yöntemleri ile ilişkisi ilave bir incelemenin yapılması gerekliliğini ortaya koymuştur. Bu doğrultuda, çıkarıma etki edebilecek faktörler düşünülmüş ve veri setinde yer alan iş yöntemleri arasında en çok kayda sahip başvuru sahipleri listelenmiştir.

Çizelge 5.3: Başvuru sayısına göre başvuru sahipleri

Başvuru Sahibi	Başvuru Sayısı
Turkcell Teknoloji Araştırma ve Geliştirme A.Ş.	336
Türk Telekomünikasyon A.Ş.	34
Vodafone Teknoloji Hizmetleri A.Ş.	25

Çizelge 5.3'te en çok başvuruya sahip olduğu görülen üç şirketin tamamı global düzeyde tanınan mobil servis ve internet sağlayıcılarıdır. Belirtilen üç şirket veri setinde yer alan 897 gözlemin yaklaşık yüzde 45'ini oluşturmuştur. Bu istatistiğin ötesinde, en yüksek başvuru sayısına sahip şirket olan Turkcell Teknoloji Araştırma ve Geliştirme A.Ş.'nin veri setinde yer alan tüm gözlemlerin yüzde 37'sini tek başına sağladığı görülmüştür.

IPC kodları ile oluşturulan SNA diyagramı ile gözlemlenen H04 sınıfına ait alt sınıfların en yüksek düğüm kuvvetine sahip teknoloji alanlarından üçünü oluşturmasının önemli bir nedeninin bu dağılım olduğu düşünülmüştür. Bununla beraber, LDA ile gerçekleştirilen modelleme sonucu üretilen temaların üçünün doğrudan doğruya “mobil teknolojiler” ile ilişkilendirmesinin de bu nedensellik varsayımını güçlendirdiği öngörülmüştür. Ayrıca, tema terimlerinde görece olarak sık gözlenen “mobil ve “cihaz” gibi sözcükler ve doğrudan mobil teknolojilerle bağdaştırılmasa dahi “müşteri hizmetleri”, “sosyal medya” gibi alakalı temalar göz önüne alındığında baskın başvuru sahiplerinin mevcut çıkarımlarda önemli etkiye sahip olduğu yorumlanmıştır.

Yukarıdaki açıklamalar değerlendirildiğinde, dominant başvuru sahiplerinin kendi faaliyet alanlarına ait teknik özellikleri geliştirdikleri iş yöntemlerine önemli ölçüde entegre ettikleri

sonucuna ulařılmıştır. LDA ile tema modelleme ve IPC kodu ađ analizinden alınan verilerin toplamsal bir deđerlendirmesi ve yorumlanması ile açıklanan görüř ortaya çıkmıştır.





6. SONUÇ

Bu çalışmada LDA ile tema modelleme yöntemi ve dosya tabanlı IPC kodu ağ analizi tekniklerinin bir kombinasyonu kullanılmış ve iş yöntemi patentlerindeki eğilimler ile iş yöntemlerinin diğer teknoloji alanları ile olan potansiyel ilişkisi ortaya konmaya çalışılmıştır. Ayrıca LDA yaklaşımının patent dokümanlarında eğilim gözlemi için kullanışlı olabileceği aktarılmıştır.

LDA çalışma içinde tema modelleme ve üretilen temalarla ilişkilendirilen sözcükleri sunarken, dosya tabanlı IPC kodu ağ analizi ile LDA ile üretilen temaların, dolayısıyla bilgisayar tabanlı iş yöntemleri ile ilişkili teknoloji alanlarının tutarlılığı kontrol edilmeye çalışılmıştır. Bu iki yaklaşım ile elde edilen bulgularla bilgisayar tabanlı iş yöntemi patentlerinin teknoloji alanları ile ilişkisini ortaya konmuştur.

Veri setinde yer alacak dosyalara uygulanan seçici algoritma ile veri güvenilirliğinin arttırılarak dinamik IPC kodu değişimi veya patentlenemeyecek buluş konularını içeren iş yöntemlerinin filtrelenmesi sağlanmıştır. Bu yaklaşımla, verilerin ancak belli kriterler altında çalışmaya dâhil olması sağlanmış, kriterleri sağlamayan dosyaların çalışma dışında bırakılarak elde edilen model çıktılarında potansiyel bir gürültü oluşturması veya yanlış yönlendirici çıktılar oluşturulması engellenmiştir.

Literatür araştırması boyunca dosya tabanlı IPC kodları ile gerçekleştirilen ağ analizlerine rastlanmamıştır. Bu yönüyle çalışmanın özgün ve orijinal çıktılar sağladığı düşünülmüştür. Patent dosyalarının bibliyografik verileri arasında yer alan IPC kodlarının önerilen algoritmayla her bir dosya özelinde düzenlenmesi ile tekil dosyaların hangi alanlarla ilişkili olduğu saptanabilmiştir. Veri setinde yer alan IPC kodları patent uzmanları tarafından belirlendiği için, çalışmanın bu bölümü uzmanların dosyalar için uygun gördükleri IPC kodları ile yapılmış, atanmış olan IPC kodları için ilave bir yorum getirilmemiştir.

İleride gerçekleştirilecek olan çalışmalar zaman penceresinin genişletilmesi yoluyla veri seti boyutunu arttırabilir, ilave kriterler belirleyerek farklı seçici algoritmaların oluşturulması vasıtasıyla veri güvenilirliğini daha da arttırabilir veya benzer veri setleri kullanılarak bu çalışmada ulaşılan sonuçları güncelleyebilir. Ayrıca, tema modelleme bölümünde açıklandığı üzere LDA dışında kalan farklı yaklaşımlar uygulanabilir ve bu yaklaşımlardan hangisinin veya hangilerinin patent verilerini değerlendirmede daha kullanışlı olabileceğine ilişkin görüşleri ortaya çıkarabilir. Bununla birlikte, farklı tema modelleme yaklaşımlarının

uygulanması sonucu LDA dışındaki diğer tema modelleme teknikleri ile LDA arasında bir model performansı mukayesesi yapılabilir.

Bu çalışmada açık kaynaklarda yer alan veriler kullanılarak gerçekleştirilmiş olup, analizlerin üçüncü kişiler tarafından yeniden yürütülebileceği ve ulaşılan sonuçların geçerliliğinin denetlenebileceği bir çalışma ortaya konmuştur.



KAYNAKLAR

- [1] Hammer, M. and Champy, J. (2009) *Reengineering the corporation: Manifesto for business revolution*. Collins Business Essentials, New York, 17-20.
- [2] Internet: Business Methods | USPTO. Available: <https://www.uspto.gov/patents/basics/types-patent-applications/utility-patent/patent-business> Erişim: 07/10/2022
- [3] Spulber, D.F. (2011) Should Business Method Inventions be Patentable? *Journal of Legal Analysis*. 3 (1), 265–340.
- [4] Internet: Alice Corp. v. CLS Bank Int'l :: 573 U.S. 208 (2014) :: Justia US Supreme Court Center. Available: <https://supreme.justia.com/cases/federal/us/573/208/> Erişim: 09/10/2022
- [5] Internet: Business Methods | USPTO. Available: <https://www.uspto.gov/sites/default/files/documents/Notices-of-Allowance-in-Business-Methods-2017-v2.pdf> Erişim: 07/10/2022
- [6] Machek, J., Triest, J.(2007) Computer related inventions, in particular business methods: Examination at the European Patent office. *2007 IEEE International Technology Management Conference (ICE)*, IEEE, pp. 1-8. Sophia Antipolis.
- [7] Internet: Espacenet – search results. Available: <https://worldwide.espacenet.com/patent/search/family/025457073/publication/US5960411A?q=US5960411A> Erişim: 11/10/2022
- [8] Internet: Espacenet – search results. Available: <https://worldwide.espacenet.com/patent/search/family/024842611/publication/US5794207A?q=US5794207> Erişim: 14/10/2022

- [9] Ernst, H. (2003) Patent information for strategic technology management. *World Patent Information*. 25 (3), 233–242.
- [10] Allison, J.R. and Tiller, E.H. (2003) The Business Method Patent Myth. *Berkeley Technology Law Journal*. 18. doi: 10.2139/ssrn.421980.
- [11] Martin, N.L. and Mykytyn, P.P. (2009) Evaluating the Financial Performance of Business Method Patent Owners. *Information Systems Management*. 26 (3), 285–301.
- [12] Internet: WIPO IPC Publication. *IPC Scheme*. Available: <https://ipcpub.wipo.int> Erişim: 05/09/2022
- [13] Makarov, M. (2006) The eighth edition of the IPC. *World Patent Information*. 28 (2), 122–126.
- [14] Makarov, M. (2000) The seventh edition of the IPC. *World Patent Information*. 22(1), 53–58.
- [15] Dereli, T. and Durmuşoğlu, A. (2009) Classifying technology patents to identify trends: Applying a fuzzy-based clustering approach in the Turkish textile industry. *Technology in Society*. 31 (3), 263–272.
- [16] Jain, R., Tripathi, M., Agarwal, V., and Murthy, J. (2020) Patent data analytics for technology benchmarking: R-based implementation. *World Patent Information*. 60 (October 2019), 101952.
- [17] Jiang, J., Goel, R.K., and Zhang, X. (2019) Knowledge flows from business method software patents: influence of firms' global social networks. *Journal of Technology Transfer*. 44 (4), 1070–1096.
- [18] No, H.J., An, Y., and Park, Y. (2015) A structured approach to explore knowledge flows through technology-based business

methods by integrating patent citation analysis and text mining. *Technological Forecasting and Social Change*. 97 181–192.

- [19] Moehrle, M.G., Wustmans, M., and Gerken, J.M. (2018) How business methods accompany technological innovations – a case study using semantic patent analysis and a novel informetric measure. *R and D Management*. 48 (3), 331–342.
- [20] Lee, W.S. and Sohn, S.Y. (2017) Identifying emerging trends of financial business method patents. *Sustainability (Switzerland)*. 9 (9). 1670
- [21] Chang, S. bin, Lai, K.K., and Chang, S.M. (2009) Exploring technology diffusion and classification of business methods: Using the patent citation network. *Technological Forecasting and Social Change*. 76 (1), 107–117.
- [22] Aggarwal, C.C. and Zhai, C.X. (2013) Mining text data. *Mining Text Data*. 9781461432234 1–522.
- [23] Blei, D.M., Ng, A.Y., and Jordan, M.T. (2002) Latent dirichlet allocation. *Advances in Neural Information Processing Systems*. 3 993–1022.
- [24] Blei, D.M. (2013) Mixed Membership Models. Columbia University. Columbia. Available: <http://www.cs.columbia.edu/~blei/fogm/2015F/notes/mixed-membership.pdf> Erişim: 18/10/2022
- [25] Bastani, K., Namavari, H., and Shaffer, J. (2019) Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. *Expert Systems with Applications*. 127 256–271.
- [26] Tang, J., Meng, Z., Nguyen, X., Edu, X., Mei, Q., and Zhang, M. (2014) Understanding the Limiting Factors of Topic Modeling via Posterior Contraction Analysis. 190–198.

- [27] Pietsch, A.-S. and Lessmann, S. (2018) Topic modeling for analyzing open-ended survey responses. *Journal of Business Analytics*. 1 (2), 93–116.
- [28] Tierney, G., Bail, C., and Volfovsky, A. (2021) Author Clustering and Topic Estimation for Short Texts. Available: <https://arxiv.org/abs/2106.09533> Eriřim: 12/09/2022
- [29] Titov, I. and Mcdonald, R. (2008) Modeling Online Reviews with Multi-grain Topic Models. Available: <https://arxiv.org/abs/0801.1063> Eriřim: 14/09/2022
- [30] Chang, V., Hall, K., Xu, Q.A., Doan, L.M.T., and Wang, Z. (2022) A social network analysis of two networks: Adolescent school network and Bitcoin trader network. *Decision Analytics Journal*. 3 100065.
- [31] Chow, J.C., Broda, M.D., Granger, K.L., Deering, B.T., and Dunn, K.T. (2021) Language skills and friendships in kindergarten classrooms: A social network analysis. *School Psychology*. 37(6). 488-500
- [32] Broda, M.D., Granger, K., Chow, J., and Ross, E. (2021) Using Social Network Analysis in Applied Psychological Research: A Tutorial. *Psychological Methods*. doi: 0.1037/met0000451
- [33] Díaz, S., Murray, L., Roberts, S.G.B., Rodway, P., Setchell, J.M., and Uk, G.A. (2020) Social Network Analysis of a Chimpanzee (Pan troglodytes) Group in Captivity Following the Integration of a New Adult Member. *International Journal of Primatology*. 41 683–700.
- [34] Kim, J., Jun, S., Jang, D., and Park, S. (2018) Sustainable technology analysis of artificial intelligence using Bayesian and social network models. *Sustainability (Switzerland)*. 10 (1). 115

- [35] Park, D., Kim, G. nam, and On, B.W. (2016) Understanding the network fundamentals of news sources associated with a specific topic. *Information Sciences*. 372, 32–52.
- [36] Ishfaq, U., Khan, H.U., and Iqbal, S. (2022) Identifying the influential nodes in complex social networks using centrality-based approach. *Journal of King Saud University - Computer and Information Sciences*. 34(10), 9376-9392
- [37] Landherr, D.A., Friedl, Dipl.-M.B., and Heidemann, J. (2010) A Critical Review of Centrality Measures in Social Networks The Authors. *Business & Information Systems Engineering*. 2. 371-385.
- [38] Lee, M.J., Choi, S., and Chung, C.W. (2016) Efficient algorithms for updating betweenness centrality in fully dynamic graphs. *Information Sciences*. 326 278–296.
- [39] Internet: R: The R Project for Statistical Computing. Available: <https://www.r-project.org/> Erişim: 19/09/2022
- [40] Internet: EPO, G-II, 2. Examination practice - Guidelines for Examination. Part G Chapter 2.2 Available: https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_ii_2.htm Erişim: 30/10/2022
- [41] Meguro, K. and Osabe, Y. (2019) Lost in Patent Classification. *World Patent Information*. 57 70–76.
- [42] Meurer, M.J. (2002) Business Method Patents and Patent Floods *Washington University Journal of Law & Policy*. 8. 309-343
- [43] Internet: EPO G-VII, 5. Problem-solution approach - Guidelines for Examination. Available: https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_vii_5.htm Erişim: 30/10/2022

- [44] Matelan, L. (2007) The Continuing Controversy over Business Methods Patents. *Fordham Intellectual Property, Media & Entertainment Law Journal*. 18. 189-223
- [45] Cockburn, I.M., Kortum, S., and Stern, S. (2002) Are All Patent Examiners Equal? The Impact of Examiner Characteristics on Patent Statistics and Litigation Outcomes. *Nber Working Paper Series*, Cambridge. 11-15
- [46] Moreno, A. and Redondo, T. (n.d.) Text Analytics: the convergence of Big Data and Artificial Intelligence. *International Journal of Interactive Multimedia and Artificial Intelligence*. 3 57-64
- [47] Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., et al. (2018) quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*. 3 (30), 774.
- [48] Uysal, A.K. and Gunal, S. (2014) The impact of preprocessing on text classification. *Information Processing and Management*. 50 (1), 104–112.
- [49] Schofield, A., Magnusson, M., and Mimno, D. (2017) Pulling Out the Stops: Rethinking Stopword Removal for Topic Models. *The Association for Computational Linguistics*. 2 432–436.
- [50] Korenius, T., Laurikkala, J., Järvelin, K., Juhola, M., and Jarvelin, K. (2004) Stemming and Lemmatization in the Clustering of Finnish Text Documents. *Proceedings of the thirteenth ACM international conference on Information and knowledge management CKIM '04*. 625-633
- [51] Zeroual, I. and Lakhouaja, A. (2017) Arabic information retrieval: Stemming or lemmatization? *Intelligent Systems and Computer Vision, ISCV 2017*, 1-6

- [52] Marcus, M.P., Santorini, B., and Marcinkiewicz, M.A. (1993) Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*. 19 (2), 313–330.
- [53] Yıldız, O.T., Solak, E., Çandır, Ş., Ehsani, R., and Görgün, O. (2016) Constructing a turkish constituency parse treeBank. *Lecture Notes in Electrical Engineering*. 363 (March 2018), 339–347.
- [54] Pölzlbauer, G. and Auer, E. (2021) Applied patent mining with topic models and meta-data: A comprehensive case study. *World Patent Information*. 67 (October 2019), 1–20.
- [55] Grün, B. and Hornik, K. (2011) Topicmodels: An r package for fitting topic models. *Journal of Statistical Software*. 40 (13), 1–30.
- [56] Griffiths, T.L. and Steyvers, M. (2004) Finding scientific topics. *Proceedings of the National Academy of Sciences*. 101 (suppl_1), 5228–5235.
- [57] Rosen-Zvi, M., Griffiths, T., Steyvers, M., and Smyth, P. (2012) The Author-Topic Model for Authors and Documents. *Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence (UAI2004)*, pp. 487–494.
- [58] Jacobi, C., van Atteveldt, W., and Welbers, K. (2016) Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*. 4 (1), 89–106.
- [59] Asmussen, C.B. and Møller, C. (2019) Smart literature review: a practical topic modelling approach to exploratory literature review. *Journal of Big Data*. 6 (1),.
- [60] Zhao, W., Chen, J.J., Perkins, R., Liu, Z., Ge, W., Ding, Y., et al. (2015) A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics*. 16 S8.

- [61] Murzintcev, N. and Nathan, C. (2022) Package “ldatuning”
Title Tuning of the Latent Dirichlet Allocation Models
Parameters. Available:
<https://cran.rproject.org/web/packages/ldatuning/ldatuning.pdf>
Erişim: 24/06/2022
- [62] Ponweiser, M. (2012) Latent Dirichlet Allocation in R,
Diploma Thesis, Wien, *Vienna University*, 2012. 33-36
- [63] Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., and Blei,
D.M. (2009) Reading tea leaves: How humans interpret topic
models. *Advances in Neural Information Processing Systems 22*
- *Proceedings of the 2009 Conference*. 288–296.
- [64] Chen, Y.L. and Chang, Y.C. (2012) A three-phase method for
patent classification. *Information Processing & Management*. 48
(6), 1017–1030.
- [65] Cassidy, C. (2020) Parameter tuning Naïve Bayes for automatic
patent classification. *World Patent Information*. 61 101968.
- [66] Blythe, J., McGrath, C., and Krackhardt, D. (1996) The effect
of graph layout on inference from social network data. *Lecture*
Notes in Computer Science (Including Subseries Lecture Notes
in Artificial Intelligence and Lecture Notes in Bioinformatics),
Springer Verlag, pp. 40–51.
- [67] Wang, E.K. and Zou, F. (2014) A New Graph Drawing Scheme
for Social Network. *Scientific World Journal*. 2014. 930314
- [68] Dwyer, T., Marriott, K., and Wybrow, M. (2007) Integrating
edge routing into force-directed layout. *Lecture Notes in*
Computer Science (Including Subseries Lecture Notes in
Artificial Intelligence and Lecture Notes in Bioinformatics),
Springer Verlag, pp. 8–19.

- [69] Fruchterman, T.M.J. and Reingold, E.M. (1991) Graph Drawing by Force-Directed Placement. *Software-Practice and Experience*. 21 (11), 1129–1164.
- [70] Hu, Y. (2006) Efficient, High-Quality Force-Directed Graph Drawing. *The Mathematica Journal*. 10. 37-71
- [71] Bastian, M., Heymann, S., and Jacomy, M. (2009) Gephi: An open source software for exploring and manipulating networks. BT - International AAAI Conference on Weblogs and Social. *International AAAI Conference on Weblogs and Social Media*. (March 2009), 361–362.
- [72] Sievert, C. and Shirley, K. (2015) LDAvis: A method for visualizing and interpreting topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, pp. 63–70.

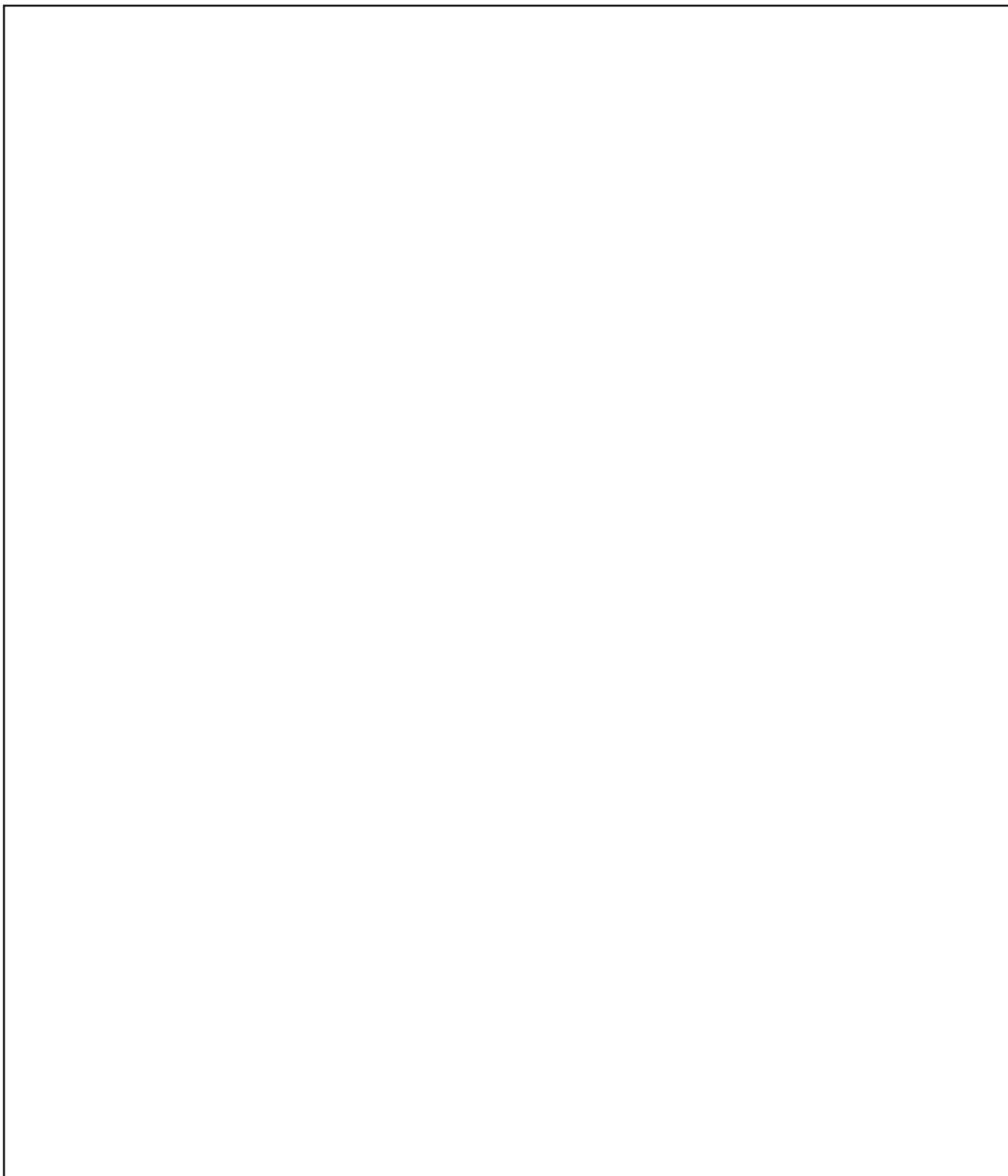
EKLER

Ek – 1: Çalışmada kullanılan veri seti

Çalışmada kullanılan veri setine aşağıda figshare veri seti paylaşım platform üzerinden aşağıda verilen DOI (digital object identifier) kullanılarak ulaşılabilir.

DOI: 10.6084/m9.figshare.21792278







GAZİLİ OLMAK AYRICALIKTIR.