

**KABA KÜME TEORİSİNİN LİTERATÜR TABANLI BİLGİ  
KEŞFİNE UYGULANMASI**

**APPLYING ROUGH SET THEORY TO LITERATURE  
BASED DISCOVERY**

**FATİH MEHMET GÜLEÇ**

**Prof.Dr. HAYRİ SEVER**  
**Tez Danışmanı**

Hacettepe Üniversitesi  
Lisansüstü Eğitim - Öğretim ve Sınav Yönetmeliğinin  
BİLGİSAYAR Mühendisliği Anabilim Dalı için Öngördüğü  
DOKTORA TEZİ  
olarak hazırlanmıştır.

2013

**Fatih Mehmet GÜLEÇ'in** hazırladığı “Kaba Küme Teorisinin Literatür Tabanlı Bilgi Keşfine Uygulanması” adlı bu çalışma jüri tarafından **BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI'nda DOKTORA TEZİ** olarak kabul edilmiştir.

Başkan  
(Prof.Dr. Şeref Sağırođlu)

Danışman  
(Prof.Dr. Hayri Sever)

Üye  
(Prof.Dr. Hakan S. Orer)

Üye  
(Yrd.Doç.Dr. Mustafa Ege)

Üye  
(Yrd.Doç.Dr. Ayça Tarhan)

Bu tez Hacettepe Üniversitesi Fen Bilimleri Enstitüsü tarafından **DOKTORA TEZİ** olarak onaylanmıştır.

Prof.Dr. Fatma SEVİN DÜZ  
Fen Bilimleri Enstitüsü Müdürü

## ETİK

Hacettepe Üniversitesi Fen Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada;

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversite veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

29 / 04 / 2013

Fatih Mehmet GÜLEÇ

## ÖZET

### KABA KÜME TEORİSİNİN LİTERATÜR TABANLI BİLGİ KEŞFİNE UYGULANMASI

**FATİH MEHMET GÜLEÇ**

**Doktora, Bilgisayar Mühendisliği Bölümü**

**Tez Danışmanı : Prof.Dr. Hayri SEVER**

**Nisan 2012, 178 sayfa**

Bilim, merak edileni açığa kavuşturmak için yapılan disiplinli akademik çalışmalar bütünü olarak düşünülebilir. Akademik çalışmaların sonuçları, diğer araştırmacılar ile paylaşımı sağlamak amacıyla bilimsel bir usul ile yayınlanırlar ve bu yayınlar akademik çalışmaların en önemli çıktısı olarak kabul edilir. Bu yayınların diğer bilim insanları tarafından incelenmesi ve kişilerdeki mevcut bilgi birikimleri ile harmanlanması sonucu yeni araştırma fikirleri ortaya çıkabilmektedir. İnsan beyninde yeni araştırma fikirlerinin nasıl oluştuğunu simule ederek, akademik yayınları bilişim sistemleri ile inceleyen çalışmalar Literatür Tabanlı Bilgi Keşfi olarak adlandırılmıştır.

Literatür Tabanlı Bilgi Keşfinde, bir yayında geçen terimlerin karşılıklı ilişkilendirilmesi ile bilgi parçaları oluşturulmaktadır. Bu bilgi parçalarının zincirleme şekilde kurulması ile yeni fikirlere ulaşılmaya çalışılmaktadır. ABC olarak isimlendirilen bu modelde, odaklanılan araştırma konusu üzerinden yola çıkılarak, “o konuyla doğrudan ilişkili olmayan” ancak “ortak terimler üzerinden dolaylı ilişkisi bulunan” terimler, yeni ve özgün fikirler olarak kullanıcıya sunulmaktadır. Bugüne kadar yapılan bir çok çalışma tıp alanındaki yayınlara odaklanmıştır. Tez çalışmasında da tıp alanında yayınlanmış makaleler, ana bilgi kaynağı olarak kullanılmıştır.

Literatür Tabanlı Bilgi Keşfi çalışmalarında bazı ortak problemler bulunmaktadır. Bu problemlerin başında, sonuç kümesinin aşırı gürültülü olması gelmektedir. Tıp alanında çok sayıda bilimsel terim bulunmaktadır ve bu terimlerin bir kısmı birbirine çok yakın anlamlar taşımaktadır. ABC modelinin terimler üzerinden kurulması sebebiyle birbirine çok benzer çıkarımlar 'kombinasyonel' olarak oluşmakta ve sonuç kümesi kontrolsüz olarak büyümektedir. Tez çalışmasında terimler yerine, birden çok terimi temsil eden 'kavramların' kullanımı sağlanmıştır. Ne çok genel, ne de çok özel olan terimler, kavram olarak işaretlenmiş ve tüm analiz kavramlar kullanılarak gerçekleştirilmiştir.

Bugüne kadar yapılan bütün çalışmalarda, ABC modeli Literatür Tabanlı Bilgi Keşfi araçlarının çekirdek işlevi olarak görülmüştür. Yapılan çalışma ile A ve C terimi arası ilişki bir B terimi olmaksızın, doğrudan yayın kümeleri üzerinden kurulmuştur. Bu amaçla önce PubMed veri tabanında bulunan 16 milyon yayın kümelenmiş ve toplam 50.000 küme oluşturulmuştur. Ortalama küme büyüklüğü 320 yayın olarak gerçekleşmiştir.

Bu yeni yaklaşımla, rasgele seçilen 80 terim için ortalama anımsama değeri 0.52, ortalama duyarlık değeri ise 0.16 olarak ölçülmüştür. Benzer çalışmalara göre hem anımsama hem de duyarlık değerinde iki kata varan bir iyileşme elde edilmiştir. Literatür Tabanlı Bilgi Keşfine yönelik temel mantık korunurken, işletim mantığının tamamen geliştirilmesi ve farkedilir bir başarının elde edilmesi, diğer araştırmacıların da farklı boyutları ile konuyu yeniden düşünmesini sağlayacaktır.

Literatür Tabanlı Bilgi Keşfi, bilimsel yayınlarda varlığını gösteren tüm insanlığa ait bilgi birikimini, harmanlayarak daha yeni keşiflere yol açmanın hayalini gütmektedir. Yapılan çalışmalar ve elde edilen ilerlemeler, bu alanda harcanan emeğin zaman içerisinde karşılığını vereceğini göstermektedir.

**Anahtar Kelimeler:** Literatür Tabanlı Bilgi Keşfi, Biyomedikal Bilişim Sistemleri, Bilgi Arama ve Erişimi, Metin Madenciliği, Belirsiz Nedenleme, Bilgi Modeli ve Çıkarım Motoru, UMLS, PubMed, Ontoloji

## **ABSTRACT**

### **APPLYING ROUGH SET THEORY TO LITERATURE BASED DISCOVERY**

**FATİH MEHMET GÜLEÇ**

**Doctor of Philosophy, Department of Computer Science**

**Supervision : Prof.Dr. Hayri SEVER**

**April 2012, 178 pages**

Science is a collection of academic studies to clarify the beeing wondered. The results of academic studies are released a scientific writing style in order to share with other researchers and these publications are considered to be the most important output of the academic studies. Examination of these publications by other scientists with their present knowledge, new research ideas may occur. The acedemic studies which are examining the publication with information systems by simulating how the new research ideas ocur in the human brain are referred to as Literature Discovery Based.

In the Literature Based Discovery studies, pieces of information are created by associating terms according to their common publications. By applying chain rule to these information pieces, establishing new ideas is aimed. The ABC model focuses on the user interested subject and finds terms “which are not directly related to the subject” but also “indirectly related through some common terms”. These disjoint but indirectly-related terms are presented as new and novel ideas. Many studies conducted to date has focused on publications in the field of medicine. Similar manner, in the thesis, published articles in the field of medicine has been used as the main source of information.

There are some common problems of Literature Based Discovery Studies. The high noise level at the result set is the biggest problem. There are a great number of terms is medical terminology and lots of these terms has also a synonymous or similar term. Because the chain rule of ABC model, similar terms are combinationally amplifies the result set in the muddy way. In the thesis, concept are used which are representing more than one terms, to reduce the noise data. Both no being very general terms, and also not being very special terms are marked as the concepts and all analyzes were carried out using these concepts.

In all studies conducted to date, the ABC model has been seen as the core function of the Literature-Based Knowledge Discovery. In the thesis, the relationship between A and C has been established by using clusters of academic publishing. The linking mission of B term are replaced by the paper clusters. For this purpose 16 million articles in PubMed database are clustered ve totaly 50.000 differens cluster are established. Average cluster size is 320 publications.

This nwe approach was test with randomly selected 80 terms. The average precision is calculated of 0.16 and avarage recall is calculated of 0.52. According to both precision and recall, two times improvement is achieved compared with similar studies. Obtaining a noticeable success by improving the core function of Literature Based discovery, will lead the other reseachers to creative ideas.

Literature Based Discovery dream to open the way to new discoveries by collating the knowledge of all humanity in scientific publications. Studies and the progress achieved in this field indicate that success of everyone senses will be achieved.

**Key Words:** Literature Based Discovery, Biomedical Information Systems, Information Retrieval, Text Mining, Fuzzy Reasoning, Information Model and Information Retrieval, UMLS, PubMed, Ontology

## TEŐEKKÜR

Prof.Dr. Hayri SEVER (tez danıřmanı), alıřmanın sonuca ulařtırılmasında ve karřılařılan güçlüklerin ařılmasında yön gösterici olmuřtur.

Prof.Dr. Hakan Sedat ORER, tez konusunda alan uzmanı olarak destek saęlamıřtır.

Tahir BIAKCI, kendi alıřmasında edindięi tecrübeleri paylařmıř, uygulamanın geliřtirilmesinde teknik destek vermiřtir.



# İÇİNDEKİLER DİZİNİ

ETİK.....	III
ÖZET .....	IV
ABSTRACT .....	VI
TEŞEKKÜR .....	VIII
İÇİNDEKİLER DİZİNİ.....	IX
ŞEKİLLER DİZİNİ.....	XI
ÇİZELGELER DİZİNİ .....	XII
KISALTMALAR.....	XIII
1. Giriş.....	1
1.1. Literatür Tabanlı Bilgi Keşfi .....	5
1.2. Önceki Çalışmalar .....	10
1.3. LTB Karşılaşılan Sorunlar .....	14
1.4. Çözüm Önerileri.....	15
2. Kapsam .....	17
2.1. Bilgi Erişim.....	17
2.2. Vektör-Uzay Modeli .....	19
2.3. Metin Madenciliği.....	22
2.4. Birleşik Medikal Dil Sistemi .....	23
2.4. Belge Kümeleme .....	27
2.5. Kaba Kümeler .....	30
3. Yöntem .....	33
3.1. Matematiksel Model .....	34
3.2. Kaba Küme Teorisi ve LTB.....	36
3.3. Uygulama Veri Tabanı Alt Yapısı .....	39
3.3.1. Yayınların temin edilmesi.....	40
3.3.2. Terim Çıkarımı .....	40
3.3.3. MeSH Ağacı ve KK-LTB için uyarlanması .....	41
3.3.4. Vektör Uzay Modeli .....	43
3.4. Belge Kümeleme .....	47

3.4.1. Kümeleme Başarım Ölçümü .....	49
3.4.2. Kümeleme Başarım Deneylei .....	50
3.4.3. Ortalama Küme Büyüklüğü.....	56
<b>3.5. LTB Sınama Yöntemi .....</b>	<b>59</b>
<b>3.6. Verilerin büyüklüğü .....</b>	<b>62</b>
<b>4. Sonuçlar .....</b>	<b>64</b>
4.1. Kontrol kümesinin genişlemesinin etkileri.....	64
4.2. Bir örnek üzerinde ilk 100 sonuç teriminin incelenmesi.....	68
4.3. Ölçme Yönteminin Bir Terim için İncelenmesi .....	69
4.4. Örnek 80 terim için başarı incelemesi .....	69
<b>5. Tartışma .....</b>	<b>71</b>
<b>Kaynaklar .....</b>	<b>77</b>
<b>ÖZGEÇMİŞ .....</b>	<b>86</b>
<b>Ek.1. Raynaud Hastalığı ve Altın Standartların Dağılımı.....</b>	<b>87</b>
<b>Ek.2. Raynaud Hastalığı ve İşletim Sonuçları .....</b>	<b>148</b>
<b>Ek.3. Ölçme Yönteminin Sınanması.....</b>	<b>153</b>
<b>Ek.4. Rasgele 80 Terim için Sonuçlar.....</b>	<b>161</b>

## ŞEKİLLER DİZİNİ

Şekil 1. Terimler Arası İlişkiler [18].....	6
Şekil 2. Hipotez Üretim Örneği.....	7
Şekil 3. LTB Standart İşletim Modeli .....	9
Şekil 4. UMLS Modülleri [51].....	26
Şekil 5. Kaba Kümeler genel görünüm [61].....	30
Şekil 6. Belge Kümeleri ve Kaba Küme Mantığı.....	36
Şekil 7. Standart LTB'dan Rough LTB'ye Geçiş.....	37
Şekil 8. Tez Çalışmasının Tüm Modülleri.....	39
Şekil 9. Terim Ağacından Kavram Ağacına Geçiş .....	42
Şekil 10. Kavramların Veri Tabanı Gösterimi .....	43
Şekil 11. Kavram Ağacından Vektör-Uzay Modeline Geçiş .....	44
Şekil 12. Vektör-Uzay Modeli ve Anlamsal Ağırlıklı TFxIDF .....	47
Şekil 13. Elbow Metodunun Uygulanması.....	57
Şekil 14. Farklı Küme Büyüklükleri için Duyarlık Grafiği.....	58
Şekil 15. Farklı Küme Büyüklükleri için Anımsama Grafiği.....	58
Şekil 16. Sistemin Sınama Yöntemi [38] .....	61
Şekil 17. Yıllara Göre Anımsama/Duyarlık Dağılımı.....	65
Şekil 18. Örnek 80 Terim için Anımsama Grafiği .....	70
Şekil 19. Örnek 80 Terim için Duyarlık Grafiği .....	70

## ÇİZELGELER DİZİNİ

Çizelge 1. LTB araçlarının karşılaştırılması [38].....	13
Çizelge 2. Metathesaurus kapsamındaki veri kümeleri [38], [54] .....	25
Çizelge 3. Örnek Bir Kaba Küme Verisi .....	32
Çizelge 4. Belge Kümeleme ve Safılık Değeri .....	52
Çizelge 5. Belge Kümeleme ve F-Ölçüm .....	53
Çizelge 6. Belge Kümeleme ve Entropy Değeri .....	54
Çizelge 7. F- Ölçüm Değerinin Karşılaştırılması .....	55
Çizelge 8. Yıllar Göre Kümülatif Duyarlık ve Anımsama Dağılımı .....	66
Çizelge 9. Yıllara Göre Parçalı Duyarlık ve Anımsama Dağılımı.....	66
Çizelge 10. Yıllara Göre Duyarlık Dağılımı .....	67
Çizelge 11. Yıllara Göre Anımsama Dağılımı .....	67
Çizelge 12. Örnek 80 Terim için Anımsama Dağılımı .....	69
Çizelge 13. Örnek 80 Terim için Duyarlık Dağılımı .....	70

## KISALTMALAR

ABC	Terimler Arası İlişki Modeli
BÇ	Bilgi Çıkarımı
BMM	Biyomedikal Metin Madenciliği
KK-LTB	Kaba Kübe ile Literatür Tabanlı Bilgi Keşfi
LTB	Literatür Tabanlı Bilgi Keşfi
MEDLINE	Tıbbi Literatür Analizi ve Erişim Sistemi
MeSH	Medical Subject Headings
MM	Metin Madenciliği
NLM	Ulusal Tıp Kütüphanesi
TF-IDF	<i>Term Frequency–Inverse Document Frequency</i>
UMLS	Bileşik Medikal Dil Sistemi
VUM	Vektör-Uzay Modeli

## 1. Giriş

Bilim, “belli bir konuyu bilme isteğinden yola çıkan, belli bir amaca yönelen bir bilgi edinme ve yöntemli araştırma süreci”<sup>1</sup> olarak tanımlanmıştır. Araştırma süreçlerinden birisi de, disiplinler arası çalışmalardır. Yenilikçi fikirlerin çoğunlukla disiplinler arası çalışmalarla ortaya çıktığı genel kabul görmektedir [1–3]. Disiplinler arası çalışmaların yapılabilmesi için bilim insanlarının, bilim dallarına yatayda etkin olması beklenmektedir. Ancak bilim insanlarının, kendi çalışma alanlarına paralel diğer bilim dallarına da tam hakim olması neredeyse imkansızdır. Bilim insanlarına, üzerinde çalıştıkları konulara ilişkin, başka hangi disiplinlerden destek alabileceğini ‘somutlaştırarak’ göstermek, bu tezin genel hedefini oluşturmaktadır.

Robert A. Day’e göre “Bilimsel araştırmanın gayesi yayındır. Bilim insanları yayınlarıyla değerlendirilir ve bilinirler” [4]. Bu varsayımın yansımalarını tıp alanında yayınları indeksleyen PubMed [5] veri tabanını inceleyerek görebiliriz. PubMed veri tabanında, yayınların indekslenmeye başlandığı 1983 yılından 2005 yılına kadar toplam 13 milyon yayın bulunmaktadır. 2010 yılı baz alındığında ise bu sayı 16 milyona yükselmektedir [6]. Diğer bir ifade ile 2010 yılında veri tabanında bulunan yayınların %18.75’i sadece son beş yıl içerisinde yayınlanmıştır. Bu büyümeyi, PubMed’de her gün 2000 yeni makalenin yayınlandığını [6] da göz önünde bulundurarak anlayabiliriz.

Bilimsel büyümenin bu derece hızlı olduğu günümüzde, bilim insanları kendi uzmanlık alanlarına ek olarak, paralel bilimlerde de ortaya çıkan yenilikleri takip etmek zorundadır. Ancak bu kadar yoğun bilgi akışının süzülmesi, analiz edilmesi ve yeni araştırma nüvelerinin oluşturulması için devamlı büyüyen bilgi kaynaklarında akıllı araçların kullanımı, giderek daha önemli hale gelmektedir [7], [8].

Bilimsel çalışmaların yönetilmesi, yönlendirilmesi ve inovasyon alanında yapılan çalışmalarda [9–11], gerçek anlamda inovasyonun, diğer araştırmacı ekiplerle

---

<sup>1</sup> “Bilim”, Türk Dil Kurumu Sözlüğü, [www.tdk.gov.tr](http://www.tdk.gov.tr)

ve/veya yayınlar ile yaşanacak etkileşimle mümkün olduğu belirtilmektedir. Böylelikle, taze yayınlardan elde edilen bilgiler ışığında, 'yeni' konular üzerine araştırma yapma fikirleri ortaya çıkmaktadır. İnsan beyninde kavramların karşılaşması (*head-on collision*) ve eşleşmesiyle; özgün fikirler, sezgisel olarak (*heuristic*) ortaya çıkmaktadır [12], [13].

Sezgisel olarak yeni fikirlerin ortaya çıkışının bilgisayar teknolojilerinde benzetim (*simulation*) ile modellenebileceği fikri, Literatür Tabanlı Bilgi Keşfi - LTB'nin temelini oluşturmaktadır. LTB ile, birden çok çalışmanın birbirini tamamlayıcı bir şekilde bir araya getirilmesi sonucu yeni ve özgün çıkarımların yapılması hedeflenmektedir [14]. Swanson, LTB'yi ilk defa ortaya atan kişi olarak, bu fikri "tamamlayıcı ama ayrık" (*complementary but disjoint*) ifadesi ile anlatmaktadır [15], [16].

LTB, bir terim için "*acaba bugüne kadar bu terimle dolaylı ilişkili olduğu halde, doğrudan çalışma yapılmamış kavramlar var mı*" sorusuna yanıt vermeye çalışır. Bu soruya yanıt vermek için yapılacak işletimde, kullanıcının araştırma yapmak istediği terim arama terimi olarak ele alınır ve bu terimle doğrudan ilişkili diğer terimler araştırılır. Doğrudan ilişkili terimler, o terimle en az bir yayında bir arada kullanılan terimleri ifade etmektedir. Ortaya çıkan terimler bağlama terimi olarak ifade edilir. İkinci aşamada ise, bağlama terimleri ile ilişkili olan ve arama terimi ile hiç bir yayında ortak kullanılmamış terimler tespit edilir. Sonuç terimleri olarak ifade edilen bu terimlerle arama terimi arasında birden çok bağlama terimi bulunabilmektedir. Sonuç terimleri ilişkili oldukları bağlama terim sayılarına göre sıralanarak kullanıcıya sunulur.

LTB çalışmalarında ortak nokta, kavramlar/terimler arası ilişkilerin aynı yayında geçmeleri üzerinden kurgulanmalarıdır. Bu çalışmalara göre, iki terimin aynı yayında geçmesi, aralarında bir ilişki olduğunu göstermektedir. Ancak bu yaklaşım, sonuç listesinde çok sayıda terim ikilisinin oluşmasına yol açmış, son kullanıcı açısından da sunulan sonuçların büyük bir gürültü taşımasına sebep olmuştur. Gürültü, istenmeyen çok sayıda hatalı verinin sonuç kümesinde yer alıyor olmasıdır. Gürültü düzeyinin yüksek olması, sonuçların son kullanıcı tarafından incelenmesini imkansız hale getirmektedir.

Tez kapsamında yapılan çalışma, LTB çalışmalarının sonuç kümesinde gürültünün yüksek olması ve başarımın düşük olması sorununa odaklanmıştır. Çözüme yönelik, yayınların içeriklerine göre kümelenmesi ve kümeler üzerinden LTB işletiminin yapılması sağlanmıştır. Böylelikle çözüm kümelerindeki hata oranı düşürülmüştür.

Ana hedef, LTB çalışmalarında gürültünün düşürülmesine yönelik çözümler üretmektir. Buna yönelik ilk adım terimlerin yerine kavramların kullanımınıdır. Terimler detay seviyede bilgi taşıyor olsa da, sadece metin analizi ile terimlerin ilişkilendiriliyor olması, detay seviyedeki bilgilerin yapısal olarak harmanlanmasına olanak tanımamaktadır. Dolayısıyla hatalı eşlemeler, detay seviyesi arttıkça daha da artmaktadır. Terimlerin daha genel bilgiler ifade eden kavramlar altında kümelenmesi çözüm olarak önerilmiştir. Ayrıca tez çalışmasında bağlama terimleri ortadan kaldırılmış ve terimler/kavramlar arası ilişkilerin kümeler aracılığı ile kurulması sağlanmıştır.

Tez çalışmasında, iki terim arasında bir ilişki kurulması için her ikisini de barındıran en az bir yayının aranması ve bunun yeter koşul olması eleştirilmektedir. Tez çalışmasında, “aynı yayında geçiyorsa ilişkilidir” önermesinin yerine bilginin kümelenmesi neticesinde ortaya çıkan kümelere yer almaları gözetilerek ilişkilerin kurulması sağlanmaktadır. Bu yeni fikir, insanların yakın disiplinler arası çalışmalar yaparak, aslında bilgi kümeleri arası ilişkileri kurdukları varsayılarak ortaya atılmıştır.

Disiplinler arası çalışmayı benzetimin (*simulation*) içine dahil etmek amacıyla birbirine yakın yayınlar gruplanmıştır. Bu gruplar içerisinde kullanıcının arama terimi olarak verdiği terimi barındıran kümeler incelemeye alınır. Yapılan incelemede, o terimi sıklıkla barındıran kümeler, nadir barındıran kümelere göre daha değerli olduğu düşüncesi ile üst sıraya çıkartılır. Bu kümeler içerisinde, arama terimini içermeyen, ama arama terimi içeren çok sayıda yayın ile ortak bir kümede bulunan yayınlar seçilir. Bu yayınlar içerisinde, daha önce arama terimi ile ortak bir çalışmada yer almamış terimler/kavramlar tespit edilerek kullanıcıya sunulur.

Bu işletimin başarılması için kümeleme büyük önem taşımaktadır. Kümelemenin yapılması için nesnelerin ve özelliklerinin tanımlanması gerekmektedir. Yayınları



nesneler, yayınların barındırdığı kavramları nitelikleri olarak tanımlayarak kümeleme yapılmıştır.

Kümelemenin gerçekleştirimi için teknik bazı sorunların çözülmesi gerekmektedir. Dikeyde 15 milyon nesnenin ve yatayda 15 bin terimin oluşturduğu bir matriste kümelemenin yapılabilmesi için bellek yönetimi güçlü algoritmaların tercih edilmesi yeterli olmamaktadır. İşlemci ve sunucunun yeterince güçlü olmasına karşın<sup>2</sup>, kümeleme işlemi haftalarca sürmektedir. Bu sürenin deneyin tekrarlanabilmesi ihtimalleri gözetilerek kısaltılması gerekmiş ve terimlerin yerine kavramların kullanımı ile matrisin yatay boyutu küçülmüştür. Kümeleme işlemleri yüksek hacimli veriler ile çalışmak üzere geliştirilmiş Cluto [17] uygulaması kullanılarak gerçekleştirilmiştir.

Tez çalışmasının gerçekleştirimi sırasında alt modüllerin başarımları ölçülmüştür. Kümelemede başarımlar F-Ölçüm değeri ile 0.92 olarak ölçülmüştür. Kümelemeden elde edilen başarımlar, tüm sistemin başarımlarını da olumlu yönde etkilemiş, ve benzer çalışmalarda 0.2 dolaylarında olan anımsama değeri, rasgele seçilen 80 terim için ortalama 0.52 olarak ölçülmüştür. Benzer şekilde diğer çalışmalarda 0.09 sınırına ancak erişen duyarlılık değeri, aynı 80 terim için ortalama 0.16 olarak ölçülmüştür.

Hazırlanan çalışmada, ilk olarak Literatür Tabanlı Bilgi Keşfi (LTB) çalışmaları anlatılmıştır. Özellikle LTB alanında yaşanan sorunlar ve bu sorunlara bugüne kadar sunulmuş çözümler belirtilmiştir. Tez metninin ilerleyen bölümlerinde LTB araçlarının geliştirilmesi için kullanılan teknolojiler anlatılmıştır. Bilgi Erişimi (*Information Retrieval*), Vektör Uzay Modeli, Metin Madenciliği, Birleşik Tıp Dil Sistemi (*Unified Medical Language System*) ve Belge Kümeleme (*Document Clustering*) kavramları anlatılmıştır. Yapılan çalışmanın özgü yönü olarak ön plana çıkan Kaba Kümelere (Rough Sets) ilişkin bilgi yine aynı başlık altında verilmiştir.

Tez raporunun ilerleyen kesimlerinde tez çalışmasında uygulanan yöntemler anlatılmıştır. Bu amaçla öncelikle tez çalışmasının Matematiksel Modeli kurulmuştur. Daha sonra Kaba Küme Teorisinin çalışmaya nasıl dahil edildiği

---

<sup>2</sup> Tez uygulaması her biri 6 çekirdekli 4 işlemciye sahip, 64 GB ana bellek barındıran, 8 adet disk ile RAID 5 yapılandırması kurulmuş bir sunucu üzerinde işletilmektedir.

anlatılmış, bu amaçla kurgulanan veri tabanı modeli anlatılmıştır. Yayınların temin edilmesi, buradan terimlerin çıkarımı, MeSH ağacı kullanım yöntemi ve işletim sırasında elde edilen bilginin veri tabanında hangi yöntemle saklandığı belirtilmiştir. Bu bilgilerden Kaba Küme Teoremine geçiş amacıyla Belge Kümelemenin tez çalışmasında hangi yöntemlerle yapıldığı anlatılmıştır. Nihai olarak tüm sistemin işleyişinde en büyük sorun olarak karşımıza çıkan veri büyüklüğü (milyonlarca veri satırı) ve çözüm yöntemleri anlatılmıştır.

Çalışmanın son kesiminde ise LTB aracı ile elde edilen bazı sonuçlar paylaşılmıştır. Bu sonuçların yapılmış diğer çalışmalar ile durumu incelenmiştir. Tartışmalar kesiminde, elde edilen başarıların sonuçları somut verilerle irdelenmiştir. Son olarak aynı alanda devam edecek çalışmalar için, olası yeni çözümlere ilişkin fikirler paylaşılmıştır.

### **1.1. Literatür Tabanlı Bilgi Keşfi**

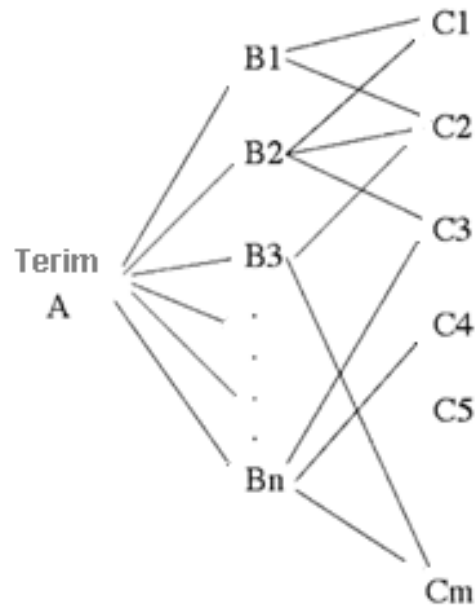
Literatür Tabanlı Bilgi Keşfi – LTB, akademik yayınların taranması ile yeni ve özgün hipotezlerin üretilmesini hedef almaktadır. Yeni hipotezler, mevcut yayınlardan daha önce hiçi ilişkilendirilmemiş, ancak birbirlerini tamamlayıcı olanlarının (*Complementary But Disjoint*) bir araya getirilmesi ile elde edilmektedir.

Literatür Tabanlı Bilgi Keşfi, Don R. Swanson'un 1986 yılında ortaya attığı Balık Yağı ve Raynaud Hastalığı arasında bulunduğu ilişki ile ilk defa tanımlanmıştır [16]. Bu çalışmada, bazı yayınlardan elde edilen "kan akışkanlığı Raynaud hastalığında artmaktadır" önermesi ile birbirinden bağımsız başka yayınlardan elde edilen "balık yağı, kan akışkanlığını azaltmaktadır" önermeleri kullanılarak, yeni bir hipotez olarak "balık yağı, Raynaud hastalığında tedavi edici olarak kullanılabilir" önermesi üretilmiştir. Çıkarımın yapıldığı 1986 yılından önce bu hipotezi deneysel olarak analiz eden ya da böyle bir ilişkiden bahseden bir yayın bulunmamaktadır. Ancak takip eden iki yıl içerisinde, hipotez klinik deneylerle ispatlanmıştır. Swanson, 1991 yılında, yukarıda bahsedilen sistematığı bir yazılım programı ile gerçekleştirmiş ve "tamamlayıcı fakat ayrık" yayınlardan analizler üretilmesinde ilk adımı atmıştır.

Literatür Tabanlı Bilgi keşfinin işleyiş sistematığının anlaşılması için, öncelikle girdi ve çıktı tanımlarının verilmesi gerekir. Sistem, kullanıcının yeni ilişkiler tespit

edilmesini istediđi arama terimini ve bu terime iliřkin anlamsal kısıtları belirlemesi ile bařlatılmaktadır. Kullanıcının beklentisi, arama terimi ile dolaylı olarak iliřkili olabilecek ancak yayınlarda dođrudan bir iliřki bulunmayan terimlerin listelenmesi ile karřılanır. Diđer bir ifadeyle, arama terimi ile daha önce hiç bir yayında iliřkilendirilmemiř ve kavramsal olarak iliřkilendirilmesi m¼mk¼n olan terimler listelenir. zetle, arama terimine iliřkin yeni hipotezler ¼retilir.

Standart LTB uygulamalarında, kullanıcı tarafından verilen bařlangıç terimi (A Terimi) ile iliřkili yayınlar temin edilir. İliřkili yayınlar, bařlangıç teriminin geçtiđi yayınların indirilmesinden ibarettir. Temin edilen makalelerde bulunan diđer terimler tek tek ıkartılır ve yayınların t¼m¼nden elde edilen bu terimler ile bađlama terim listesi (B Terimleri) elde edilir. Bylelikle iki terimin aynı yayında gemesi, aralarında bir iliřki olduđunun gstergesi olarak kabul edilmiřtir. İřletimin ikinci ařamasında, her bir bađlama terimi iin yine yayın taraması yapılır ve t¼m bađlama terimleri iin elde edilen yayınlarda yine terim ıkarsaması yapılır. İkinici ařamada elde edilen terim listesi ise sonu terimleri (C terimleri) olmaya adaydır. Bu noktaya kadar anlatılan akıř řekil 1’de gsterilmiřtir.

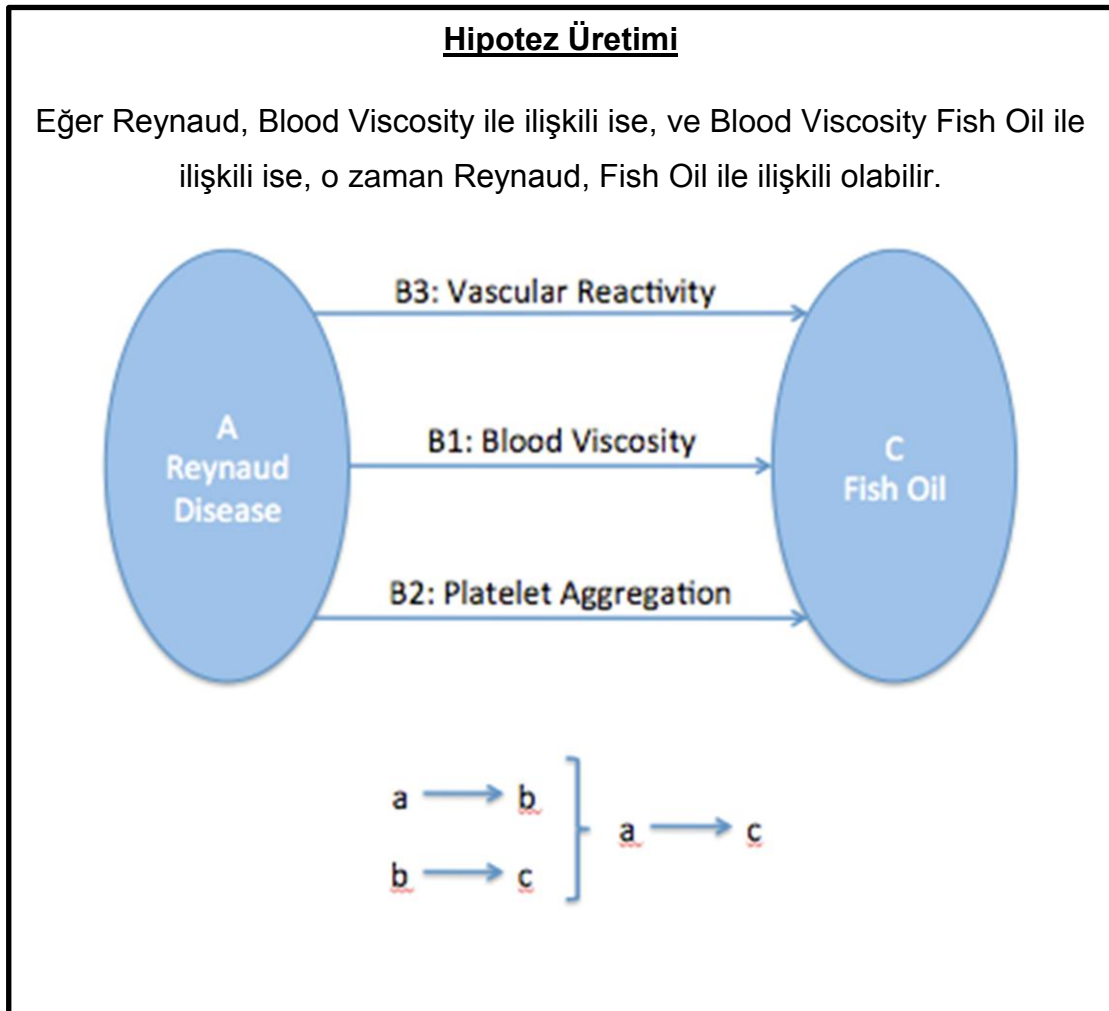


řekil 1. Terimler Arası İliřkiler [18]

Ancak LTB alıřmalarında daha önce keřfedilmiř iliřkilerin elenmesi gerekmektedir. Bu amala sonu terim listesinde yer alan terimlerden, bađlama

terim listesinde yer alanlar elenir. Böylelikle sonuç terim listesinde başlangıç terimi ile doğrudan ilişkili terim kalmamış olur.

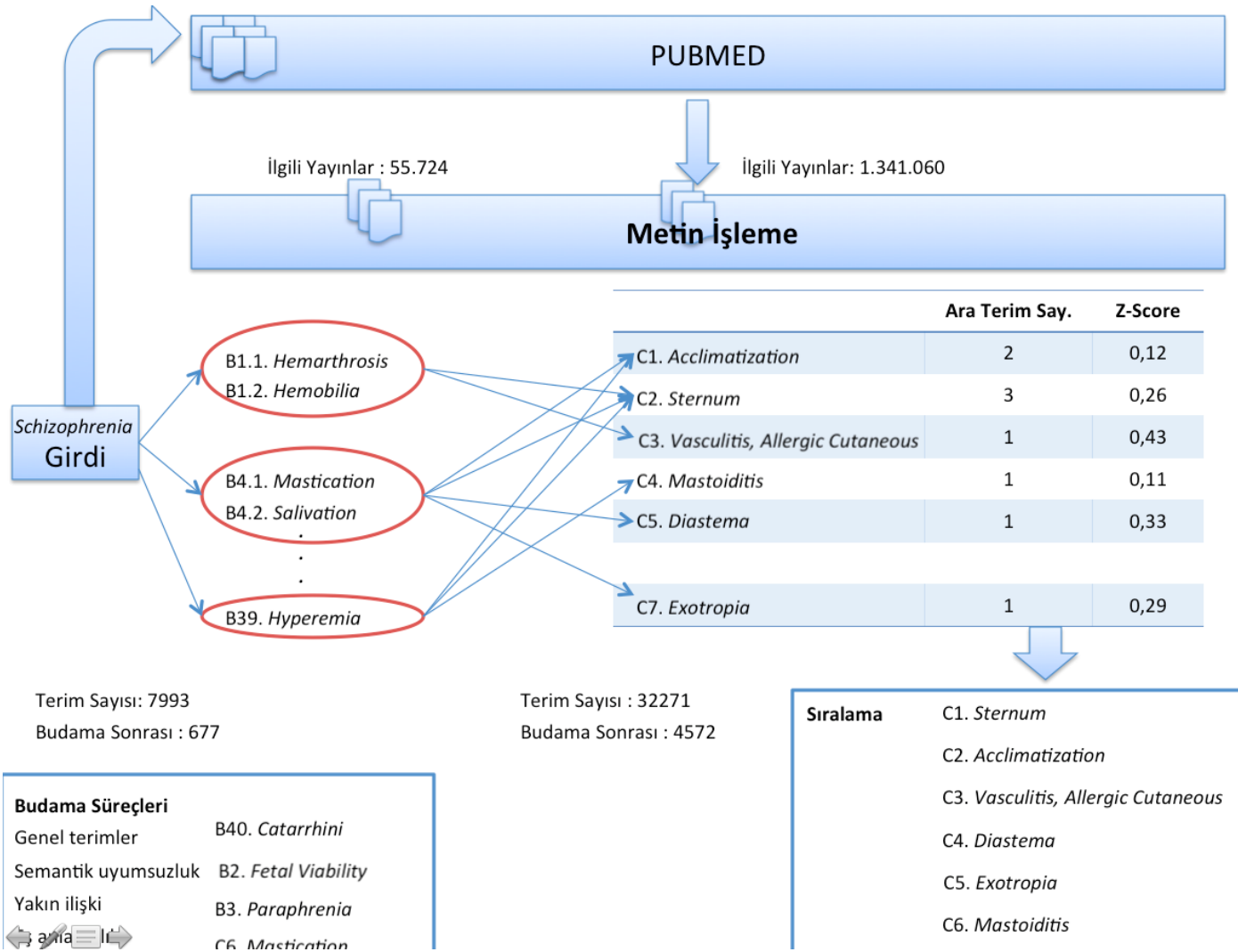
İşletimin son aşamasında sonuç terimleri kullanıcıya belirli bir sırada sunulması hedeflenir. Önem sırasına göre listelemenin avantajı, kullanıcıyı mümkün olduğunca uzun listeler içerisinde kaybolmasına engel olmak, böylelikle daha kolay kullanılabilen bir sistem sunabilmektir. Aslında hedeflenen şey, tamamı doğru bir liste üretmekten çok, ilk 100 sırada getirilen hipotezlerin (sonuç terimlerinin) olabildiğince anlamlı olmasının sağlanmasıdır. Bu amaçla bulunan ilk çözümde önem sırası, sonuç terimleri bağlama terimleri ile olan ilişki sayıları hesaba katılarak hesaplanmaktadır.



Şekil 2. Hipotez Üretim Örneği

LTB üzerine yapılan çalışmalar incelendiğinde genel bir uygulama akışı ortaya çıkarılabilir. Standart bir LTB uygulamasında, işlem kullanıcının incelemek istediği

terim ya da kavramı sisteme girdi olarak vermesi ile başlar. Terim ile birinci derecede, doğrudan ilişkili –bağlama terimleri– listesinin oluşturulması için faydalanılan kaynaktan arama terimini içeren yayınlar temin edilir. Bütün uygulamalarda yayın kaynağı olarak PubMed kullanılmaktadır. Şekil 3’de akış detaylı olarak gösterilmektedir.



Şekil 3. LTB Standart İşletim Modeli

Temin edilen yayın listesindeki terimler metin işleme teknikleri ile çıkarılmaktadır. Genelde akademik yayınların anahtar kelime listeleri üzerinden bağlama terimleri oluşturulurken, bazı çalışmalarda başlık ya da özetten metin işleme teknikleri ile terimler listelenmiştir. Bağlama terimlerinin bir kısmı ortak kavramlara ilişkin olabilir. Anlamdaş ya da yakın anlamlı terimlerin bir araya getirilmesi, bağlama terim sayısının azaltılmasını sağlar. Bazı çalışmalarda bağlama terimlerinden arama terimi ile semantik uyumsuzluk gösteren terimler elenmektedir. Ancak daha da önemli olan, arama terimi ile anlamsal olarak çok yakın ilişki gösteren terimlerin elenmesidir.

Bağlama terimleri üzerinde yapılan bir diğer eleme işlemi de, genel terimlerin budanmasıdır. Genel terimlerin budanması ile arama terimi ile sonuç terimleri arasında sadece anlamsal olarak güçlü bağlama terimleri aracılığı ile ilişki kurulması ve böylelikle sonuç terim listesinde sadece arama terimine güçlü bağları bulunan terimlerin kalması sağlanır.

Bağlama terimlerinde gerekli düzenlemelerin yapılmasının ardından, her bir bağlama terimi için yayın veri tabanından ilgili yayınların çekilmesi, bu yeni yayınlardan terimlerin benzer yöntemlerle çıkarılması sağlanır. İkinci aşamada elde edilen terim listesi üzerinde semantik uyumsuzlukların giderilmesi, eş ya da yakın anlamlıların budanması sonucu oluşan listeye sonuç listesi ismi verilir. Sonuç listesinde yer alan terimlerden hiç birisi bağlama terimleri içerisinde yer alamaz. Diğer bir ifadeyle sonuç terimlerinden hiç biri, arama terimi ile doğrudan ilişkili olamaz. Doğrudan ilişki çoğu çalışmada “aynı makalede geçmek” olarak tanımlanmıştır. Arama terimi ile sonuç terimlerinin arasında doğrudan ilişki olmaması, kurulacak hipotezlerin “yeni” olmasını temin edilmesini sağlamaktadır.

## 1.2. Önceki Çalışmalar

LTB, Swanson tarafından ilk tanımlandığı günden bugüne çok sayıda araştırmacı tarafından çalışılmış ve geliştirilmiştir. N. Smalheiser and V. Torvik, 1997 yılında Arrowsmith ismini verdikleri sistem ile önerilen yöntemin güzel bir yazılım örneğini sunmuşlardır [19], [20]. Gordon ve Lindsay yakın tarihlerde bilgi erişim tekniklerini (TF x IDF) LTB üzerine uygulamıştır [21–23]. Xie ve arkadaşları mevcut hipotezlerin üretimde kullanılması mantığında ufak bir farklılığa giderek, yerine koyma (*substitution*) ve zincirleme (*chaining*) kurallarını da eklemiştir [24].

Pratt ve Yetişgen Yıldız, uzun süre devam eden çalışmalarında bir dizi yeni kavramı LTB'ye kazandırmışlardır. 2003'de yaptıkları çalışmada, anlamsal olarak birlikte olan terimleri ontoloji kullanarak bir araya getirmiş ve böylelikle anlamsız terimlerin elenmesini sağlamaya çalışmışlardır [25]. Nihai sonuçların kullanıcıya sunulmasında, sıralama standartlarını belirledikleri çalışmalarında [26], MeSH terim ağacının kullanımı göstermişlerdir.

Ronald B. Kostoff ve arkadaşları, 2008 yılında başlama terimiyle beraber sonuç terimini alıp sadece bağlama terimlerini belirleyecek veya tüm çalışmanın belirli bir grup yayın üzerinden işletilmesini sağlayabilecek metotlarla, "Literatür İlişkili Keşif" metodolojilerini belirlemişlerdir [27]. Takiben yaptıkları bir dizi çalışma ile "Multiple Sclerosis", "Parkinson hastalığı", "Katarakt" ve "Water Purification" için uygulamalar yapmışlardır [28–31].

2011 yılında Sejoon Lee ve arkadaşları, terimleri gen, ilaç, hastalık ve semptom olarak sınıflayarak, terimler arası ilişkileri bu sınıflara göre tespit etmişler ve %15 dolaylarında bir anımsama değerine ulaşmışlardır.

Petric ve arkadaşları 2012'de, kapalı keşif olarak adlandırılan, başlangıç ve sonuç terimlerinin kullanıcıdan alınıp, aralarında bir bağ varsa, bağlama terimlerinin listelenmesini sağlayan yöntemi, istatistiksel olarak başlama ve sonuç terimlerine ait literatürlerde aykırı örneklerden yola çıkarak yeniden yorumlamıştır [32]. Aynı çalışmada bundan yaklaşık 50 yıl önce Koestler tarafından ortaya atılmış [33], insanın üretme sanatını ele alan "*bisociative reasoning*" kavramı da çalışmanın temeline oturtulmuştur. Petric ve arkadaşlarıyla aynı tarihte Jursic ve arkadaşları da [34], "*bisociative reasoning*" kavramını uyarlamaya çalışmışlardır.

Kostoff, uzunca yıllar LTB alanında çalıştıktan sonra, 2012'de edindiği tecrübeleri sıralayan yayınında[35] şu önemli noktalara temas etmektedir:

1. Sonuç terimleri içerisinde en önemli olana öncelik vermek, keşfin ilginç olması için zor bulunur olması gerçeği ile çelişmektedir. Bilinmeyen ancak çok güçlü bir şekilde LBD araçları tarafından çıkarılan bir ilişki, zaten yakın zamanda tespit edilebilecek ve büyük ilgi uyandırmayacak bir ilişkidir.
2. Alan uzmanları olmaksızın, LBD çalışmalarında en doğru ve önemli ilişkiyi tespit etmek mümkün değildir.



3. Yayınlar arası bibliyografik eşleşme, LBD çalışmalarının kalitesini artırıcı özelliği vardır.

Huang ve arkadaşlarının 2012'de yayınladıkları makaleleri, çalışmamıza bir derece yakın olması bağlamında ilgi çekicidir [36]. Ancak Huang ve arkadaşları sadece Vektör-Uzay Space Model'i kullanmakla yetinmişlerdir.

Pratt ve Yetişgen Yıldız 2009 yılında yayınladıkları çalışmaları ile [37], LTB üzerine çalışan sistemlerin sonuçlarının birbirleriyle nasıl karşılaştırılacaklarına ilişkin yöntemleri tanımlamaya çalışmışlardır.

Tahir BIÇAKCI, yapılan bu tez çalışmasının öncül çalışması olarak yaptığı Yüksek Lisans çalışmasının görünümünü şu şekilde vermiştir [38]. *“Genel işleyişte öncelikle başlangıç terimi (“A” terimi) alınır. Başlangıç terimi ile ilgili literatür kaynaklarında bulunan bütün makaleler çekilir. Makalelerin indekslemesinde kullanılmış olan ilgili MeSH terimleri ve yazarın kendi tanımladığı anahtar terimler bir küme olarak ham bağlama terimleri derlenir. Her bir ham terim incelenerek budama sürecinden geçirilir. Geriye kalan terimler arasında birbirine yakın terimler gruplanır. Her bir grubun başlangıç terimi ile ilgili bağlantısı hesaplanır. Bağlantı değeri eşik değeri altında kalanlar elenir. Bu şekilde gruplanmış (veya gruplanmaya ihtiyaç duyulmamış) terimlere bağlama terimi (“B” terimi) denir.*

*İşletimin ikinci aşamasında her bir bağlantı terimi için literatür kaynaklarında bulunan makaleler indirilir. Başlangıç terimindeki makalelerde uygulandığı gibi bu makalelerde de ham terim kümesi derlenir. İlk aşamadan farklı olarak oluşturulan ham terim kümesi işlenerek sonuç terim kümesi oluşturulmaktadır. Budama işlemi ilk aşamada yapılan budama adımlarına ek olarak bağlama terim kümesinde bulunan terimler de elenir. Daha sonra birbirine yakın terimler gruplanır ve gruplanmış terimlerin ilişki skoru hesaplanır. Gruplanan bu terimlere sonuç terimleri (“C” terimleri) denir. Sonuç terimlerinin başlangıç terimi ile ilişkileri gözetilerek sıralama işlemi yapılır ve sonuç liste kullanıcıya sunulur.”*

Tez çalışmasının sonuçlarının ve performansının anlaşılması için literatürde bulunan diğer çalışmalara göz atılması gerekmektedir. Tahir BIÇAKCI yaptığı çalışmada benzer amaçla uygulamaları ve niteliklerini belirlemiştir [38].

Özellikler	Arrowsmith <sup>3</sup>	Arrowsmith2 <sup>4</sup>	BITOLA <sup>5</sup>	Manjal <sup>6</sup>	LitLinker <sup>7</sup>
<b>Kayıt</b>	-	-	-	Gerekli	Gerekli
<b>Çevirim-içi</b>	Evet	Evet	Evet	Evet	Evet
<b>Kavram/Terim</b>	Başlıklardaki kelimeler	Başlıklardaki kelimeler başlıktaki UMLS terimleri	Mesh terimleri	Mesh Terimleri	UMLS
<b>Belgelendirme</b>	Zayıf	Zayıf	Zayıf	Güçlü	Orta
<b>Sonuçların görselleştirilmesi</b>	Terim listesi	Terim listesi ve PubMed'e bağlantılar	Terim listesi ve PubMed'e bağlantılar	Terim listesi ve PubMed'e bağlantılar	Terim listesi, PubMed'e bağlantılar ve bağların güçleri
<b>Kullanıcı arayüzü</b>	Zayıf	Orta	Orta	Orta	Gelişmiş
<b>Uygulama alanı</b>	Biyomedikal	Biyomedikal	Biyomedikal ve genomik	Biyomedikal	Biyomedikal

Çizelge 1. LTB araçlarının karşılaştırılması [38]

<sup>3</sup> ArrowSmith. 20.09.2009. <http://kiwi.uchicago.edu/>

<sup>4</sup> Arrowsmith2. 20.09.2009. <http://arrowsmith.psych.uic.edu/>

<sup>5</sup> BITOLA. 20.09.2009. <http://www.mf.uni-lj.si/bitola/>

<sup>6</sup> Manjal. 20.09.2009. <http://sulu.info-science.uiowa.edu/Manjal.html>

<sup>7</sup> LitLinker. 20.09.2009. <http://litlinker.ischool.washington.edu/>

### 1.3. LTB Karşılaşılan Sorunlar

LTB üzerinde yapılan çalışmalar ve son kullanıcılar ile yapılan görüşmeler sonucunda işleme yönelik bir dizi sorun tespit edilmiştir. Bu sorunların bir kısmı teknik, bir kısmı da kullanıma ilişkin sorunlar olarak karşımıza çıkmıştır:

1. LTB ile üretilen sonuç listesi çok kalabalık çıkmaktadır. Anımsama (*recall*)<sup>8</sup> değeri çoğu çalışmada yüksek olsa da, Duyarlık (*precision*)<sup>9</sup> değeri oldukça düşük çıkmaktadır. Diğer bir ifade ile sistem bulması gereken tüm yeni ilişkileri bulmaya çok yakın olsa da, bulduğu doğru ya da yanlış toplam ilişki sayısının yüksekliği sebebiyle, doğru olması ihtimali yüksek ilişkiler gürültülü sonuç içerisinde kaybolmaktadır.
2. Terimler arası ilişkilerin, “terimlerin aynı metinde geçmesi” mantığı üzerine kurulması, her ne kadar işlenen metinlerin akademik olması sebebiyle anlamsal bir bütünlük sunuyor olsa da, gürültünün temel kaynağı olarak görülebilir.
3. Terimler üzerinden ilişkilerin kurulduğu yöntemde en büyük sıkıntı, geniş anlam içeren genel terimlerin ilişkiler arasında yer almasıdır. Bağlama terimi olarak genel terimlere izin verilmesi durumunda, çok sayıda ilişkisiz sonuç terimi ortaya çıkmaktadır. Genel terimlerin elenmesi, bir çok çalışmada genel kabul görmüş bir yöntem olmasına karşın, hangi terimlerin genel olduğunun belirlenmesi konusunda bir uygulama birliğinden söz edilemez. Özellikle terimlerin genelliğinin kullanıldığı alana bağlı olduğu düşünüldüğünde, alanlara göre genellik tanımı yapılması ön plana çıkmaktadır.
4. Bağlama terimleri doğru ilişkilerin kurulmasında en önemli aşamayı oluşturmaktadır. Bağlama terimleri ile başlama ve sonuç terimleri arasındaki ilişkiler ilk başta aynı metinlerde yer almaları üzerinden kurulmaktadır. Dolayısıyla hatalı bağlama terimleri çok sayıda hatalı sonuç ilişkisinin kurulmasına sebep olacaktır. Bağlama terimleri üzerinde budama yöntemleri işletilerek gürültü azaltılmaya çalışılsa da,

---

<sup>8</sup> Anımsama, sistemin getirdiği ilişkili dokümanın, sistemde var olan ilişkili dokümana oranı.

<sup>9</sup> Duyarlık, sistemin getirdiği doğru ilişkili doküman sayısının toplam getirilen doküman sayısına oranı.

doğru yöntemlerin hassasiyetle uygulanmaması durumunda özel sayılabilecek ilişkilerin de yok edilmesi ihtimali bulunmaktadır.

5. Çalışmaların sonuçlarını sınanması ve diğer çalışmalar ile karşılaştırmaların yapılmasına yönelik doğruluğu genel kabul görmüş bir sına kümesi bulunmamaktadır.

#### 1.4. Çözüm Önerileri

Tez çalışmaları sonucu, tespit edilen sorunlara ilişkin bazı çözüm önerileri sunulmuştur. Tez çalışması kapsamında çözüm önerilerinin bazıları uygulanmış ve sonuçları bu belgenin ilerleyen kesimlerinde paylaşılmıştır. Ancak sonraki çalışmalara da ışık tutabilmek amacıyla belirlenen tüm çözüm önerileri burada listelenmiştir. Tartışmaların yapıldığı başlık altında bu öneriler veriler ışığında detaylandırılmıştır.

1. Başlangıç teriminin geçtiği belgeler yerine, başlangıç terimi ile ilişkili olabilecek (doğrudan metinde geçmese de) belgelerin belirlenmesi, kullanıcının arama üzere başlangıç noktasının daha iyi belirlenmesini sağlayabilir.
2. Belgelerin kümelenmesi ile bağlama terimlerinden yola çıkılarak hazırlanan sonuç terimlerinin mantıksal olarak kümelenmesine yardım edecektir.
3. Sonuç terimlerinin tespit edilmesi kadar, kullanıcıya önem sırasına göre doğru sıralanması da büyük önem taşımaktadır. Sonuç terimlerinin kıymetlerinin belirlenmesinde, kapsadıkları belge kümelerinin arama terimi ile olan yakınlıkları kullanılabilir.
4. Olası tüm çalışma alanlarının algoritmik olarak belirlenmesi, arama sonuçlarının sınırlandırılmasında faydalı olabilir. Ayrıca genel terimlerin çalışma alanlarına göre tespit edilmesi imkanı da ortaya çıkar.
5. Uzmanlar tarafından güvenilirliği testlerle kanıtlanmış ve genel kabul görmüş bilgilerin, nitelikleri belirlenerek tümü aynı standartlar içinde saklandığı ortamları, yapısal bilgi kaynakları olarak adlandırabiliriz. Yapısal bilgi kaynaklarının LTB bağlamında yoğun kullanım imkanları bulunmaktadır.

- a. Sadece yayınlar üzerinden deęil, yapısal veri tabanları üzerinden de keşiflerin yapılması imkanı doğar.
- b. Genel terimler daha iyi tespit edilebilir.
- c. Uygunsuz ilişkilerin elenmesinde doğruluęu kanıtlanmış bilgilerden faydalanılabilir. Böylelikle terimlerin semantik olarak elenmesi yerine mantıksal olarak elenmesi mümkün olabilir.

## 2. Kapsam

Tez kapsamında yapılan çalışmada Literatür Tabanlı Bilgi Keşfi tanımında yer alan “ayrık ama tamamlayıcı” temel mantığı korunmuş ancak işletim modeli tamamen yeni bir mantıkla kurgulanmıştır. Bu yeni kurguya göre, literatürde yer alan yayınlar kümelenmiştir. Arama teriminin geçtiği yayınların bulunduğu kümeler bağlama işleminin yapılmasını sağlamaktadır. Arama terimini içeren yayınlarla aynı kümelere bulunan ve arama terimini barındırmayan yayınlarda bulunan terimler eğer uygun kriterleri sağlıyorsa sonuç terimi olarak listelenmektedir. Aranılan kriter daha önce başka bir yayında arama terimi ile bir arada geçmemek ve arama terimi yakın anlamlı olmamak olarak sıralanabilir. Bu yeni kurguda bağlama terimleri gözle görünür olmaktan uzaklaşmış, modelin işletimi doğal olarak bağlama işlemini yerine getirmeye başlamıştır.

Yukarıda anlatılan modelde, yayınların temin edilmesi ve indekslenmesi işleminde bilgi erişim (*information retrieval*) ve Vektör-Uzay Modeli (*vector space model*) yöntemlerinden faydalanılmıştır. Metinlerin kümelenmesi ve kümelerin incelenmesinde Metin Madenciliği (*Text Mining*) ve Belge Kümeleme yöntemlerinden faydalanılmıştır. Nihai olarak genel işleyiş Kaba Küme Teoreminden (*Rough Set*) etkilenmiştir.

Tez çalışmasının detaylarının anlaşılabilmesi için, yukarıda bahsedilen bilgisayar bilimlerinde temel çalışma alanlarında, bir takım bilgilere sahip olunması gerekir. Tezin anlaşılabilirliğini artırmak amacıyla fayda verici bilgiler bu başlık altında anlatılmıştır.

### 2.1. Bilgi Erişim

Sürekli büyüyen veri kümeleri üzerinden etkin ve güvenilir şekilde bilgi çekmek ve analiz etmek için pek çok metod geliştirilmiştir. Verinin temizlendiği ve organize edildiği bu süreçte kullanıcının veriyi en anlamlı şekilde görebilmesi için eldeki ham veri, araçlar kullanılarak dönüştürülür[39].

Bilgi Erişim (*Information Retrieval*) ve Veri Madenciliği (*Data Mining*) büyük veri kümelerinin yönetimi ile bilgi edinme temeline dayanır. Veri madenciliği

pek çok deęişik algoritma kullanarak veriler üzerinden bilgi edinmek için örüntüler bulma işlemidir. Bilgi erişimi, veri tabanından bilginin çekilmesi ve organizasyonu üzerinde yapılan çalışmadır [40].

Metin Erişim (*Text Retrieval*) sistemleri, organize edilmemiş geniş veri kümeleri üzerinde çalışır. İlişkili verilerin kullanıcıya sunulması ve uzun vadede verilerin organizasyonu olmak üzere iki temel alana ayrılır. Eş anlamlı ve çok anlamlı kelimeler metin erişim sistemleri tarafından adreslenmiş iki önemli konudur. Bunun yanında deyimler ve nesne tanıma konuları da ele alınır.

Belirtilen problemleri çözmek için anahtar kelimeleri indeksleyen “tersine işlem dosyaları” (*inverted files*) kullanılır. Anahtar kelimelerin doküman açısından anlam taşıyıp taşımadığı Zipf’s kuralı veya TF.IDF (term frequency–inverse document frequency)[41] algoritmaları kullanılarak belirlenir. Bunların yanında aramayı daha etkili hale getiren “*stopword list*”, “*stemming*”, deyim indeksleme metotları bulunmaktadır.

En popüler sistemlerden biri kullanıcının “ve” “veya” gibi mantıksal operatörler kullanarak arama yaptığı Mantıksal Anahtar Kelime Metin Erişim (*Boolean keyword text retrieval*) sistemleridir. Bu sistem “*don’t care*” karakterleri işleyecek şekilde genişletilebilir [42]. Popüler bir model olan vektör-uzay modeli [41], dokümanı anahtar kelime sayısı kadar boyut içeren vektörler olarak gösterir. Vektörler anahtar kelimenin varlığına göre 1 veya 0 değerini alır. Karşılaştırma vektörler üzerinden yapılır.

Bilgi Erişim değerlendirmede iki önemli kriter vardır.

1. Anımsama (*Recall*): Sistemin getirdiği ilişkili dokümanın, sistemde var olan ilişkili dokümana oranı.
2. Duyarlık (*Precision*): Sistemin getirdiği doğru ilişkili doküman sayısının toplam getirilen doküman sayısına oranı.

Veri madenciliği geniş miktarda veriden bilginin çıkarılması veya keşfi olarak tanımlanabilir[39]. Bilgi Erişim ve Veri Madenciliği sistemlerinin temel farkı hedefleridir. Bilgi Erişim sistemleri, bilginin elde edilmesi için arama

yapılmasına yönelik iken [41], veri madenciliği sistemleri veriler arasındaki ilişkilerin farklı teknikler ile analiz edilmesine yöneliktir.

Veri madenciliğinin önemli bir bileşeni olan “kavram tanımlama” (*concept description*) verinin çeşitli hesaplama teknikleri ile özetlenmesidir [43], [44]. Kavram Tanımlama iki alana bölünür. Karakterizasyon analizi (*Characterization analysis*) verinin özetini oluşturmak için veri genellemesi uygular. Ayrım analizi (*Discrimination analysis*) ise veri kümeleri arasındaki ayırt edici özellikleri vurgular.

Birleşim Kural Madenciliği (*Association rule mining*) [45], [46] büyük veri kümeleri arasında ilginç ilişkiler bulma sürecidir. İlginçliğin ölçülmesi iki parametre ile sağlanır. Kuralın yararlılığı destek (*support*) değeri ile, kesinliği güven (*confidence*) değeri ile ölçülür. Priori algoritması ise veri madenciliğinin temelinde yatan birleşim kural madenciliği prensiplerini gösterir.

Sınıflandırma, bilinmeyen bir veri için, bilinen bir veri sınıfı bulma ile ilişkilidir. Karar ağacı, Bayesian sınıflandırması gibi temel modeller iki ortak adıma sahiptir: Eğitim (*training*) ve sınıflandırma (*classification*) aşaması. Eğitim aşaması önceden belirlenmiş veri setleri üzerinden yapılır. Sınıflandırma işlemi ise *supervised* öğrenme yöntemine girer.

Kümeleme işlemi herhangi bir ön bilgi olmadan verileri kümelere ayırma işlemidir [47], [48]. Eğitim süreci olmadığından çıkarım var olan bilgidен yapılır. En çok bilinen k-means algoritması ile birbirine yakın veriler gruplanmış olur.

## **2.2. Vektör-Uzay Modeli**

Büyük miktarda verilerin etkili bir şekilde işlenmesi kritik bir problemdir. Bunun için pek çok teknik geliştirilmiştir. En bilinen tekniklerden biri olan indeksleme, hangi terimlerin indeksleneceği, indekslenen bu terimlerin sorgu değerlendirmede nasıl kullanılacağı gibi soruları beraberinde getirmektedir. Burada asıl odaklanılması gereken konu dokümanın ve sorgunun birbirine ne kadar benzediğinin ölçümüdür.



Vektör uzay modeli bilgi çıkarımı, bilgi filtreleme, indeksleme gibi alanlarda kullanılan cebirsel bir modeldir. Doğal dil belgelerinin çok boyutlu uzayda özel bir anlamını simgelemektedir.

Salton tarafından önerilen vektör uzay modelinde [49], her doküman bir vektör ile temsil edilir. Bir vektörün boyutu indekslenen anahtar kelime sayısı kadardır.

t boyutlu bir doküman şu şekilde ifade edilir;

$$D_i = (d_{i1}, d_{i2}, d_{i3}, \dots, d_{it})$$

Burada ki  $d_{ij}$  ifadesi j. terimin ağırlığını belirtir. Bu ağırlıklandırma Boolean Model'e göre ikili değerler olarak da yapılabileceği gibi en çok bilinen model olan *Term-Frequency Inverse Document Frequency Model* (TF-IDF) ile yapılır.

TF-IDF, yayınların terimlerle anahtarlanmasına yönelik bilgi erişimi (*information retrieval*) çalışmalarında yaygın olarak kullanılmaktadır [50]. Kelimenin bir belge için ne kadar önemli olduğunu, diğer yayınları da göz önünde bulundurarak hesaplamak amacıyla kullanılır. Örnek olarak, bir terim bir belgede çok sık geçtiği halde, diğer yayınlarda az sayıda geçiyorsa, o terim o belge için belgede geçen diğer terimlere göre daha önemli olabilir. TF – IDF yöntemi, daha gelişmiş modelleri ile arama motorlarında sıklıkla kullanılmaktadır.

$t_i^b$  b ve l terimlerinin beraber geçtiği doküman sayısı,

n derlemdeki toplam doküman sayısı ve

$t_i$  l teriminin geçtiği doküman sayısı olmak üzere

$$TF - IDF_i^b = TF_i^b \times IDF_l = t_i^b \times \log \frac{n}{t_i} \quad (3.1)$$

Vektör Uzay Modelinde TF-IDF değeri şu şekilde ifade edilir;

$$w_{t,d} = tf_{t,d} \cdot \log \frac{|D|}{|\{d' \in D \mid t \in d'\}|} \quad (3.2)$$

Terim sıklığı (  $tf_{t,d}$  ) : t teriminin d belgesindeki terim sıklığı.

Devrik Belge Sıklığı (  $idf_t$  ) :  $|D|$  toplam doküman sayısı,  $|\{d' \in D \mid t \in d'\}|$  t teriminin geçtiği toplam belge sayısı.

İki vektör arasındaki benzerlik,  $s(D_i, D_j)$ , kosinüs benzerliği ile bulunabilir.

$$\cos \theta = \frac{D_1 \cdot D_2}{\|D_1\| \cdot \|D_2\|} \quad (3.3)$$

$D_1 \cdot D_2$  : Dokümanların vektör çarpımı ,

$\|D_1\|$  : Vektörün boyu

### Performans Değerlendirme

Bir sorgu değerlendirilirken öncelikle sorgu bir vektöre dönüştürülür. Her bir doküman ve sorgu arasındaki benzerlik bulunur. Benzerlik bulma için kosinüs benzerliğine ek olarak TF-IDF modeli kullanılabilir. Benzerlik değeri 1'e yaklaştıkça sorgu ve doküman arasındaki benzerlik artar.

$$s(d_j, q) = \frac{d_j \cdot q}{\|d_j\| \cdot \|q\|} = \frac{\sum_{i=1}^N w_{i,j} w_{i,q}}{\sqrt{\sum_{i=1}^N w_{i,j}^2} \sqrt{\sum_{i=1}^N w_{i,q}^2}} \quad (3.4)$$

Sorgu sonuçlarına göre değerlendirme de iki önemli performans kriteri olan anımsama ve duyarlık değerleri hesaplanır.

- Anımsama (*Recall*) : Çıktıdaki ilişkili dokümanın, toplam ilişkili dokümana oranı
- Duyarlık (*Precision*) : Çıktıdaki ilişkili doküman sayısının, çıktıdaki toplam doküman sayısına oranı.

Vektör uzay modelinin lineer cebir üzerine dayalı basit bir model olması, doküman ve sorgu arasındaki benzerliğin hesaplanabilmesi ve bir dokümanın diğer bir doküman ile ne kadar ilişkili olduğunun saptanabilmesi avantajları arasında sayılabilir.

### **2.3. Metin Madenciliği**

Metin madenciliği, tek tek incelenmesi mümkün olmayacak kadar çok belgenin, belirli bir amaca yönelik taranması, indekslenmesi ve bilgi çıkarımı olarak tanımlanabilir [51]. Veri madenciliği alanında yapılan çalışmaları metin madenciliğine uyarlamak mümkün olsa da, araştırmanın temelinde yatan belgenin tanımlanması ve bu tanımdan elde edilecek özelliklerin niceliği ve niteliği bağlamlarında metin madenciliği ve veri madenciliği büyük ayrım göstermektedir.

Metin madenciliğinde en önemli aşama, belgenin çözülmek istenen soruna göre talep edilecek özelliklerinin belirlenmesidir. Ortaya konacak nitelik kümesinin, çoğu durumlarda külliyattaki (*corpus*) tüm belgeler için anlamlı olamayacağı, diğer bir deyişle, ne kadar iyi çalışılsa da bazı özelliklerin bazı belgeler için boş olabileceği göz önünde bulundurulmalıdır.

Metinlerin özelliklerinin belirlenmesinde sıklıkla kullanılan bir yöntem, metin içerisinde yer alan terimlerin tespit edilmesidir. Ancak terimlerin çoğu durumda ince detay bilgileri içeriyor olması, her bir metnin diğerlerinden, detay derinliği fazla olmak üzere mutlaka ayrılmasını gerektirir. Özellikle metinlerin kümelenmesinin gerekli olduğu durumlarda, metinleri terimlerle temsil etmek, ortak yönlerin azalması bağlamında, başarımın üzerinde olumsuz etkileri olabilmektedir [51].

Metinlerin terimlerle temsil edilmesi sorunun aşılması gerektiği durumlarda, kavramlar ön plana çıkmaktadır. Kavramlar istatistiksel veya kural tabanlı üretilmiş; ya da okunarak tek tek belirlenmiş metin özellikleridir. Kavramlar her ne kadar metin içerisinde yer almak zorunda olmasa da, o metinde anlamsal düzeyde yer alması yeterli sayılabilir. Özellikle metinlerin kümelenmesinde kavramları kullanmak, niteliğin çokluğu sebebiyle bazı imkansız problemleri çözülebilir hale getirmektedir. Çalışmamızda, terimler ayrıntı düzeylerine göre elenmiş, ataları kavram olarak o belgeyi temsil edecek şekilde düzenlemeler yapılmıştır.

Metin madenciliğinde önemli bir diğer konu da, bağlam (artalan) bilgisidir. Bağlam bilgisi, metinlerden terimlerin ya da kavramların üretilmesinde kullanılabileceği gibi, doğal dil işleme gibi algoritmalara da yardımcı olabilmektedir. Metin madenciliği ile elde edilen bilginin kullanıcıya sunulma aşamasında da bağlam bilgisinden faydalanılmakta, kullanıcının sonuçları yorumlayabilmesi için bu bilgiler arayüzler üzerinden kullanıcıya sunulmaktadır.

Çalışmamız ile ilişkili olması sebebiyle, biyomedikal metin madenciliği de kapsam içine alınmıştır. Artan sayıda akademik yayınlar, ve bu yayınların PubMed üzerinde bilimsel yöntemlerle indekslenmiş olması, biyomedikal metin madenciliği için çok kıymetli bir çalışma alanı imkanı sağlamaktadır.

Biyomedikal metin madenciliği ile metinler protein, gen vb. biyolojik varlıkların tespit edilmesi, bu varlıklar arası etkileşimlerin ve işlevselliklerin çıkarımı gibi amaçlara hizmet etmektedir [52]. Ancak bu noktada, varlıkların eş anlamlı kelimelerle ifade edilmesinden kaynaklanan sorunlar karşımıza çıkmaktadır. Benzer şekilde eş seslilik de başkaca bir sorun olarak görülmektedir. Bu gibi durumlara çözüm olarak terimlerin anlamlandırılmasında, metinde geçen diğer terimler ile metin geneli anlaşılmaya çalışılmaktadır. Diğer bir ifade ile terim ancak metindeki diğerleri ile beraber bir anlam ifade etmektedir.

#### **2.4. Birleşik Medikal Dil Sistemi**

Birleşik Medikal Dil Sistemi (Unified Medikal Language System – UMLS) [53], Amerika Birleşik Devletleri'nin Ulusal Tıp Kütüphanesi (National Library of

Medicine – NLM) tarafından geliştirilmiş ve ücretsiz olarak kullanıma açılmıştır. Tıp bilimlerinde kullanılan tüm terimleri kapsamaya çalışan bir sözlük olarak hizmet vermektedir. Terimler arasında haritalamaların yapılabildiği, çeşitli terminoloji sistemleri arasında çevirim olanağı sağlayan yapıları da barındırmaktadır. UMLS, son kullanıcıların yanında bilişim sistemlerinin biyomedikal ve tıbbi yayınları, metinleri işlemelerine olanak verecek alt yapıyı da sağlamaktadır. Tez çalışması kapsamında UMLS'den biyomedikal kavramların ontolojik yapılarının temin edilmesinde büyük oranda faydalanılmıştır.

UMLS şu üç önemli modülden oluşmaktadır:

1. Metathesaurus: 1 milyondan fazla biyomedikal kavramı ve bu kavramlar arası bağlantıları, içerdiği çok sayıda sözlük ile organize eder.
2. Anlamsal Ağ: Kavramların olası tüm anlamsal türlerini barındırmak üzere organize edilmiş, 135 kavramsal tür ve bu türler arasında toplam 44 tane ilişki barındıran yapıdır.
3. Sözlük Uzmanı: NLP (Natural Language Processing – Doğal Dil İşleme) algoritmalarının kullanımına uygun olarak hazırlanmış sözlük/terim veritabanıdır.

### *Metathesaurus*

Metathesaurus TIP alanında hizmet veren sözlüklere ilişkin bir veri tabanıdır. Tıp bilimleri ile ilgili terimleri, anlamdaş/benzer ifadeleri ve diğer terimler ile aralarındaki ilişkileri yapısal olarak sunan birden çok sözlüğü barındıran veri tabanıdır [54]. Metathesaurus'ta alışıldık sözlüklerde yer alan terimler veya kelimelerin yanında bunları üstten kapatan kavramlara da yer verilmiştir. Bir kavram birden çok kelime veya terim ile ilişkilendirilebilir, farklı sözlüklerdeki terimleri gösterebilmektedir. Kavramların bilişimsel amaçlı kullanımlarını kolaylaştırabilmek amacıyla biricik bir numara (*cui - concept unique identifier – kavramsal biricik anahtar*) ile etiketlenmesi sağlanmıştır.

Metathesaurus, terimlerin analiz edildiği ve birbirleri ilişkilendirildiği tez çalışması için çok önemli bir kaynağı teşkil etmektedir. Tez çalışması

açısından en önemli özelliği, terimleri bir arada yapısal olarak sunması ve birden çok sözlükle çalışmak yerine standartları konulmuş tek bir yapıdan destek alınması bağlamında elde edilen kolaylıktır.

Metathesaurus'un birlikte çalıştığı sözlüklere göz gezdirilmesi içeriğini kavrayabilmek açısından önemlidir. Diğer taraftan Metathesaurus kavramları kategorilere göre de indekslemiştir. Bu açıdan kategorilerinde bilinmesi gerekir.

Metathesaurus'un barındırdığı kategoriler:

- Hastalık Teşhisi (*Diagnosis*) - LOINC
- İşlemler & Malzemeler (*Procedures & Supplies*) - CPT
- Hastalıklar (*Diseases*) - ICD-10
- Geniş Kapsamlı Sözlükler (*Comprehensive Vocabularies/Thesauri*) - SNOMED CT
- Vücut yapısı (*Anatomy*),
- İlaçlar (*Drugs*),
- Genetik (*Genetics*),
- Hemşirelik (*Nursing*)

Metathesaurus yapısal şekline baktığımız zaman, 40 tane veri, dizin ve üst veri dosyası ile karşılaşırız. Çizelge 2'de Metathesaurus kapsamındaki veri kümeleri verilmiştir.

Veri Kümesi	Açıklama
MRCONSO.RRF	İsimler, eşanlamlılar, terimler, terim türleri, kodlar
MRREL.RRF	Bağıntılar
MRHIER.RRF	Hiyerarşik sıralar
MRSAT.RRF	Öznitelikler
MRDEF.RRF	Tanımlar
MRMAP.RRF	Haritalar
MRSMAP.RRF	Basit haritalar
MRSTY.RRF	Anlamsal türler

Çizelge 2. Metathesaurus kapsamındaki veri kümeleri [38], [54]

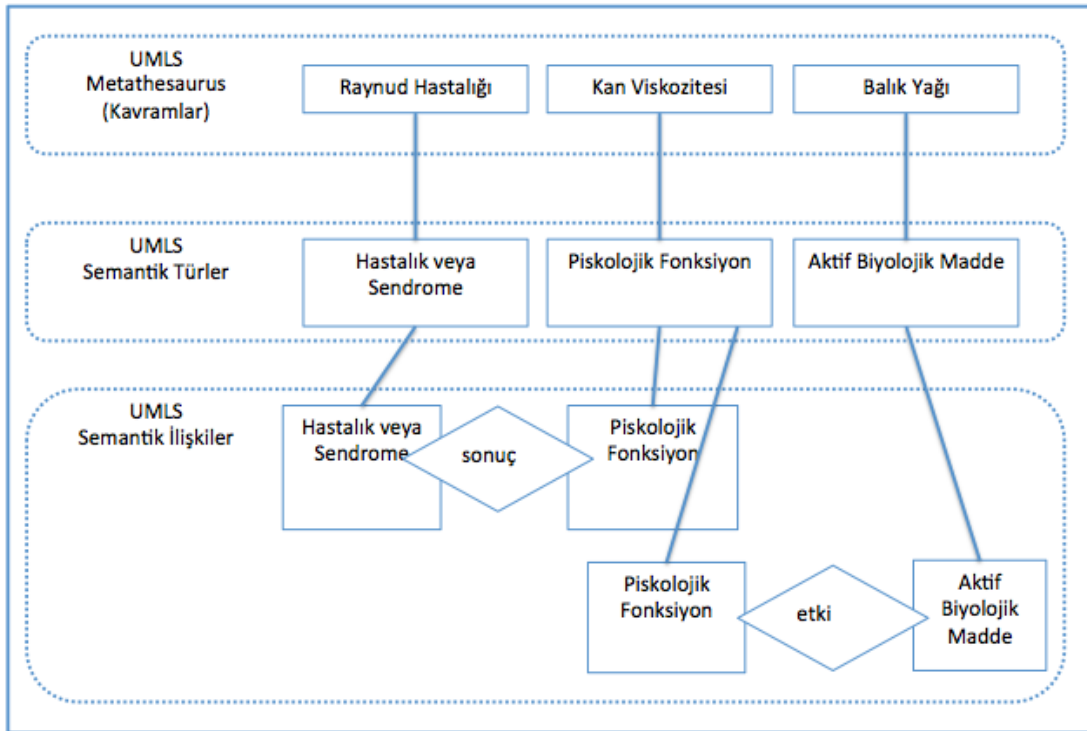
## Anlamsal türler ve anlamsal ağ

Metathesaurus sadece terimleri değil, anlamsal ifadeleri de yapısal biçimde tanımlamaktadır. Bu amaçla kurgulanmış anlamsal ağ (*semantic network*) anlamsal türleri ve aralarındaki bağları içerir. Anlamsal türleri geniş konu kategorileri (ilaçlar, semptomlar, hastalıklar vb.) olarak ifade edebiliriz [54].

Metathesaurus'taki her bir kavram, bir veya birden fazla anlamsal tür indekslendirilmiştir. MRSTY.RRF veri dosyasında bütün anlamsal türler listelenmiştir. 135 tane anlamsal tür ve 54 tane anlamsal bağlantı vardır [54].

Her bir anlamsal türler kendine özgü numaraları, hiyerarşideki yerini gösteren ağaç konum numaraları, anlamsal tanımları ve ataları ve çocukları ile kurgulanmıştır. Her bir anlamsal bağlantı ise, kendine özgü numaraları, ağaç konum numaraları, tanımları, örneklerini ve bu bağlantıyla bağlanabilecek anlamsal türleri ile kurgulanmıştır.

Şekil 6'da LTB keşfi için sıklıkla örneklenen "Raynaud Disease", "Blood Viscosity" ve "Fish Oils" kavramları, anlamsal türleri ve bu anlamsal türlerin arasındaki anlamsal bağıntıları gösterilmektedir.



Şekil 4. UMLS Modülleri [51]

Diğer BMDS araçları

UMLS projesi, metin madenciliği içinde önemli, erişimi düzenleyen ya da çıkarsamalar için referans alınabilecek uygulama modülleri sunmaktadır [54].

Bunlardan başlıcaları

- UMLS Bilgi Kaynak Sunucusu (UMLSKS) : Metathesaurus'a yerel bilgisayarlardan, daha hızlı sürücüler ile internet tabanlı erişim için kullanılır.
- MetamorphoSys : Metathesaurus'u özel uygulamalarda, uygulamalara özel alt kümeler oluşturmak için hazırlanmış bir araçtır.
- MetaMap: Bir yazı verildiğinde, yazıda geçen Metathesaurus kavramlarını bulmak için kullanılır. Metnin parse edilmesi, terimlerin çekim hallerine ve köklerine göre incelenmesi sonucu, metin içerisindeki UMLS terimlerini işaretlenmesini sağlar. Ancak terimin kullanım anlamı veya metin ile ilişkisi hakkında bilgi sunmaz.

## 2.4. Belge Kümeleme

Belge kümelenme (aynı zamanda metin kümeleme olarak da adlandırılır) veri kümelenme kavramı ile yakından ilgilidir. Belge kümeleme denetimsiz belge organizasyonu, otomatik başlık çıkarma ve hızlı bilgi alma ya da filtreleme için özelleşmiş bir tekniktir [55].

Metinleri kategoriler halinde kümeleme, indeksleme, bilgi geri alma, yönetim ve WEB üzerinde ya da bir kurumda çok büyük miktarda metnin veri madenciliği ile işlenmesi için çok önemli bir adımdır. Metin kümelemede, bilindik en büyük sorun, büyük hacimli, yüksek boyutluluk ve karmaşık semantiklere sahip metinler üzerinde çalışılmasıdır [56].

Bir web arama motoru genellikle geniş bir sorguya yanıt olarak binlerce sayfa döndürür. Bu durum kullanıcılar açısından, gerekli bilgilerin araştırılması ve tanımlanmasında büyük zorluklar oluşturabilmektedir. Kümelenme yöntemleri, sorgu sonucu listelenecek belgelerin, otomatik olarak anlamlı bir kategori listesi içine gruptandırmasında kullanılmaktadır.



Belge kümeleme tanımlayıcıların belirlenmesini ve tanımlayıcıların metin içerisinden çıkarımını gerekli kılar. Tanımlayıcılar, kümenin içeriğini açıklayan kelime setleridir. Diğer taraftan standart belge kümeleme yöntemlerinde öncelikle belgelerin doğru kelimelerle nitelendirilmesi gerekir. Belgeler nesne olmak üzere, belge içerisinde geçen bazı kelimelerde nitelikleri olduğu durumda belge kümeleme problemi, standart kümeleme problemine dönüşmektedir. Sıklıkla kullanılan kümeleme algoritmaları, çok büyük eleman sayıları üzerinde çok sayıda nitelik ile çalışacak şekilde geliştirilmesi durumunda, belge kümeleme üzerine uygulanabilmektedir.

Hotho ve arkadaşlarının, belge kümeleme işlemlerinde sözlüklerin katkısını, İngilizce sözcük veri tabanı olan Wordnet<sup>10</sup> bağlamında inceleyen çalışmalarında [57], ontoloji kullanmanın yayınları kümeleme üzerine etkisini göstermişlerdir. Yayınlar, kelimeler ile değil ontolojideki kavramları ile tanımlanmıştır. Ontolojide kavramların frekansları, atası ve çocukları bilgilerini kavramları budamak için kullanılmıştır. Kavram vektörünün terim vektörüne göre a) anlamdaş sözcüklerin teke indirilmesi b) daha fazla genel kavramlar üzerinde çalışıldığından, yayınların hangi konu üzerine yazıldığına anlaşılmasına yönelik başarı artışı sağlanmıştır.

Ontolojilerin kümeleme üzerine etkisini gösteren bir diğer çalışmada da yazarlar [58], *Term Mutual Information* (TMI) ile iki terimin birbirine olan bağımlılığı ontoloji üzerinden hesaplanarak klasik Vektör-Uzay Modeli de terimin belge üzerindeki ağırlığını farklı bir yöntemle ifade etmişlerdir. Makalede geçen bir terimin TFxIDF değerinin üstüne, ontolojiden yararlanarak o terimin makalede geçen diğer terimler ile bağımlılık değerleri eklenerek, makale için daha anlamlı olması amaçlanmıştır. Tez kapsamında yapılan çalışmada da ontoloji olarak mesh ağacını kullanılmış ve benzer şekilde elde edilen bu bilgi TFxIDF değerinin üzerine eklenerek Vektör-Uzay Modeli oluşturulmuştur.

MeSH ontolojisi ve MEDLINE belgeleri üzerine Jing ve arkadaşlarının yaptığı bir çalışmada [59], MESH ontolojisinin MEDLINE belgelerin kümeleneğinde

---

<sup>10</sup> <http://wordnet.princeton.edu/>

nasıl yararlanılabileceği araştırılmıştır. Yapılan çalışmada, belgelerde geçen mesh terimleri, mesh ontolojisindeki kavramlar ile eşleştirip, anlamdaş olanlar ve genel terimler elenmiştir.

MeSH ontolojisi ve MEDLINE üzerine Zhu ve arkadaşlarının yaptığı bir diğer çalışmada [60], MeSH sözlüğünü kullanan yöntemlerin yaptığı hatalar ve (I) Otomatik olarak işaretlenen MeSH kavram vektörleri orijinal metindeki semantik anlamı doğru vermeyebilecek olması, (II) orijinal metnin içerik bilgisi önemli olabileceği ele alınmıştır. MeSH sözlüğünü kullanarak iki belge arasındaki benzerliği ölçen yeni bir yöntem geliştirmişlerdir (MeSH kümelerinin benzerliği). Semantik ve içerik benzerliklerini birleştirerek belgeler arası benzerlik matrisi üzerinde spektral yaklaşım ile belgeleri kümelemişlerdir.

Vektör-Uzay Modeli (VUM) üzerine Jing, Ng ve Huang'ın çalışmasında [56], terim bazlı ve bilgi bazlı VUM yapılarını karşılaştırılmıştır. Bilgi bazlı modelde WordNet ve MeSH ontolojileri ile iki terim arasındaki ortalama uzaklık, kenar-sayma (edge-counting) tekniğine göre hesaplanarak yeni benzerlik değeri kullanılmıştır. Ortalama uzaklık, bir terim birden fazla kavrama karşılık geldiği için ontoloji ağacında farklı yerlerde olabilir. İki terim arasındaki uzaklık hesaplanırken kavramlarının birbiriyle olan uzaklıklarının ortalaması alınmıştır.

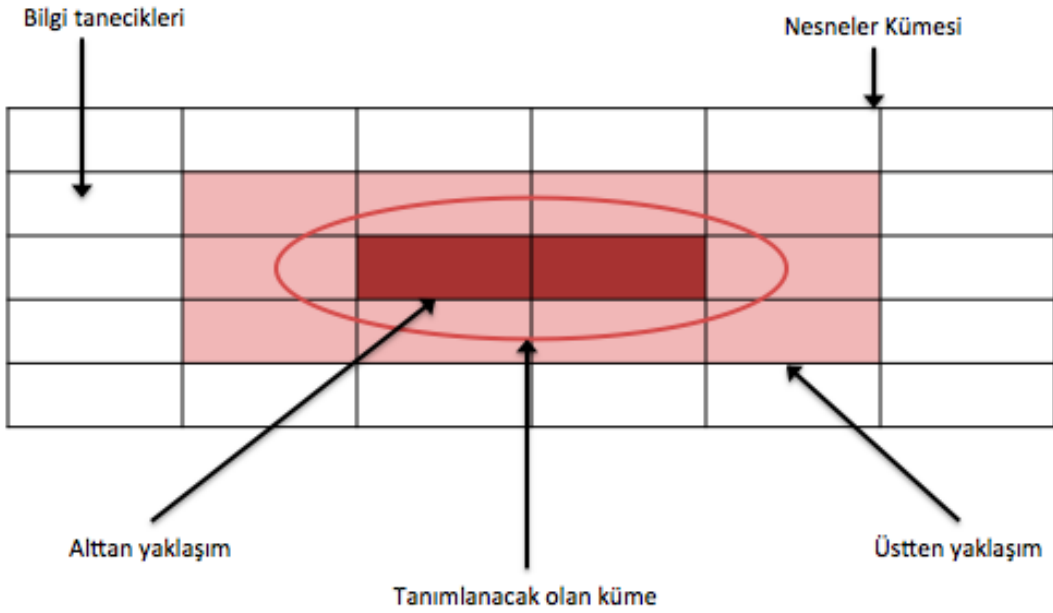
Bader Aljaber ve arkadaşları, 2009 yılında yayınladıkları çalışmalarında [61], belirsizlik (bir kelime kullanıldığı yerde farklı anlam içermesi) ve benzerlik (birbirinin aynı anlamına sahip kelimelerin farklı belgelerde farklı anlamda olması) içeren kelimeleri kullanan *bag-of-word* temelli modeller yerine kelimenin anlamını alıntının (*citation*) geçtiği yere göre çıkararak bir model önermişlerdir.

Zhang ve arkadaşlarının 2010 yılında yayınlanmış bir çalışmalarında da [62], terim benzerliklerinin ölçülmesinde kullanılan yöntemlerin karşılaştırılması yapılmıştır. Terimlerin yeniden ağırlıklandırılmasının nedenleri (I) genel terimlerin etkilerini kaldırılmaması, fakat genel terimin belgeleri çok da fazla etkilememesi, (II) hiyerarşik olarak terim kümeleme sayısal olarak çok fazla işlem gerektiriyor ve terim ağırlıklandırmaya göre ölçeklikten yoksun olması gösterilmiştir. Sonuç olarak olası yöntemlerin birbirleri üzerine önemli bir farkla üstün gelmedikleri görülmüştür. Terimlerin yeniden ağırlıklandırılması,

kümeleme yöntemine destek vermiş ve ağırlıklandırma ilişkin yöntemlerden “Mao yöntemi” bu yayından esinlenerek tez çalışmasında kullanılmıştır.

## 2.5. Kaba Kümeler

Bilgisayar bilimlerinde ilk defa Polonyalı bilgisayar bilimcisi Zdzisław I. Pawlak tarafından açıklanan kaba kümeler (“*rough sets*”) [63], alttan ve üstten yaklaşan net kümelerle (yani alışıldık kümeler) biçimsel olarak tanımlanabilen kümeyi ifade etmektedir. Kümenin tanımı net bir şekilde ortaya konamasa da, mutlaka içinde barınan elemanlardan oluşan alttan yaklaşım kümesi ve tüm elemanlarını (ve belki başkaca elemanları da) içeren üstten yaklaşım kümeleri ile tanımlanmaya çalışılmaktadır. Şekil 8’de Kaba Kümelerin genel işletim mantığı gösterilmiştir.



Şekil 5. Kaba Kümeler genel görünüm [61]

### Kaba Kümeler Biçimsel Tanım

$$I = (U, A)$$

$$B = \{b \mid b \in A\}$$

$$IND_I(B) = \{(x, x') \in U^2 \mid \forall a \in B, a(x) = a(x')\}$$

$$\underline{B}X = \{x \mid [x]_B \subseteq X\}$$

$$\overline{B}X = \{x \mid [x]_B \cap X \neq \emptyset\}$$

$$BR_B(X) = \overline{B}X - \underline{B}X$$

I : Bilgi Sistemi

U : Boş küme olmayan sınırlı nesne kümesi

A : Boş küme olmayan sınırlı özellik kümesi

B : Boş küme olmayan sınırlı özellik kümesi, koşul özellikleri

X : Tanımlanmak istenen küme

IND<sub>I</sub> : Bilgi kümesi üzerinde ayırdedilmezlik

$\underline{B}X$  : X'i B özellik alt kümesi ile alttan tanımlayan küme

$\overline{B}X$  : X'i B özellik alt kümesi ile üstten tanımlayan küme

BR<sub>B</sub>(X) : X'in B özellik alt kümesi ile tanımlanan sınır kümesi

Örnek bir kümeyle temel kavramları daha rahat açıklanabilir. Bilgi uzayı olarak nesnelere ve özellikleri ile tanımlanan / yapısı aşağıdaki tabloda verilmiştir.

Nesne	P <sub>1</sub>	P <sub>2</sub>	P <sub>3</sub>	P <sub>4</sub>	P <sub>5</sub>
O <sub>1</sub>	1	0	1	1	0
O <sub>2</sub>	1	0	0	1	1
O <sub>3</sub>	1	0	1	1	0
O <sub>4</sub>	1	0	0	1	1
O <sub>5</sub>	1	1	0	0	0

Çizelge 3. Örnek Bir Kaba Küme Verisi

Nesne kümesi :  $U = \{O_1, O_2, O_3, O_4, O_5\}$

Özellik kümesi :  $A = \{P_1, P_2, P_3, P_4, P_5\}$

Ayırılmazlık kümeleri :  $IND_1 = \{O_1, O_3\}, \{O_2, O_4\}, \{O_5\}$

Rasgele verilmiş bir X kümesi için kesin tanım vermek mümkün değilse, X kümesi alttan ve üstten yaklaşan sınırlamalar ile tanımlanabilir.

Tanımlanacak küme :  $X = \{O_1, O_2, O_3\}$

Altan yaklaşım :  $\underline{P}X = \{O_1, O_3\}$

Üstten yaklaşım :  $\underline{\underline{P}}X = \{O_1, O_2, O_3, O_4\}$

Sınır Bölgesi :  $BR_B(X) = \underline{\underline{P}}X - \underline{P}X = \{O_2, O_4\}$

Kaba küme teorisi belirsizlik içeren bir veri kümesinden bilgi çıkarımını hedeflemektedir. Bu bağlamda temel çalışma konusu olan veri madenciliği ile Kaba küme teorisi benzer sorulara yanıt aramaktadır. Ancak kaba küme teorisi ile özelliklerin bağımlılığı ve sınıflandırma konularında daha iyi sonuçlar elde edilebilmektedir. Hipotez testleri ve bir öznitelik değerleri arasındaki ilişkiler kaba küme teorisi ile kolaylıkla çözülebilmektedir [64].

### 3. Yöntem

Tez kapsamında Literatür Tabanlı Bilgi Keşfi için uygulanan klasik model geliştirilmiş, literatür üzerinden elde edilen bilgiler ışığında yeni hipotezlerin üretilmesi mantığı korunurken, bir çok yapısal soruna yol açan ABC modeli küme mantığı üzerine oturtulmuş ve sınırları mantıksal düzenlemelerle daha az keskin hale getirilmiştir.

Uygulama metinlerin kümelenmesine dayanmaktadır. Kullanıcı tarafından verilen başlangıç terimi, bir kümede en az bir belgede geçiyorsa, o küme bağlama kümesi olarak işaretlenmektedir. Kümede yer alan diğer metinlerde geçen terimler, eğer daha önce başlama terimi ile ilişkisi tespit edilmemişse sonuç kümesine dahil edilmektedir. Sonuç terimlerinin ağırlıklandırılması ise, geçtiği kümelerin arama terimi ile olan yakınlığının toplamından elde edilmektedir. Tüm bu anlatılan model, temelde kaba küme teorimini çok yakından anımsatmaktadır. Bu yüzden matematiksel model kaba küme teoremi üzerine kurulmuştur.

Yukarıda kısaca özetlenen yöntemin matematiksel modeli anlatıldıktan sonra, modelin gerçekleştirimi için geçilen aşamalar, kullanılan algoritmalar ve gerçekleştirilen modüller ayrı ayrı anlatılmıştır.

### 3.1. Matematiksel Model

Yapılan tez çalışmasının genel görünümünün anlaşılması için önce matematiksel model kurulmuştur. Matematiksel model, başlangıç terimi ( $S_T$ )'yi içeren yayınların ( $L$  kümesi) alttan ve üstten yaklaşımla ( $\underline{C}L$  ve  $\overline{C}L$ ) tanımlanması ve böylelikle ortaya çıkan sınır bölgesindeki ( $BR_C(L)$ ) yayınların incelenmesiyle elde edilen yeni terimlerin ( $R$ ) sunulmasını göstermektedir.

#### Matematiksel Model – Tanımlar

LTB : Bilgi Sistemi

D : boş küme olmayan sınırlı yayın kümesi

T : boş küme olmayan sınırlı terim kümesi, MeSH terimleri

C : Kavramların kümesi olmak üzere

$S_T$  : Başlangıç terimi

L :  $S_T$  içeren yayınlar

$IND_{LTB}$  : Bilgi kümesi üzerinde ayırdedilmezlik

$\underline{C}L$  : L'yi alttan tanımlayan küme

$\overline{C}L$  : L'yi üstten tanımlayan küme

$BR_C(L)$  : Boundary Region of L using C concepts as attribute subset

R : Sistemin çıktısı, sonuç terimler kümesi

Ancak sınır bölgesine ulaşılabilmesi için, alttan ve üstten yaklaşım kümelerinin çıkarılması, bunun için de bilgi taneciklerinin, diğer bir ifade ile ayırdedilmezlik kümelerinin elde edilmesi gerekir. Bu işlem belgelerin kümelenmesi ile sağlanmıştır. Modeli karmaşıktırmamak amacıyla standart bir işlem olan kümeleme (*clustering*) işlemi modele dahil edilmemiş, adıyla anılarak geçilmiştir.

### Matematiksel Model - Çözümleme

$$LBD = (D, T)$$

$$C = \{t \mid t \in T, k_{1,2} \in N, |Child_t| < k_1 \wedge |Docs_t| < k_2\}$$

$$L_s = \{d \mid d \in D, \exists t(d) = S_T\}$$

$$IND_L(C) = \text{Document Clustering}$$

$$IND_{LBD}(C) = \{(d, d') \in D^2 \mid \forall t \in C, t(d) = t(d')\}$$

$$\underline{C}L_s = \{d \mid [d]_C \subseteq L_s\}$$

$$\overline{C}L_s = \{d \mid [d]_C \cap L_s \neq \emptyset\}$$

$$BR_C(L_s) = \overline{C}L_s - L_s$$

$$R = \{r \mid r \in T, d \in BR_C(L_s), r \in t(d) \wedge r \notin t(L_s)\}$$

Verilen matematiksel çözümde, R sonuç kümesi, sınır kümesindeki yayınlarda geçen terimlerden oluşur. Sınır kümesi, L'yi üsten tanımlayan kümeden L'deki yayınların çıkarılması ile tanımlanmıştır. L ise başlangıç teriminin geçtiği yayınları ifade etmektedir. Ayırdedilmezlik sınıfları ile aynı kavramlarla işaretlenmiş yayınları ifade etmektedir.

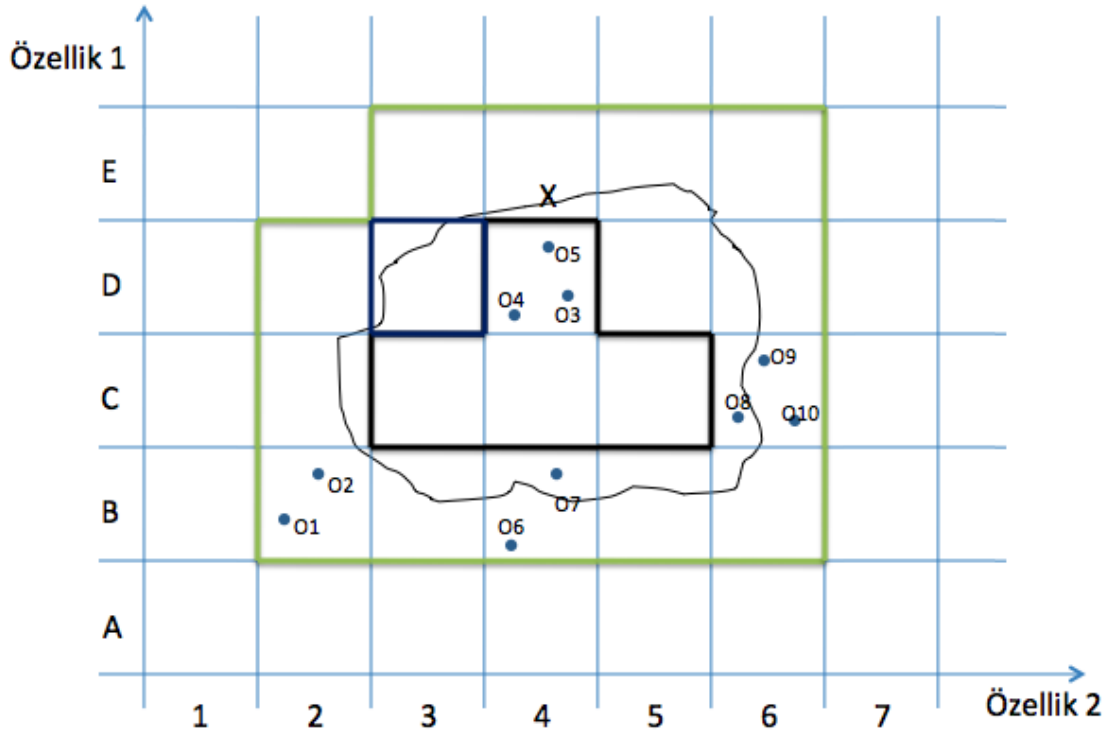
Bu noktada tekrar hatırlatılması gereken önemli bir yaklaşım karşımıza çıkar. Yayınları eğer terimleri kullanılarak kümelemiş olsaydık, çok küçük kümeler ortaya çıkacak ve küme sayısı çok fazla olacaktır. Bu durum ise sınıf kümelerinin daralmasına, yakın olduğu halde bir çok yeni hipotezin yakalanamamasına sebep olacaktır. Bunu engellemek için yayınlar, indekslemelerinde kullanılan terimlerle değil, o terimlerin dahil olduğu kavramlarla nitelenmiş ve kümeleme bu nitelermeler üzerine kurulmuştur. Bu konuyu ayrıntıları, uygulama modelinde belge kümelemenin nasıl uygulandığının anlatıldığı başlıkta ayrıntıları ile verilmiştir.



### 3.2. Kaba Küme Teorisi ve LTB

LTB, terimler arasında ortak geçtikleri yayınları kullanarak ilişkilerin kurulmasını esas almaktadır. Kaba kümelerin tez çalışmasına uyarlanması çıkış fikri, kullanıcıdan alınan terimle belirlenen yayın kümesinin sınırlarının biçimsel olarak çizilememesinden kaynaklanmaktadır. Bu bakış açısı ile kaba kümeler ise, terimlerin geçtiği yayınların toplam kümesinin birbirleri ile olan benzerlikleri üzerinden terimler arası ilişki kurmaktadır.

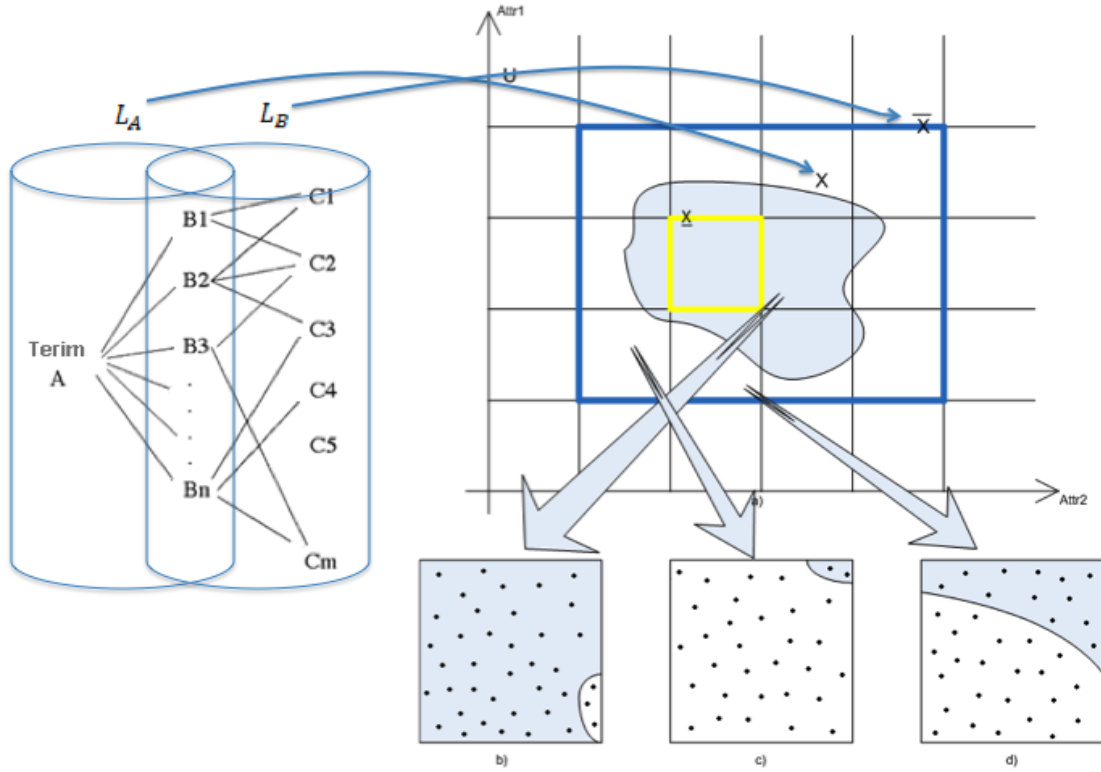
Bir terimin ya da kavramın geçtiği yayınları, daha önceden belirlenmiş yayın kümelerindeki dağılımı, ancak yukardan yaklaşım yöntemi ile sınırlanabilir. Diğer bir ifade ile, başlangıç teriminin geçtiği yayınlar L olarak ifade edilirse, hali hazırda belirlenmiş kümeler ile L'nin tanımının yapılmasında bazı kümeler L'nin tam olarak içerisinde yer alırken, bazı kümeler ile L ile kesiştikleri halde ortak olmayan elemanları da olacaktır. L'nin alt kümesi olmayan ancak L ile kesişimleri de boş küme olmayan bu kümeler kaba küme teorisindeki sınır kümeleri ile tanımlanmıştır.



Şekil 6. Belge Kümeleri ve Kaba Küme Mantiği

Kaba küme teoreminin uygulanması için LTB üzerinde bazı tanımların uyarlanması gerekmektedir.

Yayınların kümelenmesi ile ayırdedilmezlik kümelerin oluşturulması sağlanmıştır. Kümeleme işleminin yapılabilmesi için varlıkların ve niteliklerinin tespit edilmesi gerekmektedir. Yayınlar varlıkları oluşturmaktadır. Yayınların nitelendirilmesinde ise, diğer benzer çalışmalarda yapıldığı gibi o yayında geçen terimlerin kullanılması bir dizi sorunu beraberinde getirmektedir. Bu sorunların çözümü için terimler yerine, terimlerin bir alt kümesi olan kavramların kullanılması uygun görülmüştür. Böylelikle kümeleme algoritmalarının daha etkili çalışabileceği nitelik sayısına inilmiş olur. Diğer taraftan, terimlerin nitelik olarak kullanıldığı durumlarda ayırdedilmezlik kümeleri çok daralacak hatta tek elemanlı kümeler ortaya çıkacaktır. Kavramların nitelik olarak kullanılması ile ayırdedilmezlik kümeleri keskin hatlarından kurtulmuş ve sınırları biraz daha bulutumsu bir konuma kavuşmuştur.



Şekil 7. Standart LTB'dan Rough LTB'ye Geçiş

Standart LTB uygulamalarından Kaba Küme temelinde LTB'ye geçişte başlangıç terimini (A Terimi) barındıran yayınlara  $L_A$ , ve bu yayınlarda geçen diğer tüm terimlerin geçtiği terimleri bağlama terimi (B terimleri) olarak kullandığımızda elde edilen yayınlara  $L_B$  olarak adlandırdığımız zaman, Kaba küme teorisinde  $L_A$  X kümesine tam olarak karşılık gelir.  $L_B$  ise mantıksal zeminde X üstten tanımlayan kümeye karşılık gelmektedir. Buradaki fark,  $L_B$  bağlama terimlerinin geçtiği diğer yayınlarken, X üstten tanımlayan küme ise  $L_A$ 'ya yakın yayınlardan oluşmaktadır. Her ikisi de sonuç kümesi ile başlangıç teriminin birleştirilmesini sağlamaktadır. Ancak yenilikçi yaklaşımımız sayesinde, terimlerin bir metinde geçmesi yeterli sayılmayıp, yayının başlangıç terimini barındırmadığı halde, başlangıç teriminin literatürüne çok yakın olması üzerine gidilmektedir.

Şekilde gösterilen önemli bir özellik de, sınır bölgesinde bulunan sınıfların yorumlanmasıdır. Sınırdaki bulunan sınıfların değeri, X kümesi yani  $L_A$  ile sahip olduğu ortak eleman sayısı ile değerlendirilmiştir. Örnek olarak, kümede bulunan yayınlardan %95'inin arama terimini barındırdığı durumda, arama terimini barındırmayan %5'lik yayın büyük önem taşımaktadır. Ancak  $L_A$  ile çok düşük bir yakınlık gösteren örneğin ortak eleman oranı %2 olan bir kümede bulunan ve %98'lik kesimde yer alan yayınların  $L_A$ 'ya yakınlığının ve dolayısıyla değerinin düşük olduğu düşünülmüştür.

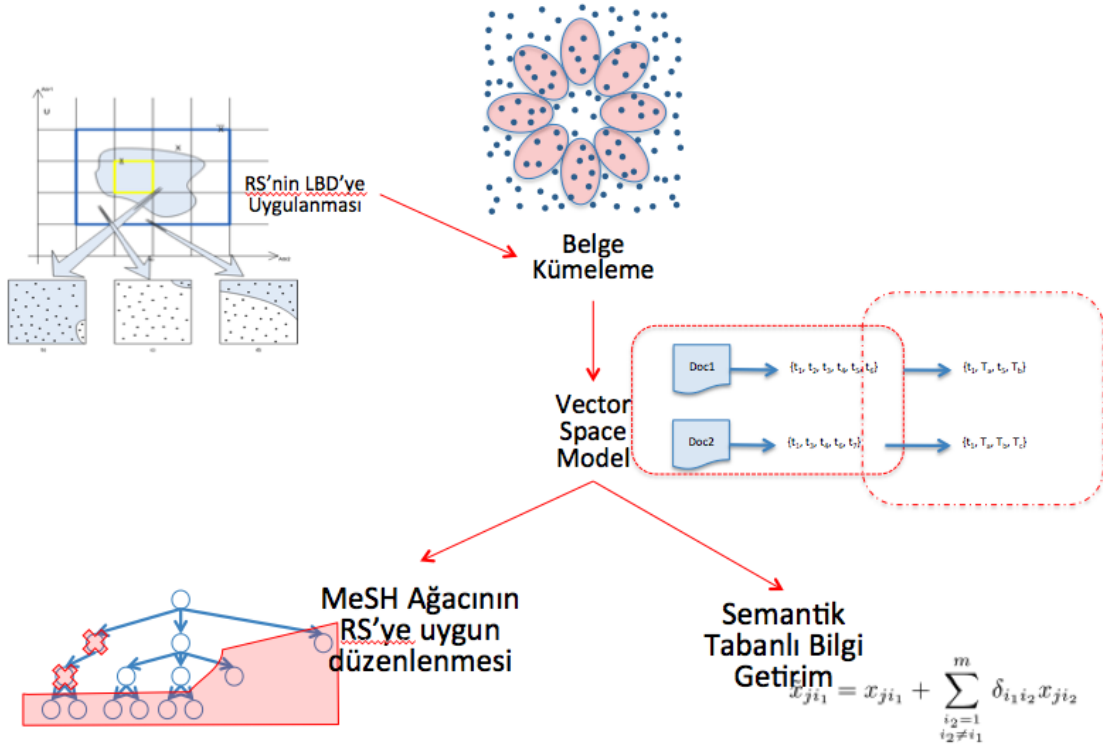
LTB'ye Kaba Küme Teoremi uygulanırken dikkat edilen konular şu şekilde sıralanabilir:

1. Yayınlar içerdikleri terimler ile tanımlanıp, eşitlik sınıflarının belirlenmesi için kümeleme algoritmaları işletilirse, çok küçük ve sınırları çok keskin eşitlik sınıfları ortaya çıkmaktadır.
2. Eşitlik sınıflarının hatları keskinleştikçe, başlangıç terimini barındıran yayınların kümesi  $L_A$ 'yı tanımlayan sınır bölgesi küçülmekte, ya da yok olmaktadır.
3. Sınır bölgesinin yok olması durumunda elde edilecek sonuçlar kümesi boş küme olacaktır.

### 3.3. Uygulama Veri Tabanı Alt Yapısı

Tez çalışması kapsamında geliştirilen uygulama, verilerin çalışma ortamına uygun hale getirilmesine yönelik modüller ve son kullanıcının işletimine yönelik modüller olmak üzere temel düzeyde ikiye ayrılabilir.

Uygulamanın çalışmasına yönelik veriler, PubMed yayın veri tabanından ve MeSH terim kütüphanesinden oluşmaktadır. Bu verilerin Kaba Küme Teorisine uygun LTB için uygun hale getirilmesine yönelik yapılan işlemler şekilde gösterilmiştir.



Şekil 8. Tez Çalışmasının Tüm Modülleri

MeSH ağacında tıp alanında yer alan bütün terimler yer almaktadır. PubMed'de bulunan yayınlar MeSH'de yer alan terimler aracılığı ile indekslenmiştir. Ancak bu indekslerin doğrudan yayınların niteliği olarak kullanılması durumunda daha önceki başlıklarda bahsedilmiş çok sayıda problemle karşılaşmaktadır. Bu amaçla MeSH üzerinde yapılan çalışmada,

terim ağacı budanarak kavram ağacına geçilmiş ve yayınlar kavramlarla niteliklendirilmiştir.

Yayınların içerdikleri kavramlara göre bir matrise yerleştirilmesi ile Vektör-Uzay Model kurulmuştur. Bu modelde yayın içerisinde geçen kavramların ağırlıkları ile orantılı şekilde yayını nitelemeleri sağlanmıştır.

Yayın-Kavram matrisi kullanılarak belgelerin kümelenmesi sağlanmıştır. Böylelikle Kaba Kümeler teorisindeki eşitlik sınıfları kurulmuştur. Bu noktadan itibaren temel veri tabanı kurulmuş ve artık son kullanıcıdan gelen sorguların işletilebileceği alt yapı hazır hale gelmiştir. Son kullanıcıdan gelen sorgular Kaba Küme teorisine uygun olarak işletilmektedir.

### 3.3.1. Yayınların temin edilmesi

Literatür tabanlı bilgi keşfi, işleme mantığı olarak akademik yayınlara güvenmektedir. LTB'in ilk aşaması, kullanıcıdan alınan arama teriminin yayınlarla ilişkilendirilmesidir. Arama terimini barındıran yayınların bulunması, ve bu yayınlar üzerinden sonuca ulaşılacak üzere farklı diğer yayınların temin edilmesi gerekmektedir. Tez kapsamında yapılan çalışmada, yayınlar kullanıcıdan gelen sorgular sonucu değil, önceden hazır bir halde kümelenecek olarak saklanmaları üzerine modellenmiştir.

LTB çalışmalarında ortak yayın kaynağı olarak MEDLINE kullanılmıştır. MEDLINE üzerinde hizmet veren PubMed ile kullanıcılar anahtar kelimeler, yazar isimleri ve tarih kısıtlamaları ile sorgulamalar yapabilmektedir. PubMed, son kullanıcılara hizmet etmenin yanında, bilgisayar yazılımlarının iletişim kurmasına yönelik web servis hizmetleri de sağlamaktadır.

Yapılan tez çalışmasında, PubMed kullanılarak MEDLINE'daki tüm yayınlar indirilmiş ve anahtar kelimeleri ile indekslenerek veri tabanında saklanmıştır.

### 3.3.2. Terim Çıkarımı

MEDLINE'da barındırılan yayınlar, MeSH terimleri kullanılarak indekslenmiştir. Bir yayın PubMed ile indirilirken, bağımsız kontrol görevlileri tarafından atanmış anahtar kelimeleri olarak MeSH terimleri de temin edilebilmektedir.

MEDLINE üzerine metin madenciliği denemeleri yapılan çalışmalarda genelde PubMed'den elde edilen bu anahtarlar ve yayınlarda yazarların kendi belirlediği anahtar terimler yeterli görülmüştür [26][65][66].

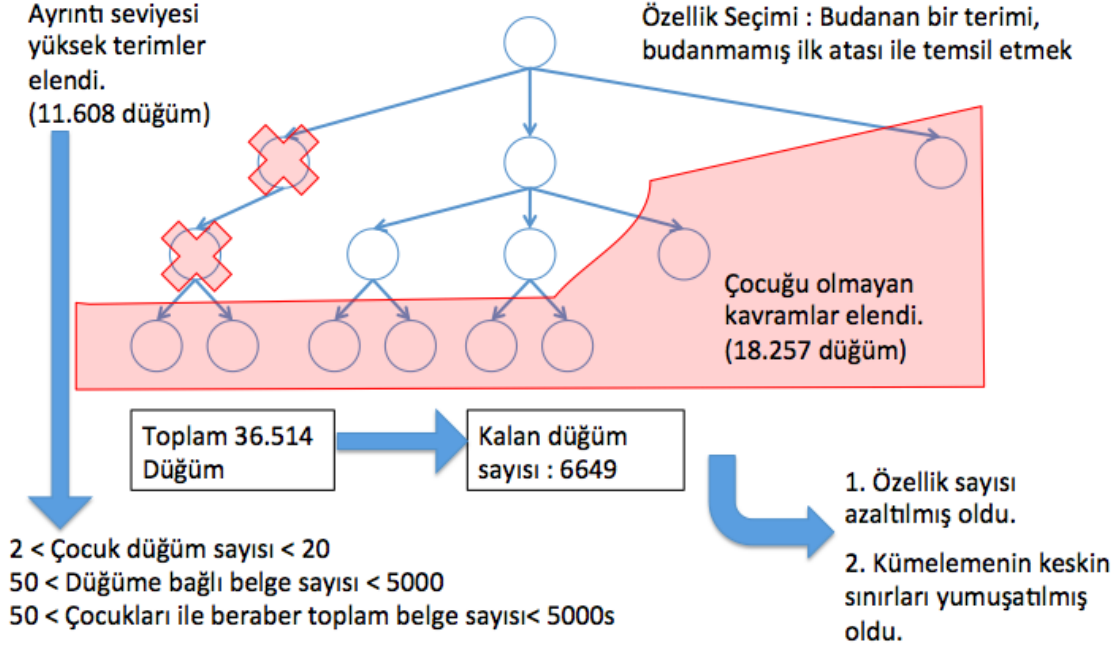
Ancak özellikle iki terimin daha önce beraber kullanımının incelenmesinde özellikle yayın başlıklarının çok önemli olduğu, tez kapsamında yapılan analiz çalışmaları sırasında uzman kullanıcılar tarafından belirlenmiştir. Bu sebeple tez çalışmasında, yayınların başlıklarında bulunan MeSH terimleri de belirlenmiş, o yayının özellikleri arasında o terimlerde kullanılmıştır.

### 3.3.3. MeSH Ağacı ve KK-LTB için uyarlanması

MeSH ağacı 36.514 adet terimden oluşmaktadır. Terimler arası ilişkiler, genelden detaya doğru bir ağaç şeklinde verilmiştir. MeSH ağacı uygulamada, barındırdığı terimler, bilimsel yayınların niteliği olacak şekilde kullanılmıştır. Ancak MeSH ağacının direk kullanımı, Belge-Terim matrisinin sütun ekseninde boyutunu çalışılmaz düzeyde büyütmemektedir.

Belge-Terim matrisi, yayınların kümelenmesini temin etmek amacıyla kurulmaktadır. Bu matriste yayınların çok ince detaylarda nitelendirilmesi, yayınların bir arada kümelenmesini zorlaştırmakta, kümelerin çok küçük hatta tek yayından oluşur hale gelmesi durumunu ortaya çıkarmaktadır.

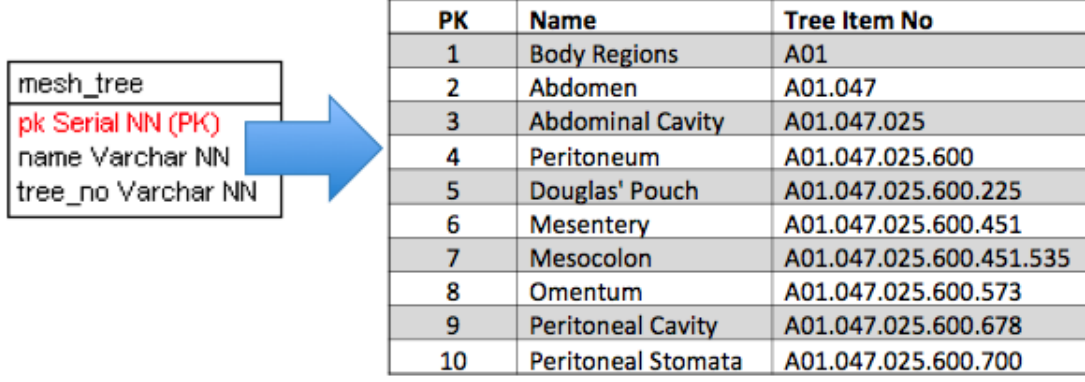
MeSH ağacının kavramsal düzeyde küçültülmesi mümkün görünmese de, uygulama içerisinde yayınlarda geçen terimler yerine o terimlerin bağlı olduğu daha genel terimlerin, diğer bir ifade ile kavramların kullanılmasına karar verilmiştir. Böylelikle matrisin sütun büyüklüğü azalırken, kümeleme için daha ideal bir ortam yakalanmaya çalışılmıştır.



Şekil 9. Terim Ağacından Kavram Ağacına Geçiş

Ağacın küçültülmesi amacıyla elenecek terimlerin belirlenmesinde şu kurallar uygulanmıştır:

1. Çocuk düğüm sayısı 20'den az olan terimler elenmiştir. Bu kural, altında yeterince terim bulunmayan terimlerin, detay bilgi taşıyor olması varsayımı ile kurulmuştur.
2. Kendine ve alt çocuklarına bağlı yayın sayısı, diğer bir ifadeyle nitelediği yayın sayısı 5000 altında olan terimler elenmiştir.



Şekil 10. Kavramların Veri Tabanı Gösterimi

MeSH Ağacı veri tabanında alışıldık indeksleme yöntemlerinden farklı olarak yapılandırılmıştır. Bu yöntemde ağaç veri tabanında olduğu üzere sadece bir üst nesne (atası) değil, köke kadar yolun tümü gösterecek şekilde indekslenmiştir. Terimlerin tez çalışmasında çok temelde olduğu, ve algoritmaların sıklıkla terimlere eriştiği düşünüldüğünde, milyonlarca terime erişim zamanının azaltılması büyük önem taşımaktadır. Böylelikle, klasik yabancı anahtar (“foreign key”) yönetiminde elde edilemeyecek şu özellikler kazanılmıştır:

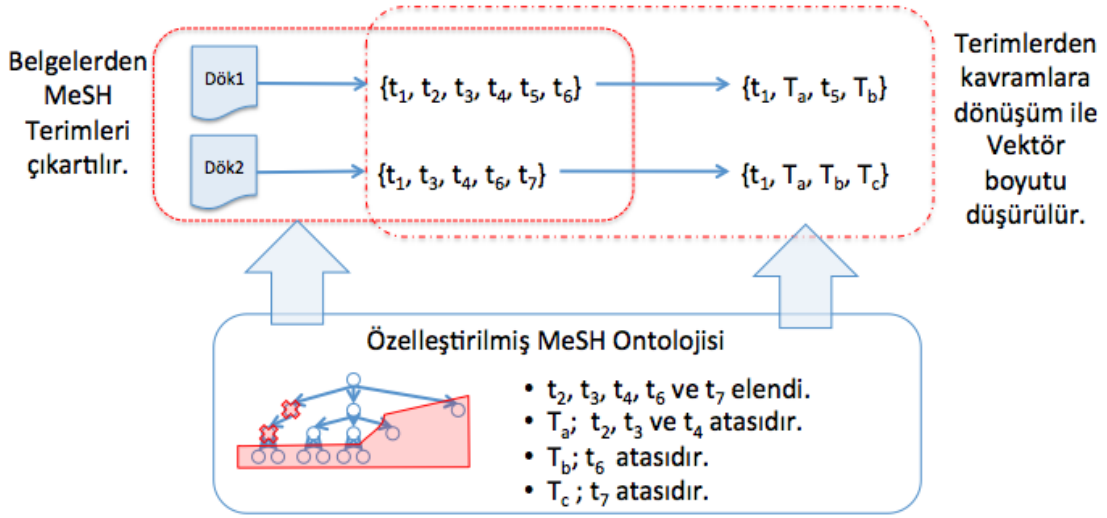
1. Bir terimin köke olan uzaklığı sadece bir sorgulama ve tek bir satırın getirilmesi ile hesaplanabilmektedir.
2. Bir kavramın ya da terimin, toplam çocuk sayısı daha kolay bir şekilde, sadece bir tarama işlemi ile sayılabilmektedir. Klasik yöntemde ise mutlaka her bir çocuk kadar tabloya erişim yapılması gerekli olacaktır.
3. Her hangi iki terimin ya da kavramın, hangi noktada atalarının ortaklaştığı çok hızlı bir şekilde bulunabilir.
4. Klasik yöntemde B++ tree kurulmuş olsa, erişim zamanı ortalama 5 kat daha fazla olacaktır.

### 3.3.4. Vektör Uzay Modeli

Uygulama kapsamında yayınlar Vektör Uzay Modeli kullanılarak nitelendirilmiştir. Her bir yayının, barındırdığı terimlerin bağlı olduğu kavramlar kullanılarak bir vektöre dönüştürülmüştür. Bir kavramın o yayının için değeri



sadece yayında kaç defa geçtiği ile değil, öncelikle TF-IDF yöntemiyle ve bunun üzerine o yayındaki diğer terimler ile ne derece ilintili olduğu da gözetilerek belirlenmiştir. TF, terim frekansı ifade ederken, IDF ise o terimin tüm yayınlar içerisinde ne kadar önemli olduğunun bir göstergesidir. Bu iki değere ek olarak, o terimin, o belge ile ne kadar uyumlu olduğu, belgede geçen diğer terimlerle yakınlığı gözetilerek hesaplanmıştır. Böylelikle kavramın bir belge için ne kadar önemli olduğu belirlenmeye çalışılmıştır.



Şekil 11. Kavram Ağacından Vektör-Uzay Modeline Geçiş

Standart TF-IDF yöntemleri ile terim ve belgeler arasında ilişki kurulmasında, bir yayında o kelimenin hangi anlamıyla kullanıldığı, eş anlamlı, eşsesli seçimi gibi sorunlar ortaya çıkmaktadır. Ancak terimin semantik türü ile ikili halde yayınlarda araştırılması daha verimli sonuçlar üretecektir [67], [68].

Yayınlara kümelenebilmesinde, belge-kavram matrisinde değer olarak  $TF \times IDF$  değerine ek olarak, yayınlara anlamsal içeriklerinden çıkarsama yapıldığı düşünüldüğünde, terimlerin semantik değerlerinin de göz önünde bulundurulması gerekir. Semantik değerlendirme, o terimin/kavramın, yayın içerisindeki diğer terimler/kavramlar ile ilişkisinin toplamı alınarak hesaplanabilir. Diğer kavramlar ile olan ilişki değerlerinin hesaplanması için mesh ontolojisi kullanılmaktadır.

Bir kelimenin, geçtiği yayın açısından ne kadar önemli olduğunu belirlemek üzere o yayındaki diğer terimlerle ilişkisini inceleyen, semantik TF-IDF olarak isimlendirebileceğimiz 6 farklı yöntemden, literatürde desteklendiği görülmüş 3 tanesi denenmiştir.

Bu yöntemler altı yöntem şu şekilde sıralanabilir :

1. Cos benzerliği : İki terim için TF-IDF değerlerinin çarpımının, tüm belgelerde geçme sıklığına bölümünden elde edilen değerdir. Böylelikle iki terimin belge içerisinde varlıklarının, diğer tüm literatürdeki varlığına oranı tespit edilmiştir olur.

$$\cos((t_1, t_2)) = \frac{t_1 \cdot t_2}{\|t_1\| \cdot \|t_2\|} = \frac{\sum_{j=1}^n \tilde{x}_{j1} \tilde{x}_{j2}}{\sqrt{\sum_{j=1}^n \tilde{x}_{j1}^2} \sqrt{\sum_{j=1}^n \tilde{x}_{j2}^2}} \quad (4.1)$$

2. Ağaç dolaşımı : Bu yöntemde, iki terim arasındaki mesafe, bir terimden diğerine yol alırken, ağaç üzerinde kaç düğümden geçildiğini hesaplamaktadır. Böylelikle iki terimin birbirine olan uzaklığı, kavramsal ağaç üzerinden hesaplanmaktadır [56][69].

3. Sw&P Yöntemi : İki terimin ortak babalarının kök dizine olan uzaklığını da göz önünde bulundurarak, terimlerin ağaç üzerinde uzaklıkları hesaplanmaktadır. Zira ortak ataları köke uzaksa, terimler ağaç içerisinde göreceli olarak diğer terimlere göre daha yakın demektir [70].

$c_1, c_2$  : kavram1 ve kavram2

H: iki kavramın ortak babalarının kök dizine olan uzaklığı

$N_1$ :  $C_1$ 'in köke olan uzaklığı

$N_2$ :  $C_2$ 'in köke olan uzaklığı

$$S_{W\&P}(c_1, c_2) = \frac{2H}{N_1 + N_2 + 2H} \quad (4.2)$$

4. Li : Ontolojideki iki terim arası en kısa yol uzunluğu ile derinlik değerlerinin lineer olmayan bir fonksiyon ile benzerliği ölçülmektedir [71]. 0 ile 1 arasında sonuç üretmektedir. 1'e yakınlık benzerliğinin arttığını gösterir.

$c_1, c_2$  : kavram1 ve kavram2

$\alpha$ : katsayı 0.2 (en kısa yol uzunluğunun ölçeklendirilmesi için kullanılmaktadır)

$\beta$ : katsayı 0.6 (derinliğinin ölçeklendirilmesi için kullanılmaktadır)

e: matematikteki "e" değeri

L: iki kavram arası en kısa yolun uzunluğu

H: iki kavramın en yakın babalarının köke olan uzaklığı

$$S_{Li}(c_1, c_2) = e^{-\alpha L} \frac{e^{\beta H} - e^{-\beta H}}{e^{\beta H} + e^{-\beta H}} \quad (4.3)$$

5. Leacock and Chodorow : Benzerlik değerlerinin normalize değerler olabilmesi için, iki terim arası en kısa yolun, hiyerarşideki en uzun derinlik değerinin 2 katı ile ölçeklendirip, sonucun logaritması alınmaktadır.

$c_1, c_2$  : kavram1 ve kavram2

$d(C_1, C_2)$ : iki kavram arası en kısa yol

D: hiyerarşideki en uzun derinlik

$$S_{L\&C}(c_1, c_2) = -\log(d(c_1, c_2)/2D) \quad (4.4)$$

6. Mao: Bu yöntemde iki kavram arası en kısa yol ile birlikte kavramların tüm çocuk sayıları da hesaba katılarak, kavramların genelliği ölçülmüştür [72]. Genel kavramların bu yöntemle elenmesi ve doküman benzerliklerinin detaydaki kavramlarla ölçülmesi sağlanmaktadır.

$c_1, c_2$  : kavram1 ve kavram2

$d(C_1, C_2)$ : iki kavram arası en kısa yol

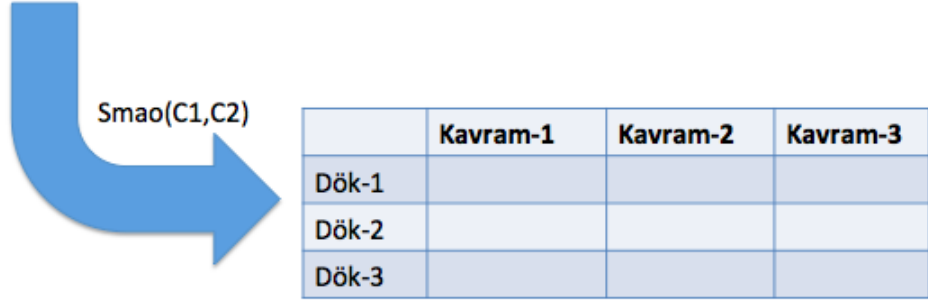
$d(C1)$ : C1'in tüm çocuklarının sayısı

alpha: sabit 0.9

$$S_{Mao}(c_1, c_2) = \frac{\delta}{d(c_1, c_2) \log_2(1 + d(c_1) + d(c_2))} \quad (4.5)$$

	Kavram-1			Kavram-2			Kavram-3		
	M1	M2	M3	M1	M2	M3	M1	M2	M3
Dök-1									
Dök-2									
Dök-3									

M1:  $S_{w\&p}(C1,C2)$   
M2:  $S_{li}(C1,C2)$   
M3:  $S_{mao}(C1,C2)$



Şekil 12. Vektör-Uzay Modeli ve Anlamsal Ağırlıklı TFxIDF

### 3.4. Belge Kümeleme

Kaba Kümeler teorisinin LTB alanına uygulanması için öncelikle kümeleme mantığının, PubMed yayınları üzerine uygulanması gerekmektedir. Sistemin genel başarımı düşünüldüğünde, belge kümeleme sürecinin en önemli modül olduğu düşünülmektedir. Belge kümeleme ile yayınların ortaklıkları ve dolayısıyla yakınlıkları ortaya konulmaktadır. Böylelikle kullanıcının araştırma yapmak istediği terimin, kümeler üzerinde dağılımı incelenebilir. Bir kümede yoğun olarak geçen bir terimin, o kümede hiç yer almadığı yayınlarda bir yenilik bir fikri ortaya çıkabilir. Zira burada bakış açısı, bir kümede arama terimini barındırmayan yayın, o kümede arama terimini barındıran yayınlara çok yakın olduğu halde arama terimini çalışması içerisinde göstermemiştir.

Çalışma mantığı tekrar gözden geçirildiği zaman, kümelemenin sistemin genel başarımı için ne derece önemli olduğu görülmüş olur.

Kümeleme sistemini etkileyen çok sayıda faktör bulunmaktadır. Kümelemenin yapılması için öncelikle yayınların niteliklerinin belirlenmesi gerekir. Standart bir çok belge kümeleme çalışmasında, yayınlar içerdikleri terimler/kelimeler ile nitelendirilmiştir. Kelimelerin nitelik olarak kullanılması durumunda, çok sayıda gürültü olarak adlandırabileceğimi veri ile baş etmek zorunluluğu ortaya çıkacaktır. Terimlerin doğrudan nitelik olarak kullanılması durumunda da, yayınlar çok detayda nitelendirilmesi sonucu benzer terimlerle nitelendirilen yayınların farklı kümelerde yer aldığı görülmüştür. Bu durumu düzeltebilmek için terimlerin yerini kavramlar almıştır. Eğer yayın içerisinde geçen bir terim kavram ağacında yer almıyorsa, o terimin kavram ağacında yer alan ilk atası yayını nitelemek üzere kullanılmıştır.

Yaptığımız çalışmada, bir kavramın ya da o kavramın kapsamındaki terimlerin yayının içerisinde geçmesi o yayını nitelemesinde yeterli görülmemiş, öncelikle TF-IDF ile terimin o yayın için frekans bakış açısı ile önemi belirlenmiş, sonrasında ise semantik ilişkiler gözetilerek o terimin o yayın için ne derece anlamsal yakınlığa sahip olduğu belirlenmiştir.

Kümeleme için yayınların niteliklerinin çıkarılmasını takiben, ortalama küme büyüklüğü veya toplam küme sayısının tespit edilmesi gerekmektedir. Bu amaçla toplam küme sayısı üzerinde deneyerek sonuç elde edilmeye çalışılmıştır. Ortaya çıkan sonuç, küme sayısının artırılmasından fayda elde edildiğini göstermiş olsa da, yayınların çok fazla oluşu, eldeki en iyi kümeleme algoritmasının dahi sınırlarını zorlamış ve ancak 50.000 küme sayısına kadar sınama yapılabilmektedir. 50.000 küme üzerinde sonuçları incelemek için yaptığımız denemeler süre kısıtlamaları sebebiyle başarılı olmamıştır. Dolayısıyla en iyi küme sayısını tespit ettiğimizi ifade edemesek de, imkanlar dahilindeki en iyi küme sayısı ile çalışma sonuçları elde edilmiştir.

Kümeleme algoritmalarının sınanmasına yönelik genel kabul görmüş üç yöntem bulunmaktadır. İlk yöntemde, kullanıcıların kümeleri teker teker değerlendirmeleri öngörülmüştür. Böylelikle kümelemenin ne kadar tatmin edici olduğu doğrudan alan uzmanları tarafından belirlenmektedir. İkinci

yöntemde ise kümeleme algoritmalarının hedef fonksiyonları (“*objective function*”) değerlendirme amaçlı kullanılmaktadır. Örneğin K-Means algoritmasındaki toplam hatanın karesi, kümelemenin başarımının ölçümü için kullanılabilir. Son olarak sınıflandırılmış veri üzerinde kümeleme yapılarak, sınıflarla kümelerin ne kadar örtüştüğü incelenebilir.

Tez çalışması kapsamında kullanılan kümeleme modülü, doğrudan LTB uygulaması ile sınılanması mümkün değildir. LTB uygulamaları için hali hazırda bir sınıama ortamı kurulmamıştır. Ancak kümeleme algoritmasının ve bu algoritmanın işletiminde kullanılan parametrelerin belirlenmesi ve değerlendirilmesine yönelik bilinen başka bir platformda sınıama işlemi yapılmıştır.

#### 3.4.1. Kümeleme Başarım Ölçümü

Kümeleme sonuçlarının karşılaştırılmasında Duyarlık (*Presicion* - doğru getirilen sonuçların tüm sonuçlara oranı), Anımsama (*Recall* - doğru getirilen sonuçların, getirilmesi gerekenlere oranı), Miisclassification indeks (hatalı belgelerin doğru belgelere oranı), F-Score (2 x anımsama x duyarlık / (anımsama + duyarlık)) gibi değerlendirme formülleri kullanılabilir. F-Score ile sistemin entropisi de ölçülebilmektedir. Eğer bir sistemin tamamı düzenli ise entropisi sıfır olacaktır.

n: döküman sayısı

k: küme sayısı ve sınıf sayısı (sınıflar uzmanlar tarafından belli)

Nh: h. sınıftaki döküman sayısı

Nl: l. kümedeki döküman sayısı

Nhl: hem sınıfta hem de kümedeki döküman sayısı

$$\text{entropy} : \sum_{l=1}^k \frac{n_l}{n} \left( -\frac{1}{\log k} \sum_{j=1}^k \frac{n_{jl}}{n_l} \cdot \log \frac{n_{jl}}{n_l} \right) \quad (4.6)$$

$$Fscore: \sum_{l=1}^k \frac{n_j}{n} \cdot \max_{1 \leq l \leq k} \left\{ \frac{2 \cdot n_{jl} / n_j \cdot n_{jl} / n_l}{n_{jl} / n_j + n_{jl} / n_l} \right\} \quad (4.7)$$

$$MI: \text{misclassification index} : \frac{\sum_{j,l} n_{jl} \log \left( \frac{n \cdot n_{jl}}{n_j n_l} \right)}{\sqrt{(\sum_j n_j \log \frac{n_j}{n}) (\sum_l n_l \log \frac{n_l}{n})}} \quad (4.8)$$

$$\text{Saflık: } \sum_{l=1}^k \frac{1}{n} X(\max_{1 \leq l \leq k} \{n_{jl}\}) \quad (4.9)$$

### 3.4.2. Kümeleme Başarım Deneyleri

Tez çalışmasının, başarımı etkileyen en önemli modülü belge kümeleme olması sebebiyle, burada kullanılan algoritma ve kütüphanelerin başarımı ayrıca incelenmiştir. Bu amaçla PubMed verileri üzerine kümeleme yapan iki çalışmada izlenen yöntem takip edilmiş [73][74], sonuçların incelenmesi için altın standart (*gold standard*) olarak OHSUMED [75][76] verisi kullanılmıştır.

OHSUMED verisi TREC-9 (*Text REtrieval Conference-9 Filtering Track*) kısmında kullanılmıştır. Bu test koleksiyon, 1987 - 1991 yılları arası, MEDLINE veri tabanından alınan 348566 makaleden oluşmaktadır. Makalelerin başlık, özet, mesh terimleri, yazarları ve yayın türleri test koleksiyonuna alınmıştır. Test koleksiyonu web sitesinden <sup>11</sup> temin edilebilmektedir.

Bu koleksiyonda MEDLINE'in her bir makaleye vermiş olduğu MUI (*Medline Unique Identifier*) bilgisi de mevcuttur. Test koleksiyonunda MUI bilgisi birincil anahtar olarak kullanılarak özet, mesh terim bilgileri bu anahtara göre verilmiştir. Fakat 2001 yılından sonra PubMed, MUI bilgisini desteklemediği için, test koleksiyonundan MUI bilgisi kullanılarak PubMed'den makale çekilememektedir. Yapılan çalışmada makalelerin başlık, özet, mesh terim bilgileri PubMed'den değil test koleksiyonundan alınmıştır.

---

<sup>11</sup> [http://trec.nist.gov/data/t9\\_filtering.html](http://trec.nist.gov/data/t9_filtering.html)

Cluto kümeleme uygulamasının Pubmed yayınları üzerinde “metin kümeleme” bağlamında başarımı incelenmiştir. Böylelikle tez uygulamasının temel modüllerinden olan kümeleme mantığının başarımı üzerine net veriler sunulması sağlanmıştır. Bu veri kümesi üzerinde kümeleme çalışması yapan 2008 yılına ait doktora tezinin [74] sonuçları üzerinde karşılaştırma yapılmıştır. Karşılaştırma amaçlı incelenen bu doktora tezinde, veri kümesi üzerinde yapılan sonuçlarda F-Ölçüm değeri sunulmuştur. Yaptığımız çalışmada karşılaştırma yaparken sadece F-Ölçüm değeri üzerinden karşılaştırma yapılacaktır. F-Ölçüm sonuçları ile çalışmanın sonuçları Çizelge 5’de karşılaştırılmıştır.

İndirilen veride ohsu-trec -> trec9-test klasörünün altında qrels.mesh.88-91 dosyası kontrol kümeleri (uzmanlar tarafından gruplandırılmış) test verisi olarak ele alınmıştır. Karşılaştırma amaçlı odaklanan doktora tezi [74], bu test verisinde bulunan 4904 gruptan sadece 8 grubu ele almıştır. Karşılaştırma yapabilmek için doktora tezinde kullanılan 8 grubun aynısını çalışmamızda kullanılmıştır. Her bir grupta 100 adet MEDLINE dokümanı bulunduğundan 800 makale üzerinde çalışma yapılmıştır. "trec9-test" klasörünün altında "ohsumed.88-91" dosyasında da makalelerin başlıkları ve mesh terimleri bulunmaktadır.

Deney setinin işletim yönteminde, makaleler ve mesh terimler kullanılarak, önerilen mesh ağacı üzerinden benzerlik hesaplanıp Vektör Space Model kuruldu. Bu Vektör Space Model’deki veriler, “cluto” kümeleme uygulamasına girdi olarak verildi. Küme sayısı parametresi, doktora tezi ile karşılaştırma yapılabilmesi için 8 olarak verildi. Cluto programı sonucu oluşan kümeler ile kontrol kümeleri arasında karşılaştırma yapılarak tablolar elde edildi.



GRUP NO	Döküman Sayısı	MESH								Max*Döküman Sayısı	Toplam Döküman Sayısı	Purity
		MSH1058	MSH1262	MSH1473	MSH1486	MSH1713	MSH2025	MSH2030	MSH2235			
1	102	0,02	0,00	0,01	0,00	0,95	0,02	0,00	0,00	96,9		
2	102	0,02	0,00	0,01	0,00	0,00	0,01	0,95	0,01	96,9		
3	99	0,02	0,98	0,00	0,00	0,00	0,00	0,00	0,00	97,02		
4	105	0,03	0,01	0,01	0,91	0,01	0,00	0,02	0,01	95,55		
5	109	0,11	0,00	0,02	0,02	0,00	0,84	0,01	0,00	91,56		
6	103	0,03	0,02	0,00	0,01	0,00	0,00	0,00	0,94	96,82		
7	95	0,05	0,00	0,91	0,00	0,01	0,03	0,00	0,00	86,45		
8	84	0,85	0,00	0,11	0,01	0,00	0,02	0,00	0,01	71,4		
										732,6	799	0,92
	Döküman Sayısı	100	100	100	100	100	100	100	100			
	Max*Döküman Sayısı	85	98	91	91	95	84	95	94	733		
	Toplam Döküman Sayısı									799		
	<b>Inverse Purity</b>									<b>0,92</b>		

Çizelge 4. Belge Kümeleme ve Saflık Değeri

	Döküman Sayısı	OHSU								
GRUP NO		MSH1058	MSH1262	MSH1473	MSH1486	MSH1713	MSH2025	MSH2030	MSH2235	
1	102	0,02	0,00	0,01	0,00	0,97	0,02	0,00	0,00	
2	102	0,02	0,00	0,01	0,00	0,00	0,01	0,96	0,01	
3	99	0,02	0,97	0,00	0,00	0,00	0,00	0,00	0,00	
4	105	0,03	0,01	0,01	0,94	0,01	0,00	0,02	0,01	
5	109	0,11	0,00	0,02	0,02	0,00	0,88	0,01	0,00	
6	103	0,03	0,02	0,00	0,01	0,00	0,00	0,00	0,96	
7	95	0,05	0,00	0,88	0,00	0,01	0,03	0,00	0,00	
8	84	0,77	0,00	0,10	0,01	0,00	0,02	0,00	0,01	
Döküman Sayısı		100	100	100	100	100	100	100	100	
Max*Döküman Sayısı		77,17391	97,48744	88,20513	93,65854	96,55172	88,03828	96,0396	95,5665	732,7211231
Toplam Döküman Sayısı										799
<b>F measure</b>										<b>0,92</b>
Max. F-measure değeri		0,77	0,97	0,88	0,94	0,97	0,88	0,96	0,96	

Çizelge 5. Belge Kümeleme ve F-Ölçüm

	Döküman Sayısı	OHSU								
GRUP NO		MSH1058	MSH1262	MSH1473	MSH1486	MSH1713	MSH2025	MSH2030	MSH2235	
1	102	0,02	0,00	0,01	0,00	0,95	0,02	0,00	0,00	
2	102	0,02	0,00	0,01	0,00	0,00	0,01	0,95	0,01	
3	99	0,02	0,98	0,00	0,00	0,00	0,00	0,00	0,00	
4	105	0,03	0,01	0,01	0,91	0,01	0,00	0,02	0,01	
5	109	0,11	0,00	0,02	0,02	0,00	0,84	0,01	0,00	
6	103	0,03	0,02	0,00	0,01	0,00	0,00	0,00	0,94	
7	95	0,05	0,00	0,91	0,00	0,01	0,03	0,00	0,00	
8	84	0,85	0,00	0,11	0,01	0,00	0,02	0,00	0,01	
	Döküman Sayısı	100	100	100	100	100	100	100	100	
	Precision(i,j)*log2 Precision(i,j)	-140,78	-20,79	-78,63	-36,96	-20,32	-65,53	-24,96	-28,32	-416,29
	Toplam Döküman Sayısı									799
	<b>Entropy</b>									<b>0,52</b>
	<b>Normalized Entropy</b>									<b>0,44</b>

Çizelge 6. Belge Kümeleme ve Entropy Değeri

Kategori	Dök. Sayısı	Mesh No.	1. Çalışma F-Ölçümü	2. Çalışma F-Ölçümü
1	100	1058	0,486	0,772
2	100	1262	0,898	0,975
3	100	1473	0,916	0,882
4	100	1486	0,899	0,937
5	100	1713	0,844	0,966
6	100	2025	0,810	0,880
7	100	2030	0,871	0,960
8	100	2235	0,922	0,956

Çizelge 7. F- Ölçüm Değerinin Karşılaştırılması

Safılık, Karşıt Safılık ve F-Ölçüm değerleri 0,92 olarak bulunduğundan kümeleme başarısının yüksek olduğu söylenebilir. Entropy ve Normalized Entropy değerleri ne kadar 0'a yaklaşırsa kümelemenin o kadar iyi olduğu söylenebilmektedir.

Doktora tezi ile çalışmanın karşılaştırılmasına bakıldığında yapılan çalışmanın başarısı açıkça görülebilmektedir. Sadece 1. küme hariç, diğer 7 kümede F-Ölçüm değeri, doktora tezindeki sonuçlardan yüksek çıkmıştır. 1. kümenin sonuçlarına bakıldığında referans alınan doktora tezine karşın %30 daha iyi bir sonuç elde edilmiştir.

Sonuç olarak Cluto uygulamasının MEDLINE belgelerinde, kümeleme için etkinliğinin yeterli düzeyde olduğu düşünülmektedir.

### 3.4.3. Ortalama Küme Büyüklüğü

Kümeleme algoritmaları, istenen küme büyüklüğüne göre bölme işlemi yapabilmektedir. Bu durum, "istenen küme büyüklüğünün" belirlenmesini gerektirmektedir. Bu amaçla, literatürde kullanılan 3 yöntem uygulanmıştır.

#### a. Başparmak Kuralı

$$k \approx \sqrt{n/2} \quad (4.10)$$

Yapılan çalışmada 8341536 yayın bulunmaktadır. Bu sayı üzerinden küme büyüklüğü:

$$k \approx \sqrt{\frac{8,341,536.00}{2}} = \sqrt{4,170,768.00} = 2042.24582261783559$$

#### b. Sıfır Olmayan Girişler

D = doküman matrisi

m = doküman sayısı

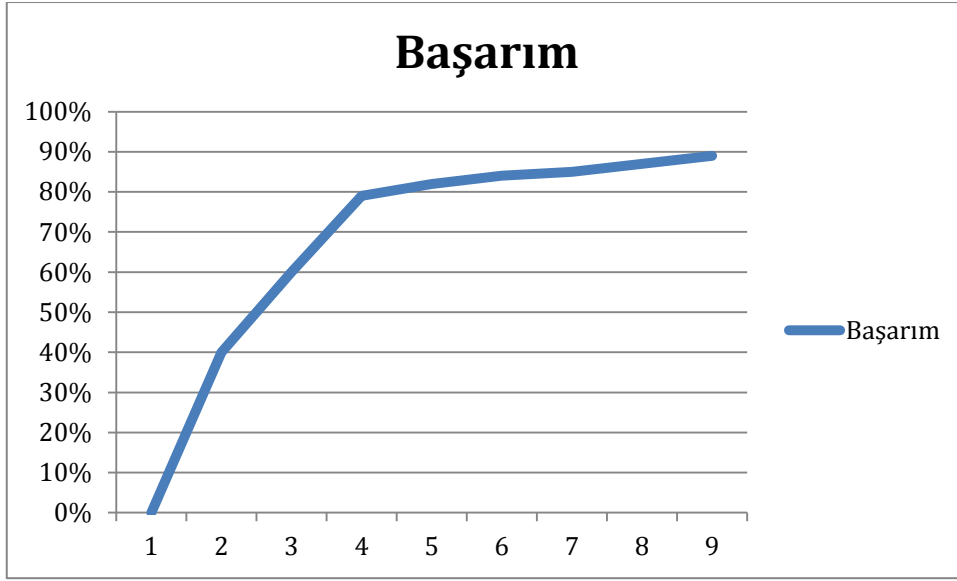
n = terim sayısı

t = non-zero giriş sayısı

$$k = (m \times n) / t \rightarrow m = 8341536 \quad n = 6649 \quad t = 63152162 \quad k = 878$$

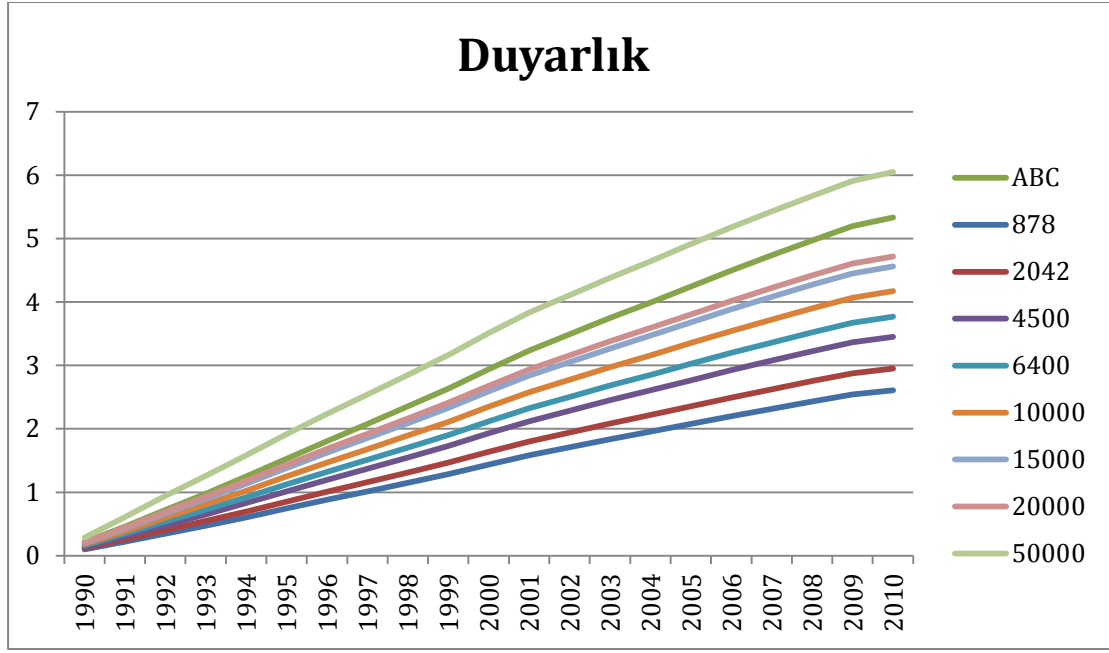
#### c. Elbow Metodu

Deneyerek bulma metodu olarak ifade edilebilir. Küme büyüklüğü olabilecek mantıklı en küçük değerden başlanarak, yavaş yavaş büyütülür ve başarımlar ölçülür. Belli bir noktadan sonra küme büyüklüğünün artırılmasının başarımlar üzerine etkisinin azaldığı görülmektedir. Dolayısıyla hem başarımlar, hem de performans açısından optimum bir değer bulunabilir. Örnek olarak verilen grafikte, en uygun küme büyüklüğü 4 olarak bulunmuştur.

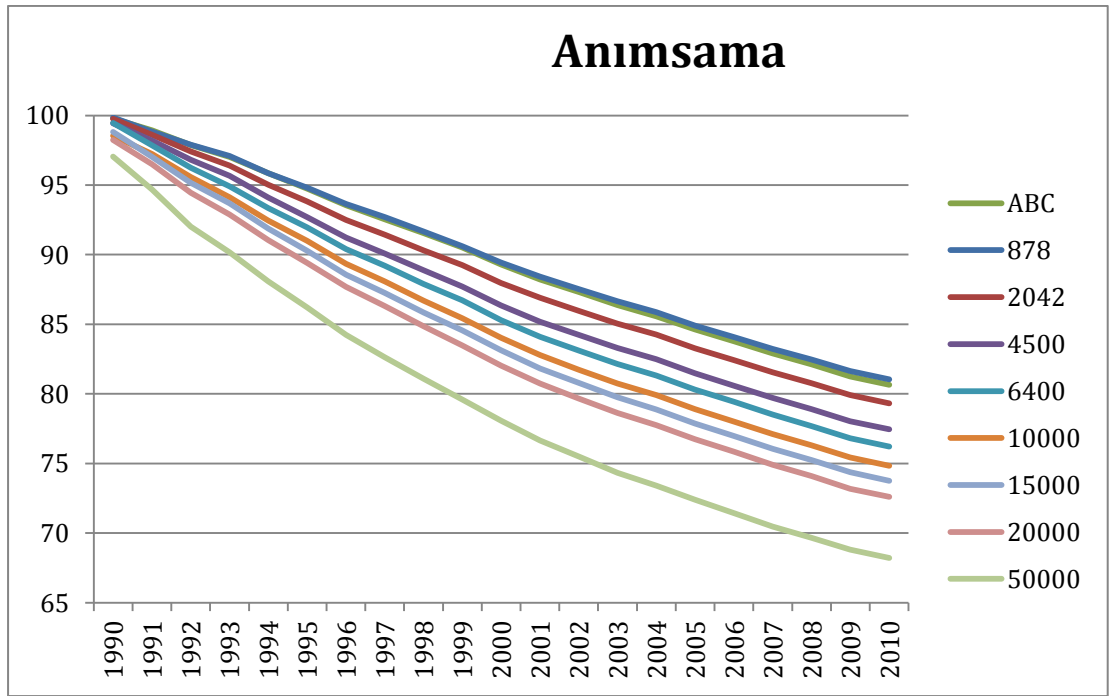


Şekil 13. Elbow Metodunun Uygulanması

Tez kapsamında, küme büyüklüğünün belirlenmesi için 1990 yılına kadar ki yayınların tümü değişken büyüklüklerde kümeleneş, her bir çevirim için duyarlık-anımsama değerleri hesaplanmıştır. Duyarlık ve anımsama değerleri, 878 (non-Zero Entries), 2042 (Rule of Thumb), 4500, 6400, 10000, 15000, 20000 ve 50000 küme büyüklüklerine göre hesaplanmıştır. Küme büyüklüğünün 50000 olduđu deneme, diđer küme büyüklüklerine daha iyi sonuç üretmiştir, Standart ABC modeli ile karşılaştırma yapıldığı zaman, küme büyüklüğünün 50,000 olduđu durumlarda daha yüksek duyarlık değeri elde edilmiştir. Fakat anımsama bağlamında en kötü sonuç 50,000 büyüklüğünde görülmüştür. Bu iki grafiğe bakıldığında ortalama küme büyüklüğünün 50,000 olması, bize anlamlı ve az sayıda sonucu göstermektedir. Az sonuç gösterdiğinden dolayı anımsama değeri düşüktür, anlamlı sonuçlar gösterdiğinden dolayı duyarlık değeri yüksektir.



Şekil 14. Farklı Küme Büyüklükleri için Duyarlık Grafiği



Şekil 15. Farklı Küme Büyüklükleri için Anımsama Grafiği

### 3.5. LTB Sınama Yöntemi

LTB çalışmalarında zorlu kesimlerden birisi de, yapılan çalışmanın başarımlarının analizinin yapılması için sistematik olarak oluşturulmuş bir deney kümesinin bulunmamasıdır [77]. Yapılan çalışmanın doğruluk ve yeterlik analizi için diğer uygulama ortamları için geliştirilmiş altın standart kümelerine benzer, kalite test veri tabanı ("*Benchmark Data Set*") yapılarının LTB için oluşturulması çok uzun çalışmalar gerektirmekte ve güçlü bir alan bilgisine sahip uzman ekibe ihtiyaç duymaktadır. Bu uzman ekibin, milyonlarca yayını incelemesi, yayınlar içerisinde bir başlangıç terim kümesi için çıkarsamalar yapması neredeyse imkansızdır.

Bununla beraber, LTB çalışmalarının bilimsel bir zeminde götürülebilmesi için, ortaya konan fikirlerin, geliştirilen algoritmaların tartışmalara açık olmayacak bir şekilde doğrulanması, sınanması vazgeçilemez bir gereksinimdir. Bu gereksinim, mevcut çalışmalarda genelde Swanson'un ilk çıkarımı olan Raynoud hastalığına referansla çözülmeye çalışılmıştır. Burada uygulanan yöntem, Swanson tarafından doğrulanmış hipotezin, aynı veri kümesi ile yeniden üretilmesini hedeflemektedir. Bunun için hipotezin ilk üretildiği tarih esas alınarak literatür kısıtlanmakta ve bu hipotezin üretilme şekli ve sıralanma başarımları tartışılmaktadır.

Ancak ortaya konması beklenen karşılaştırma yönteminin bilimsel yönü de büyük önem taşımaktadır. Kostoff bu konuda belirli kriterleri şu şekilde sıralamaktadır [37]:

1. Akademik çalışmaların değerlendirme yöntemleri, tüm hedeflerin kalitesini belirlenmesinde çözüm üretmelidir. LTB çalışmalarında uygulanan, daha önceden bulunmuş hipotezlerin yeniden üretilmesine yönelik işletilen yöntemde, diğer olası hipotezler hakkında net ifadeler kullanmak mümkün değildir.
2. Sonuç kümesinin çok az olması durumunda (LTB bağlamında genelde 3-4 sonuç üzerinde doğrulama yapılabilmektedir), araştırmacıların uygulama parametrelerini sadece bu çözümlere göre ayarlamaları engellenemez, diğer bir ifadeyle, uygulama genel amaçlı değil, dar kapsamlı bir başarıya odaklanmış olur.



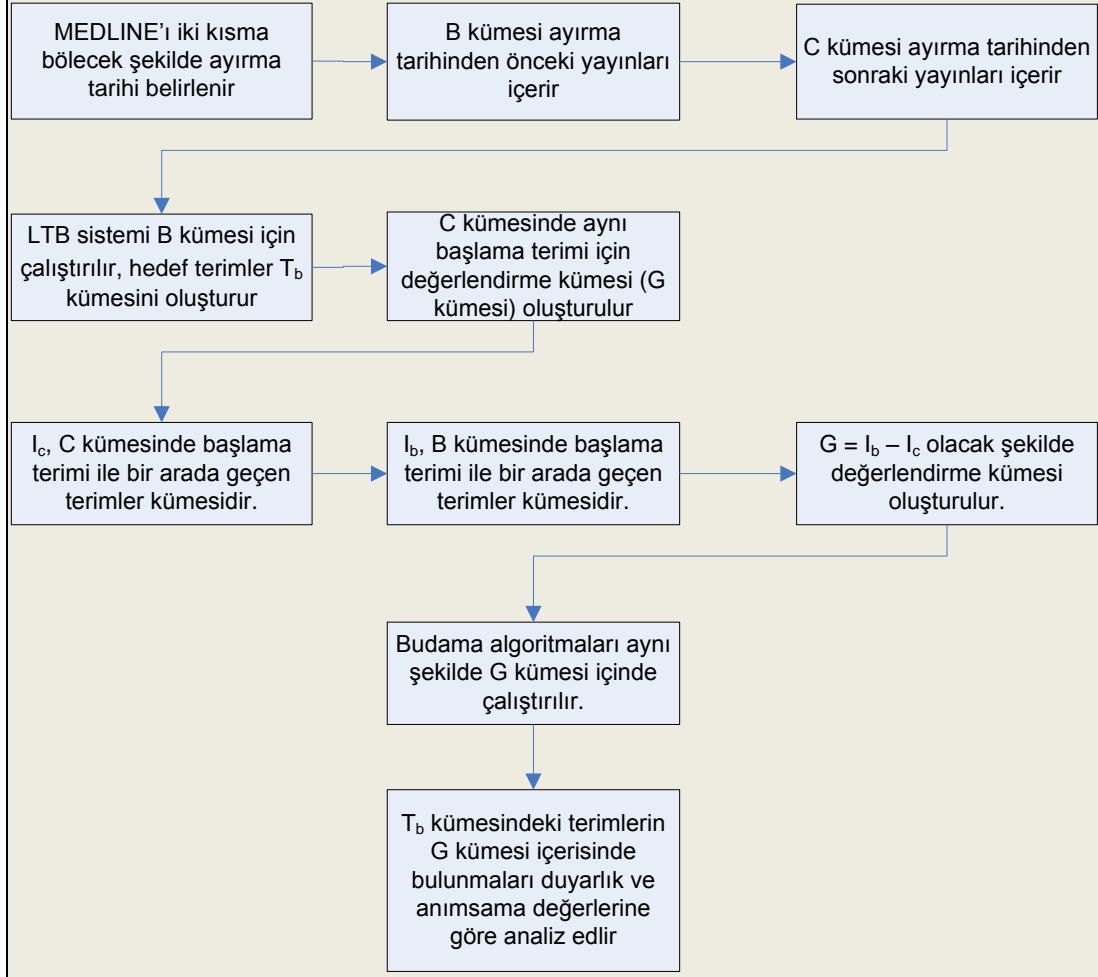
3. Karşılaştırma yöntemi, sadece doğru hedefe ulaşılmasını değil, bilgi getirim ("*information retrieval*") algoritmalarında olduğu gibi, sonuçların sıralanması konusunda da karşılaştırma imkanı sağlamalıdır. Böylelikle sistemlerin doğruluğunun yanında, kullanılabilirliği de ölçülebilir ve karşılaştırılabilir olmalıdır.

Bu bahsedilen sorunların bir kısmına yanıt verebilen bir yöntem olarak Pratt ve arkadaşlarının [37] ortaya attığı sına algoritması dikkate değerdir. Bu yöntemde belirli bir tarihe kadar kısıtlanan literatür üzerinde işletilen LTB sisteminin sonuçları, o tarihten sonra ortaya çıkan yeni hipotezlerle karşılaştırılmaktadır. Kendi ifadeleri ile geçmişin geleceğe ışık tutacağı ("*using history as a guide to the future*") varsayımıyla, PubMed yayınları belirli bir tarihe göre öncesi ve sonrası olarak ikiye ayrılmaktadır. İkiye ayrıldığı tarihe kesim tarihi ("*cut off date*") denir. Kesim tarihi öncesinde uygulama sonuçları temin edilir. Kesim tarihi sonrasındaki yayınlardan ise, başlama terimi ile kesim tarihinden önce ilişkili olmayan terimler üretilir. Böylelikle kesim tarihi öncesi "yeni" olarak üretilen sonuç terimleri ile kesim tarihi sonrası ilk defa ortaya çıkan bağlama terimleri karşılaştırılarak ölçüm yapılmış olur.

Pratt'in ortaya attığı bu karşılaştırma/doğrulama yöntemi, diğer yöntemlere göre büyük bir fark oluştursa da, bilimsel kriterler açısından kesin ve doğru bir yöntem olarak kabul edilemez. Pratt'in yöntemine yöneltilebilecek eleştiriler şu şekildedir:

1. Terimler arası ilişkiler üzerinden hipotez üretimi yapılmaktadır. Doğrulama kesiminde (kesim tarihinden sonraki yayınlar) ilişkiler yine aynı mantıkla üretilmektedir. Bu durumda algoritmaların ilişki çıkarma yöntemleri test edilememektedir.
2. Kesim sonrası yayınlar, hali hazırda yenilerinin eklenmesi sebebiyle hiç bir zaman tamamlanmayacaktır. Dolayısıyla altın standart olarak üretilen küme hiç bir zaman "tam" olmayacaktır.
3. Kesim tarihi için en uygun zamanın tüm çalışmalar için ortak olarak belirlenmesi gerekmektedir.

Tüm bu eleştirilerin yanında, şu anda LTB çalışmaları için en uygun sınaama/karşılaştırma yönteminin Pratt'in yöntemi olduğu görülmektedir. Bu bağlamda tez çalışmasının sonuçları bu yöntem kullanılarak sınaanmıştır.



Şekil 16. Sistemin Sınaama Yöntemi [38]

Pratt yönteminin uygulanmasında standart duyarlılık ve hatırlama yöntemleri ölçme yöntemi olarak kullanılabilir. Bununla beraber, sıralamanın öneminin gösterilmesi için Voorhees [78], olası yöntemleri bir araya getirmiştir. Genel yaklaşım, ilk 'n' sıradaki hedef terimlerin duyarlılık değerlerinin hesaplanmasına dayanmaktadır.

### 3.6. Verilerin büyüklüğü

Yapılan çalışma kapsamında üzerinde önemli durulması gereken bir diğer nokta da, çalışmanın yapıldığı alanın genişliğinin ve alanda yer alan nesnelere arası ilişkilerin toplam büyüklüğünün zorlu bir ortam oluşturmasıdır.

Terimlerin ve aralarındaki ilişkilerin yarı semantik bir yaklaşımla ortaya konduğu MeSH ağacından 52.546 adet düğüm (terim) bulunmaktadır. MeSH ağacı tez çalışması kapsamında yayınların indekslenmesinde kullanılmıştır. Basit anlamda bir yayını olası nitelik sayısı 52.546'dır. Terimlerden kavramlara geçiş fikri, bu sayının çok olmasından doğmuştur. Yapılan çalışma ile MeSH ağacında budamalar yapılmış ve nihayetinde 6.649 kavram geriye kalmış, ve yayınların indekslenmesinde 6.649 terim/kavram kullanılmıştır.

Toplam yayın sayısı da çalışma ortamını çok zorlaştırmaktadır. Çalışmanın algoritmalarının işletildiği ve sınındığı literatür kesiminde (1990 yılına kadar PubMed'de yayınlanmış makaleler) toplam 9.054.813 yayın bulunmaktadır. Diğer bir ifadeyle, tez çalışması diğer çalışmalarla aynı karşılaştırma imkanını sağlamak amacıyla 1990 yılına kadarki yayınları göz önünde bulundurmuş, bu tarihten sonraki yayınları sınama ortamı olarak değerlendirmiştir. Bu bağlamda 9 milyon yayın üzerinde parse işlemi yapılmıştır.

Kavramlar ve yayınlar arasında ilişki kurulduğu zaman toplam 63.152.162 adet ilişki ortaya çıkmıştır. Diğer bir ifade ile veri tabanında tek bir tabloda 63.152.162 satır bulunmaktadır. Kümeleme algoritması bu kadar satırı işlemek zorundadır. Kümeleme algoritmasının çalıştığı Vektör-Uzay Model'in boyutları ise sınırları zorlamaktadır. 16.000.000 x 6649 boyutların bir matris üzerinde kurulan modelde, en ufak bir algoritmanın dahi test edilmesi, kimi zaman haftalarca sürmektedir.

Bu kadar büyük bir veri tabanında kolaylıkla çalışabilmek için bazı önlemler alınması gerekiyor. Bunların bir kısmını şu şekilde sıralayabiliriz:

- Her bir terim için, geçtiği yayınlarda semantik TF-IDF değeri hesaplanacaktır. TF değeri o yayın için geçerli olsa da, IDF değeri terime ve literatürdeki diğer yayınlara bağlıdır. Dolayısıyla her bir terim için IDF değeri, TF-IDF'nin hesaplanacağı yayından bağımsız olarak

hesaplanabilir. Bu yaklaşımla, sistemde kayıtlı tüm terimlerin IDF değeri ilk işletimden önce hesaplanmış ve ihtiyaç oldukça direk bu değer kullanılmıştır.

- Terimler arası semantik benzerlik ihtiyaç duyulduğunda hesaplanmıştır. Zira her terimin mutlaka her terimle ikili olarak uzaklığının hesaplanmasına ihtiyaç duyulmayacaktır. Ancak en az bir yayında ikili olarak bir araya gelen iki terim, yüksek ihtimalle birden çok yayında tekrar bir arada kullanılacaktır. Her defasında semantik yakınlıklarının hesaplanması, işletim zamanını olumsuz yönde etkileyeceği için, iki terim ilk karşılaştıklarında hesaplanan semantik yakınlıkları, veri tabanında saklanmış ve sonraki karşılaştırmalarda direk hesaplanmadan bu değer kullanıma alınmıştır.
- Terim ağacı sıklıkla kullanılan bir yapıdır. Terim ağacında, terimlerin ataları ve dolaylı olarak çocukları saklanmaktadır. Bu ağaç üzerinde, bir terimin toplam çocuk sayısı, bir terimin köke uzaklığı, iki terimin birbirine olan uzaklığı ya da ortak atalarının mesafesi gibi hesaplamalar yapılmaktadır. Bu hesaplamalar standart ağaç veri yapısı ile çözülmek istense, her terim için bilinmeyen sayıda ata terimin çekilmesi ve sayılması gerekecektir. Ancak kullanılan veri yapısında, terimin köke kadar ki atalarının numaraları (örn. A01.B07.C03.D15) sırayla saklanmaktadır. Böylelikle yukarıda anlatılan tüm hesaplamalar çok hızlı bir şekilde gerçekleştirilebilmektedir.

Tüm bu optimizasyon işlemlerinden önce yaklaşık 1200 saat süren tek bir işletim döngüsü, 60 saate düşürülmüştür.

## 4. Sonular

Tez alıřmasında LTB'nin geliřtirilmesine ynelik kullanılan yntemlerin etkinlięi, mevcut yayınların son 10 yıla ait olanlarının kontrol kmesi olarak ayrılması ve daha eskiler zerinde sistemin iřletilmesi ile llmüřtr. 2000 yılına kadar yayınlanmış alıřmalar ile sistem iřletilmekte ve elde edilen sonular, 2000-2012 aralıęında keřfedilme durumuna gre deęerlendirilmektedir. Elbette kontrol kmesinin daha byk tutulması ile bařarı oranı artacaktır. Ancak alıřmanın somut ve yenilebilir bir yntemle etkinlięinin llmesi ve dięer alıřmalarla karřılařtırma yapılabilmesi amacıyla bu yntem yeterli grlmüřtr.

### 4.1. Kontrol kmesinin geniřlemesinin etkileri

Sistemin bařarımının llmesine ynelik nerilen yntemin yetkinlięi de test edilmelidir. Sorgulanması gereken en nemli nokta, eęer kontrol kmesi yeterince byk olsaydı, deęerlendirme sonucu olarak verilen anımsama ve duyarlık deęerlerinin ne kadar etkileneceęidir.

Anımsama ve duyarlık deęerleri sistemin getirdięi iliřkili nesne sayısını pay almaktadır. Sistemin getirdięi nesnelere hangilerinin iliřkili olduęu ise, kesim yılı (*cut-off date*) sonrasındaki yıllarda arama terimi ile doęrudan iliřkilendirilmiş yeni terimler zerinden belirlenmektedir.

1.Anımsama : Sistemin getirdięi iliřkili dokmanın, sistemde var olan iliřkili dokmana oranı.

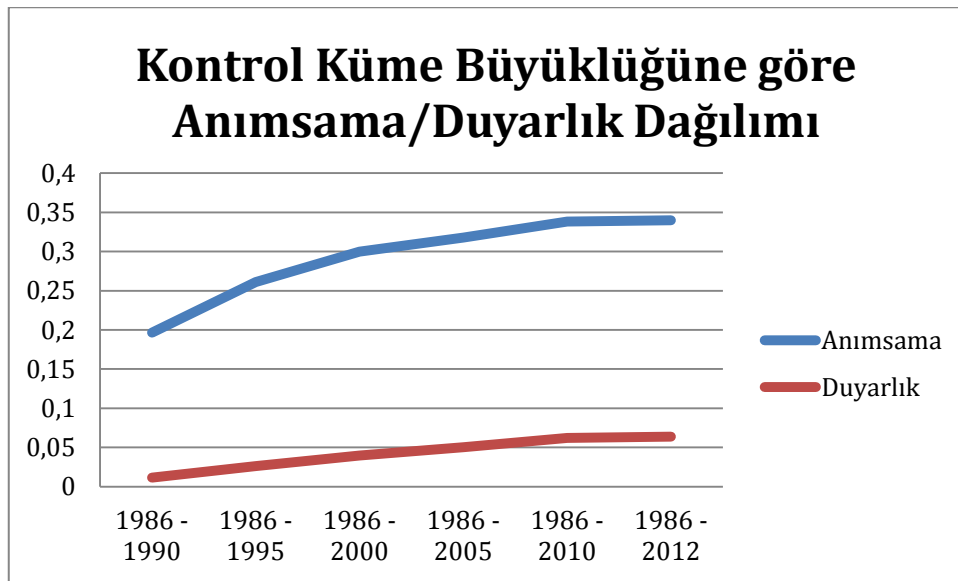
$$\text{Anımsama (Recall)} = \frac{\text{Doęru getirilen nesnelere}}{\text{Getirilmesi gereken tm doęru nesnelere}}$$

2.Duyarlık : Sistemin getirdięi doęru iliřkili dokman sayısının toplam getirilen dokman sayısına oranı.

$$\text{Duyarlık (Precision)} = \frac{\text{Doęru getirilen nesnelere}}{\text{Getirilen tm nesnelere}}$$

Kontrol kümesinin başlangıcı *cutt-off date* olmak üzere, bitişi değiştirilerek, kontrol kümesinin daha büyük olmasının sistemin başarısının değerlendirilmesi üzerine etkileri incelenmiştir.

Çalışmada “*Raynaud Disease*” terimi arama terimi olarak alınmıştır. Kesim yılı olarak 1986 yılı seçilmiş ve beş yıl aralıklarla kontrol kümesi büyütülmüştür. Ayrıca belirli yıl aralığında yapılan yayınların sonuca etkisinin daha rahat görebilmesi amacıyla, her bir beş yıl da bir o yayınlar kontrol kümesi olarak işletilmiştir. Diğer bir ifadeyle, ikinci çalışmada kontrol kümesinin bitiş tarihi ilerletilirken, başlangıç tarihi de beraberinde ilerletilmiştir.



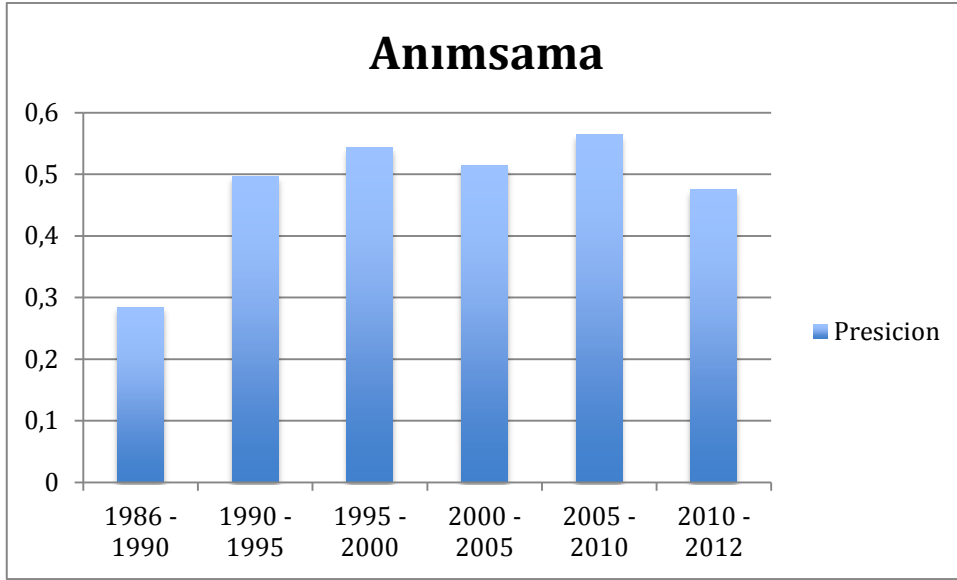
Şekil 17. Yıllara Göre Anımsama/Duyarlık Dağılımı

Kontrol Yılları		Anımsama	Duyarlık
1986 - 1990	Pay	190	190
	Payda	967	16246
	<b>Sonuç</b>	<b>0.20</b>	<b>0.01</b>
1986 - 1995	Pay	423	423
	Payda	1621	16246
	<b>Sonuç</b>	<b>0.26</b>	<b>0.03</b>
1986 - 2000	Pay	644	644
	Payda	2146	16246
	<b>Sonuç</b>	<b>0.30</b>	<b>0.04</b>
1986 - 2005	Pay	816	816
	Payda	2566	16246
	<b>Sonuç</b>	<b>0.32</b>	<b>0.05</b>
1986 - 2010	Pay	1010	1010
	Payda	2984	16246
	<b>Sonuç</b>	<b>0.34</b>	<b>0.06</b>
1986 - 2012	Pay	1039	1039
	Payda	3057	16246
	<b>Sonuç</b>	<b>0.34</b>	<b>0.06</b>

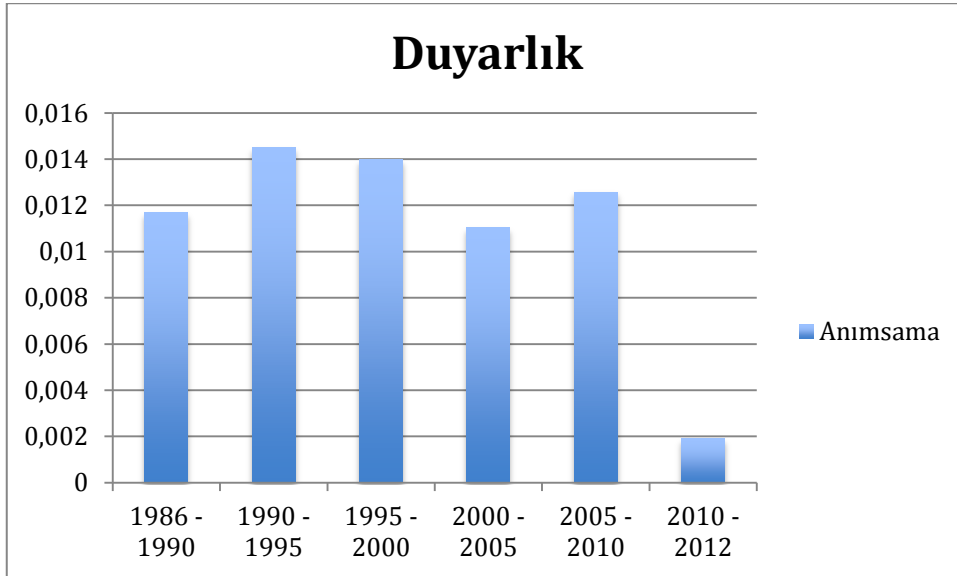
Çizelge 8. Yıllar Göre Kümülatif Duyarlık ve Anımsama Dağılımı

		Anımsama	Duyarlık
1986 - 1990	Pay	190	190
	Payda	967	16246
	<b>Sonuç</b>	<b>0.20</b>	<b>0.01</b>
1990 - 1995	Pay	233	233
	Payda	470	16056
	<b>Sonuç</b>	<b>0.50</b>	<b>0.01</b>
1995 - 2000	Pay	221	221
	Payda	407	15823
	<b>Sonuç</b>	<b>0.54</b>	<b>0.01</b>
2000 - 2005	Pay	172	172
	Payda	335	15602
	<b>Sonuç</b>	<b>0.51</b>	<b>0.01</b>
2005 - 2010	Pay	194	194
	Payda	344	15430
	<b>Sonuç</b>	<b>0.56</b>	<b>0.01</b>
2010 - 2012	Pay	29	29
	Payda	61	15236
	<b>Sonuç</b>	<b>0.48</b>	<b>0.00</b>

Çizelge 9. Yıllara Göre Parçalı Duyarlık ve Anımsama Dağılımı



Çizelge 10. Yıllara Göre Anımsama Dağılımı



Çizelge 11. Yıllara Göre Duyarlık Dağılımı



## 4.2. Bir örnek üzerinde ilk 100 sonuç teriminin incelenmesi

Literatür Tabanlı Bilgi Keşfi alanında yapılmış diğer bir çok çalışmada, önerilen yöntemlerin etkinliği, ortak bir işletim alanı olarak Raynaud Hastalığı üzerinden incelenmiştir. Tez çalışması sonuçlarının diğerleri ile karşılaştırılabilmesine imkan tanımak amacıyla, Raynoud Hastalığı ayrıca ele alınmıştır.

Bu amaçla, 1960-1986 yılları arasında yayınlanmış makale özetleri temel alınarak, başlangıç terimi “Raynaud Disease” olmak üzere uygulama işletilmiş, sonuç terimleri ortak küme sayısının çokluğuna göre sıralanmış ve ilk yüz terim listelenmiştir. Ek.1 ile Raynaud Hastalığı ve Altın Standartların Dağılımı verilmiştir, Ek.2’de ise İşletim Sonuçları verilmiştir.

Verilen listede “İlk Bulunma Tarihi”, başlangıç terimi ile sonuç teriminin ilk defa bir arada bulunduğu yayının tarihini göstermektedir. Bu veri, iki terimin ilk defa ne zaman çalışıldığını göstermektedir. “Ortak Küme Sayısı”, sonuç terimi ile arama teriminin bir arada geçtiği toplam küme sayısını göstermektedir. Bu veri ile iki terim arasında ilişkinin gücü gösterilmek istenmiştir. Küme yakınlığı, küme içerisinde başlangıç terimi ile ilgili makale sayısının kümedeki toplam makale sayısına oranını göstermek üzere, toplam küme yakınlığı ve ağırlıklı toplam küme yakınlıkları verileri tabloda gösterilmiştir.

Toplam küme yakınlığı, ilgili terim ile başlangıç teriminin ortak olduğu tüm kümelerin yakınlıklarının toplamıdır. Ağırlıklı Toplam ise, kümenin yakınlığının ilgili terimin o kümede geçme sıklığı ile çarpımları ağırlıklandırılması ve bu değerlerin toplanması ile elde edilir. Ağırlıklı Toplam Küme Yakınlığı’nın Ortak Küme Sayısına bölünmesi ile de o terimin başlangıç terimi ile ilişki katsayısı belirlenmiş olur.

Aşağıda verilen listede 39 tane terim True-Positive olarak işaretlenmiştir. İlk 100 terime göre duyarlık değeri 0.39 olmaktadır.

### 4.3. Ölçme Yönteminin Bir Terim için İncelenmesi

Ölçme yöntemi olarak kullanılan kontrol kümesi mantığı, diğer çalışmalarla karşılaştırma imkanı vermek amacıyla yine Raynaud hastalığı kullanılarak işletilmiştir. Ek.3. Ölçme Yönteminin Sınanması başlığı altında verilen sonuç kümesinde 1986'dan önce Raynaud hastalığı ile ilişkisi olmadığı halde, sonraki yıllarda en az bir çalışmada Raynaud hastalığı ile bir arada çalışılmış terimler ve beraber geçtikleri yayın sayıları verilmiştir.

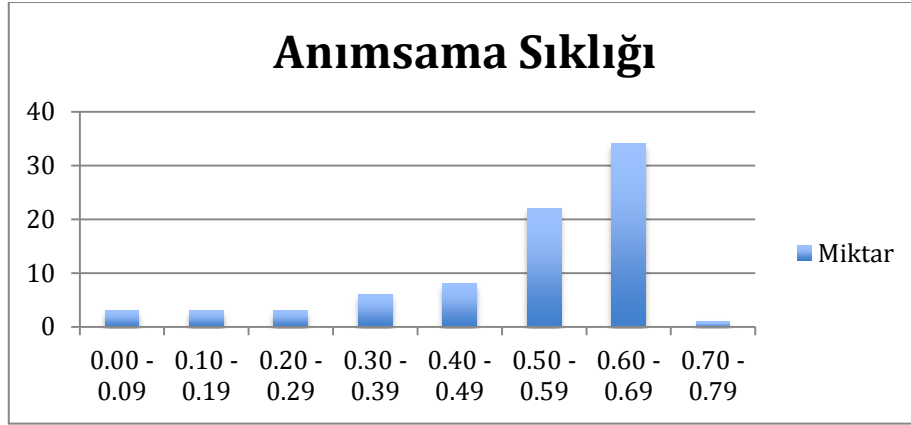
### 4.4. Örnek 80 terim için başarı incelemesi

Tamamen rasgele seçilmiş 80 terim ile sistemin ortalama başarımı ölçülmüştür. Budanmış Mesh ağacında bulunan terimlerden (yani kavramlardan) 1990 yılından önce en az bir yayında geçmiş, rasgele seçilmiş 80 terim için, 1990 yılı kesim yılı olmak üzere LTB uygulaması işletilmiştir. Buna göre, her bir terim için ayrı ayrı 1990 yılı öncesi yayınlar esas olmak üzere işletim yapılmış, yine her terim için çıkan sonuçlar 2012 yılına kadar yayınlanmış metinler üzerinde doğrulukları incelenmiştir. Değerler Ek.4. Rasgele 80 Terim için Sonuçlar başlığı altında paylaşılmıştır.

Bu çalışma ile ortalama anımsama değeri 0.52 çıkarken, duyarlık değeri 0.16 çıkmıştır. Anımsama dağılımına bakıldığı zaman, elde edilen değerlerin %42'sinin 0.60-0.69 aralığına toplandığı görülebilir. Ancak beklendiği gibi Duyarlık değeri, anımsama değerinin artması ile düşüş göstermektedir.

Anımsama	Miktar	Dağılım %	Duyarlık Ort.
0.00 - 0.09	3	3.75	0.47
0.10 - 0.19	3	3.75	0.52
0.20 - 0.29	3	3.75	0.22
0.30 - 0.39	6	7.50	0.24
0.40 - 0.49	8	10.00	0.22
0.50 - 0.59	22	27.50	0.11
0.60 - 0.69	34	42.50	0.1
0.70 - 0.79	1	1.25	0.06

Çizelge 12. Örnek 80 Terim için Anımsama Dağılımı

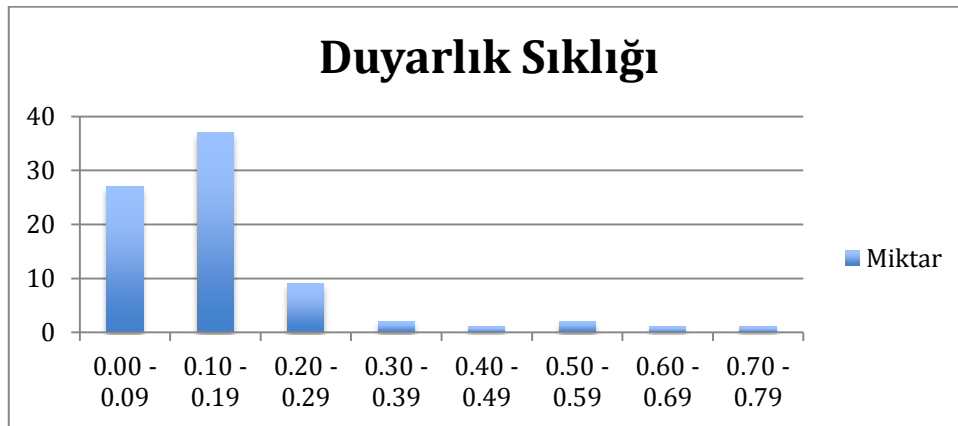


Şekil 18. Örnek 80 Terim için Anımsama Grafiği

Çalışmada ele alınan 80 terim için Duyarlık gruplaması yapıldığı zaman, %79'unun 0.2'den düşük değerler aldığı görülmektedir. Ancak Duyarlık yüksek olduğu değerlerde dahi, anımsama oranında büyük bir düşüş olmamıştır.

Duyarlık	Miktar	Dağılım %	Anımsama Ort.
0.00 - 0.09	27	33.75	0.62
0.10 - 0.19	37	46.25	0.55
0.20 - 0.29	9	11.25	0.39
0.30 - 0.39	2	2.50	0.3
0.40 - 0.49	1	1.25	0.32
0.50 - 0.59	2	2.50	0.11
0.60 - 0.69	1	1.25	0.01
0.70 - 0.79	1	1.25	0.16

Çizelge 13. Örnek 80 Terim için Duyarlık Dağılımı



Şekil 19. Örnek 80 Terim için Duyarlık Grafiği

## 5. Tartışma

### Literatür Tabanlı Çalışmalarda Sorunlara Yeni Bakış

Literatür Tabanlı Bilgi keşfi çalışmalarında genel olarak ortaya konulan sıkıntılar düşük Duyarlık rakamlarıyla ilgilidir. Diğer bir ifadeyle elde edilen sonuçlar tatmin edici düzeyde doğru sonuçları içeriyor olsa da, listeler son kullanıcının incelemesine olanak vermeyecek şekilde kalabalık olmaktadır. Bu sorunu gidermek amacıyla çok sayıda önlem alınabilir. ABC modelinde genelde görülen çözümler, (I) bağlama terimlerinin daha nitelikli ve doğru seçilmesi, (II) terimlerin ilişkilerinde anlamsa tutarlılığın zorlanması, (III) mevcut bilgi alt yapıları ile bazı ilişkilerin elenmesi ve (IV) kullanıcının bağlama terim türlerini sınırlandırabilmesi gibi işletimsel mantıklar üzerine kurgulanmıştır.

Ancak yaptığımız çalışmada, bu sorunu çok daha temel düzeyde çözümlenmesi planlanmıştır. Bu amaçla sunulacak yöntemin, soruna destekleyici unsurlarla çözüm getirmesi yerine; temel işletimde değişiklik yapılarak, kalıcı ve yapısal çözümler getirmesi gerektiği öngörülmüştür. Somut olarak, terimler arası ilişkilerin “aynı yayında geçmeleri” ile kuruluyor olması, gürültülü verinin temel sebebidir. Çözüm önerisi olarak bilgi kümelerinden esinlenerek, yayın kümeleri ön plana çıkarılmıştır. Bir bilim insanının, okuduğu yayınları zihninde kümelediği ve kavramların aynı kümede önemli oranda görünmeleri gözetilerek ilişkilendirildikleri varsayımı ortaya atılmıştır. Bu varsayımın bilgisayarlı sistemler üzerinden benzetimi amacıyla, kaba küme teoremi kullanılarak, yayınlar kümelenebilir ve bu kümeler üzerinden ilişkilerin kurulması sağlanmıştır.

### Metinlerin Kümelenmesi

Metinlerin kümelenmesi tez çalışmasının başarımını etkileyen en önemli modüldür. Metin kümelenimin başarımını incelemek amacıyla yaptığımız öncül çalışma önemli sonuçlar içermektedir. PubMed verileri üzerinde, altın standart olarak OHSUMED verilerinin kullanıldığı çalışmanın sonuçlarında F-Ölçüm

değeri 0.92 olarak bulunmuştur. F-Ölçüm, duyarlık ve hatırlama değerlerinin birleşiminden oluşmaktadır. Bu noktada en yüksek değerin 1.00 olduğu ve benzer çalışmalarda başarımın 0.85-0.95 aralığında olduğu gözetildiğinde, 0.92 değeri tatmin edici bir başarı olduğu düşünülmektedir. Böylelikle kümeleme modülü sistemin genel başarısını destekleyecek şekilde yapılandırıldığı görülmüştür.

### Terimler Yerine Kavramların Kullanımı

Tez kapsamında yapılan bir diğer önemli gelişme de, terimlerden bir kısmının daha az detaylı bilgi içeren terimlerle temsil edilmesinin sağlanmasıdır. Diğer bir ifade ile, detaylı bir bilgiyi temsil eden bir terim, terim ağacındaki atalarından daha kavramsal düzeyde bilgi temsil eden bir terim ile temsil edilmektedir. Böylelikle detay düzeyde kurulan bir çok ilişki, aslında aynı şeyi ifade ettikleri düşüncesi ile kavramsal düzeyde birleştirilmiştir.

Terimler yerine kavramlarla işleme yönelik yapılan çalışmalarda (I) çok genel olanlar, (II) detay seviyesi yüksek olanlar, (III) çocuğu olmayanlar elenmiştir. Eleme işleminde bilgi kaybı olmaması amacıyla, elenen terimleri içeren yayınlar, o terimin elenmemiş atası ile indekslenmiştir. Böylelikle 36.514 adet terim, 6649 adet kavrama indirgenmiştir. Bu indirgemenin en temel katkısı, kümeleme algoritmasına verilecek matrisin bir boyutunun düşürülmesidir. Ancak 36.514 terim ile kurulacak matris üzerinde kümeleme işlemi teknik sorunlar sebebiyle yapılamadığı için, eleme öncesi ve sonrası arasında kümeleme bağlamında yaşanan değişimlere ilişkin bir karşılaştırma verilememiştir.

### Sınama Yöntemleri

Genel kabul görmüş bir sınama yönteminin bulunmaması tez çalışmasının doğrulanmasında büyük bir sorun olarak karşımıza çıkmıştır. Problemin doğası gereği, henüz keşfedilmemiş ilişkilerin keşfedilmesine yönelik bir listenin bulunması imkansızdır. Geçmişe dönük çalışma yapılsa bile, geçmişteki yayınlar içerisinde dahi tüm keşiflerin yapıldığı iddia edilemez.

Bu soruna yönelik çözüm, bir yıldan önceki yayınlar üzerinde işletimin yapılması ve sonraki yayınlar ile elde edilen sonuçların karşılaştırılmasıdır. Bu yöntem, tez çalışmasının diğer çalışmalarla karşılaştırılmasına imkan sağlamak amacıyla gerçekleştirilmiştir.

Kontrol kümesinin genişlemesinin etkilerini incelediğimiz çalışmada, 1986-1990 yılları kontrol kümesi alındığında sırasıyla 0.20 ve 0.01 olan anımsama ve duyarlık değerleri, 1986 – 2012 aralığında 0.34 ve 0.06 değerlerine yükselmektedir. Bu yükseliş, 1990-2012 aralığında 1990'dan önce bulunmamış yeni ilişkilerin çıkarıldığını ve bu yeniliklerin hiç de küçümsenmeyecek düzeyde olduğunu göstermektedir.

Kontrol kümesine ilişkin yapılan aynı çalışmada elde edilen bir diğer sonuç ise, günümüze yakın zamanlarda elde edilen başarımın çok daha yüksek olduğudur. 1986-1990 yıllarında elde edilen başarım ile sadece 2005-2010 yıllarında yayınlanmış yayınların kontrol grubu olarak alınması karşılaştırıldığında, 0.20 olan anımsama değeri 0.56 olarak ölçülmüştür. Aslında günümüze yaklaştıkça yayın sayısının artması bu sonucun temeli gibi görünse de, dikkat edilmesi gereken en önemli çıkarım, yıllar geçtikçe başarım değerleri – algoritma ve veriler değişmediği halde – kendiliğinden yükselecektir.

Burada elde edilen sonuç, LTB çalışmalarında görülen düşük anımsama ve duyarlık değerleri ile diğer çalışma alanlarındaki yüksek değerlerle karşılaştırma yapmanın mantıksızlığını göstermektedir. LTB çalışmalarında elde edilen anımsama ve duyarlık değerleri ancak LTB çalışmalarını kendi içlerinde karşılaştırmak için kullanılabilir, bir çalışmanın diğerlerine göre ne kadar başarılı olduğunu kanıtlamak amacıyla bir gösterge olabilir. LTB çalışmalarında, daha önce anlatılan sına yöntemi kullanılarak elde edilen anımsama ve duyarlık değerleri, bilgi erişim alanındaki gerçek karşılığını göstermemektedir. Problemin doğası gereği gerçek değerler asla hesaplanamayacaktır.

Diğer taraftan iki kavramın ilişkili olduğunu tıp alanında doğrulamak ya da ispatlamak çoğu zaman oldukça uzun zaman almaktadır. Hatta bazı örneklerde, o gün için doğru olduğu varsayılan ifadeler, zaman içerisinde

anlamını yitirmekte ya da hatalı olduđu gör÷lmektedir. Bu şartlar altında LTB için anımsama ve duyarlık sonuçlarının, alışlageldik arama motorları başarımları ile karşılaştırmak hatalı olacaktır. LTB algoritması ile kurulmuş bir ilişki hali hazırda keşfedilmemiştir, gelecekte keşfedilmesi mümkündür. Dolayısıyla LTB sonuçlarını mevcut bilgi birikimi ile doğrulamak felsefi olarak hatalıdır. Bu yaklaşım sadece LTB alanında yapılan teknik çalışmaların, diğerleri ile karşılaştırılmasına yönelik kabul edilebilir.

### Uygulamanın Sonuçları

Tez kapsamında yapılan çalışmada, Kaba küme teorisinin LTB üzerine uygulanması ile daha önce bahsedilen bir dizi soruna çözüm üretilmesi hedeflenmiştir. Bu amacın başarımına yönelik yaptığımız deneylerden ilkinde diğer çalışma grupları ile karşılaştırma imkanı vermek amacıyla, 1960-1986 yılları arasındaki yayınlar üzerinde “Raynaud Disease” teriminin araştırılması yapılmıştır. İlk yüz sonuca bakıldığı zaman 39 tane terim, o zamana kadar ilişkili olmadığı halde daha sonra “Raynaud Disease” terimi ile ilişkilendirilmiştir.

Yapılan bir diğer işleminde ise budanmış Mesh ağacında bulunan terimlerden (yani kavramlardan) 1990 yılından önce en az bir yayında geçmiş, rasgele seçilmiş 80 terim için, 1990 yılı kesim yılı olmak üzere LTB uygulaması işletilmiştir. Ortalama anımsama değeri 0.52 çıkarken, duyarlık değeri 0.16 çıkmıştır.

### Diğer Çalışma Grupları ile Karşılaştırma

Tahir BIÇAKCI'nın LTB'nin standart gerçekleştirimi ile elde ettiği anımsama değeri 0.42 olarak verilmiştir [38]. Yetişgen Yıldız ise benzer bir çalışmasında [37], hatırlama değerlerini 0,20 ile 0,25 arasında değiştirken şekillerde sunmuştur. Aynı çalışmada da duyarlılık değerleri 0,01 ile 0,06 arasında değişmektedir. Tez çalışmamızda ise ortalama anımsama 0.52 ve ortalama duyarlık değeri 0.16 olarak ölçülmüştür.

Diğer çalışmalarda genelde test yöntemi ortaya konulmamış, anımsama ve duyarlık değerleri ölçülmemiştir. Bu bağlamda ancak daha önce yapılmış iki çalışma ile karşılaştırma yapma imkanı bulunmaktadır.

### Yapılabilecek Olası İyileştirmeler

Terimlerin elenmesine ilişkin daha detaylı çalışma yapılması gerekmektedir. Terimlerin elenmesinin çalışma üzerine etkilerinin analiz edilmesi için tez çalışmasının süresinin yetmeyeceği görülmüş, ayrıca eldeki sunucu donanımının yeterli olmadığı tespit edilmiştir.

Terimlerin “genel olmaları” sebebiyle elenmesine yönelik, uygulama alanları gözetilerek bir çalışma yapılmalıdır. Bir terimin sıklığı, genelliği konusunda fikir veriyor olsa da, aslında genel bir terim olmadığı halde, dar bir çalışma alanında genel terim olarak kabul edilecek terimler, bu elemelerde göz ardı edilmektedir. Benzer şekilde, aslında çok genel bir terim olmadığı halde, çok sayıda yayının bulunduğu bir alanda eleme sayısına ulaşan terimler de, duruma göre hatalı olarak elenebilmektedir. Eleme işleminin sadece sıklığa göre yapılmasının yerine, alan uzmanlarınca işlenmesi daha doğru bir yaklaşım olacaktır.

Tez çalışması, araştırma odaklı yapılmıştır. Ancak son kullanıcılara sunulması için, işletiminin kullanıcı tarafından yönlendirilebiliyor olması gerekir. Örneğin, genel terim kümeleri kullanıcı tarafından belirlenebilmeli, sadece bir başlangıç terimi ile değil, birden çok terim ya da bir alan belirtilerek işletim yapılabilmelidir.

Diğer taraftan bir çok araştırmacı bilim insanının ilgisini çeken konular “hot topic” olarak tarif edilmektedir. Akademik dikkatini çekme konusunda daha da ön plana çıkan konularda ortaya çıkabilecek sonuçlar, nihai kullanıcıların daha çok ilgisini çekecektir. Tüm alt yapı gözden geçirilerek, “hot topic” konuların yakalanması, faydalı bir geliştirim olacaktır.

LBD sistemleri, Doğal Dil İşleme algoritmalarının gelişmesi ile daha başarılı çalışmalarla güçlenecektir. Metinler üzerinde nesnelere anlamsal değerleri ile



diğer nesnelere arasındaki ilişkilerin kurulması sağlanabilir. Bu nokta anlamda ya da eş anlamlı kelimelerin cümlede taşıdığı ifadelerle göre belirlenmesi gerekecektir. Ancak özellikle İngilizce Tıbbi metinler için, hem doğal dil işleme imkanları gelişmiş, hem de tıp alanında güçlü sözlükler üretilmiştir.

Diğer taraftan mevcut bilgilerin düzenli ve yapısal olarak saklandığı ilaç veri tabanı, gen veri tabanı gibi yapısal bilgi kaynaklarının kullanımı da büyük önem taşımaktadır. Yapısal bilgi kaynakları ile anlamsız sonuçlar elenebileceği gibi, kümeleme süreçleri daha iyi yönetilebilecektir.

Tez çalışması kümelemenin başarıyla LTB alanına uyarlanmasını sağlamıştır. Ancak kümeleme çalışmaları 16 milyon satır ve 65 bin sütundan oluşan bir matrisi, yeterince büyük küme hacimleri olarak şekilde analiz etme konusunda yetersiz kalmıştır. Ortaya çıkacak kümelerin hacimleri arttıkça başarımın yükseldiği görülmüştür. Ancak nihai olarak 20.000 yayından oluşan kümelemeler üretilebilmiş, daha büyük kümeler kullanılan algoritmaların/uygulamaların daha fazlasını kaldıramaması sebebiyle sınırlanmıştır. Daha büyük kümelemelerin yapılabilmesi amacıyla, ön kümeleme yapılabilir, 16 milyon yayın 10 parçaya ayrılabilir ve her parça üzerinde ayrıca kümeleme algoritması işletilebilir. Ancak bu çalışmanın yapılabilmesi için ana kümenin nasıl bölünebileceğine ilişkin detaylı bir çalışmanın yapılması gerekmektedir.

## Kaynaklar

- [1] Canadian Academy of Health Science - Framework Document, "The Benefits and Barriers to Interdisciplinary Research in the Health Sciences in Canada," Jan. 2005.
- [2] A. Yıldırım, "Disiplinlerarası Öğretim Kavramı ve Programlar Açısından Doğurduğu Sonuçlar," *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, vol. 12, pp. 89–94, 1996.
- [3] F. Valentin, M. Larsen, and L. Alkærsig, "Breakthroughs And Interdisciplinary Research - The Case Of Research Behaviour In Translational Medical Science," in *Proceedings from the DRUID Summer, 2008*.
- [4] R. A. Day, *How to Write & Publish a Scientific Paper: 5th Edition*. Oryx Press, 1998, p. 296.
- [5] US National Library of Medicine National Institutes of Health, "PubMed," 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/>. [Accessed: 06-May-2013].
- [6] USA National Library of Medicine, "Medline/PubMed Baseline Repository," 2012. [Online]. Available: <http://mbr.nlm.nih.gov/>. [Accessed: 05-Jun-2013].
- [7] J. D. Wren, "Extending the mutual information measure to rank inferred literature relationships," *BMC Bioinformatics*, vol. 5, p. 145, Oct. 2004.
- [8] J. D. Wren, R. Bekeredjian, J. A. Stewart, R. V. Shohet, and H. R. Garner, "Knowledge discovery by automated identification and ranking of implicit relationships," *Bioinformatics (Oxford, England)*, vol. 20, no. 3, pp. 389–98, Feb. 2004.
- [9] P. M. Blau, *The Organization of Academic Work*, 2nd ed. New Jersey, USA: John Wiley & Sons, 1994.

- [10] E. Enkel and O. Gassmann, "Open R&D and open innovation: exploring the phenomenon," *R&D Management*, vol. 39, no. 4, pp. 311–316, 2009.
- [11] A. Hidalgo, "Innovation management techniques and tools: a review from theory and practice," *R&D Management*, 2008.
- [12] M. A. White and G. D. Bruton, *The management of technology and innovation: a strategic approach*. Toronto, Ontario: Thomson Higher Education, 2007.
- [13] R. A. Burgelman, C. M. Christensen, and S. C. Wheelwright, *Strategic Management of Technology and Innovation*. New York: McGraw-Hill, 2004.
- [14] P. Bruza and M. Weeber, *Literature-based discovery*. Springer-Verlag New York Inc, 2008.
- [15] D. R. Swanson, "Undiscovered Public Knowledge," *Library Quarterly*, vol. 56, no. 2, pp. 103–118, 1986.
- [16] D. R. Swanson, "Fish Oil, Raynauds Syndrome, and Undiscovered Public Knowledge," *Perspectives in Biology and Medicine*, vol. 30, no. 1, pp. 7–18, 1986.
- [17] G. Karypis, "Techincal Report: CLUTO-a clustering toolkit," University of Minnesota, Department of Computer Science, 2002.
- [18] M. Weeber and H. Klein, "Using Concepts in Literature-Based Discovery: Simulating Swanson's Raynaud – Fish Oil and Migraine – Magnesium Discoveries," *Journal of the*, vol. 52, no. 7, pp. 548–557, 2001.
- [19] "Arrowsmith Tool," 2012. [Online]. Available: [http://arrowsmith.psych.uic.edu/arrowsmith\\_uic/index.html](http://arrowsmith.psych.uic.edu/arrowsmith_uic/index.html).

- [20] N. Smalheiser and V. Torvik, "Collaborative development of the Arrowsmith two node search interface designed for laboratory investigators," *Journal of Biomedical Discovery and Collaboration*, vol. 1, p. 8, Jan. 2006.
- [21] R. K. Lindsay and M. D. Gordon, "Literature-based discovery by lexical statistics," *Journal of the American Society for Information Science*, vol. 50, no. 7, pp. 574–587, 1999.
- [22] M. D. Gordon and S. Dumais, "Using latent semantic indexing for literature based discovery," *Journal of the American Society for Information Science*, vol. 49, no. 8, pp. 674–685, 1998.
- [23] M. Gordon, R. K. Lindsay, and W. Fan, "Literature-based discovery on the World Wide Web," *ACM Transactions on Internet Technology*, vol. 2, no. 4, pp. 261–275, Nov. 2002.
- [24] Y. Xie, J. Katukuri, and V. Raghavan, "Conceptual Biology Research Supporting Platform: Current Design and Future Directions," *Intelligence in Biology*, pp. 1–17, 2008.
- [25] W. Pratt and M. Yetisgen-Yildiz, "LitLinker: Capturing Connections Across the Biomedical Literature," in *Proceedings of the 2nd International Conference on Knowledge Capture*, 2003, pp. 105–112.
- [26] M. Yetisgen-Yildiz and W. Pratt, "Using statistical and knowledge-based approaches for literature-based discovery.," *Journal of biomedical informatics*, vol. 39, no. 6, pp. 600–11, Dec. 2006.
- [27] R. N. Kostoff, M. B. Briggs, J. L. Solka, and R. L. Rushenberg, "Literature-related discovery (LRD): Methodology," *Technological Forecasting and Social Change*, vol. 75, no. 2, pp. 186–202, Feb. 2008.
- [28] J. R. W. Ronald N. Kostoff, Jeffrey L. Solka, Robert L. Rushenberg, "Literature-related discovery (LRD): Water purification," *Technological Forecasting and Social Change*, vol. 75, no. 2, pp. 256–275, 2008.

- [29] R. N. Kostoff, J. A. Block, J. A. Stump, and D. Johnson, "Literature-Related Discovery (LRD): Potential Treatments for Raynaud's Phenomenon," *Technological Forecasting and Social Change*, vol. 75, no. 2, pp. 203–214, Feb. 2008.
- [30] R. N. Kostoff, M. B. Briggs, and T. J. Lyons., "Literature-related discovery (LRD): Potential treatments for multiple sclerosis," *Technological Forecasting and Social Change*, vol. 75, no. 2, pp. 239 – 255, 2008.
- [31] R. N. Kostoff and M. B. Briggs, "Literature-related discovery (LRD): Potential treatments for Parkinson's disease," *Technological Forecasting and Social Change*, vol. 75, no. 2, pp. 226– 238, Feb. 2008.
- [32] I. Petrič, B. Cestnik, and N. Lavrač, "Bisociative Knowledge Discovery by Literature Outlier Detection," *Bisociative Knowledge*, pp. 313–324, 2012.
- [33] A. Koestler., *The Act of Creation*. New York: The Macmillan Co., 1964.
- [34] B. Cestnik, "Cross-domain Literature Mining: Finding Bridging Concepts with CrossBee," in *International Conference on Computational Creativity*, 2012, no. P2 0103, pp. 33–40.
- [35] R. N. Kostoff, "Literature-related discovery and innovation — update," *Technological Forecasting and Social Change*, vol. 79, no. 4, pp. 789–800, May 2012.
- [36] S. Huang, L. He, B. Yang, and M. Zhang, "A compound correlation model for disjoint literature-based knowledge discovery," *Aslib Proceedings*, vol. 64, no. 4, pp. 423–436, 2012.
- [37] M. Yetisgen-Yildiz and W. Pratt, "A new evaluation methodology for literature-based discovery systems.," *Journal of biomedical informatics*, vol. 42, no. 4, pp. 633–43, Aug. 2009.

- [38] T. Bıçakcı, "Using Biomedical Concepts for Drug Discovery," Hacettepe University, 2009.
- [39] J. Han and M. Kamber, *Data mining: concepts and techniques*, Second Edi. San Francisco: Morgan Kaufmann, 2006.
- [40] L. A. Riloff, E. and Hollaar, "Text databases and information retrieval," *ACM Computing Surveys (CSUR)*, vol. 23, no. 1, pp. 133 – 135, 1996.
- [41] R. Baeza-Yates and B. Ribeiro-Neto, "Modern information retrieval," B. Baeza-Yates, R. and Ribeiro-Neto, Ed. New York: ACM Press, Addison-Wesley, 1999.
- [42] D. Shasha and J. Wang, "Algorithmics and Applications of Tree and Graph Searching," *Proceedings of the twenty-first ACM*, 2002.
- [43] G. Chang, M. Healey, J. A. M. McHugh, and T. L. Wang, *Mining the World Wide Web: An Information Search Approach*. Boston: Kluwer Academic, 2001.
- [44] W. S. Cleveland, *Visualizing data*. New Jersey, USA: AT&T, 1993.
- [45] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th Int. Conf. Very Large Data Bases*, 1994.
- [46] M. Heikki and T. Hannu, "Efficient Algorithms for Discovering Association Rules," in *Workshop on Knowledge Discovery in Databases*, 1994, pp. 181–192.
- [47] L. Kaufman and P. Rousseeuw, *Finding groups in data: an introduction to cluster analysis*, Second Edi. New Jersey, USA: John Wiley & Sons, 2005.
- [48] J. Wen and J. Nie, "Clustering user queries of a search engine," *of the 10th international conference on*, no. 49, pp. 162–168, 2001.

- [49] G. Salton, A. Wong, and C. S. Yang, "A Vector Space Model for Automatic Indexing," *Communications of the ACM*, vol. 18, no. 11, 1975.
- [50] K. S. Jones, "A statistical interpretation of term specificity and its application in retrieval," *Journal of Documentation*, vol. 60, no. 5, pp. 493–502, 2004.
- [51] R. Feldman and J. Sanger, *The Text Mining handbook: Advanced Approaches to Analyzing Unstructured Data*. New York: Cambridge University Press, 2007.
- [52] M. Krallinger and R. Erhardt, "Text-mining approaches in molecular biology and biomedicine," *Drug discovery today*, vol. 10, no. 6, 2005.
- [53] "UMLS Official WEB Site." [Online]. Available: [www.nlm.nih.gov/research/umls](http://www.nlm.nih.gov/research/umls). [Accessed: 20-Oct-2012].
- [54] "UMLS Online Learning." [Online]. Available: [http://www.nlm.nih.gov/research/umls/new\\_users/online\\_learning/](http://www.nlm.nih.gov/research/umls/new_users/online_learning/). [Accessed: 20-Oct-2012].
- [55] M. Steinbach, "A Comparison of Document Clustering Techniques," in *KDD Workshop on Text Mining*, 2000, pp. 1–20.
- [56] L. Jing, M. K. Ng, and J. Z. Huang, "Knowledge-based vector space model for text clustering," *Knowledge and Information Systems*, vol. 25, no. 1, pp. 35–55, Oct. 2009.
- [57] A. Hotho, S. Staab, and G. Stumme, "Wordnet improves Text Document Clustering," in *Proc. of the SIGIR 2003 Semantic Web Workshop*, 2003, pp. 541–544.
- [58] L. Jing, L. Zhou, M. K. Ng, and J. Z. Huang, "Ontology-based distance measure for text clustering," in *Proc. of SIAM SDM Workshop on Text Mining*, 2006.

- [59] I. Yoo and X. Hu, "A comprehensive comparison study of document clustering for a biomedical digital library MEDLINE," in *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries - JCDL '06*, 2006, p. 220.
- [60] S. Zhu, J. Zeng, and H. Mamitsuka, "Enhancing MEDLINE document clustering by incorporating MeSH semantic similarity.," *Bioinformatics (Oxford, England)*, vol. 25, no. 15, pp. 1944–51, Aug. 2009.
- [61] B. Aljaber, N. Stokes, J. Bailey, and J. Pei, "Document clustering of scientific texts using citation contexts," *Information Retrieval*, vol. 13, no. 2, pp. 101–131, Aug. 2009.
- [62] X. Zhang, L. Jing, X. Hu, M. Ng, and X. Zhou, "A Comparative Study of Ontology Based Term Similarity Measures on PubMed Document Clustering," *Advances in Databases: Concepts, Systems and Applications*, pp. 115–126, 2007.
- [63] Z. Pawlak, "Rough Sets," Springer Netherlands, Warsaw, 1991.
- [64] V. V. Raghavan and H. Sever, "The state of rough sets for database mining applications," in *San Jose State University*, 1995, pp. 1–11.
- [65] P. Srinivasan, "Text mining: Generating hypotheses from MEDLINE," *Journal of the American Society for Information Science and Technology*, vol. 55, no. 5, pp. 396–413, Mar. 2004.
- [66] C. C. van der Eijk, E. M. van Mulligen, J. A. Kors, B. Mons, and J. van den Berg, "Constructing an associative concept space for literature-based discovery," *Journal of the American Society for Information Science and Technology*, vol. 55, no. 5, pp. 436–444, Mar. 2004.
- [67] X. Hu, G. Li, I. Yoo, and X. Zhang, "A semantic-based approach for mining undiscovered public knowledge from biomedical literature," *Granular Computing*, pp. 22–27 Vol. 1, 2005.



- [68] I. Yoo, X. Hu, and I.-Y. Song, *Integration of Semantic-Based Bipartite Graph Representation and Mutual Refinement Strategy for Biomedical Literature Clustering*. New York, USA: ACM Press, 2006, p. 791.
- [69] G. Hirst and D. St Onge, "Lexical Chains as Representation of Context for the Detection and Correction of Malapropisms," *WordNet*, 1995.
- [70] Z. Wu, "Verbs semantics and lexical selection," *32nd. Annual Meeting of the Association for Computational Linguistics.*, 1994.
- [71] Z. A. Bandar and D. McLean, "An approach for measuring semantic similarity between words using multiple information sources," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 4, pp. 871–882, Jul. 2003.
- [72] W. Mao and W. W. Chu, "Free-text medical document retrieval via phrase-based vector space model," in *AMIA Annual Symposium*, 2002, pp. 489–93.
- [73] V. Dobrynin, D. Patterson, M. Galushka, and N. Rooney, "SOPHIA: An Interactive Cluster-Based Retrieval System for the OHSUMED Collection," *IEEE Transactions on Information Technology in Biomedicine*, vol. 9, no. 2, pp. 256–265, Jun. 2005.
- [74] H. Chim, "Effective document clustering approach based on phrase feature extraction," City University of Hong Kong, 2008.
- [75] W. Hersh, C. Buckley, T. Leone, and D. Hickam, "OHSUMED: an interactive retrieval evaluation and new large test collection for research," *Conference on Research ...*, 1994.
- [76] W. Hersh and D. Hickam, "Use of a multi-application computer workstation in a clinical setting.," *Bulletin of the Medical Library Association*, vol. 82, no. 4, pp. 382–9, Oct. 1994.

- [77] R. N. Kostoff, *Where is the Discovery in Literature-Based Discovery?*, vol. 15. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 57–72.
- [78] E. M. Voorhees and D. K. Harman, Eds., *TREC - Experiment and Evaluation in Information Retrieval*. Cambridge: The MIT Press, 2006.

## ÖZGEÇMİŞ

Adı Soyadı : Fatih Mehmet GÜLEÇ

Doğum Yeri : ANKARA

Doğum Yılı : 1982

Medeni Hali : Evli

Eğitim ve Akademik Durumu:

Lise 1993-2000 Ankara Gazi Anadolu Lisesi

Lisans 2000-2004 Hacettepe Üniversitesi Bilgisayar Mühendisliği  
Bölümü, ANKARA

Y.Lisans 2004-2007 Hacettepe Üniversitesi Bilgisayar Mühendisliği  
Bölümü, ANKARA

Yabancı Dil : İngilizce

İş Tecrübesi :

09.2004-12.2012 : Hacettepe Üniversitesi, Bilgisayar Mühendisliği Bölümü,  
Araştırma Görevlisi

01.2013-... : Algan Yazılım ve Bilişim – Genel Müdür

### Ek.1. Raynaud Hastalığı ve Altın Standartların Dağılımı

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
2576	109	109	0	76	76	0	0	0
29449	269	59	210	445	138	396	210	148
9740	46	46	0	30	30	0	0	0
30883	189	42	147	320	85	281	147	109
15234	29	29	0	50	50	0	0	0
13309	24	24	0	55	55	0	0	0
21275	267	23	244	369	56	345	244	142
11402	37	23	14	40	15	29	14	7
49977	8407	19	8388	5939	54	5934	4615	1063
161	1093	19	1074	1495	51	1484	814	437
1672	19	19	0	22	22	0	0	
49795	1330	18	1312	2200	64	2188	1307	509
29067	286	17	269	423	51	402	209	149
26790	306	14	292	874	54	852	292	240
23960	121	13	108	347	87	313	107	117
45020	88	13	75	281	68	248	74	102
41	7534	12	7522	6049	42	6047	4128	1038
49953	7306	11	7295	6331	63	6326	5014	1165
11027	922	11	911	2261	37	2252	911	512

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
49645	772	11	761	1438	42	1430	578	430
172	3265	10	3255	4881	62	4872	3175	1017
49902	1222	10	1212	2310	39	2301	1212	544
11981	422	10	412	736	57	711	316	243
44772	171	10	161	399	17	391	161	121
11236	102	10	92	315	103	265	92	44
49836	647	9	638	1143	40	1133	463	277
15814	306	9	297	416	32	403	284	149
45763	213	9	204	624	34	613	204	166
49851	3858	8	3850	4027	26	4026	3828	824
72	3447	8	3439	3889	14	3888	3218	778
31	1173	8	1165	2066	31	2054	692	462
11	590	8	582	737	17	731	582	220
35145	191	8	183	385	34	366	120	109
33670	106	8	98	249	26	240	98	72
49999	3562	7	3555	4925	38	4915	3379	983
13	2360	7	2353	3523	38	3520	2353	810
60	2298	7	2291	3352	46	3348	2284	777
217	2192	7	2185	3360	35	3351	1493	766
49868	1257	7	1250	1921	50	1906	1247	414
221	1040	7	1033	1566	33	1561	1029	448
11145	944	7	937	1165	27	1153	651	340
45152	785	7	778	1702	33	1689	777	400

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
49677	727	7	720	1281	31	1275	684	389
49706	689	7	682	1383	36	1375	518	412
49776	663	7	656	1497	58	1481	559	427
12949	333	7	326	736	23	728	315	195
41263	290	7	283	668	25	654	283	172
45919	140	7	133	790	108	746	112	117
49980	4790	6	4784	4449	16	4445	2381	823
49909	3730	6	3724	4444	34	4441	2961	917
80	3413	6	3407	5008	30	5006	3355	993
66	1717	6	1711	3114	25	3108	1711	717
49669	1477	6	1471	2213	29	2207	1059	595
49724	1397	6	1391	2548	21	2542	904	534
8151	926	6	920	696	23	685	248	95
8635	476	6	470	653	27	646	284	188
8633	406	6	400	914	62	906	365	308
26604	352	6	346	857	27	850	344	258
20547	349	6	343	733	23	727	252	231
10821	305	6	299	708	33	696	272	207
10173	299	6	293	504	39	488	293	162
41072	290	6	284	412	23	405	217	135
28600	287	6	281	570	43	558	281	175
42630	283	6	277	741	36	726	277	224
32306	209	6	203	431	27	427	203	133

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
42064	193	6	187	509	52	492	187	170
41978	95	6	89	341	29	334	89	99
39576	92	6	86	223	26	217	81	79
35809	52	6	46	180	38	167	46	72
37643	46	6	40	285	34	268	40	75
53	6345	5	6340	4686	18	4678	4125	819
79	2841	5	2836	4056	35	4054	2730	961
49898	1871	5	1866	3116	21	3115	1353	719
49979	1796	5	1791	3339	35	3331	1730	829
1	1186	5	1181	2289	27	2282	1181	584
702	1045	5	1040	1593	23	1589	849	454
49481	869	5	864	2531	34	2522	862	562
49738	678	5	673	1349	42	1334	646	374
49745	674	5	669	1039	27	1035	570	340
8896	508	5	503	1608	41	1598	503	434
9867	496	5	491	1037	35	1029	491	301
19481	305	5	300	684	29	678	300	214
44237	288	5	283	890	36	878	227	243
44298	149	5	144	417	36	407	144	150
34881	123	5	118	456	24	450	115	164
7508	75	5	70	136	27	114	70	30
18998	65	5	60	115	19	106	59	49
30191	63	5	58	161	24	148	58	58

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
47082	61	5	56	251	27	243	47	58
13200	47	5	42	49	8	46	42	13
49936	8595	4	8591	6798	35	6797	6745	1208
49885	3140	4	3136	4190	21	4186	3136	906
1332	2078	4	2074	3732	21	3729	1744	831
49733	2074	4	2070	3501	23	3498	2070	730
9537	2025	4	2021	1477	20	1468	386	162
49734	1411	4	1407	3029	34	3026	1407	694
49679	1395	4	1391	3043	21	3040	1391	669
17573	1380	4	1376	1610	10	1608	1372	157
49834	1280	4	1276	1738	18	1731	819	439
50	1159	4	1155	2398	29	2394	1035	559
844	1086	4	1082	1900	18	1893	892	398
4156	1066	4	1062	685	20	678	223	84
3456	1021	4	1017	2744	17	2738	1017	587
49434	793	4	789	1578	36	1566	624	412
49845	628	4	624	1179	22	1173	624	282
49893	615	4	611	1533	28	1527	550	413
39	583	4	579	1220	27	1215	579	310
5360	469	4	465	361	24	349	106	53
8560	456	4	452	1024	26	1015	448	268
15160	421	4	417	597	11	592	231	149
12602	415	4	411	800	22	796	321	233



Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
9522	406	4	402	953	27	951	395	295
3336	347	4	343	265	20	250	337	37
29416	335	4	331	814	25	809	324	265
44628	244	4	240	747	19	739	192	223
35926	202	4	198	472	23	460	198	110
38116	202	4	198	908	26	897	198	232
38660	196	4	192	504	24	498	192	178
23122	193	4	189	754	54	740	106	101
19327	185	4	181	299	37	272	157	75
28633	178	4	174	461	33	449	174	171
39478	164	4	160	546	28	534	160	136
41984	138	4	134	391	24	383	131	118
26273	134	4	130	509	34	496	116	164
7197	125	4	121	121	11	116	119	13
26630	121	4	117	406	19	398	98	132
34486	105	4	101	275	24	265	101	106
23836	104	4	100	271	39	247	100	90
30644	92	4	88	272	35	261	88	94
22751	91	4	87	234	18	228	87	77
23354	78	4	74	234	30	224	74	102
18414	69	4	65	210	35	198	65	82
49961	7866	3	7863	6485	13	6484	5743	1122
186	5683	3	5680	5383	16	5375	1656	352

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
49930	5517	3	5514	4404	6	4403	2575	787
7643	4878	3	4875	1912	12	1905	447	180
9033	4034	3	4031	3536	12	3528	4029	233
49942	3654	3	3651	5094	18	5092	3566	912
22	2936	3	2933	4554	26	4551	2915	945
67	2867	3	2864	3011	36	2999	1357	641
87	2816	3	2813	4932	33	4929	2347	1000
11160	2463	3	2460	1892	21	1887	2038	192
49827	1940	3	1937	3861	27	3858	1928	911
10600	1852	3	1849	1297	14	1291	290	160
49674	1784	3	1781	2156	21	2145	404	180
2379	1598	3	1595	2497	25	2491	1585	584
6072	1543	3	1540	3276	21	3274	1540	708
49621	1460	3	1457	2818	30	2812	1454	674
49703	1383	3	1380	1746	16	1743	769	487
49804	1319	3	1316	2519	19	2516	961	603
49825	1280	3	1277	3054	31	3048	1277	709
49911	1239	3	1236	2660	22	2656	1236	655
6690	1203	3	1200	664	13	660	1199	71
148	1072	3	1069	2203	34	2197	973	581
205	1047	3	1044	2441	31	2437	1044	682
5179	973	3	970	1442	19	1438	952	363
9164	949	3	946	1874	16	1872	648	473

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
6689	896	3	893	1964	31	1955	756	535
41410	856	3	853	2071	23	2070	853	529
49562	844	3	841	1522	14	1520	556	361
44162	833	3	830	1887	24	1879	811	505
49530	816	3	813	1160	9	1158	568	324
5272	775	3	772	451	13	444	97	54
49438	770	3	767	1159	23	1152	667	362
12465	728	3	725	798	12	790	118	62
49862	632	3	629	1567	34	1560	467	363
49958	604	3	601	1789	26	1786	601	443
481	549	3	546	1132	21	1127	451	319
178	539	3	536	1540	22	1536	536	334
6385	512	3	509	1304	24	1301	437	358
6294	476	3	473	785	14	782	473	181
9774	403	3	400	756	33	748	372	236
12884	379	3	376	873	26	867	329	285
19469	338	3	335	706	18	701	324	209
15633	324	3	321	1334	37	1323	321	350
20815	322	3	319	919	23	910	250	233
13925	303	3	300	556	19	553	297	185
21879	235	3	232	576	30	567	232	180
30336	227	3	224	356	11	353	136	102
38041	227	3	224	692	21	690	224	205

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
31592	221	3	218	371	20	365	218	158
18694	214	3	211	636	36	623	207	206
35241	197	3	194	462	31	456	193	175
27726	178	3	175	234	9	230	175	61
21104	173	3	170	315	28	303	168	59
10723	159	3	156	267	16	261	156	111
36410	159	3	156	443	15	440	121	113
7831	150	3	147	114	5	111	13	9
41609	138	3	135	351	16	347	89	84
23890	136	3	133	189	16	186	123	80
42981	136	3	133	275	16	270	114	101
11336	132	3	129	169	9	167	129	66
20800	129	3	126	346	23	336	122	118
35372	129	3	126	347	19	339	126	105
26063	124	3	121	200	16	195	111	67
42327	124	3	121	303	24	291	121	89
45367	120	3	117	439	22	432	117	128
21799	118	3	115	199	11	195	115	63
47098	108	3	105	333	22	325	95	92
32065	105	3	102	418	24	411	102	147
26652	100	3	97	120	14	116	94	51
48201	97	3	94	278	29	262	94	106
11807	81	3	78	139	5	136	17	15

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
21845	78	3	75	184	28	175	75	68
28070	76	3	73	193	26	183	71	78
40353	69	3	66	225	26	217	66	76
36021	65	3	62	216	19	211	53	74
27841	64	3	61	183	18	175	61	55
38908	58	3	55	143	16	136	55	58
34116	49	3	46	102	14	96	46	43
17484	42	3	39	170	19	164	39	44
11616	34	3	31	68	22	54	29	10
1158	5904	2	5902	2420	3	2419	1106	424
83	5667	2	5665	5137	16	5134	4216	944
49985	5376	2	5374	5793	9	5792	3730	1041
33	4666	2	4664	5167	20	5163	2320	894
82	4480	2	4478	5208	30	5204	4478	964
111	4479	2	4477	5010	26	5007	4400	959
4327	3916	2	3914	2157	5	2155	442	226
13056	3283	2	3281	3412	10	3409	634	254
30	2923	2	2921	3761	13	3759	1774	789
49769	2669	2	2667	4103	20	4101	2203	951
49847	2591	2	2589	4751	23	4745	1893	873
6441	2262	2	2260	1502	19	1492	2248	179
179	2152	2	2150	4121	29	4116	2134	845
6118	2125	2	2123	1026	5	1023	248	111

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
49864	1910	2	1908	2836	12	2833	1280	676
13883	1878	2	1876	1819	8	1814	338	167
49946	1832	2	1830	3184	28	3182	1245	710
49983	1790	2	1788	2834	27	2833	1352	619
49997	1781	2	1779	3658	18	3655	1779	783
21	1764	2	1762	3606	20	3602	1727	810
22472	1728	2	1726	3055	20	3043	562	321
86	1701	2	1699	3127	15	3124	1699	789
49705	1379	2	1377	1693	15	1690	1373	473
49730	1363	2	1361	1980	16	1974	700	450
49782	1317	2	1315	2541	16	2538	1313	554
49875	1255	2	1253	3021	35	3017	1253	640
49901	1246	2	1244	1926	16	1924	1243	526
49932	1224	2	1222	1436	20	1427	694	400
49954	1215	2	1213	1767	25	1758	1098	432
49965	1202	2	1200	2672	11	2669	1039	624
49968	1200	2	1198	2189	10	2186	768	461
37	1168	2	1166	3017	17	3015	1165	718
88	1138	2	1136	1487	9	1486	1128	385
122	1115	2	1113	3443	21	3436	758	661
140	1080	2	1078	2312	23	2304	1078	590
2385	1079	2	1077	2483	10	2482	916	521
1658	988	2	986	1629	20	1627	755	474

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
8072	967	2	965	2513	23	2507	964	605
15629	962	2	960	1664	8	1661	622	352
8055	935	2	933	1331	8	1327	933	273
21650	919	2	917	2103	11	2102	894	507
48646	846	2	844	1980	13	1977	844	458
47350	837	2	835	1391	20	1387	822	383
3441	826	2	824	424	40	401	804	33
49587	807	2	805	2031	32	2020	804	541
44629	803	2	801	2147	7	2143	308	232
49655	767	2	765	1574	20	1571	573	416
49634	738	2	736	1435	32	1428	736	427
3384	715	2	713	312	8	308	708	28
49748	681	2	679	1303	11	1297	381	284
49701	679	2	677	1580	25	1572	677	391
49762	679	2	677	1714	9	1713	677	393
7872	678	2	676	547	40	514	658	60
8450	650	2	648	549	6	547	98	47
49858	632	2	630	1569	12	1568	527	406
49878	629	2	627	1305	27	1298	528	387
49943	605	2	603	1765	8	1763	561	486
49974	600	2	598	1143	11	1140	380	283
49750	594	2	592	538	4	536	93	57
27	588	2	586	1616	19	1611	388	322

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
154	555	2	553	1489	22	1483	547	426
130	554	2	552	643	9	637	316	160
796	546	2	544	1501	18	1497	544	359
163	540	2	538	502	10	494	236	97
489	526	2	524	947	17	942	365	255
7178	524	2	522	1011	10	1008	417	250
17100	511	2	509	1437	14	1430	409	291
17784	495	2	493	1697	19	1690	493	378
7278	493	2	491	1563	20	1559	412	364
5806	488	2	486	542	17	535	486	145
16053	467	2	465	962	10	958	465	214
6545	464	2	462	827	14	824	343	238
15475	435	2	433	726	12	719	433	86
12304	427	2	425	700	18	692	389	205
16669	416	2	414	1027	19	1021	414	264
46659	412	2	410	2645	44	2633	295	321
5660	396	2	394	339	7	336	49	28
17093	388	2	386	840	16	837	373	223
14890	386	2	384	1000	18	991	265	210
20689	385	2	383	864	32	852	383	251
30331	385	2	383	1303	8	1300	359	355
11679	383	2	381	1378	19	1372	381	320
11864	375	2	373	640	28	634	371	228



Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
16278	374	2	372	729	8	728	359	224
14694	360	2	358	1042	17	1038	354	224
24718	358	2	356	933	15	927	294	225
12940	351	2	349	450	7	447	349	97
14603	348	2	346	617	9	612	70	66
22870	348	2	346	833	16	831	327	283
12701	346	2	344	755	17	752	344	207
18028	342	2	340	1070	15	1065	340	258
14533	335	2	333	680	17	673	292	205
5104	330	2	328	300	11	295	27	26
24082	326	2	324	674	17	665	321	187
20519	322	2	320	660	14	656	319	189
19435	321	2	319	563	12	560	261	170
15735	318	2	316	1057	28	1052	301	332
22633	317	2	315	742	22	737	315	204
13078	314	2	312	376	13	372	312	139
28811	313	2	311	841	23	836	311	238
10827	308	2	306	512	36	505	305	205
17355	306	2	304	931	20	926	292	307
29942	304	2	302	962	20	958	282	288
35184	302	2	300	802	14	797	293	210
35446	302	2	300	701	21	697	299	220
11713	298	2	296	662	28	654	221	187

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
34837	295	2	293	607	16	595	105	50
44717	295	2	293	806	16	802	228	215
5564	291	2	289	287	9	281	156	40
16453	289	2	287	859	16	855	287	224
44727	279	2	277	893	14	890	240	231
36420	272	2	270	635	16	632	227	205
43982	272	2	270	587	14	579	156	140
5916	269	2	267	354	66	305	250	31
19123	269	2	267	416	23	411	267	162
39392	266	2	264	517	15	512	198	149
23593	257	2	255	775	18	772	232	250
42871	255	2	253	841	20	839	252	250
47945	254	2	252	634	10	628	252	152
8862	250	2	248	324	22	311	248	46
40558	247	2	245	502	13	499	200	174
31011	245	2	243	639	17	634	241	209
46370	238	2	236	681	12	679	235	219
28819	235	2	233	773	21	768	232	222
42526	231	2	229	672	13	669	227	213
39782	229	2	227	537	14	534	208	172
40683	224	2	222	913	20	906	220	249
27490	221	2	219	771	17	767	173	215
6615	219	2	217	317	30	309	217	52

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
28698	207	2	205	345	9	341	176	64
32599	205	2	203	424	20	421	196	151
36277	189	2	187	500	29	488	187	152
28835	188	2	186	483	21	476	145	169
32620	187	2	185	419	10	417	185	148
19016	186	2	184	305	22	301	181	104
27534	186	2	184	541	25	531	182	174
44873	178	2	176	495	14	490	172	143
22224	176	2	174	512	26	503	158	173
4817	157	2	155	169	35	149	152	21
31833	157	2	155	287	21	279	96	75
47919	156	2	154	653	32	641	154	151
3663	154	2	152	128	4	126	9	6
38461	154	2	152	265	10	262	150	62
42047	153	2	151	457	23	454	151	149
46558	147	2	145	553	38	541	142	175
28253	146	2	144	667	18	661	144	190
33540	144	2	142	203	6	202	82	64
37569	142	2	140	435	13	433	126	152
38288	142	2	140	364	11	358	140	120
47364	142	2	140	600	15	591	102	138
27328	141	2	139	331	19	325	103	81
42454	141	2	139	374	26	368	137	132

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
28477	139	2	137	471	28	463	136	159
24627	136	2	134	291	18	280	119	106
47549	130	2	128	325	8	321	128	100
37813	128	2	126	472	23	467	126	146
28781	124	2	122	733	32	715	122	95
17824	122	2	120	152	4	151	120	57
29899	121	2	119	324	19	323	119	121
41124	121	2	119	449	13	447	102	122
20868	120	2	118	362	22	358	118	99
27471	118	2	116	270	21	257	113	96
31209	115	2	113	241	16	235	113	101
23970	114	2	112	295	12	293	112	102
33320	112	2	110	213	21	204	102	75
34185	112	2	110	327	24	315	110	123
28959	110	2	108	388	20	382	92	127
32487	108	2	106	298	34	286	106	107
25993	107	2	105	299	15	293	105	101
30252	106	2	104	282	7	281	85	75
26079	103	2	101	252	13	249	98	97
30588	103	2	101	336	13	334	98	95
27103	102	2	100	196	12	190	63	49
21255	101	2	99	237	21	230	99	91
22657	93	2	91	273	33	257	91	100

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
47968	93	2	91	284	18	276	86	112
22863	88	2	86	153	15	141	86	63
40849	88	2	86	181	9	177	58	50
29126	87	2	85	210	14	206	47	45
34269	81	2	79	200	20	191	78	66
25943	80	2	78	138	6	135	36	38
23032	78	2	76	258	17	257	76	89
26383	78	2	76	153	14	148	58	54
45098	78	2	76	310	10	306	56	79
31911	77	2	75	271	18	266	75	72
28275	75	2	73	256	34	239	73	83
39385	75	2	73	190	12	186	73	64
48164	75	2	73	229	15	222	72	83
20822	74	2	72	212	25	205	72	66
21745	72	2	70	238	26	233	70	87
28622	72	2	70	225	17	218	70	85
45604	72	2	70	303	28	293	67	99
28818	67	2	65	178	17	171	65	68
35417	67	2	65	199	15	192	65	83
14554	65	2	63	91	10	88	63	26
28560	65	2	63	198	16	187	19	13
10340	64	2	62	117	10	115	62	44
35244	64	2	62	239	18	231	62	70

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
23232	63	2	61	194	14	188	61	82
26730	63	2	61	161	15	155	59	58
41587	63	2	61	192	12	187	61	57
41914	63	2	61	210	11	206	61	62
29695	61	2	59	236	12	233	59	73
49173	61	2	59	261	14	256	49	63
22643	60	2	58	88	7	83	56	25
21996	58	2	56	127	12	124	54	50
39335	57	2	55	143	13	138	37	41
25053	56	2	54	120	6	118	54	12
37876	56	2	54	246	20	244	51	50
23849	53	2	51	116	23	108	48	56
25580	50	2	48	160	18	150	48	53
25550	49	2	47	132	19	121	34	46
16717	47	2	45	78	4	77	45	16
36582	46	2	44	211	15	203	43	61
38407	46	2	44	151	11	144	44	66
38467	43	2	41	173	18	164	41	38
13486	40	2	38	75	26	55	16	6
11054	39	2	37	42	5	41	37	11
43045	33	2	31	169	29	152	31	61
4204	30	2	28	48	12	42	28	5
30544	24	2	22	99	14	93	21	32

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
49978	11892	1	11891	5963	6	5961	2676	478
6375	6085	1	6084	1569	5	1567	265	115
49952	5354	1	5353	5969	6	5965	1151	428
49892	5009	1	5008	4944	8	4940	2575	872
7244	4566	1	4565	3304	12	3301	469	213
49929	4287	1	4286	4383	10	4382	4034	820
17769	4244	1	4243	3245	7	3243	525	257
11349	3920	1	3919	3255	4	3254	507	259
8836	3822	1	3821	1883	4	1881	3382	169
16941	3700	1	3699	3703	4	3701	653	285
14904	3602	1	3601	2454	8	2451	321	151
46	3470	1	3469	5085	20	5082	3469	919
7688	3395	1	3394	2470	4	2468	431	185
11243	3390	1	3389	3660	3	3658	471	262
49993	2976	1	2975	3870	20	3858	2323	827
99	2856	1	2855	4241	10	4239	2855	799
11634	2835	1	2834	2827	6	2825	465	216
5990	2761	1	2760	1514	6	1512	365	143
49779	2671	1	2670	3898	10	3895	1588	796
5757	2514	1	2513	1153	7	1149	186	111
16731	2500	1	2499	2813	10	2808	1486	217
32708	2466	1	2465	3758	20	3756	483	267
49935	2449	1	2448	4245	7	4244	1721	806

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
12490	2447	1	2446	2110	6	2106	2334	166
2	2372	1	2371	3382	13	3380	1725	801
5177	2286	1	2285	1504	4	1502	2279	131
10063	2256	1	2255	3515	10	3512	586	305
21018	2251	1	2250	3519	10	3515	1190	419
112	2207	1	2206	4176	14	4174	2206	881
139	2205	1	2204	3750	17	3747	2204	828
7661	2170	1	2169	1496	18	1484	2168	136
49765	2041	1	2040	3711	13	3710	2033	779
49736	2010	1	2009	3225	22	3221	2006	684
49723	1998	1	1997	3609	19	3608	1979	785
49784	1977	1	1976	2670	20	2663	1934	617
18543	1935	1	1934	3353	12	3348	1900	323
49850	1915	1	1914	3432	18	3431	1660	826
49879	1900	1	1899	3593	7	3592	1185	776
49882	1885	1	1884	3577	12	3576	1883	801
49873	1884	1	1883	2463	4	2462	1771	437
9719	1871	1	1870	1221	10	1215	328	143
49918	1851	1	1850	3455	11	3454	1846	743
14056	1844	1	1843	2008	4	2006	185	109
49949	1823	1	1822	1975	14	1974	856	441
2953	1621	1	1620	3253	12	3251	1614	681
151	1607	1	1606	3157	20	3153	1475	730



Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
2629	1583	1	1582	2445	16	2439	1538	557
10800	1516	1	1515	1690	4	1688	239	155
22377	1504	1	1503	2715	4	2713	434	287
11659	1500	1	1499	1681	4	1679	232	124
11448	1454	1	1453	1668	5	1665	1439	128
2360	1451	1	1450	443	8	437	78	42
11755	1446	1	1445	1635	8	1633	269	142
49632	1442	1	1441	2853	12	2851	1092	704
49474	1421	1	1420	2873	12	2871	1410	612
49680	1404	1	1403	2417	12	2414	1019	612
49719	1367	1	1366	2983	19	2982	1363	679
11368	1366	1	1365	1372	4	1370	207	103
49591	1311	1	1310	4322	14	4318	744	422
49813	1295	1	1294	2378	8	2377	842	524
49905	1244	1	1243	1953	9	1949	801	424
49915	1229	1	1228	2353	4	2352	827	527
12930	1223	1	1222	1379	8	1373	198	113
49951	1212	1	1211	2734	16	2733	1211	630
9135	1198	1	1197	1068	28	1053	205	102
49696	1194	1	1193	1498	23	1483	274	139
6540	1157	1	1156	948	6	946	84	51
8609	1147	1	1146	985	4	983	1136	66
114	1127	1	1126	3317	16	3314	904	574

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
76	1126	1	1125	2446	22	2442	1125	500
106	1114	1	1113	2764	13	2762	1113	676
127	1109	1	1108	1126	7	1124	1031	298
6450	1101	1	1100	913	5	911	798	83
117	1093	1	1092	2487	19	2483	929	616
13795	1062	1	1061	1291	6	1289	241	112
168	1050	1	1049	2861	15	2858	1045	694
4829	1042	1	1041	2702	10	2698	1040	672
3086	1033	1	1032	2421	17	2418	1023	601
235	1026	1	1025	2204	16	2202	752	551
10562	1008	1	1007	1928	18	1926	1006	404
44656	1004	1	1003	2128	9	2122	351	243
3089	983	1	982	1340	9	1336	519	334
9872	976	1	975	2235	16	2233	974	572
10069	973	1	972	1219	12	1211	173	108
2296	966	1	965	1438	13	1436	964	369
8278	961	1	960	832	4	830	102	50
13716	951	1	950	1790	9	1788	711	461
34600	949	1	948	2435	7	2433	948	493
8377	937	1	936	634	4	632	935	84
15244	931	1	930	2210	11	2209	624	431
7201	907	1	906	1160	10	1157	528	296
12536	905	1	904	1924	22	1914	904	384

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
11058	902	1	901	1899	12	1897	884	463
17106	902	1	901	1893	6	1892	901	453
27361	897	1	896	2682	7	2681	896	581
6115	896	1	895	553	7	549	881	54
47469	889	1	888	2017	14	2015	696	473
49351	886	1	885	1558	13	1553	588	359
49470	884	1	883	1576	11	1574	725	404
39741	875	1	874	1693	15	1690	848	438
48955	866	1	865	1236	10	1233	865	343
49602	865	1	864	2532	20	2530	859	570
47138	864	1	863	1135	11	1133	725	312
49546	854	1	853	1614	9	1613	574	386
49584	845	1	844	2668	10	2665	693	480
49475	839	1	838	1440	9	1438	487	319
49334	834	1	833	1206	6	1205	476	268
49312	822	1	821	877	7	875	417	246
49127	809	1	808	1191	14	1186	471	279
49610	795	1	794	1369	15	1365	770	310
49615	794	1	793	1496	7	1495	514	365
48471	787	1	786	1891	8	1889	757	533
49385	783	1	782	1437	11	1435	675	438
6330	781	1	780	644	22	630	165	62
49154	780	1	779	1848	17	1845	642	436

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
49598	776	1	775	1881	11	1879	576	468
48266	770	1	769	1073	16	1066	499	311
49311	759	1	758	1959	24	1949	737	473
48748	749	1	748	1896	18	1893	609	407
42866	740	1	739	1366	4	1364	462	221
49224	730	1	729	2440	21	2433	729	481
49648	730	1	729	1610	16	1609	729	357
49631	716	1	715	903	9	900	714	263
49689	714	1	713	1958	9	1957	593	456
49735	710	1	709	2302	10	2300	576	523
49704	709	1	708	1391	11	1390	541	379
49693	708	1	707	1164	10	1161	707	340
49685	704	1	703	923	9	920	524	207
7163	699	1	698	633	7	629	132	55
49726	688	1	687	1568	6	1566	473	398
9776	682	1	681	877	16	862	681	76
9107	675	1	674	715	6	713	196	86
49743	673	1	672	1726	12	1720	668	469
49740	671	1	670	1242	24	1233	520	282
6342	666	1	665	498	4	496	663	47
49772	663	1	662	1082	9	1078	465	298
49789	653	1	652	602	6	598	597	112
49833	647	1	646	1748	10	1746	646	436

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
49821	643	1	642	1091	12	1086	602	256
49838	643	1	642	1694	14	1692	642	451
49831	640	1	639	1186	6	1184	525	288
49843	639	1	638	2037	11	2033	633	483
20610	638	1	637	1317	6	1313	194	126
49844	629	1	628	1555	5	1554	527	414
49802	627	1	626	1902	10	1898	626	474
16508	624	1	623	931	6	927	98	66
49883	623	1	622	1180	6	1178	622	217
33880	620	1	619	743	14	734	619	119
49900	619	1	618	1451	20	1449	618	346
49887	618	1	617	1384	9	1382	570	340
49904	613	1	612	918	14	916	383	250
49920	613	1	612	1012	5	1009	430	218
49934	613	1	612	1324	13	1322	385	352
49926	611	1	610	1196	8	1194	355	276
49944	611	1	610	912	12	909	399	264
49947	608	1	607	821	6	820	344	209
49963	603	1	602	1811	20	1806	600	478
49964	601	1	600	766	7	764	600	236
7857	583	1	582	729	16	719	83	67
55	582	1	581	1460	10	1458	580	361
6717	575	1	574	395	4	393	574	42

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
69	570	1	569	870	7	869	364	185
84	570	1	569	1364	16	1361	397	347
141	564	1	563	1671	17	1666	484	292
149	559	1	558	1601	14	1598	558	376
120	556	1	555	955	10	950	419	282
153	555	1	554	767	9	761	554	181
34893	555	1	554	2068	8	2066	438	259
119	552	1	551	690	7	689	547	213
197	551	1	550	1265	5	1264	542	292
185	549	1	548	1763	9	1759	548	305
36400	547	1	546	2753	12	2751	377	280
6184	544	1	543	764	8	761	208	130
192	543	1	542	856	5	855	541	197
640	543	1	542	1384	13	1380	412	320
171	542	1	541	1358	10	1357	453	366
1106	538	1	537	823	5	822	537	168
485	537	1	536	965	7	963	432	281
4825	532	1	531	367	14	357	48	36
760	523	1	522	938	7	937	489	232
6837	521	1	520	853	4	851	89	68
224	517	1	516	1675	31	1662	426	357
1925	517	1	516	942	9	938	305	245
2112	515	1	514	1005	5	1004	356	252

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
5995	513	1	512	726	6	724	90	62
6797	512	1	511	255	4	253	59	24
2781	510	1	509	836	7	834	509	246
4347	506	1	505	1116	20	1109	406	275
10478	505	1	504	1231	15	1225	503	349
8837	500	1	499	557	8	551	56	47
5640	496	1	495	1079	13	1075	391	308
5958	492	1	491	446	4	444	56	43
11139	487	1	486	1397	13	1393	374	340
5328	486	1	485	969	21	963	410	259
11714	486	1	485	907	7	903	389	276
10380	482	1	481	602	4	600	74	56
7815	481	1	480	1729	17	1727	476	358
42977	481	1	480	2383	20	2381	385	309
3424	478	1	477	1126	16	1121	477	348
1701	476	1	475	1435	19	1427	475	282
8499	474	1	473	791	11	789	473	193
11599	473	1	472	836	11	834	472	215
3097	467	1	466	332	6	329	45	30
13267	465	1	464	685	8	683	464	204
19193	462	1	461	1252	9	1249	326	268
12750	461	1	460	1307	12	1305	360	297
11597	457	1	456	959	9	956	447	308

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
21765	450	1	449	1014	6	1011	430	260
9491	449	1	448	589	8	588	200	149
19883	449	1	448	1443	11	1440	371	270
19559	448	1	447	1173	24	1170	447	181
18055	444	1	443	740	4	738	107	81
9688	437	1	436	795	12	793	302	156
7933	436	1	435	908	16	901	370	303
5469	434	1	433	314	4	312	432	26
9928	434	1	433	1485	17	1479	433	379
4989	433	1	432	334	4	332	414	36
13470	432	1	431	879	5	878	427	218
9286	430	1	429	1476	14	1472	320	352
11414	429	1	428	666	6	663	319	195
11372	428	1	427	837	10	836	419	242
25839	428	1	427	1013	18	1005	371	290
5772	426	1	425	326	6	322	62	44
24318	426	1	425	882	6	881	340	228
3695	425	1	424	184	6	180	49	21
17973	423	1	422	1128	8	1127	355	284
30293	421	1	420	1169	6	1167	416	320
7032	420	1	419	1427	14	1424	419	420
19026	415	1	414	821	9	820	290	163
20037	413	1	412	692	9	689	349	208



Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
12018	411	1	410	1224	21	1219	307	296
12985	408	1	407	849	13	846	269	209
9458	406	1	405	830	22	827	403	237
44855	406	1	405	2125	30	2113	285	263
11119	405	1	404	928	27	921	404	293
11197	405	1	404	676	6	672	264	65
20084	405	1	404	919	13	915	278	252
20353	404	1	403	899	8	896	328	244
18134	402	1	401	816	18	810	300	249
11989	401	1	400	1160	11	1158	295	310
31567	401	1	400	985	21	982	399	263
2190	393	1	392	150	4	148	16	5
6612	393	1	392	436	5	435	192	123
10866	392	1	391	476	9	472	391	144
22846	392	1	391	1085	12	1079	169	121
14832	391	1	390	456	3	455	337	140
18510	390	1	389	561	10	555	264	168
10700	384	1	383	519	5	518	174	131
11307	381	1	380	807	10	805	380	215
16964	377	1	376	700	12	699	264	135
13289	376	1	375	768	16	765	363	211
6530	375	1	374	547	21	541	374	173
18880	375	1	374	514	14	508	248	137

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
6649	372	1	371	834	21	828	288	201
28764	372	1	371	788	9	786	371	189
16346	369	1	368	641	7	640	279	173
13636	367	1	366	728	15	725	365	195
12916	362	1	361	847	12	843	361	227
22349	362	1	361	836	14	830	322	191
30639	361	1	360	905	19	901	360	138
48076	360	1	359	1222	9	1220	288	211
20487	359	1	358	989	16	988	358	274
15828	357	1	356	755	6	753	216	189
22098	356	1	355	907	17	898	355	215
38552	355	1	354	1471	7	1469	354	341
29478	353	1	352	881	11	879	351	219
15142	350	1	349	1341	16	1334	349	361
39005	350	1	349	1391	30	1375	250	153
13663	347	1	346	904	16	900	261	206
10686	345	1	344	611	15	608	344	167
13755	345	1	344	442	12	438	334	82
16770	345	1	344	737	17	735	288	210
6136	344	1	343	425	4	423	89	58
25792	344	1	343	824	10	819	255	181
37702	340	1	339	918	9	916	305	256
15765	338	1	337	511	9	506	216	136

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
22682	332	1	331	675	13	671	241	200
11412	331	1	330	1251	20	1243	330	340
12659	331	1	330	982	12	980	330	236
22498	330	1	329	1113	10	1106	224	221
16982	325	1	324	1050	13	1046	324	254
7399	324	1	323	873	11	872	306	227
29599	321	1	320	631	10	629	320	161
18167	320	1	319	1266	11	1263	319	315
23690	320	1	319	1241	14	1236	276	298
24248	320	1	319	703	22	698	252	217
30498	320	1	319	1002	13	997	257	256
11217	319	1	318	660	12	657	260	208
14081	317	1	316	1144	13	1141	314	294
21118	316	1	315	1027	27	1012	233	225
12680	315	1	314	1024	23	1018	314	292
15007	314	1	313	686	17	685	231	211
19151	314	1	313	874	25	869	284	240
31968	313	1	312	765	11	759	275	202
23790	311	1	310	1032	12	1028	310	278
27017	311	1	310	644	8	640	263	195
44072	311	1	310	926	11	918	310	178
9731	308	1	307	433	13	428	204	127
17031	308	1	307	969	15	966	307	239

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
22115	307	1	306	672	11	669	266	195
28993	307	1	306	782	8	780	243	205
12132	306	1	305	512	9	508	165	139
16341	306	1	305	682	14	678	247	171
17958	303	1	302	813	12	812	302	169
22414	303	1	302	527	11	525	238	162
33365	302	1	301	607	12	601	301	47
9799	300	1	299	786	14	784	275	227
22106	298	1	297	629	7	624	149	106
20861	294	1	293	376	11	372	159	124
33961	293	1	292	550	16	545	292	157
17056	291	1	290	1018	10	1016	285	278
24298	288	1	287	446	14	445	227	138
19086	286	1	285	396	14	390	280	117
21643	286	1	285	553	11	549	239	156
36049	286	1	285	588	14	582	284	166
30242	284	1	283	895	13	890	212	229
34426	284	1	283	639	6	638	170	169
33451	281	1	280	1036	11	1033	280	283
38957	280	1	279	568	8	564	265	177
34444	279	1	278	660	4	658	71	61
45931	276	1	275	686	11	684	272	196
41759	275	1	274	665	10	664	265	217

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
42408	275	1	274	603	11	600	274	195
47243	275	1	274	707	13	704	274	219
21895	271	1	270	509	15	508	256	173
25755	270	1	269	421	5	420	169	115
16245	269	1	268	797	40	779	268	77
22000	269	1	268	331	19	325	254	91
43795	269	1	268	631	7	630	268	224
44090	269	1	268	544	6	540	257	131
21867	268	1	267	476	4	474	41	23
38894	267	1	266	1232	22	1219	219	290
7873	266	1	265	278	6	275	32	22
29389	266	1	265	448	21	438	265	109
35220	266	1	265	942	10	940	265	248
2940	265	1	264	123	4	121	9	8
30004	265	1	264	608	15	604	264	155
30036	265	1	264	340	6	339	264	86
30313	265	1	264	517	22	505	264	176
18947	264	1	263	259	9	258	256	50
43015	264	1	263	1017	16	1006	198	222
22139	262	1	261	402	9	400	261	114
44164	261	1	260	836	16	829	260	240
44487	261	1	260	653	12	652	238	142
14741	260	1	259	1037	14	1036	259	289

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
21240	260	1	259	508	13	506	258	171
31261	259	1	258	538	12	535	238	132
21877	257	1	256	737	18	734	255	185
14087	256	1	255	913	28	902	251	238
19376	256	1	255	954	22	949	229	239
45057	256	1	255	782	14	779	224	187
36788	255	1	254	670	22	658	238	150
24759	254	1	253	822	18	821	247	214
38626	254	1	253	538	8	536	212	133
23848	253	1	252	379	19	373	167	98
44692	253	1	252	728	9	726	252	236
43479	252	1	251	400	6	397	251	100
23156	251	1	250	673	16	667	172	172
29267	250	1	249	524	13	522	249	158
28656	249	1	248	353	4	351	220	95
11231	248	1	247	298	4	296	193	26
25675	248	1	247	657	9	656	223	221
30295	248	1	247	581	24	572	184	157
45606	247	1	246	832	16	830	246	270
36715	246	1	245	716	16	709	172	171
45877	246	1	245	937	15	933	245	268
49241	246	1	245	887	16	883	241	250
1767	244	1	243	149	6	145	14	8

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
21886	243	1	242	760	18	753	242	232
29710	243	1	242	498	14	495	230	148
47489	242	1	241	797	12	793	238	212
20727	241	1	240	718	13	715	239	181
16549	239	1	238	909	16	906	208	217
23883	238	1	237	536	10	533	237	163
35940	236	1	235	340	8	338	235	73
35993	236	1	235	530	15	526	231	105
36961	236	1	235	686	16	685	235	211
17225	232	1	231	405	16	401	231	118
26578	232	1	231	528	16	526	218	165
43157	232	1	231	586	9	582	174	139
39913	231	1	230	360	5	359	187	109
40856	230	1	229	512	14	510	229	177
16663	229	1	228	538	23	528	186	137
22720	229	1	228	413	22	402	184	143
26625	229	1	228	256	7	252	228	90
31442	228	1	227	704	14	703	209	216
43549	228	1	227	423	4	422	107	82
48142	227	1	226	1607	40	1585	188	187
27348	225	1	224	494	17	492	187	145
39230	225	1	224	305	4	304	135	77
45644	225	1	224	570	6	569	192	158

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
21298	223	1	222	547	11	546	191	187
22977	222	1	221	346	10	343	221	98
46215	222	1	221	750	8	745	193	197
42390	221	1	220	580	10	577	220	219
44234	220	1	219	1091	18	1089	218	140
36741	218	1	217	696	14	693	216	218
44127	218	1	217	1283	12	1281	167	163
44518	218	1	217	732	8	729	216	143
48019	217	1	216	725	11	724	158	154
32432	216	1	215	477	11	473	198	169
28762	215	1	214	428	15	425	164	130
31280	215	1	214	892	11	889	214	239
35223	215	1	214	521	13	514	214	134
39746	215	1	214	459	11	454	174	100
41268	215	1	214	675	9	673	214	203
44272	215	1	214	772	24	763	212	146
45313	215	1	214	655	8	653	214	172
46430	215	1	214	569	5	566	214	129
23572	214	1	213	340	8	338	211	81
31606	214	1	213	477	17	475	213	155
28695	213	1	212	846	6	844	117	119
49265	212	1	211	782	18	776	201	191
27094	211	1	210	346	6	345	210	133



Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
34121	210	1	209	624	16	619	209	189
47913	210	1	209	655	5	653	207	143
48059	210	1	209	878	7	876	209	221
22030	208	1	207	458	16	455	207	142
23167	208	1	207	926	18	918	207	218
39373	208	1	207	434	7	433	112	85
17102	207	1	206	587	11	584	205	120
30353	206	1	205	438	17	431	126	115
23107	205	1	204	660	18	654	173	160
19905	204	1	203	382	9	381	173	127
25429	204	1	203	401	15	394	192	112
32482	204	1	203	475	9	474	203	140
39080	204	1	203	623	14	618	203	174
1868	203	1	202	87	14	77	202	14
27323	203	1	202	324	7	323	97	74
30622	203	1	202	554	11	553	191	173
46607	203	1	202	894	17	889	141	167
3197	201	1	200	132	4	130	21	11
16908	200	1	199	706	17	702	193	221
40119	200	1	199	569	11	566	158	200
42271	200	1	199	429	8	425	198	97
34684	199	1	198	556	16	555	191	171
23596	198	1	197	804	12	801	197	196

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
30713	198	1	197	300	9	298	174	76
16166	197	1	196	266	11	260	196	78
45251	195	1	194	547	16	542	194	139
28400	194	1	193	543	13	539	173	142
29023	194	1	193	507	10	505	192	156
34319	194	1	193	438	5	437	138	153
18038	192	1	191	242	3	241	190	47
32188	192	1	191	752	16	749	190	178
39953	192	1	191	442	9	438	117	95
20461	191	1	190	202	5	199	190	67
31818	191	1	190	638	12	633	190	172
15204	190	1	189	231	4	229	64	36
47405	190	1	189	715	21	712	183	192
28522	189	1	188	839	17	832	188	234
30815	188	1	187	459	14	456	168	148
23423	187	1	186	267	7	264	186	58
29654	187	1	186	227	8	222	161	73
46463	187	1	186	695	10	691	185	199
18696	186	1	185	675	14	672	183	188
26010	186	1	185	410	11	408	181	104
44601	185	1	184	575	15	569	137	118
33529	184	1	183	314	16	304	183	78
37759	184	1	183	327	6	323	69	58

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
20849	183	1	182	582	14	577	180	186
32646	183	1	182	462	22	460	175	141
34638	183	1	182	653	11	651	182	215
40028	183	1	182	1280	22	1270	133	121
26165	181	1	180	698	20	692	157	173
32904	181	1	180	323	17	310	177	67
18668	180	1	179	430	16	426	143	124
31248	180	1	179	504	14	501	179	172
32266	180	1	179	398	13	395	144	145
22364	179	1	178	393	20	384	102	95
25239	177	1	176	518	11	515	141	163
27905	177	1	176	328	11	325	176	78
3425	176	1	175	134	4	132	173	24
39784	175	1	174	452	10	448	174	155
23197	174	1	173	432	15	427	153	136
27287	174	1	173	803	31	779	87	100
39863	174	1	173	558	11	555	138	160
41535	174	1	173	521	17	511	173	136
23486	173	1	172	365	13	360	124	91
43577	173	1	172	535	7	534	165	165
25796	172	1	171	259	8	255	171	75
28782	172	1	171	508	12	504	154	140
38553	172	1	171	481	12	480	171	169

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
24935	171	1	170	409	11	406	153	111
29136	170	1	169	628	12	625	154	143
30405	170	1	169	433	12	430	136	148
16169	169	1	168	511	22	503	82	83
44752	168	1	167	548	14	544	166	187
36390	167	1	166	444	15	439	166	146
21945	166	1	165	415	14	411	115	90
28795	165	1	164	380	8	377	84	75
16378	164	1	163	716	12	714	163	192
33020	164	1	163	291	12	288	163	95
48251	164	1	163	638	13	634	163	159
7396	163	1	162	171	6	167	18	20
30338	162	1	161	434	10	432	134	126
30612	161	1	160	411	14	407	142	138
28268	160	1	159	423	11	421	159	97
29392	160	1	159	609	23	600	159	174
45316	160	1	159	461	9	459	57	56
24608	158	1	157	290	5	289	111	80
28362	158	1	157	448	15	442	144	131
46154	158	1	157	536	18	535	154	139
49128	158	1	157	456	12	452	117	97
28559	157	1	156	340	13	334	156	84
36604	157	1	156	396	14	389	139	134

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
41947	156	1	155	356	18	346	154	100
9062	155	1	154	252	16	239	29	18
29439	155	1	154	429	12	428	129	156
32246	155	1	154	523	14	517	150	134
37470	154	1	153	590	20	580	113	108
5672	153	1	152	121	4	119	17	7
22420	153	1	152	256	11	252	149	81
26870	153	1	152	330	8	329	134	94
39822	153	1	152	378	8	376	105	98
18839	152	1	151	454	15	453	151	131
20190	152	1	151	477	13	474	98	110
22326	151	1	150	554	15	549	98	140
24833	151	1	150	410	13	405	150	129
28462	151	1	150	401	8	398	150	88
31391	151	1	150	785	13	780	150	181
31914	151	1	150	624	16	619	147	152
36156	151	1	150	375	9	373	150	119
39553	151	1	150	349	13	345	144	135
42599	151	1	150	713	10	709	150	187
44002	151	1	150	519	13	516	150	159
17889	150	1	149	536	21	527	149	166
25259	149	1	148	262	8	261	122	66
42305	149	1	148	481	21	468	127	104

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
26102	148	1	147	863	26	854	126	200
22225	147	1	146	401	16	396	114	127
27079	147	1	146	438	11	435	146	126
38842	147	1	146	888	22	881	121	187
20703	146	1	145	181	4	179	66	50
21424	145	1	144	532	12	529	104	138
23192	145	1	144	381	15	376	144	90
28994	145	1	144	865	16	859	142	180
36722	145	1	144	339	16	333	144	101
44213	145	1	144	817	26	804	121	177
33671	144	1	143	442	26	422	132	121
9495	142	1	141	278	27	267	141	41
27731	142	1	141	324	14	320	141	103
29951	142	1	141	362	11	361	136	140
19886	141	1	140	267	13	263	130	81
33839	141	1	140	625	10	619	108	92
37461	141	1	140	459	14	454	130	125
35403	140	1	139	423	8	422	139	119
36991	140	1	139	242	14	237	139	55
9593	139	1	138	205	4	203	4	4
37021	139	1	138	251	17	244	83	79
38855	139	1	138	488	10	484	106	107
48616	138	1	137	631	22	625	122	130

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
49277	138	1	137	763	12	757	102	93
24656	137	1	136	162	4	161	45	33
35449	137	1	136	456	24	455	136	126
45375	137	1	136	444	18	430	136	56
21666	134	1	133	301	12	299	107	107
28357	133	1	132	412	12	411	131	143
28441	133	1	132	482	11	480	130	178
26290	131	1	130	380	19	368	130	101
37906	131	1	130	385	14	379	130	111
39119	131	1	130	346	13	342	127	82
40770	130	1	129	556	16	549	129	143
49030	130	1	129	587	21	582	112	127
24368	129	1	128	108	5	106	128	34
28923	129	1	128	269	9	268	128	80
30686	129	1	128	420	12	418	102	147
32582	129	1	128	391	14	387	128	125
49437	129	1	128	518	16	510	120	129
23043	128	1	127	390	7	389	105	102
43910	128	1	127	425	6	423	103	114
44556	128	1	127	368	14	363	127	111
15106	127	1	126	319	24	311	126	52
22437	127	1	126	234	12	229	126	79
28750	127	1	126	311	11	306	126	110

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
30739	127	1	126	269	24	260	126	101
39692	127	1	126	446	20	440	122	110
30897	125	1	124	402	13	398	115	109
34437	125	1	124	407	14	405	116	144
34627	125	1	124	303	16	299	119	81
46165	125	1	124	355	4	353	62	66
46405	125	1	124	519	19	509	123	145
18884	124	1	123	210	9	207	61	53
30154	124	1	123	422	12	418	123	126
19846	123	1	122	484	17	479	117	135
21204	123	1	122	164	6	161	121	55
30352	123	1	122	181	12	175	120	54
31075	123	1	122	302	16	298	122	89
40000	123	1	122	352	12	349	122	94
18754	122	1	121	294	15	292	117	89
41178	122	1	121	219	5	218	71	55
42073	122	1	121	322	9	320	103	85
24974	121	1	120	140	6	137	59	32
28172	121	1	120	340	5	339	120	103
33663	121	1	120	327	9	324	120	63
37686	121	1	120	427	15	425	120	133
45435	121	1	120	396	15	392	120	125
8704	120	1	119	310	19	303	35	46



Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
22948	120	1	119	318	8	316	54	24
38568	120	1	119	251	8	247	89	81
40241	119	1	118	431	10	426	105	107
49428	119	1	118	390	15	384	101	74
19903	118	1	117	468	23	455	100	112
40493	118	1	117	304	9	302	114	109
48575	118	1	117	418	7	417	117	103
12873	117	1	116	243	17	241	116	77
31832	117	1	116	404	11	402	116	116
38624	117	1	116	430	15	425	102	100
40622	117	1	116	323	6	322	112	116
19840	116	1	115	200	10	195	115	51
27783	116	1	115	242	16	239	115	69
39229	116	1	115	384	16	376	115	101
42407	116	1	115	360	18	352	112	133
43135	116	1	115	292	5	289	50	48
23763	115	1	114	328	16	320	114	84
23923	115	1	114	381	14	379	89	89
41510	115	1	114	313	9	311	99	112
44231	115	1	114	340	14	338	103	108
49251	115	1	114	406	19	394	114	125
27651	114	1	113	283	15	277	109	83
45635	114	1	113	413	13	408	78	104

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
18928	113	1	112	144	8	141	49	42
32753	113	1	112	526	17	518	112	124
41197	112	1	111	325	10	319	85	94
48797	112	1	111	356	17	346	79	84
836	111	1	110	38	6	35	1	1
18921	111	1	110	401	17	396	82	125
36717	111	1	110	342	17	338	110	96
19160	110	1	109	362	15	358	106	120
37714	110	1	109	331	7	328	106	105
46537	110	1	109	856	24	840	92	102
29726	109	1	108	521	17	515	108	125
29921	109	1	108	301	9	300	85	96
41243	109	1	108	395	14	391	108	123
43987	109	1	108	373	13	366	105	108
44287	109	1	108	327	10	320	108	80
48105	109	1	108	405	8	403	102	116
24523	108	1	107	164	13	161	106	45
38620	108	1	107	328	12	321	105	89
42186	108	1	107	298	18	291	107	84
46859	108	1	107	358	14	354	107	101
23079	107	1	106	220	16	215	76	65
26912	107	1	106	393	10	392	90	138
31892	107	1	106	286	19	283	92	99

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
35820	107	1	106	346	20	337	106	92
40834	107	1	106	455	14	448	82	112
44739	107	1	106	521	16	517	106	81
2123	106	1	105	78	6	76	105	11
11822	106	1	105	196	22	191	47	33
12758	106	1	105	146	4	144	91	28
37055	106	1	105	442	6	440	101	86
49011	106	1	105	513	16	508	82	107
19473	105	1	104	162	5	160	104	59
23787	105	1	104	202	15	192	58	61
26435	105	1	104	299	13	297	95	91
29931	105	1	104	337	7	335	104	89
31872	105	1	104	318	11	313	74	89
33813	105	1	104	318	12	313	102	95
45279	105	1	104	488	19	480	104	158
31497	104	1	103	589	11	585	103	132
38278	104	1	103	288	16	284	98	101
45834	104	1	103	343	19	335	100	99
19391	103	1	102	117	8	116	63	44
26587	103	1	102	442	24	432	101	129
28330	103	1	102	189	8	185	102	67
34700	103	1	102	303	15	298	98	91
47530	103	1	102	263	12	261	98	84

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
21066	101	1	100	334	14	326	87	41
22901	101	1	100	231	11	228	100	73
25963	101	1	100	281	9	278	65	85
42682	101	1	100	253	9	252	97	84
44398	101	1	100	378	18	370	95	127
45717	101	1	100	388	12	381	89	84
27967	100	1	99	182	6	180	76	54
36698	100	1	99	346	11	344	98	105
39424	100	1	99	318	13	315	97	111
17349	99	1	98	146	14	136	98	48
21589	99	1	98	131	10	125	77	44
34081	99	1	98	259	8	258	93	93
35287	99	1	98	327	17	321	91	95
9821	98	1	97	267	18	262	94	53
16439	98	1	97	279	9	278	77	92
33860	98	1	97	226	9	224	94	72
34868	98	1	97	338	13	333	96	98
31029	97	1	96	188	8	184	41	34
34410	97	1	96	252	12	248	96	86
43225	97	1	96	274	9	273	96	96
46195	97	1	96	453	18	448	96	94
26403	96	1	95	565	21	557	95	105
34841	96	1	95	249	10	244	41	53

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
34888	96	1	95	419	13	417	89	125
48948	96	1	95	305	25	292	86	88
44600	95	1	94	421	26	412	94	77
45304	95	1	94	332	9	327	91	79
27370	94	1	93	269	18	265	85	102
37105	94	1	93	306	17	294	93	80
1630	93	1	92	34	7	29	91	4
30892	93	1	92	334	21	328	92	126
31973	93	1	92	193	18	186	32	39
33401	93	1	92	237	8	235	92	68
37305	93	1	92	207	13	204	89	60
43685	93	1	92	281	10	276	85	105
15967	92	1	91	150	7	149	91	54
21340	92	1	91	251	9	249	90	86
24279	92	1	91	179	8	178	80	60
36337	92	1	91	245	10	242	91	73
43967	92	1	91	537	20	527	91	91
13733	91	1	90	178	13	176	78	57
42918	91	1	90	194	7	192	60	44
47100	91	1	90	328	16	317	80	95
20400	90	1	89	221	14	218	82	89
20826	90	1	89	314	14	308	89	63
22565	90	1	89	159	10	155	89	48

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
39044	90	1	89	226	8	224	89	95
46254	90	1	89	396	12	391	89	71
32948	89	1	88	215	12	210	88	61
37820	89	1	88	259	11	252	57	69
42759	89	1	88	335	15	334	88	139
49483	89	1	88	524	24	515	77	116
5251	88	1	87	103	8	98	22	11
20696	88	1	87	286	15	281	45	68
20821	88	1	87	206	10	203	87	83
26733	88	1	87	201	9	198	87	72
30059	88	1	87	269	9	268	86	94
40918	88	1	87	265	6	263	67	71
43759	88	1	87	357	8	356	76	103
19605	87	1	86	295	16	290	81	93
19805	87	1	86	164	11	159	86	61
20755	87	1	86	127	6	125	59	34
32478	87	1	86	278	12	273	86	62
45187	87	1	86	202	7	199	35	33
27816	86	1	85	229	10	226	85	56
30874	86	1	85	303	12	299	85	120
32012	86	1	85	787	30	771	71	109
46367	86	1	85	338	14	334	85	58
25407	85	1	84	283	5	282	84	61

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
25969	85	1	84	165	6	163	69	49
32926	85	1	84	333	12	329	84	75
48218	84	1	83	453	18	444	83	135
19837	82	1	81	207	12	203	59	57
39992	82	1	81	216	16	210	55	59
42810	82	1	81	268	9	265	81	101
46470	82	1	81	348	10	345	80	89
9958	81	1	80	303	27	287	75	40
48512	81	1	80	370	13	366	80	142
48669	81	1	80	315	10	313	60	59
4699	80	1	79	60	8	56	79	7
16697	80	1	79	217	11	213	79	83
34019	80	1	79	171	14	164	51	47
36930	80	1	79	232	8	230	67	81
38594	80	1	79	299	5	297	76	107
26530	79	1	78	239	15	237	78	75
28670	79	1	78	137	11	134	78	46
36491	79	1	78	146	9	144	78	45
48178	79	1	78	328	10	326	78	94
10954	78	1	77	141	5	140	38	42
23609	78	1	77	146	12	141	69	27
24872	78	1	77	245	12	239	77	72
33827	78	1	77	234	10	231	77	73

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
31464	77	1	76	193	7	192	76	67
36053	77	1	76	160	10	158	75	52
45091	77	1	76	431	29	420	75	140
25260	76	1	75	285	8	282	70	104
29616	76	1	75	249	16	243	66	67
43653	76	1	75	269	26	257	75	109
13300	75	1	74	103	11	98	74	34
15649	75	1	74	299	23	290	74	86
16960	75	1	74	226	11	224	74	61
33498	75	1	74	283	11	281	61	84
38289	75	1	74	262	15	258	74	68
18773	74	1	73	274	15	265	72	64
24323	74	1	73	206	12	201	64	63
28815	74	1	73	178	7	176	50	52
37629	74	1	73	302	10	299	73	85
38255	74	1	73	241	17	234	73	88
41018	74	1	73	237	9	235	73	85
41330	74	1	73	255	6	253	51	28
6554	73	1	72	214	18	202	69	29
10437	73	1	72	221	8	217	72	34
18308	73	1	72	161	12	156	72	57
20189	73	1	72	230	14	225	52	69
20408	73	1	72	180	11	173	72	39



Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
30028	73	1	72	248	9	245	72	61
31585	73	1	72	307	16	301	72	105
34578	73	1	72	326	16	319	54	75
34964	73	1	72	141	5	140	44	36
37078	73	1	72	191	9	187	59	33
37332	73	1	72	393	10	388	72	101
38200	73	1	72	444	40	422	72	60
24787	72	1	71	300	13	299	66	113
38151	72	1	71	222	9	215	47	55
24275	71	1	70	265	14	261	70	92
24765	71	1	70	177	12	175	64	86
44662	71	1	70	249	6	248	65	90
44573	70	1	69	280	14	276	67	87
47594	70	1	69	290	12	285	65	88
15708	69	1	68	247	20	235	36	44
36029	69	1	68	182	8	178	35	32
36967	69	1	68	231	15	221	51	53
38196	69	1	68	460	26	446	50	68
17321	68	1	67	208	18	204	65	66
18788	68	1	67	83	3	82	55	33
20286	68	1	67	107	11	102	67	37
35323	68	1	67	183	26	161	67	32
45496	68	1	67	275	15	266	67	35

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
12103	67	1	66	194	17	190	51	48
21016	67	1	66	242	16	237	66	68
27327	67	1	66	145	10	142	66	45
7285	66	1	65	49	10	46	65	19
10197	66	1	65	172	14	163	21	19
17138	66	1	65	262	13	261	65	69
32998	66	1	65	168	6	166	39	46
9397	65	1	64	109	6	105	64	22
19422	65	1	64	149	6	145	28	27
31346	65	1	64	199	13	193	53	66
35853	65	1	64	155	10	152	64	60
35956	65	1	64	212	8	208	55	68
47985	65	1	64	200	10	196	59	68
48722	65	1	64	261	5	260	57	87
17899	64	1	63	105	14	96	63	51
18856	64	1	63	144	14	138	55	57
19718	64	1	63	92	10	90	63	27
21294	64	1	63	160	10	158	32	42
40861	64	1	63	264	10	257	53	77
44641	64	1	63	220	13	211	57	67
20574	63	1	62	126	5	125	33	29
43344	63	1	62	232	6	229	50	51
46001	63	1	62	228	12	222	55	81

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
11381	62	1	61	255	36	233	61	44
47078	62	1	61	239	15	230	61	84
17407	61	1	60	55	5	52	50	26
27907	61	1	60	176	8	175	60	54
28192	61	1	60	195	15	188	60	55
29941	61	1	60	284	12	281	60	96
30359	61	1	60	211	8	208	34	57
31776	61	1	60	166	7	164	60	36
37295	61	1	60	239	13	234	60	79
39038	61	1	60	176	8	174	60	58
41017	61	1	60	202	10	198	60	78
13072	60	1	59	124	7	123	59	35
26393	60	1	59	140	12	136	45	42
35932	60	1	59	238	8	236	47	75
15029	59	1	58	55	6	53	58	5
26666	59	1	58	270	12	266	58	69
47524	59	1	58	436	12	430	50	56
31265	58	1	57	174	13	170	57	71
14214	57	1	56	200	10	196	14	14
30901	57	1	56	208	10	202	47	64
7825	56	1	55	79	5	77	54	11
22844	56	1	55	255	19	243	55	48
25546	56	1	55	203	31	185	19	22

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
28259	56	1	55	198	9	194	36	31
36863	56	1	55	256	11	252	43	66
47738	56	1	55	251	12	246	55	86
48050	56	1	55	229	12	222	47	53
20207	55	1	54	125	11	122	54	51
26031	55	1	54	270	14	266	54	52
30065	55	1	54	114	14	111	54	52
39947	55	1	54	222	17	209	43	49
41588	55	1	54	207	17	201	54	62
42638	55	1	54	208	9	207	53	75
44331	55	1	54	263	10	258	49	77
6552	54	1	53	115	6	113	53	20
12652	54	1	53	107	6	105	53	18
26987	54	1	53	127	10	122	44	36
27600	54	1	53	114	6	111	51	28
22471	53	1	52	136	8	133	52	43
24657	53	1	52	69	10	66	49	23
34755	53	1	52	174	8	172	45	56
15985	52	1	51	109	9	106	39	37
27922	52	1	51	122	6	121	42	26
33868	52	1	51	293	16	291	51	49
46506	52	1	51	260	16	254	48	62
18171	51	1	50	245	12	238	28	40

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
18209	51	1	50	128	14	121	50	48
19767	51	1	50	92	5	91	21	24
30439	51	1	50	171	15	159	50	50
33019	50	1	49	210	9	206	48	87
34370	50	1	49	97	3	96	46	30
19745	49	1	48	101	4	100	48	43
23950	49	1	48	190	11	185	48	76
38440	49	1	48	168	14	160	48	63
20309	48	1	47	71	5	70	41	24
22113	48	1	47	125	10	124	47	44
26888	48	1	47	142	8	140	47	54
33430	48	1	47	137	14	129	30	35
38442	48	1	47	184	8	181	36	38
44202	48	1	47	163	9	161	42	39
1882	47	1	46	52	24	32	46	4
14682	47	1	46	68	5	67	38	17
26396	47	1	46	194	20	189	40	63
30247	47	1	46	138	7	134	46	62
35371	47	1	46	154	9	149	35	39
17965	46	1	45	141	7	140	37	54
30898	46	1	45	122	7	120	45	40
15714	45	1	44	99	6	98	44	47
23549	45	1	44	187	20	177	44	80

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
27221	45	1	44	103	9	99	40	36
18948	44	1	43	151	24	133	40	18
20327	44	1	43	188	12	183	43	52
25388	44	1	43	140	12	138	43	21
30983	44	1	43	132	9	130	43	43
31037	44	1	43	127	16	115	17	10
34136	44	1	43	211	12	206	30	52
35661	44	1	43	150	13	142	43	71
42793	44	1	43	136	8	132	41	33
20076	43	1	42	110	18	97	42	16
23785	43	1	42	132	12	129	32	48
24786	43	1	42	133	15	124	42	33
35328	43	1	42	123	9	122	39	39
47645	43	1	42	179	18	168	40	45
23410	42	1	41	119	12	113	32	25
26437	42	1	41	117	6	116	41	29
34518	42	1	41	132	7	130	34	34
15978	41	1	40	128	4	126	39	32
12042	40	1	39	42	6	41	7	7
20791	40	1	39	185	17	180	31	56
29382	40	1	39	115	10	113	39	41
32276	40	1	39	128	6	127	29	23
48402	40	1	39	227	21	214	36	59

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
17223	38	1	37	131	30	105	14	15
28329	38	1	37	130	18	116	13	16
18662	37	1	36	62	6	60	32	17
27461	37	1	36	216	16	204	20	31
29976	37	1	36	132	17	122	36	51
42889	37	1	36	210	6	208	36	34
11309	36	1	35	44	6	41	35	17
14752	36	1	35	48	6	46	29	20
19588	36	1	35	108	13	103	28	24
42374	36	1	35	162	24	149	35	59
45392	35	1	34	337	14	329	24	38
5824	34	1	33	43	6	41	22	5
39298	33	1	32	68	8	64	24	22
10803	30	1	29	96	20	80	7	11
15043	30	1	29	54	5	53	18	16
23460	30	1	29	168	14	158	29	13
8728	29	1	28	31	6	29	27	6
16920	29	1	28	138	20	122	28	29
20775	29	1	28	142	26	130	28	20
11505	28	1	27	54	7	53	12	12
17662	28	1	27	105	16	93	8	7
7913	27	1	26	70	16	58	5	6
13877	27	1	26	78	6	76	9	9

Grup No	Toplam Doküman Sayısı	A Teriminin Geçtiği Doküman Sayısı	Geçmediği Doküman Sayısı	Toplam Terim Sayısı	A Teriminin Geçtiği Dokümanlarda Toplam Terim Sayısı	Geçmediği Dokümanlarda Toplam Terim Sayısı	Altın Terimin Geçtiği Makale Sayısı	Altın terim Sayısı
10920	26	1	25	73	10	68	25	23
12507	26	1	25	68	7	65	25	26
27180	26	1	25	106	7	104	14	15
45852	23	1	22	110	10	102	6	9
27641	22	1	21	61	6	59	21	24
9908	19	1	18	48	8	46	18	21



## Ek.2. Raynaud Hastalığı ve İşletim Sonuçları

	C terimi	Doğru - Olumlu	İlk Bulunma Tarihi	Ortak Küme Sayısı	Toplam Küme Yakınlığı	Ağırlıklı Toplam Küme Yakınlığı	C'nin Bağlantılık Değeri
1	L-LACTATE DEHYDROGENASE			421	2.373	4.031	0.010
2	TUBERCULOSIS, PULMONARY			418	2.179	7.617	0.018
3	BRAIN INJURIES			401	2.063	4.917	0.012
4	LEUKEMIA, MYELOID	*	1997	386	1.961	4.529	0.012
5	PARAPLEGIA	*	1995	369	2.226	3.952	0.011
6	VIRUS DISEASES			350	1.793	3.394	0.010
7	RUPTURE, SPONTANEOUS			349	2.097	5.193	0.015
8	ANTIBODIES, VIRAL	*	1999	347	1.745	5.513	0.016
9	IODINE ISOTOPES			340	1.971	3.258	0.010
10	ANEMIA, APLASTIC			337	1.619	3.622	0.011
11	HEART ARREST			327	1.479	3.098	0.009
12	BRONCHIAL NEOPLASMS			324	1.575	3.486	0.011
13	STREPTOCOCCAL INFECTIONS			324	1.506	3.036	0.009
14	PHENYTOIN			323	1.489	2.807	0.009
15	CARDIAC SURGICAL PROCEDURES			322	1.639	3.966	0.012
16	NEOPLASM RECURRENCE, LOCAL	*	1987	321	1.486	5.224	0.016
17	GRAVES DISEASE	*	1997	319	1.674	3.616	0.011
18	ACCIDENTS, TRAFFIC			317	1.613	5.659	0.018
19	LEUKEMIA, MYELOID, ACUTE	*	1996	316	1.542	3.818	0.012
20	BLOOD CHEMICAL ANALYSIS	*	2003	315	1.629	2.874	0.009
21	GASTRIC JUICE			313	1.515	3.055	0.010

	<b>C terimi</b>	<b>Dođru - Olumlu</b>	<b>İlk Bulunma Tarihi</b>	<b>Ortak K¼me Sayısı</b>	<b>Toplam K¼me Yakınlığı</b>	<b>Ađırlıklı Toplam K¼me Yakınlığı</b>	<b>C'nin Bađlantılık Deđeri</b>
22	ANESTHESIA, INHALATION	*	1994	312	1.727	3.275	0.010
23	RESPIRATION, ARTIFICIAL	*	2009	311	1.420	3.260	0.010
24	EUROPEAN CONTINENTAL ANCESTRY GROUP	*	1993	308	1.583	2.937	0.010
25	RADIOTHERAPY DOSAGE	*	2009	307	1.397	3.950	0.013
26	INTRACRANIAL ANEURYSM	*	1993	304	1.647	5.127	0.017
27	TRANQUILIZING AGENTS			298	1.410	2.600	0.009
28	COSTS AND COST ANALYSIS			296	1.335	2.500	0.008
29	AORTA, THORACIC	*	1990	292	1.754	4.413	0.015
30	MYCOSES	*	2007	292	1.340	2.422	0.008
31	AEROSOLS	*	1990	286	1.358	3.788	0.013
32	VENA CAVA, INFERIOR			283	1.524	4.133	0.015
33	NUTRITION DISORDERS			283	1.169	1.824	0.006
34	BODY HEIGHT			282	1.230	2.595	0.009
35	ATTITUDE TO HEALTH	*	1995	281	1.164	2.830	0.010
36	STOMACH DISEASES	*	2010	280	1.083	2.689	0.010
37	MITRAL VALVE INSUFFICIENCY			278	1.313	3.183	0.011
38	DIETARY PROTEINS			277	1.003	1.955	0.007
39	NASAL MUCOSA			276	1.429	2.572	0.009
40	PERIODICITY	*	1993	276	1.491	2.214	0.008
41	ABDOMINAL INJURIES			275	1.246	3.036	0.011
42	OXYGEN INHALATION THERAPY	*	2007	275	1.357	2.247	0.008
43	PNEUMONECTOMY			274	1.097	2.507	0.009
44	ANTITUBERCULAR AGENTS			274	0.958	2.277	0.008
45	ANTIBODIES, MONOCLONAL	*	1999	271	1.468	4.030	0.015

	<b>C terimi</b>	<b>Dođru - Olumlu</b>	<b>İlk Bulunma Tarihi</b>	<b>Ortak Kme Sayısı</b>	<b>Toplam Kme Yakınlığı</b>	<b>Ađırlıklı Toplam Kme Yakınlığı</b>	<b>C'nin Bađlantılık Deđeri</b>
46	AUDIOMETRY	*	2009	270	1.382	4.014	0.015
47	WOUNDS, GUNSHOT			270	1.375	2.810	0.010
48	PATIENT CARE PLANNING	*	1994	270	1.010	1.880	0.007
49	ADAPTATION, PSYCHOLOGICAL	*	1995	267	1.039	2.652	0.010
50	INTUBATION, INTRATRACHEAL	*	2009	267	1.182	2.407	0.009
51	DRUG RESISTANCE, MICROBIAL			265	0.976	2.528	0.010
52	PITUITARY NEOPLASMS			264	1.070	2.368	0.009
53	ANTIBODIES, BACTERIAL	*	1990	263	1.264	3.130	0.012
54	MOUTH NEOPLASMS	*	2002	263	1.044	2.654	0.010
55	BARIUM SULFATE			263	1.306	2.397	0.009
56	ABDOMEN, ACUTE			262	1.304	2.938	0.011
57	ANTHROPOMETRY			262	1.367	2.794	0.011
58	NUTRITIONAL PHYSIOLOGICAL PHENOMENA			262	0.938	1.927	0.007
59	CARCINOMA, HEPATOCELLULAR	*	2006	260	1.206	2.872	0.011
60	17-HYDROXYCORTICOSTEROIDS			260	1.369	2.357	0.009
61	EYELIDS			258	1.022	2.719	0.011
62	AGRANULOCYTOSIS	*	1991	258	1.098	2.410	0.009
63	KIDNEY TUBULES			258	1.148	1.925	0.007
64	URINARY BLADDER NEOPLASMS	*	2003	257	1.187	3.722	0.014
65	RETINAL DISEASES	*	1987	257	1.485	2.865	0.011
66	SUPPURATION			257	1.413	2.506	0.010
67	AMYLASES			257	0.906	1.596	0.006
68	PRECANCEROUS CONDITIONS	*	2007	256	1.192	2.534	0.010
69	RADIATION EFFECTS			256	1.252	2.006	0.008

	<b>C terimi</b>	<b>Doğru - Olumlu</b>	<b>İlk Bulunma Tarihi</b>	<b>Ortak Küme Sayısı</b>	<b>Toplam Küme Yakınlığı</b>	<b>Ağırlıklı Toplam Küme Yakınlığı</b>	<b>C'nin Bağlantılık Değeri</b>
70	MYELOGRAPHY			255	1.066	3.456	0.014
71	GINGIVA			254	1.193	3.234	0.013
72	SALICYLATES			254	1.226	1.969	0.008
73	OUTPATIENT CLINICS, HOSPITAL			254	0.868	1.692	0.007
74	ANTICONVULSANTS	*	2000	252	0.962	1.720	0.007
75	GERMANY, EAST			251	1.439	3.485	0.014
76	TOMOGRAPHY, X-RAY			251	1.096	2.406	0.010
77	CHROMATOGRAPHY, GAS	*	1996	250	1.304	2.889	0.012
78	HEPATIC ENCEPHALOPATHY	*	1998	250	1.049	2.064	0.008
79	BONE MARROW DISEASES			250	0.995	1.698	0.007
80	CYTODIAGNOSIS			248	1.186	2.337	0.009
81	NEOPLASMS, MULTIPLE PRIMARY	*	1989	247	1.132	2.843	0.012
82	THORACIC INJURIES			247	1.095	2.560	0.010
83	DIGOXIN			247	1.169	2.275	0.009
84	OPHTHALMOSCOPY			247	1.297	1.923	0.008
85	CARBOHYDRATE METABOLISM			247	1.106	1.810	0.007
86	LIVER CIRCULATION	*	2006	246	1.569	3.346	0.014
87	URINARY CATHETERIZATION			246	1.005	1.853	0.008
88	EXUDATES AND TRANSUDATES			246	0.965	1.358	0.006
89	NYSTAGMUS, PATHOLOGIC			245	1.192	2.054	0.008
90	CYTOPLASMIC GRANULES	*	1996	245	1.160	1.878	0.008
91	ALBUMINS			245	1.231	1.695	0.007
92	MENIERE DISEASE	*	1994	244	1.142	2.078	0.009
93	BRONCHIAL DISEASES			244	1.057	2.001	0.008

	<b>C terimi</b>	<b>Dođru - Olumlu</b>	<b>İlk Bulunma Tarihi</b>	<b>Ortak Küme Sayısı</b>	<b>Toplam Küme Yakınlığı</b>	<b>Ağırlıklı Toplam Küme Yakınlığı</b>	<b>C'nin Bağlantılık Deđeri</b>
94	MEDIASTINAL NEOPLASMS	*	2008	243	1.164	2.603	0.011
95	NEURILEMMOMA			243	1.193	2.330	0.010
96	HEART CONDUCTION SYSTEM			242	1.096	3.469	0.014
97	MURAMIDASE			242	0.980	1.657	0.007
98	HISTOLOGICAL TECHNIQUES			241	1.064	2.170	0.009
99	HYPOKALEMIA			240	1.184	2.178	0.009
100	BODY FLUIDS			240	1.068	1.620	0.007

### Ek.3. Ölçme Yönteminin Sınanması

Sıra	C terimi	İlk Bulunma Tarihi	Toplam	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
1	RETINAL DISEASES	1987	8	1	1		1	1			1				1			1				1							
2	THYROIDITIS, AUTOIMMUNE	1993	7							1					1	1	2					1					1		
3	ANTIBODIES, MONOCLONAL	1999	6													1	1					1		2		1			
4	ANESTHETICS, LOCAL	2001	6															1		1				1		1		1	1
5	WOUNDS, NONPENETRATING	1987	5	1											1			1					1						
6	DIABETES MELLITUS, TYPE 2	1987	5	1						1	1																		
7	VISUAL FIELDS	1987	5	1	1				1																1	1			
8	REGISTRIES	2002	5																1						1	2	1		
9	INTRACRANIAL ANEURYSM	1993	4							1		1												1	1				
10	ADAPTATION, PSYCHOLOGICAL	1995	4									1											1						
11	ANTIFUNGAL	1996	4										1	1										1					

Sıra	C terimi	ilk Bulunma Tarihi	Toplam	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
	AGENTS																												
12	NEOPLASM RECURRENCE, LOCAL	1987	3	1		2																							
13	EUROPEAN CONTINENTAL ANCESTRY GROUP	1993	3							1												1					1		
14	PERIODICITY	1993	3							1					1				1										
15	PATIENT CARE PLANNING	1994	3								1	1		1															
16	ANTIBODIES, BACTERIAL	1990	3				1								1			1											
17	HEPATIC ENCEPHALOPAT HY	1998	3												1							1					1		
18	NEOPLASMS, MULTIPLE PRIMARY	1989	3			2													1										
19	ARTERIOVENOU S SHUNT, SURGICAL	1991	3					1		1				1															
20	ANTIDEPRESSIVE AGENTS	1992	3						1							1										1			
21	ANESTHESIA, EPIDURAL	1991	3					1									1							1					

Sıra	C terimi	ilk Bulunma Tarihi	Toplam	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
22	INFERTILITY, MALE	1991	3					1											1		1								
23	PATIENT CARE TEAM	2002	3																1				1	1					
24	INTERFERONS	1993	3							1			1					1											
25	PULMONARY VENTILATION	1987	3	1													1								1				
26	ANTIBODIES, VIRAL	1999	2													1	1												
27	GRAVES DISEASE	1997	2											1											1				
28	BLOOD CHEMICAL ANALYSIS	2003	2																	1					1				
29	ATTITUDE TO HEALTH	1995	2									1													1				
30	ANTICONVULSA NTS	2000	2														1						1						
31	CHROMATOGRA PHY, GAS	1996	2										1										1						
32	POSITIVE- PRESSURE RESPIRATION	1994	2								1					1													
33	BIOLOGICAL TRANSPORT	1991	2					1							1														
34	LIPOPROTEINS, LDL	1995	2									1				1													



Sıra	C terimi	ilk Bulunma Tarihi	Toplam	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
35	FIBER OPTIC TECHNOLOGY	2002	2																1						1				
36	BONE TRANSPLANTATI ON	1999	2													1									1				
37	SOCIAL CLASS	1992	2						1						1														
38	PSYCHIATRIC STATUS RATING SCALES	1989	2			1																		1					
39	COLORING AGENTS	2009	2																						1		1		
40	SPECTRUM ANALYSIS	1987	2	1																					1				
41	PRIMARY MYELOFIBROSIS	2003	2																	1				1					
42	TELEVISION	1987	2	1		1																							
43	CROSS REACTIONS	1991	2					1										1											
44	SPECIMEN HANDLING	2005	2																			2							
45	DUODENAL DISEASES	2002	2																1							1			
46	ANTI-ANXIETY AGENTS	2007	2																					1			1		
47	THYMUS NEOPLASMS	2007	2																					1	1				

Sıra	C terimi	ilk Bulunma Tarihi	Toplam	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
48	AEROSPACE MEDICINE	2000	2														1											1	
49	SULFHYDRYL COMPOUNDS	1992	2						1																1				
50	LEUKEMIA, MYELOID	1997	1											1															
51	PARAPLEGIA	1995	1									1																	
52	LEUKEMIA, MYELOID, ACUTE	1996	1										1																
53	ANESTHESIA, INHALATION	1994	1								1																		
54	RESPIRATION, ARTIFICIAL	2009	1																							1			
55	RADIOTHERAPY DOSAGE	2009	1																							1			
56	AORTA, THORACIC	1990	1				1																						
57	MYCOSES	2007	1																				1						
58	AEROSOLS	1990	1				1																						
59	STOMACH DISEASES	2010	1																							1			
60	OXYGEN INHALATION THERAPY	2007	1																				1						
61	AUDIOMETRY	2009	1																							1			

Sıra	C terimi	ilk Bulunma Tarihi	Toplam	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
62	INTUBATION, INTRATRACHEAL	2009	1																							1			
63	MOUTH NEOPLASMS	2002	1																1										
64	CARCINOMA, HEPATOCELLULA R	2006	1																				1						
65	AGRANULOCYTO SIS	1991	1					1																					
66	URINARY BLADDER NEOPLASMS	2003	1																	1									
67	PRECANCEROUS CONDITIONS	2007	1																					1					
68	LIVER CIRCULATION	2006	1																				1						
69	CYTOPLASMIC GRANULES	1996	1										1																
70	MENIERE DISEASE	1994	1								1																		
71	MEDIASTINAL NEOPLASMS	2008	1																					1					
72	ESOPHAGOSCO P Y	1989	1			1																							
73	NEOPLASM STAGING	2002	1																1										

Sıra	C terimi	ilk Bulunma Tarihi	Toplam	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
				74	RURAL POPULATION	2009	1																						
75	SHOCK, SEPTIC	1988	1		1																								
76	HYDRONEPHROS IS	2010	1																								1		
77	EAR DISEASES	2002	1																1										
78	MUSCLE SPASTICITY	2001	1														1												
79	ANTIGENS, BACTERIAL	1998	1												1														
80	COGNITION DISORDERS	2007	1																					1					
81	LEAD POISONING	1993	1							1																			
82	SLEEP STAGES	1990	1				1																						
83	PSYCHOTROPIC DRUGS	1992	1						1																				
84	MILITARY MEDICINE	2002	1																1										
85	LUNG DISEASES, FUNGAL	1996	1										1																
86	SAMPLING STUDIES	2000	1														1												
87	PHYSICIAN- PATIENT	2009	1																						1				

Sıra	C terimi	ilk Bulunma Tarihi	Toplam	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
	RELATIONS																												
88	ANGIOPLASTY, BALLOON	2011	1																									1	
89	ANTIPSYCHOTIC AGENTS	2011	1																									1	
90	CASTS, SURGICAL	2009	1																						1				
91	COST-BENEFIT ANALYSIS	2012	1																										1
92	HEARING TESTS	2007	1																				1						
93	VAGINAL SMEARS	2007	1																				1						
94	SCLEROSING SOLUTIONS	2006	1																			1							
95	UNCONSCIOUSN ESS	2001	1														1												
96	MACULA LUTEA	1992	1						1																				
97	COLLOIDS	2007	1																				1						
98	STERNUM	1990	1				1																						
99	CHROMATOGR APHY, HIGH PRESSURE LIQUID	1999	1													1													
100	OPTIC ATROPHY	2008	1																					1					

#### Ek.4. Rasgele 80 Terim için Sonuçlar

Terim İsmi	Anımsama	Duyarlık	Doğru Getirilen	Getirilmesi Gereken	Getirilenler
SPECIFIC GRAVITY	0.70	0.06	827	1184	14823
THERAPEUTIC EQUIVALENCY	0.67	0.13	2038	3044	16217
IODOBENZENES	0.66	0.07	1026	1543	14790
LACERATIONS	0.66	0.12	1812	2731	15017
PSYCHOLINGUISTICS	0.66	0.05	603	911	13115
NIGHT CARE	0.66	0.05	627	952	12228
FOOT DERMATOSES	0.65	0.08	1047	1604	13693
SPOUSE ABUSE	0.65	0.09	1134	1742	12508
SULFASALAZINE	0.65	0.09	1591	2456	16893
PARAQUAT	0.65	0.11	1690	2619	16072
GASTROINTESTINAL CONTENTS	0.65	0.14	1957	3033	13825
NONPRESCRIPTION DRUGS	0.64	0.11	1744	2704	15448
LINDANE	0.64	0.06	1022	1585	17771
RADIOALLERGOSORBENT TEST	0.64	0.06	797	1242	14226
BONE WIRES	0.64	0.07	936	1466	13901
GASTRIC FUNDUS	0.63	0.09	1145	1809	13382
CARCINOGENICITY TESTS	0.63	0.18	2215	3518	12192
MASSETER MUSCLE	0.63	0.09	1373	2182	14730
SCIURIDAE	0.63	0.08	1091	1741	13061
PREALBUMIN	0.63	0.11	1610	2571	14564
POLYGLYCOLIC ACID	0.63	0.17	2546	4066	15044
IMPULSE CONTROL DISORDERS	0.62	0.07	803	1289	11343

Terim İsmi	Anımsama	Duyarlık	Doğru Getirilen	Getirilmesi Gereken	Getirilenler
DENTAL IMPRESSION TECHNIQUE	0.62	0.03	472	758	15955
FERRETS	0.62	0.16	2021	3246	13008
ELIGIBILITY DETERMINATION	0.62	0.10	1234	1990	12607
IMMUNOGLOBULIN IDIOTYPES	0.62	0.04	504	814	14240
SOUND SPECTROGRAPHY	0.62	0.11	1401	2264	12609
BENZENESULFONATES	0.62	0.14	2153	3488	15875
SATIATION	0.61	0.06	795	1296	14319
MARIJUANA ABUSE	0.61	0.10	1173	1915	12164
GLUCURONOSYLTRANSFERASE	0.61	0.15	2107	3446	14515
TRANSGLUTAMINASES	0.61	0.18	2477	4056	13734
FAT NECROSIS	0.61	0.05	581	956	12645
CONGENITAL HYPOTHYROIDISM	0.60	0.05	800	1326	17041
MITOXANTRONE	0.60	0.11	1581	2642	13822
UTERINE CERVICAL DYSPLASIA	0.59	0.06	777	1306	13971
PERIAQUEDUCTAL GRAY	0.59	0.09	1067	1805	12450
INHIBINS	0.59	0.13	1704	2889	13620
PNEUMOCOCCAL VACCINES	0.59	0.10	1358	2303	13248
BETA RHYTHM	0.59	0.04	497	843	11329
ARCUATE NUCLEUS	0.59	0.08	1086	1847	12824
DACARBAZINE	0.59	0.11	1562	2660	13653
SPINOCEREBELLAR DEGENERATIONS	0.59	0.07	995	1698	15012
GHANA	0.58	0.12	1824	3129	15407
TRIGEMINAL GANGLION	0.58	0.10	1476	2534	15366
EPIRUBICIN	0.58	0.11	1504	2588	13510
EXTRACORPOREAL MEMBRANE OXYGENATION	0.58	0.17	1727	2978	9889
SUPERANTIGENS	0.58	0.12	1596	2764	13228

Terim İsmi	Anımsama	Duyarlık	Doğru Getirilen	Getirilmesi Gereken	Getirilenler
LEUKOCYTE ELASTASE	0.57	0.19	2220	3879	11564
PREDATORY BEHAVIOR	0.57	0.13	1586	2783	11753
MIFEPRISTONE	0.57	0.18	2505	4418	14001
HORNS	0.57	0.03	330	583	11516
DENGUE VIRUS	0.56	0.09	1464	2625	15915
SYNAPTOPHYSIN	0.54	0.17	1979	3661	11565
BUNYAVIRIDAE INFECTIONS	0.53	0.03	330	619	12251
AZITHROMYCIN	0.51	0.23	2179	4242	9660
BARRETT ESOPHAGUS	0.50	0.11	1296	2610	11671
DEPSIPEPTIDES	0.48	0.21	1553	3246	7285
POLYMETHYL METHACRYLATE	0.48	0.23	1955	4109	8389
TROPONIN T	0.48	0.19	1767	3718	9499
GASTRIC BYPASS	0.44	0.24	1353	3043	5756
RYANODINE RECEPTOR CALCIUM RELEASE CHANNEL	0.43	0.26	1522	3560	5806
OSTEOPONTIN	0.42	0.35	2235	5289	6402
VETERINARIANS	0.41	0.12	739	1791	5997
ROTATOR CUFF	0.40	0.14	736	1818	5119
PARATHYROIDECTOMY	0.38	0.15	979	2595	6659
THYROID NODULE	0.37	0.14	733	1995	5382
EPITHELIOID CELLS	0.35	0.15	512	1475	3318
SUPERIOR CERVICAL GANGLION	0.34	0.25	927	2712	3684
INCIDENTAL FINDINGS	0.32	0.47	1498	4685	3160
INTERNAL FIXATORS	0.30	0.27	706	2339	2615
IXODIDAE	0.28	0.14	437	1548	3105
HYPEROXIA	0.26	0.24	831	3258	3509
PASSERIFORMES	0.23	0.27	379	1657	1418



<b>Terim İsmi</b>	<b>Anımsama</b>	<b>Duyarlık</b>	<b>Doğru Getirilen</b>	<b>Getirilmesi Gereken</b>	<b>Getirilenler</b>
VENTRICULAR PREMATURE COMPLEXES	0.18	0.36	306	1692	846
EXTERNAL FIXATORS	0.16	0.70	360	2241	516
INHIBITORY POSTSYNAPTIC POTENTIALS	0.15	0.50	255	1720	509
SOLITARY NUCLEUS	0.07	0.52	174	2615	335
ILIAC ANEURYSM	0.06	0.19	58	921	305
FUNDOPLICATION	0.01	0.69	27	1948	39

Notlar :