



---

TEKNİK NOT

---

**BİR METİN MADENCİLİĞİ UYGULAMASI: VOSVIEWER**

**Murat ARTSIN**<sup>1,\*</sup> 

<sup>1</sup> Uzaktan Eğitim Departmanı, Bahçeşehir Üniversitesi, İstanbul, Türkiye

**ÖZET**

Geçmişten günümüze kadar veri ve veri yığınlarından anlamlı bilgilerin üretilme çabası artarak devam etmektedir. Büyük veriden anlamlı bilginin üretilme sürecinde araştırmacıların üzerinde durduğu yaklaşımlardan birisi olarak görselleştirme, son yıllarda oldukça ilgi toplamış ve bir çok farklı alanda yaygınlaşmaya başlamıştır. Durum böyle olunca, veri madenciliğinde karmaşık ilişkilerin ortaya çıkarıldığı bibliyometri çalışmaları da elde edilen verinin görselleştirilmesi ve görselleştirme yaklaşımlarının incelenmesi önemli bir konu haline gelmiştir. Bu çalışmada, bibliyometrik ağların görselleştirilmesinde kullanılan ve aynı zamanda metinler içerisindeki kelime ve kelime öbeklerinin ilişkisini ortaya çıkarmaya yarayan başka bir ifade ile metin madenciliği uygulaması olan VOSviewer programının kullanımı incelenecektir. Bu teknik not ile veri madenciliğinin kullanımına yönelik bilgilendirici yönlendirmelere yer verilecektir.

**Anahtar Kelimeler:** Veri madenciliği, Metin madenciliği, Bibliyometrik analiz, VOSviewer

**A TEXT MINING APPLICATION: VOSVIEWER**

**ABSTRACT**

From past to present, efforts to produce meaningful information from data and data stacks increasingly continue. Visualization, as one of the approaches emphasized by researchers in the process of producing meaningful information from big data, has gained great interest in recent years and has become widespread in many different fields. As such, the visualization of data obtained in bibliometric analysis studies revealing complex relationships in data mining and examination of visualization approaches have become an important issue. In this study, the use of VOSviewer program, being a text mining application, used to visualize bibliometric networks and also used to reveal the relationship between words and phrases in the texts, will be examined. This technical note will include informative directions for the use of data mining.

**Keywords:** Data mining, Text mining, Bibliometric analysis, VOSviewer

---

**1. GİRİŞ**

Veri madenciliği: bilgisayar bilimleri, veri tabanı teknolojileri, derin öğrenme, makine öğrenmesi ve veri bilimi gibi birçok farklı disiplini içinde barındırmaktadır [1]. Ancak en genel ifade ile anlamsız veri yığınlarının belirli işlemlerden geçerek anlamlı hale getirilmesi olarak ifade edilebilir. Birbirinden farklı disiplinlerin bir arada kullanılması ile ortaya çıkan bu yöntem, günümüzde mühendislik çalışmalarının yanı sıra sağlık [2, 3, 4], eğitim [5, 6, 7], pazarlama [8, 9, 10], sosyal medya [11, 12] ve tarım [13, 14, 15] gibi birçok farklı alanda kullanılmaktadır. Yukarıdaki belirtilen farklı disiplinlere ait veri madenciliğinin kullanım amaçlarındaki ortak nokta, veri yığınları arasından bir “bilgi keşfi” sürecinin işe koşulmasıdır.

Veri madenciliğinin bir bilgi keşfi süreci olarak ifade edilmesinin en temel nedeni, aşamalı olarak gerçekleştirilen bir girdi ve sonucun bulunduğu bir yapıya sahip olmasıdır. Bu yapıyı; problemin tanımlanması, verinin anlaşılması, verinin hazırlanması, modelleme, modelin değerlendirilmesi ve modelin kullanılması [16] şeklinde betimlemek mümkündür. Bu süreç her ne kadar ortak aşamalardan

---

\*Sorumlu Yazar: [artsinm@gmail.com](mailto:artsinm@gmail.com)

Geliş: 09.11.2019 Yayın: 31.08.2020

oluşsa da, veri madenciliği modelleri tanımlayıcı ve tahmin edici olmak üzere iki ana başlıkta toplanmaktadır. Tanımlayıcı modeller ilişki analizi ve kümeleme analizlerini içerirken, tahmin edici modeller sınıflama ve istatistiksel modelleri içermektedir.

Tanımlayıcı modeller bir karar verme ve rehberlik süreci iken, tahmin edici modeller sonuçları belli olan verilerden tahminlerde bulunma sürecidir. Tanımlayıcı modellerde, birliktelik kuralları ve ardışık zaman örüntülerini içerisine alan *ilişki analizi* ve *kümeleme analizi* yer almaktadır. Tahmin edici modellerde ise, karar ağaçları, yapay sinir ağları ve genetik algoritmaları içerisine alan *sınıflama* ve regresyon analizi, diskriminant analizi ve lojistik regresyon analizini içerisine alan *istatistiksel tahmin tekniklerini* içermektedir [16, 1].

Veri madenciliğinin yukarıda belirtilen özellikleri, kullanım amaçları, kullanım alanları ve modelleri temelde, ses, görüntü, numerik veya alfanumerik gibi veri çeşidi ayrımı yapmaksızın araştırma verisi olarak ele almaktadır. Araştırma verilerinin sadece metin yığınlarından yola çıkılarak gerçekleştirildiği çalışmalar metin madenciliği alanında sınıflandırılmaktadır.

## 2. METİN MADENCİLİĞİ

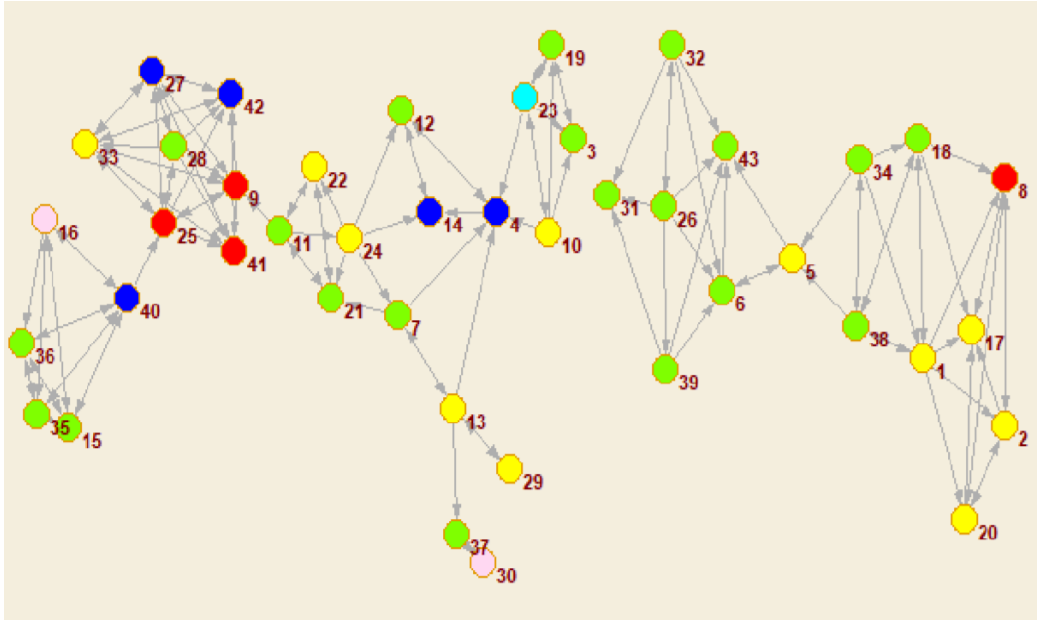
Metin madenciliği uygulamalarında, teknolojinin sahip olduğu yüksek hız ve güvenilirlik [17] terimiyle birlikte “bilgisayarlı metin analizi” [18] terimi de kullanılmaktadır. Bunun temel nedeni, bilgi ve iletişim teknolojilerin sahip olduğu yüksek hız ve güvenilirlik ile [17] gerçekleştirilen çalışmalarda bilgisayar uygulamalarının alana hakim olması olarak ifade edilebilir.

Veri madenciliği çalışmalarında yer verilen konulardan biri olarak metin madenciliği, web üzerinde yer alan büyük verinin kullanımı ile gerçekleştirilen çalışmalara ev sahipliği yapan bir uygulama alanı olarak karşımıza çıkmaktadır [19]. Metin madenciliği, hacimsel olarak büyük verilerin bulunduğu [18], büyük veri yığınları arasından anlamlı yapıların ortaya çıkarıldığı [19] ve bu anlamlandırma süreçlerinde birçok farklı uygulama ve analizin kullanıldığı uygulamaların bütünüdür.

Genel itibariyle metin madenciliği bir veri yığını içindeki verilerin özelliklerinin ortaya çıkarılması, kümelenmelerinin sağlanması, sınıflandırmalarının gerçekleştirilmesi, özetlenmelerinin sağlanması, trend analizinin ortaya çıkarılması ve görselleştirmelerinin yapılması amacıyla kullanılan [20] temeli anahtar kelimelerin kaç kez tekrar ettiği bağlamında hesaplanan frekansın sunulduğu uygulamalardır. Metin madenciliği bir alanda ya da herhangi bir konuda değişen fenomenin belirlenmesi amacıyla [18] ve literatür analizlerinde ortaya çıkan değişimin sunulması [21, 22, 23, 24] gibi amaçlarla kullanılmaktadır. Durum böyle olunca metin madenciliği ile ilgili gerçekleştirilen çalışmalardan biri olarak bibliyometrik analiz, araştırma bulgularının sunulması açısından bir hayli öneme sahip olduğu karşımıza çıkmaktadır.

Bibliyometri analizi; elde edilen belgelerin veya makalelerin belirli özelliklerine göre analiz edilmesi ve bulguların ortaya çıkarılmasına dayanmaktadır [25]. Belirli bir alana özgü çalışmalar veya spesifik bir dergide yayınlanan araştırmalar, farklı parametrelere göre incelenmekte ve bu parametrelere göre bulguların sunulmasını sağlamaktadır [26]. Bu bulgular, anahtar kelimeler, yazar sayısı, atıf ve araştırma konusu gibi verileri içerebilmektedir. Metin madenciliği uygulamalarının daha iyi anlaşılması için, bibliyometrik analize ilişkin detaylı bir tanımlama yapmak ve özelliklerini belirlemek önemlidir. Bibliyometrik analiz çok geniş amaçla gerçekleştirilen çalışmalardır: genel olarak alanlar hakkındaki bulguların okuyuculara sunulması ya da özel bir konu hakkında derinlerde bulunan bilginin ortaya çıkarılması için kullanılan çalışmalardır. Bibliyometri analizleri, çalışmalar içerisindeki atıfların analizi ve araştırmaların değerlendirilmesinde kullanılmaktadır [25]. Bibliyometrik analizlerde farklı uygulamalarda bulunmaktadır.

Bibliyometrik analizlerde görselleştirme yaklaşımı bağlamında kullanılan mesafe tabanlı, grafik tabanlı ve zaman çizelgesi tabanlı yaklaşımlara dair haritalar kullanılırken [27], mesafe tabanlı ve grafik tabanlı yaklaşım harita türleri daha fazla dikkat çekicidir [28]. Grafik tabanlı haritalarda, analiz sonucunda elde edilen çıktılardaki öğelerin birbirleriyle ilişkileri ortaya konulurken, ilişki gücüne ait bir bulgu ifade edilmemektedir. Bununla birlikte mesafe tabanlı haritalarda ise, elde edilen çıktılardaki öğelerin birbirlerine olan uzaklık ve yakınlıkları öğeler arasındaki ilişkinin gücünü ortaya koymaktadır [29]. Burada en önemli nokta, araştırmacının kendi araştırma sorularındaki hedeflenen analizlere yönelik grafik tabanlı veya mesafe tabanlı haritalardan faydalanmasıdır. Bir başka ifade ile, öğeler arasındaki ilişkilerin araştırma problemi olduğu durumlarda mesafe tabanlı haritalar kullanılabilecek iken, sadece mevcut öğelerin ortaya konulmasının amaçlandığı çalışmalarda da grafik tabanlı haritaların kullanıldığı bibliyometrik analizlerden yararlanılabilir.



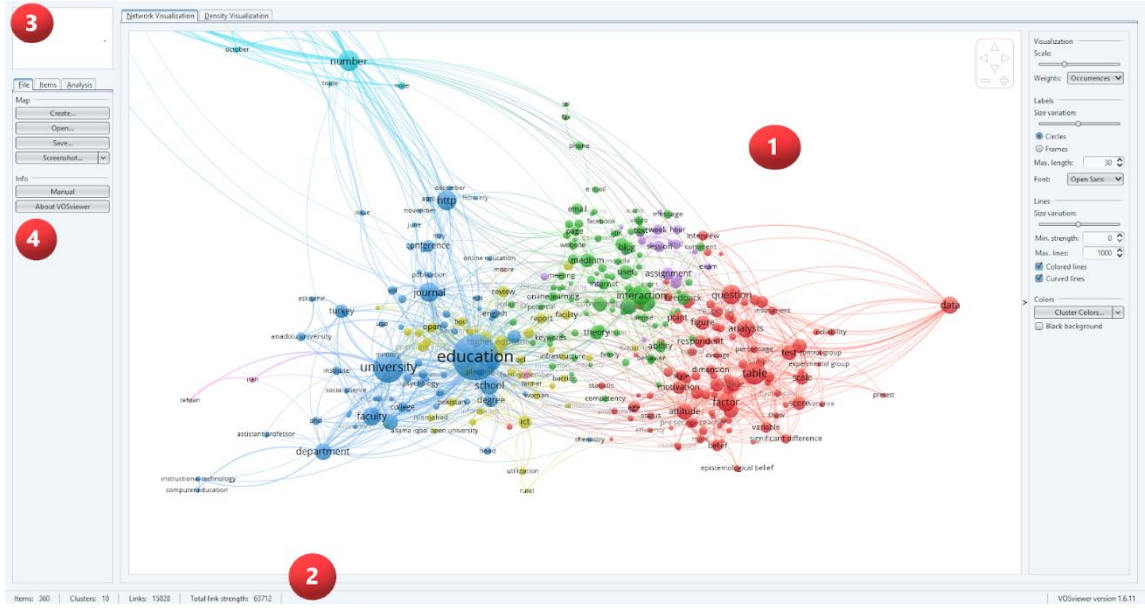
Görsel 1: Kamada–Kawai tekniği ile grafik tabanlı haritalama [30].

Mesafe tabanlı ve grafik tabanlı haritalar oluşturmak için farklı haritalandırma teknikleri bulunmaktadır. Grafik tabanlı haritalama için Kamada–Kawai, Fruchterman–Reingold ve Pathfinder ağları gibi teknikler bulunurken, mesafe tabanlı haritalama için çok boyutlu ölçekleme, VOS, VxOrd, Kopcsa–Schiebel gibi teknikler bulunmaktadır. Bu teknikler içerisinde en iyi performansa sahip olan VOS tekniği [31], alanda gerçekleştirilen çalışmalarda dikkat çekici olarak yer almaktadır. Genel olarak VOS, öğeler arasındaki benzerliklerin görselleştirilmesi için çok boyutlu ölçeklendirme süreci olup ideal koordinatların sunumu, dolaylı benzetme ve çok boyutlu ölçekleme gibi özellikleri ile dikkat çekicidir. Daha açık bir ifade ile iki nesne arasındaki benzerlik arttıkça, toplamda kare uzaklıklarının ağırlığı da artmaktadır. Bu bağlamda da tüm nesnelerin aynı koordinatlarda bulunduğu çözümleri önlemek için tüm mesafelerin toplamının bir miktar pozitif sabite eşit olması gerektiği şeklinde temellendirilen teknik bir metottur [32].



VOSviewer programı birçok farklı panellerden oluşmaktadır. Bunlar; ana panel, bilgi paneli, genel bakış paneli ve eylem panelidir.

**Ana panel;** aktif olarak kullanılan haritaların sunulduğu ve navigasyon butonlarının bulunduğu alandır. **Bilgi paneli;** ana panel içerisinde aktif olarak hangi öge hakkında bilgi görüntüleniyorsa o öge hakkında detaylı bilginin bulunduğu kısımdır. **Genel bakış paneli;** ana panel içerisinde aktif olan haritanın genel bir görünümünün bulunduğu alandır ve bu panel ana panele göre oldukça küçük yer kaplayarak haritalar üzerindeki renklerin görüntülenmesini sağlar. **Eylem paneli;** yeni harita oluşturma, haritayı kaydetme, ekran görüntüsü alma ve var olan bir haritayı geri yükleme işlemlerinin gerçekleştirildiği alandır. Görsel 3’de ana panel (1), bilgi paneli (2), genel bakış paneli (3) ve eylem panelinin (4) program içerisindeki konumları bulunmaktadır.



Görsel 3: VOSviewer programına ait panel ve sekmeler

**Ana panel:** Bu panel içerisinde VOSviewer programında gerçekleştirilen bibliyometrik analizin sonucu görüntülenebilmektedir. Ana panel sahip olduğu esnek yapısı sayesinde panel üzerinde bulunan tüm öğelere yakınlaşılabilir özelliğine sahiptir. Bu sayede bir ögenin diğer öğelerle detaylı ilişkilerinin görselleştirilmesine olanak sağlar. Bu özellik öğeler arasındaki yeni bağlamların tespiti ve görünürlüğü açısından oldukça dikkat çekici bir özelliktir.

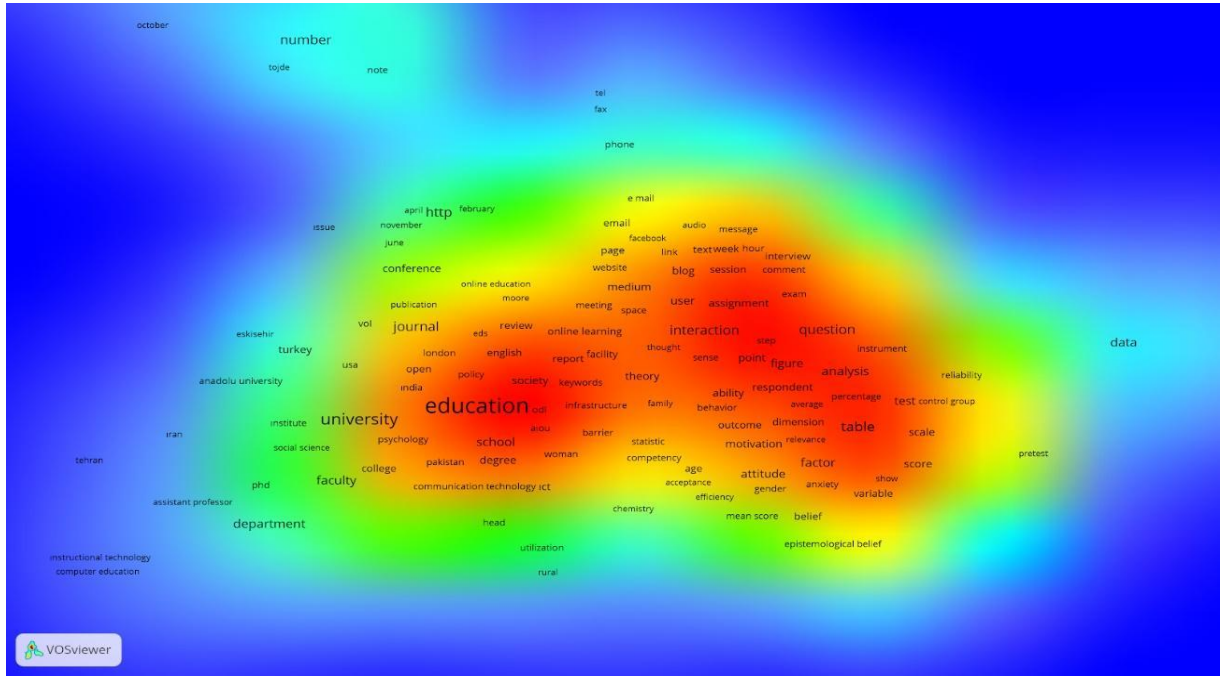
**Bilgi paneli:** Genel sonuçların listelenmesi sonucunda bu panel içerisinde öge sayısı, küme sayısı, bağlantı sayısı ve diğer tüm bağlantıların güçlülük değeri yer almaktadır. Yine bu panelde herhangi bir öğeye odaklanılması durumunda ilgili ögenin hangi kümenin içerisinde bulunduğu, sahip olduğu bağlantı sayısı, sahip olduğu bağlantıların toplam ağırlığına dair bilgiler bulunmaktadır.

**Genel bakış paneli:** Bu alan içerisinde gerçekleştirilen analizin sonucunda ortaya çıkan haritanın daha küçük bir görüntüsü bulunmaktadır. Ana panel içerisinde gerçekleştirilecek yakınlaştırma veya uzaklaştırma süreçlerinde bile genel bakış paneli içerisinden mevcut genel durumun görüntülenmesi mümkündür. Araştırmacının ana panel üzerinde yaptığı yakınlaştırma ve uzaklaştırma işlemlerinin nerede olduğuna dair konum bilgisi yine bu panel ara yüzünde görüntülenebilir.

**Eylem Paneli:** Burada yer alan dosya, nesne ve analiz sekmelerinde farklı amaçlar doğrultusunda kullanılacak fonksiyonlar bulunmaktadır. Dosya sekmesi içerisinde yeni analiz oluşturma, daha önce gerçekleşen analizin tekrar ana panel içerisinde görüntülenmesinin sağlanması, analizin kayıt edilmesi

işlevleri yer almaktadır. Bunun yanı sıra araştırmalarda kullanılabilir şekilde ekran görüntüsünün alınması amacıyla kullanılan işlevler de bulunmaktadır. Nesne sekmesi analiz sonucunda oluşan kümelerin ve kümelere ait öğelerin görüntülenebildiği sekmedir. Analiz sekmesi ise oluşacak olan kümelerin boyutu ve yapısı hakkında düzenleme yapılması istendiği zaman kullanılacak bir alandır. Bu alan içerisinde farklı analiz metotlarının kullanılabilceği seçenekleri barındırmaktadır.

VOSviewer programında farklı görünümler bulunmaktadır. Bunlar; yoğunluk görünümü, küme yoğunluk görünümü ve dağılım görünümüdür. Yoğunluk görünümünde, harita üzerindeki her nokta, ilgili öğenin yoğunluğuna göre bir renge sahip olmakla birlikte [33], genel bir bakış açısının sunulması ve önemli alanlara dikkat çekilmesi açısından önemlidir. Bir öğenin çevresindeki öğe sayısı artmaya başladıkça, öğenin renginde de yoğunluk artarak kırmızılaşmaktadır. Tam tersi bir durum ise, öğenin etrafındaki öğe sayısının azalması ile rengin yoğunluğunun azalmaya başladığı şeklindedir. Böyle bir durumda da öğenin mavi renge yaklaştığı görüntülenebilir. Görsel 4’de ana panel içerisinde yer alan yoğunluk görünümü bulunmaktadır.



Görsel 4: VOSviewer yoğunluk görünümü

VOSviewer programında küme yoğunluk görünümü elde edilebilmesi için öğelerin kümelere ayrılmış olması gerekmektedir. Küme yoğunluğu görünümünde bir noktanın rengi; noktanın çevresinde kümeye ait çok sayıda öğe varsa kümenin rengine yakındır. Ağırlığı yüksek olan öğeler, düşük olan öğelere göre daha yüksek yoğunluğa sahip kabul edilmektedir. Bu bağlamda yüksekliğe göre değişen yoğunluklar haritada kolaylıkla görüntülenebilir. Dağılım görünümü ise, öğelerin her birinin daireler içerisinde sunulduğu görünümdür. Eğer öğeler için renk ataması yapılmış ise her bir öğenin dairesi içerisinde bulunduğu kümenin rengi ile görüntülenebilir. Burada en dikkat çekici nokta ise etiketler ile karşılaşmıyor olmaktır. Dağılım görünümü bir harita için genel bir bakışa sahip olmamız için önemlidir. Çünkü kavramların genel özetini sunmaktadır.

Bunların yanı sıra VOSviewer programı birçok farklı kullanım kolaylığı sunmaktadır. Ana panel üzerinde bulunan navigasyon butonları, sunulan haritalarda belirli noktalara odaklanılarak, daha detaylı görünüm için yakınlaşma işlevi yapılmasını sağlamaktadır. Öte yandan VOSviewer’da gerçekleştirilen çalışmalarda veri seti farklı dosya formatlarında yüklenebilir. Txt dosya uzantısının yanı sıra excel, spss,

matlab ve r dosya uzantılarını da desteklemektedir. Web of Science ve Scopus gibi dizinlerden elde edilen veriler kullanılmak istendiğinde *Read data from bibliographic database files* seçeneği, kaynak dosyanın mevcut bilgisayarda olduğu durumda da *Read data from reference manager files* veya doğrudan VOSviewer programında bulunan Api kaynakları aracılığıyla veri elde edilmek istenirse *Download data through API* seçeneği kullanılabilir.

Veri dosyası program içerisine yükledikten sonra program, otomatik olarak veri seti içerisinde en sık kullanılan kelimelerin sayısını sunmaktadır. *Choose threshold* penceresinde, araştırmacıların oluşturulacak olan harita üzerinde bir kelime öbeğinin görünür hale gelebilmesi için veri dosyası içerisinde en az kaç kez tekrar edilmesi gerektiğini belirtmesi gerekmektedir. Bu belirtilen rakama *eşik değeri* ismi verilmektedir. Burada araştırmacının verdiği rakamın yüksek olması durumunda oluşturulacak olan öğeler daha az, verilen rakamın düşük olması durumunda ise oluşturulacak olan öğeler daha fazla olacaktır. Bu noktada araştırmacıların veri setinden anlamlı bilgilerin ortaya çıkarabilmeleri için en doğru eşik sayılarını belirtmeleri oldukça önemlidir. Eşik sayısı belirlendikten sonra *Choose Number of Terms* penceresinde, eşik sayısı ile öğelerin alaka düzeylerini puanlandırılır. Bu alaka puanına göre en alakalı öğelerin seçilmesi faydalı olacaktır. Bu pencere otomatik olarak hesaplanarak sunulan alakalı puanın, eşik sayısında belirtilen rakam ile ortaya çıkan öğelerin ne kadar bir kısmını ifade ettiğini ve genel olarak tüm harita için yeterli olan öğelerin program tarafından belirtildiği alandır. Araştırmacılar buradan isterler ise sadece alakalı puan ile oluşan öğeler ile devam edebilirler iken isterlerse eşik sayısı ile meydana gelen öğe sayısı ile de devam edebilirler.

Bir analiz süreci olması bağlamında, VOSviewer analizi gerçekleştirmek isteyen araştırmacıların yapması gereken bazı öncelikler bulunmaktadır. Araştırmacılar ilk olarak analiz yapacağı veriyi temizlemeli, ardından VOSviewer programının kabul ettiği uygun formata dönüştürmelidirler. Araştırmacının bu veri temizleme sürecinden sonra izlemesi gereken adımlar şu şekildedir;

**Adım 1:** VOSviewer programı işletim sistemine uygun bir şekilde [www.vosviewer.com/download](http://www.vosviewer.com/download) web sayfasından indirilir.

**Adım 2:** VOSviewer uygulaması açılır.

**Adım 3:** *Create* butonuna tıklanarak yeni proje açılır.

**Adım 4:** Metin madenciliği analizi gerçekleştirilmesi için açılan *Choose type of data* penceresinden *Create a map based on text data* seçeneği seçilerek devam edilir.

**Adım 5:** Araştırmacı tarafından hazırlanan veri dosyasının eklenebilmesi için *Choose data source* penceresinden *Read data from VOSviewer files* alanı seçilerek devam edilir.

**Adım 6:** *Select files* penceresinde, daha önceden oluşturulan veri dosyasının *VOSviewer corpus file (required)* alanına eklenmesi gerçekleştirilerek devam edilir.

**Adım 7:** Açılan *Choose counting method* penceresinden *Full counting* seçilerek devam edilir.

**Adım 8:** *Choose threshold* penceresinden eşik değeri belirtilerek devam edilir.

**Adım 9:** *Choose number of terms* penceresinden tüm harita için yeterli olduğu belirtilen eşik değeri bulunmakta olup istenirse araştırmacı tarafından değiştirilebilir veya doğrudan programın sunduğu eşik değeri ile devam edilebilir.

**Adım 10:** *Verify selected terms* penceresinden araştırmacı isterse araştırma bağlamı dışında bulunan kelimelerin harita üzerinde yer verilmemesi amacıyla öğe veya öğelerinin dahil edilmemesi sağlayarak sürece devam edebilir.

**Adım 11:** *Bitir* butonuna tıklanır.

Analiz sonucunda ana panel üzerinde sunulan dağılım görünümü haritasının yanı sıra *Density Visualization* alanından da aynı haritanın yoğunluk görünümünü haritası da görüntülenebilir. Öte yandan ana panel içerisinde ilgili harita görünümüne göre değişen özellikler alanında da her haritaya özgü işlemler gerçekleştirilebilir. Araştırmacı *size variation* alanın dağılım görünümündeki haritadaki ilişkilerin harita üzerinde daha kalın veya daha ince gösterilebilmesi için gerekli ayarlamaları yapabilir. Aynı özellik yoğunluk görünümünde de *Density* alanından harita üzerindeki yoğunluğun daha belirgin

olması veya daha az belirgin olması için kullanılabilir. Zira doğrudan program tarafından tanımlanan sabit renklerin yanı sıra yine araştırmacılar kendi konuları bağlamındaki renkleri *Colors* sekmesi altından belirleyerek kullanabilirler. *Labels* alanından ise harita üzerinde görüntülenen tüm öğelerin yazı boyutunun büyütülmesi gerçekleştirilebilir. Yine istenirse aynı alandan *Circles* ve *Frames* olacak şekilde öğelerin nasıl görünebileceklerine karar verebilirler.

VOSviewer’da gerçekleştirilen metin dosyalarındaki metinlerin sunulmasının yanı sıra bibliyometrik analizde gerçekleştirilmektedir. Bu bağlamda VOS tekniği temeline dayanan VOSviewer’da bir bibliyometrik analiz için izlenmesi gereken adımlar şu şekildedir;

**Adım 1:** VOSviewer programı işletim sistemine uygun bir şekilde [www.vosviewer.com/download](http://www.vosviewer.com/download) web sayfasından indirilir.

**Adım 2:** VOSviewer uygulaması açılır.

**Adım 3:** *Create* butonuna tıklayarak yeni proje açılır.

**Adım 4:** Bibliyometrik analiz gerçekleştirilmesi için açılan *Choose type of data* penceresinden *Create a map based on bibliographic data* seçeneği seçilerek devam edilir.

**Adım 5:** Veri dosyasının program içerisine eklenebilmesi için *Choose data source* penceresinden *Read data from reference manager files* seçeneği ile devam edilir.

**Adım 6:** *Read data from bibliographic database files* seçeneği ile devam edildiğinde *Select files* penceresi açılır. Burada *RIS*, *EndNote* ve *RefWorks* gibi kaynaklardan herhangi birinin eklenmesi ile analize devam edilir.

**Adım 7:** *Choose type of analysis and counting method* penceresinden analiz türü, analiz birimi ve hesaplama metodu gibi ayarların gerçekleştirilmesi sağlanır.

**Adım 8:** *Choose threshold* penceresinden eşik değeri belirtilerek devam edilir.

**Adım 9:** *Choose number of authors* penceresinden tüm harita için yeterli olduğu belirtilen eşik değeri istenirse araştırmacı tarafından değiştirilebilir veya doğrudan programın sunduğu şekliyle devam edilebilir.

**Adım 10:** *Verify selected authors* penceresinden araştırmacı analize dahil edilmeyecek olan çalışmaları belirleyerek sürece devam edebilir.

**Adım 11:** *Finish* butonuna tıklanır.

VOSviewer’da bir bibliyometrik analiz sonucunda çalışma gerçekleştiren yazarların birbirleri ile ilişkileri, yıllara göre değişen kümeleme renklerinin görüntülenmesi söz konusudur. Araştırmacılar ana panel üzerinde bunları yoğunluk görünümü, küme yoğunluk görünümü ve dağılım görünümü ile görüntüleyebilirler. Yine araştırmacılar ana panelde görüntülenen özellikler alanına göre de haritanın görsel sunumu için farklı değişiklikler meydana getirebilirler.

VOSviewer’da analiz sonucunda birçok farklı küme ve öge meydana gelmektedir. VOSviewer programının kendine has bir kümeleme tekniğinin olduğu da belirtmiştir [38]. Meydana gelen kümeler, metin madenciliği ile gerçekleştirilen çalışmalar için oldukça önemlidir. Kümeleme işlemi bibliyometrik araştırmalarda en önemli sorunlardan birisi olarak görülmektedir [39]. VOSviewer programında gerçekleştirilen kümeleme yöntemi, öğelerin toplam düzeyine odaklanmaktadır. Kümeleme işlemi süreci, bibliyometrik analizlerde yayınların birbirleriyle ilişkilendirilmesi ve her kümeye yayınların atanması ile başlar; bu sayede ataması yapılmayan veya küme çakışması yaşayan bir yayın meydana gelmez [39]. VOSviewer’da kümelerin veya öğelerin birbirleri ile yakın konumda olması aralarındaki ilişkiden kaynaklıdır.

#### 4. SONUÇLAR

Bu teknik notta yer alan metin madenciliği çalışmalarında kullanılan bibliyometrik ağların görselleştirilmesi ile birçok alandan araştırmacıya bilgilendirici yönlendirmeler bulunmaktadır.



Mesafe tabanlı ve grafik tabanlı haritaların temeli aktarılmış ve mesafe tabanlı haritalama tekniklerinden VOS'un yapısı tüm detayları ile sunulmuştur. Bunun beraberinde araştırmacıların gerçekleştirecekleri bibliyometri analizi çalışmalarında kullanabilecekleri, VOS tekniği temeli ile oluşturulan VOSviewer programı tüm detayları ile aktarılmıştır. VOSviewer programının arayüzü, kullanımı, altyapısının kullanımına dair bilgiler sunulmuştur. Son olarak literatürde veri madenciliği ve metin madenciliği bağlamında VOSviewer programı ile gerçekleştirilen çalışmalar hakkında genel bilgi verilmiştir.

Bir metin madenciliği uygulaması olan VOSviewer programı ile dergilerin bibliyometrik analizi, belirli konu alanlarının analizi, çalışmalardaki kelime yoğunluklarının belirlenmesi amacıyla analizler, web sitelerin içeriklerinin analizi, öğrenci performanslarının analizi, tezlerin analizi ve ortak yazarlık ile ilgili analizler gerçekleştirilebilir. VOSviewer ile dergilerin yıllara göre değişen konu dağılımlarının belirlendiği çalışmalar oluşturulabilir. Yazar/Yazarların atıflarına dair analizler gerçekleştirilebilir. Mesafe tabanlı ve grafik tabanlı haritalardan faydalanılarak konular ve alanlar hakkında genel bir profilin ortaya çıkarılması sağlanabilir. Mesafe tabanlı haritalar ile analiz gerçekleştirilecek olan çalışmalardaki kelimelerin yakınlıkları tespit edebilir. Bu gibi çalışmalar ile alanyazında kullanılan kelimelerin hangi kelimeler ile daha yakın koordinatlara sahip olduğu tespit edilebilir. Alanda ilişkili kelimelerin tespiti sağlanabilir. Öte yandan büyük verilerin içerisindeki anlamlı ilişkilerin tespit edilebilmesi daha olanaklı hale getirilebilir. VOSviewer programında bulunan küme yoğunluk görünümü ile bir alan veya derginin sahip olduğu konular kümeler biçiminde gösterilebilir. Dergilerin belirledikleri konular ile yayınladıkları konular hakkında detaylı analizler sağlanabilir. Derinlemesine gerçekleştirilecek olan analizler ve kümeler ile disiplinler arası çalışmaların oluşturulması sağlanabilir.

## **KAYNAKLAR**

- [1] Han J, Pei J, Kamber M. Data mining: concepts and techniques. Elsevier, 2011.
- [2] Koh HC, Tan G. Data mining applications in healthcare. Journal of healthcare information management 2011; 65: 19-2.
- [3] Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang JF, Hua L. Data mining in healthcare and biomedicine: a survey of the literature. Journal of medical systems 2012; 36: 2431-2448.
- [4] Tomar D, Agarwal S. A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology 2013; 5: 241-266.
- [5] Romero C, Ventura S. Educational data mining: A survey from 1995 to 2005. Expert systems with applications 2007; 33: 135-146.
- [6] Baker RSJD. Data mining for education. International encyclopedia of education 2010; 7: 112-118.
- [7] Romero C, Ventura S. Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 2013; 3: 12-27.
- [8] Shaw MJ, Subramaniam C, Tan GW, Welge ME. Knowledge management and data mining for marketing. Decision support systems 2001; 31: 127-137.
- [9] Berry MJ, Linoff GS. Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons, 2004.

- [10] Chiang WY. Applying data mining for online CRM marketing strategy: An empirical case of coffee shop industry in Taiwan. *British Food Journal* 2018; 120: 665-675.
- [11] Monkman GG, Kaiser MJ, Hyder K. Text and data mining of social media to map wildlife recreation activity. *Biological conservation* 2018; 228: 89-99.
- [12] Stieglitz S, Mirbabaie M, Ross B, Neuberger C. Social media analytics–Challenges in topic discovery, data collection, and data preparation. *International journal of information management* 2018; 39: 156-168.
- [13] Cunningham SJ, Holmes G. Developing innovative applications in agriculture using data mining. In *The proceedings of the Southeast Asia regional computer confederation conference*; 1999; pp. 25-29.
- [14] Mucherino A, Papajorgji P, Pardalos PM. *Data mining in agriculture*. New York, NY, USA: Springer Science & Business Media, 2009.
- [15] Rajesh D. Application of spatial data mining for agriculture. *International Journal of Computer Applications* 2011; 15: 7-9.
- [16] Şimşek Gürsoy UT. *Veri Madenciliği ve Bilgi Keşfi*. Ankara, TR: Pegem Akademi, 2009.
- [17] Short JC, Broberg JC, Coglisier CC, Brigham KH. Construct validation using computer-aided text analysis (CATA) an illustration using entrepreneurial orientation. *Organizational Research Methods* 2010; 13: 320-347.
- [18] Kobayashi VB, Mol ST, Berkers HA, Kismihók G, Den Hartog DN. Text mining in organizational research. *Organizational research methods* 2018; 21: 733-765.
- [19] Salloum SA, AlHamad AQ, Al-Emran M, Shaalan K. A Survey of Arabic Text Mining. In: Shaalan K, Hassanien A, Tolba F. (eds) *Intelligent Natural Language Processing: Trends and Applications*. Studies in Computational Intelligence. Springer, Cham, 2018. p. 417-431.
- [20] Karanikas H, Theodoulidis B. *Knowledge discovery in text and text mining software*. Manchester: Centre for Research in Information Management, Department of Computation, UMIST, 2002.
- [21] Singh N, Hu C, Roehl WS. Text mining a decade of progress in hospitality human resource management research: Identifying emerging thematic development. *International Journal of Hospitality Management* 2007; 26: 131-147.
- [22] Delen D, Crossland MD. Seeding the survey and analysis of research literature with text mining. *Expert Systems with Applications* 2008; 34: 1707-1720.
- [23] Moro S, Cortez P, Rita P. Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications* 2015; 42: 1314-1324.
- [24] Scherf M, Epple A, Werner T. The next generation of literature analysis: integration of genomic analysis into text mining. *Briefings in bioinformatics* 2005; 6: 287-297.

- [25] Al U, Coştur R. Türk Psikoloji Dergisi'nin bibliyometrik profili. Türk kütüphaneciliği 2007; 21: 142-163.
- [26] Temizkan SP, Çiçek D, Özdemir C. Bibliometric profile of articles published on health tourism. International Journal of Human Sciences 2015; 12: 394-415.
- [27] Van Eck NJ, Waltman L. Visualizing bibliometric networks. In Measuring scholarly impact; 2014; Springer, Cham. pp. 285-320.
- [28] Garfield E. From the science of science to Scientometrics visualizing the history of science with HistCite software. Journal of Informetrics 2009; 3: 173-179.
- [29] Van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2009; 84: 523-538.
- [30] Zvereva OM, Shams SR. Software Support for Team Engineering: Educational Case for IT Students. In 2018 IV International Conference on Information Technologies in Engineering Education; 2018; IEEE. pp. 1-5.
- [31] Van Eck NJ, Waltman L, Dekker R, Van den Berg J. An experimental comparison of bibliometric mapping techniques. In 10th international conference on science and technology indicators; 2008; Vienna.
- [32] Van Eck NJ, Waltman L. VOS: A new method for visualizing similarities between objects. In Advances in data analysis; 2007; Springer, Berlin: Heidelberg. pp. 299-306.
- [33] Van Eck NJ, Waltman L. Text mining and visualization using VOSviewer. ISSI Newsletter 2011; 7: 50-54.
- [34] Perianes-Rodriguez A, Waltman L, Van Eck NJ. Constructing bibliometric networks: A comparison between full and fractional counting. Journal of Informetrics 2016; 10: 1178-1195.
- [35] Yu D, Wang W, Zhang W, Zhang S. A bibliometric analysis of research on multiple criteria decision making. Current Science 2018; 114: 747-758.
- [36] Nascimento KRDS, Alencar MH, Management of risks in natural disasters: A systematic review of the literature on NATECH events. Journal of Loss Prevention in the Process Industries 2016; 44: 347-359.
- [37] Doleck T, Lajoie S. Social networking and academic performance: A review. Education and Information Technologies 2018; 23: 435-465.
- [38] Waltman L, Van Eck NJ, Noyons ECM. A unified approach to mapping and clustering of bibliometric networks. Journal of Informetrics 2010; 4: 629-635.
- [39] Van Eck NJ, Waltman L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. Scientometrics 2017; 111: 1053-1070.