

Development of a Web-Based Application for the Automatic Creation of Concept Maps

Şeyhmus AYDOĞDU*^a, Tolga GÜYER^b

Article Info

DOI: 10.14686/buefad.395735

Article History:

Received: 16.02.2018

Accepted: 14.04.2018

Published: 30.06.2018

Keywords:

Concept Map,
Knowledge Map,
Term Frequency,
Inverse Document Frequency,
Text Mining

Article Type:

Research Article

Abstract

Concept maps are two-dimensional visual tools providing the definition of the relationships between the concepts in a certain subject area. Although various studies indicate that the graphical display of information has a positive contribution to the process of learning, the creation of concept maps in contents especially with a high text density brings some difficulties with it. It is observed that since studies are conducted on the automatic creation of concept maps for the purpose of overcoming these difficulties, but the algorithms used in these kinds of studies are generally developed according to the language structure, the products introduced do not adjust to the language structure of Turkish. From this viewpoint, a web-based application was developed for the creation of concept maps for Turkish texts by using the term frequency and inverse document frequency algorithms and the application developed was published at www.kavramharitasiolustur.com. The concept maps created by the application can be downloaded in a format that can be used directly in web-based learning environments.

Kavram Haritalarının Otomatik Oluşturulmasına Yönelik Web Tabanlı Bir Uygulama Geliştirilmesi

Makale Bilgisi

DOI: 10.14686/buefad.395735

Makale Geçmişi:

Geliş: 16.02.2018

Kabul: 14.04.2018

Yayın: 30.06.2018

Anahtar Kelimeler:

Kavram Haritası,
Bilgi Haritası,
Terim Frekansı,
Devrik Doküman Frekansı,
Metin Madenciliği.

Makale Türü:

Araştırma Makalesi

Öz

Kavram haritaları, belirli bir konu alanında bulunan kavramlar arasındaki ilişkilerin tanımlanmasını sağlayan iki boyutlu görsel araçlardır. Bilgilerin grafiksel olarak gösteriminin öğrenme sürecine olumlu yönde katkısının bulunduğu çeşitli araştırmalarda belirtilmesine karşın, özellikle metin yoğunluğu yüksek olan içeriklerde kavram haritalarının oluşturulması beraberinde bazı güçlükleri getirmektedir. Bu güçlüklerin üstesinden gelmek amacıyla, kavram haritalarının otomatik oluşturulmasına yönelik araştırmaların gerçekleştirildiği, ancak bu tür çalışmalarda kullanılan algoritmalar genellikle dil yapısına özel olarak geliştirildiğinden ortaya konan ürünlerin Türkçe dil yapısı ile uyum göstermediği görülmektedir. Bu noktadan hareketle bu çalışmada, terim frekansı ve devrik doküman frekansı algoritmaları kullanılarak Türkçe metinler için kavram haritalarının oluşturulmasına yönelik web tabanlı bir uygulama geliştirilmiş ve geliştirilen uygulama www.kavramharitasiolustur.com adresinden yayınlanmıştır. Uygulama aracılığıyla oluşturulan kavram haritaları, web tabanlı öğrenme ortamlarında doğrudan kullanılabilir formatta indirilebilmektedir.

*Corresponding Author: aydogduseyhmus@gmail.com

^a Asst. Prof. Dr. Nevşehir Hacı Bektaş Veli University, Nevşehir/Turkey, <https://orcid.org/0000-0002-9075-8055>

^b Prof. Dr., Gazi University, Ankara/Turkey, <https://orcid.org/0000-0001-9175-5043>

Introduction

Concept maps can be defined as a visual presentation of information in a certain area using textual and graphical elements (Novak, 1990). In these maps, there are labels of the relationships between the nodes. For instance, the label between the two nodes in the form of Turkey and Ankara should have a feature that can characterize the relationship between these two concepts. In this case, the relationship expression between these two concepts can be "the capital". Then, the expression of "Ankara is the capital of Turkey." can be stated in this part of the concept map.

The knowledge maps, which were firstly stated by Holley and Dansereau (1984), are created for the purpose of expressing the concepts such as concept maps and the relationships between them. The difference between the knowledge map and concept map is derived from the relationships defined between the nodes. As it is stated above, although the expressions between nodes in the concept maps are certain expressions with a supplementary nature, these relationships in knowledge maps have special expressions such as "example", "part", and sometimes these expressions can be shown by abbreviations.

Graphical displays can be preferred as learning materials in cases when table of contents or index list is inadequate, it is necessary to acquire basic concepts of a subject in a short time and the relationship between concepts in complex issues should be comprehended, as well as the realization of effective learning (Lee & Segev, 2012). In hypermedia, concept maps can be used to support the mental structuring of knowledge and to guide students during navigation (Amadiou & Salmerón, 2014). It is stated in many studies that the use of concept maps positively contributes to learning (Nesbit & Adesope, 2006; Horton, ve diğerleri, 1993) but many difficulties are encountered in creating these maps. The most important of these difficulties are (McNeese & Ayoub, 2011):

- 1. Domain expert level:** It is the level of expertise of the person who will prepare the map about the subject for which concept map will be prepared.
- 2. Smooth designing of the scheme:** It is the designing of the concept map to be prepared to provide a quick understanding of it.
- 3. Integration of Maps:** It is the integration of concept maps developed by different experts.
- 4. Reduction of visual complexity:** It is to simply specify the important concepts and the relationships between them in the prepared concept map.

There are studies on the automatic creation of concept maps and the efficiency of created concept maps to overcome these difficulties in creating concept maps (Tseng, Sue, Su, Weng, & Tsai, 2007; Bai & Chen, 2008; Chen, Wei, & Chen, 2008; Lee, Lee, & Leu, 2009; Lee & Segev, 2012; Watthananon & Mingkhwan, 2012; Arbizu, 2014; Lee, Park, & Yoon, 2015). When automatic creation studies are examined, it is observed that these studies are usually carried out for English texts. In particular, it is important to develop an application for the automatic creation of concept maps in order to analyze Turkish texts, to reveal important concepts of a certain subject area and the relations between these concepts, and to spread the use of concept maps in online learning environments. Besides, it is thought that the outputs produced by the developed application will provide great convenience in the examination of the effect of concept maps in different contexts. From this viewpoint, in this study, an application was developed for the automatic creation of concept maps in Turkish texts. Within the framework of this purpose, studies on the automatic creation of concept maps were examined, a method was developed to analyze Turkish contents, and the application of this developed method was published at www.kavramharitasiolustur.com.

Method

Process of Creating Concept Maps

The following assumptions were created by examining the assumptions determined to create a concept map using the academic articles of Chen et al. (2008) and the process of creating the knowledge map by Lee and Segev (2012) while developing the application for automatic creation of the concept map:

1. A term which is frequently used in a document is an important concept.

2. The presence of two concepts in a document indicates that there is a relationship between these two concepts.
3. The high frequency of the presence of two concepts in a sentence indicates that the relationship between these two concepts is high.
4. The close use of two concepts in a sentence indicates that the relationship between these two concepts is high.

The following steps were followed for the automatic creation of the concept map by taking into consideration the assumptions mentioned above:

1. **Detection of word roots:** At this step, the roots of the derived words such as "araba (car), arabaya (to the car), arabanın (of the car), arabacı (car-man)" used in the document are detected. Thus, the derived words are provided to be calculated under a single root while calculating the significance values of the concepts. The morphological analysis component of the application of ITU Turkish Natural Language Processing Software Chain (Şahin, Sulubacak, & Eryiđit, 2013; Eryiđit, 2014) was used to detect the word roots in the developed application. The texts included in the files in index uploaded by the user are sent to the application in the form of sentences, and the root of the word is determined by separating the responses coming from the application. The text with the shortest letter length is taken as a basis for the words that are detected by the application to possibly have more than one word roots. For instance, the words of "toplum" (society) and "toplu" (collective) are specified as the root for the word "toplum" (society) by the application. At this point, the expression of "toplu" (collective) is taken as a basis as the word root since the number of letters of the word "toplu" (collective) is less compared to the number of letters of the word "toplum" (society). After determining the word roots, sentence cleaning is performed.
2. **Sentence cleaning process:** At this step, the words such as "ve, veya, ki, gibi, de, ol, bu, dolaylı, birçok" ("and, or, that, like, also, be, this, because of, many") included in the sentence are removed from the sentence. In this way, the elimination of the words that are frequently used in the text document but are not important for the subject is performed. A word cancellation module was added to the application as there is no certainty about the content of the texts uploaded by the users in the developed application. Along with this module, the words that are not desired to be included in the analysis process can be determined from among the words in the sentence, and thus more meaningful results can be obtained.
3. **Determination of concept weights:** After determining the word roots, concept weight values are determined to detect the significance of the words determined. The term frequency / inverted document frequency (TF/IDF) algorithm developed by Salton and Buckley (1988) and proposed by Pirrone, Cossentino, Pilato and Rizzo (2003) was used to determine the concept weight values. They proposed that the TF/IDF algorithm could be used to create a concept map. The term frequency can be expressed as the usage percentage of a term or concept in the document. The inverted document frequency is a variable depending on the number of documents containing a term or concept. In other words, while term frequency expresses the concept frequency, the inverted document frequency expresses the distinctiveness of the concept. Depending on the TF/IDF algorithm, the term weighting is performed using the formula (1).

$$a_{ik} = \frac{tf_{ik} \log(N/n_k)}{\sqrt{\sum_{k=1}^t (tf_{ik})^2 [\log(N/n_k)]^2}} \quad (1)$$

a_{ik} : weight of term k in document i

tf_{ik} : term frequency value of term k in document i

N : total number of documents

n_k : number of documents containing term k

t : number of terms

- 4. Determination of the maximum weight value for the concept:** Since each calculated concept will have different weight values in different documents, the maximum value that the concept weight takes in different documents at this step is determined as the weight value of the concept.

$$A_k = \text{Max}(a_{ik}) \quad (2)$$

A_k : weight value of term k

a_{ik} = weight of term k in document i

After calculating the weight values of the concepts, sorting is performed according to the significance level along with the correct order of the concepts from big to small according to the calculated values. At the next step, the relationship values between the concepts will be calculated according to the number of concepts specified at the beginning.

- 5. Calculation of the relationship value between concepts:** The following method was used for the calculation of the relationship value between important concepts specified based on the assumption that "The close use of two concepts in a sentence indicates that the relationship between these two concepts is high" (Lee & Segev, 2012). In this method, as it is understood from the formula (3), the use of concept pairs in short sentences leads to the fact that the relationship is higher compared to long sentences.

$$R_{i,j} = \sum_{D_m} \sum_{S_n} \frac{2}{N_{D_m S_n}} \quad (3)$$

i,j: concept pairs

$R_{i,j}$: relationship value of concepts i and j

m: number of documents

n: number of sentences in the document no. m

S_n : nth sentence

D_m : mth document

$N_{D_m S_n}$: number of words included in mth document and nth sentence

Only the relationship values are indicated in the labels between the concepts as no semantic deduction is made in the application developed within the scope of the study. Furthermore, the thickness of the joining lines between the concepts also varies according to the calculated relationship value.

Structure of the Developed Application

Within the scope of the study, a web-based application was developed for the automatic creation of the above-mentioned concept maps. In the application developed, users were allowed not only to automatically create concept maps but also to be able to save these maps and download them to their computers. In line with these purposes, the application consists of the following modules.

1. **User Management:** This module is needed to allow users to access the system and gain access to the concept maps they have created previously. This module consists of user login, user registration and password retrieval components. The user login screen of the application is presented in Figure 1.

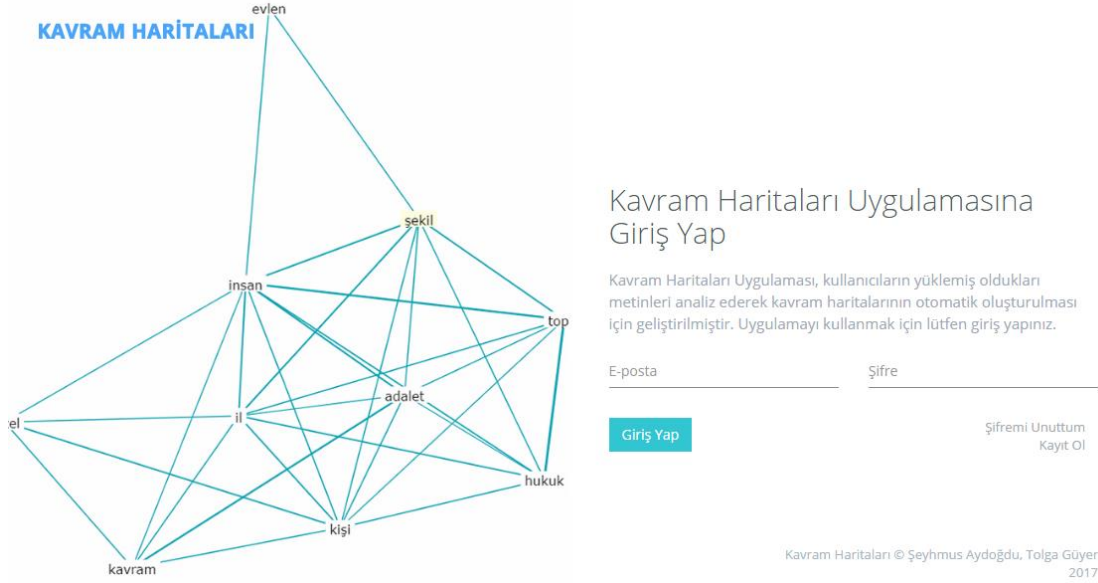


Figure 1. Application login screen

2. **Map Creation Module:** With this module, users can create the concept map by uploading the compressed files in zip format containing the files with txt extension to the system. The screen shot of the map creation module is presented in Figure 2.

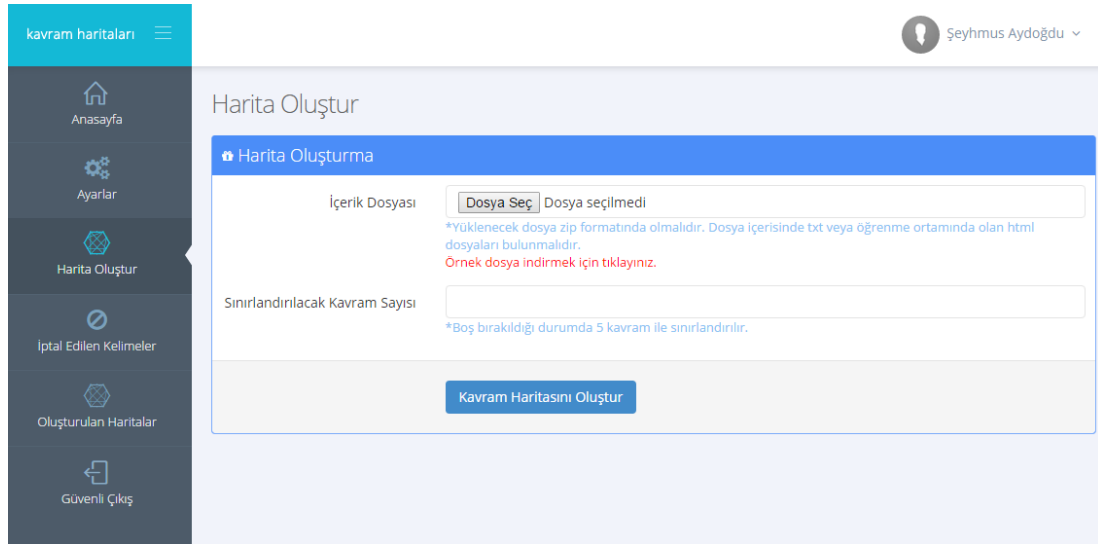


Figure 2. Content uploading screen to create a map

In this module, it is also possible to determine the number of concepts to be limited and how many concepts will be included in the concept map. At this point, sorting is performed according to the concept weight, which is the largest according to the term weighting algorithm.

3. **Cancelled Words Module:** With this module, conjunctions, suffixes or the other words in the sentence that are not intended to be analyzed in the sentence cleaning process determined in the process of creating

a concept map. The screenshot of the module is presented in Figure 3. Here, it is enough to write the bare infinitive of the word which is intended to be cancelled to exclude the word from the analysis.

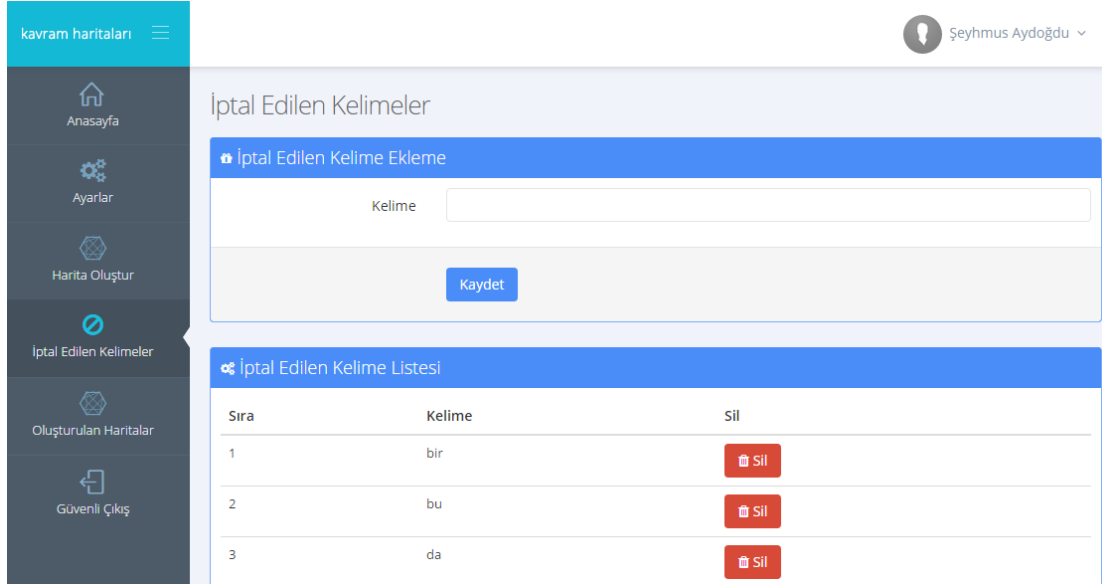


Figure 3. Canceled words screen

4. **Created Maps Module:** With this module, users have the option to display, download and delete the concept maps they have created previously. The screenshot of the created maps module is presented in Figure 4.

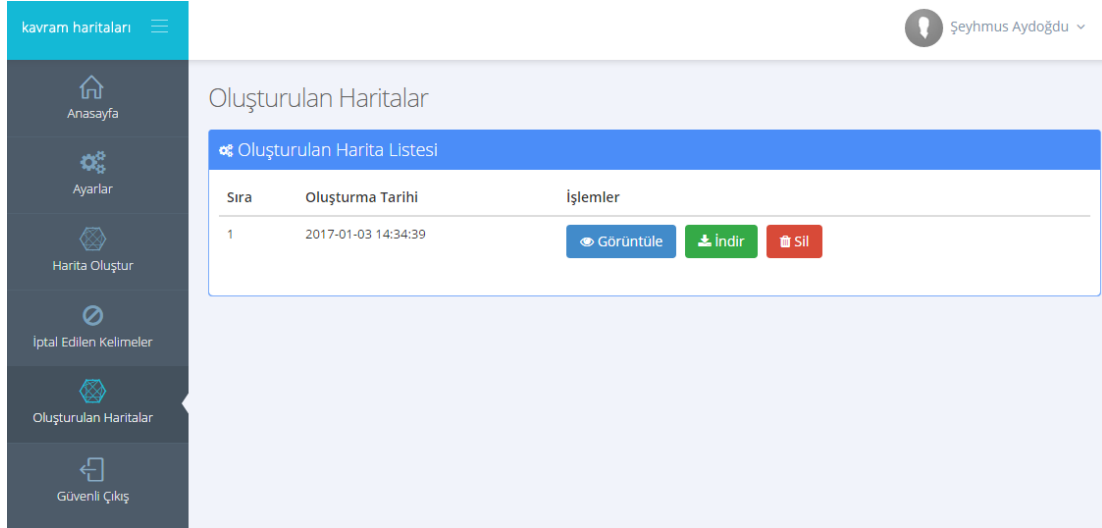


Figure 4. Created maps screen

The display of a created concept map is as shown in Figure 5. A concept map which was created with the contents for the legal field from Wikipedia web page and was limited by 6 concepts is illustrated in this example.

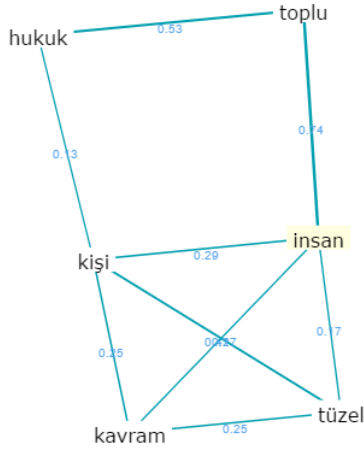


Figure 5. A sample concept map view

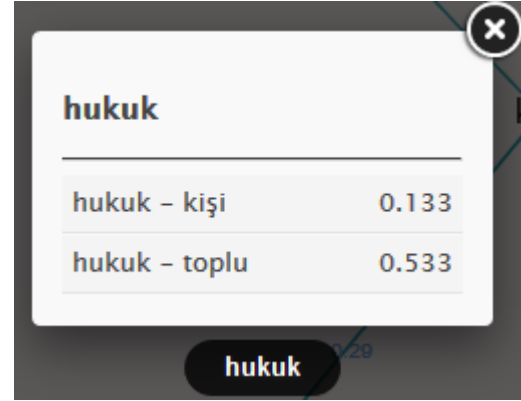


Figure 6. A sample concept map view

When the concept is clicked, the list of the clicked concept with the other concepts is displayed on the screen (Figure 6). Then, when these relationships are clicked, the texts containing these concepts and the information of in which file they are located are displayed on the screen (Figure 7).

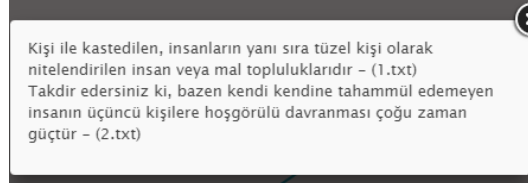


Figure 7. List of the texts containing the concepts

Thus, the basic concepts of a subject, the relationships between them and the use of these concepts in the text can be presented to the reader.

Conclusion and Suggestions

In this study, a web-based application was developed for the automatic creation of a concept map by analyzing Turkish texts in a certain subject area. In the application developed, the steps of detection of word roots, sentence cleaning, the weighting of concepts and the determination of relationship values between concepts were followed in the process of creating a concept map.

The morphological analysis component of the application of ITU Turkish Natural Language Processing Software Chain application developed by Şahin, Sulubacak and Eryiđit (2013) and Eryiđit (2014) was used to determine the word roots. In the application, the word with the shortest text length is determined as the word root for the multiple word roots proposed. To decide the word root by taking into account the semantic use of the word in the sentence at the next steps may allow for the more accurate creation of the concept map.

The frequency of the concepts is taken as a basis in the method followed in the application which was developed within the scope of the study. Therefore, the sentences that are not used in the sentence but in which any concept is specified are excluded from the analysis. In language knowledge, the text integrity is expressed by the concept of coherence. Ensuring integrity of the relationships between the words is defined as lexical coherence. The types of ensuring lexical coherence are presented below (Keçik & Uzun, 2003):

- Repetition of a word
- Use of different word types derived from the same root
- Use of synonyms or near-synonymous words
- Use of words involving upper term-low significance relationship

- Use of words from the same concept area

The words that are repeated in terms of lexical coherence and different words derived from the same root can be included within the scope of the analysis as the concept roots are taken into account in the application developed. In the subsequent studies, it is possible to ensure creating concept maps more effectively by taking into account the use of words in the field of synonymousness, upper term-low significance and same concept specified in lexical coherence types.

In the developed application, the concept map is created by analyzing the text files with the txt extension uploaded by the users. In the study carried out by Chen, Wei and Chen (2008), it was analyzed by text mining methods in academic articles. In subsequent studies, the files with pdf and doc extension which are commonly used in electronic media can be processed and Turkish articles in a certain subject area can be analyzed.

Acknowledgments

This study was produced from doctoral dissertation titled “The effect of digital concept maps on students’ achievements and disorientation in online learning environments”.

Kavram Haritalarının Otomatik Oluşturulmasına Yönelik Web Tabanlı Bir Uygulama Geliştirilmesi

Giriş

Kavram haritaları (concept maps), belirli bir alandaki bilgilerin metinsel ve grafiksel elemanlar kullanılarak görsel olarak sunulması şeklinde tanımlanabilir (Novak, 1990). Bu haritalarda düğümler arasında yer alan ilişkilere ait etiketler bulunmaktadır. Örneğin, Türkiye ve Ankara şeklinde iki düğüm arasındaki etiket bu iki kavram arasındaki ilişkiyi nitelendirebilecek özellikte olmalıdır. Bu durumda bu iki kavram arasındaki ilişki ifadesi “başkentidir” olabilir. Bu durumda kavram haritasının bu bölümünde “Ankara, Türkiye'nin başkentidir.” ifadesi belirtilebilir.

İlk olarak Holley ve Dansereau (1984) tarafından belirtilen bilgi haritaları (knowledge maps), kavram haritaları gibi kavramlar ve bunlar arasındaki ilişkilerin belirtilmesi amacıyla oluşturulmaktadır. Bilgi haritası ile kavram haritası arasındaki farklılık, düğümler arasındaki tanımlanan ilişkilerden kaynaklanmaktadır. Yukarıda belirtildiği gibi kavram haritalarında düğümler arasındaki ifadeler tamamlayıcı nitelikte herhangi bir ifade olmasına karşın bilgi haritalarında bu ilişkiler “örnek”, “parçası” gibi özel ifadelerle sahiptir ve bazen bu ifadeler kısaltmalar ile gösterilebilir.

Etkili öğrenmenin gerçekleşmesinin yanısıra içindekiler veya indeks listesinin yetersiz olduğu, kısa zamanda bir konuya ilişkin temel kavramların edinilmesi ve karmaşık konularda kavramlar arasındaki ilişkinin kavranması gerektiği durumlarda grafiksel gösterimler öğrenme materyali olarak tercih edilebilmektedir (Lee & Segev, 2012). Hiper ortamlarda kavram haritaları bilgilerin zihinsel olarak yapılandırılmasını desteklemek ve gezinmede öğrencilere rehberlik amacıyla kullanılabilir (Amadiou & Salmerón, 2014). Kavram haritaları kullanımının öğrenmeye olumlu yönde katkısı olduğu birçok çalışmada belirtilmektedir (Nesbit & Adesope, 2006; Horton, ve diğerleri, 1993) fakat bu haritaların oluşturulmasında birçok zorlukla karşılaşmaktadır. Bu zorluklardan en önemlileri (McNeese & Ayoub, 2011):

- 1. Alan uzmanı düzeyi:** Kavram haritası hazırlanacak konuda, haritayı hazırlayacak kişinin uzmanlık düzeyidir.
- 2. Şemanın akıcı şekilde tasarlanması:** Hazırlanacak kavram haritasının hızlı bir şekilde anlaşılmasını sağlayacak şekilde tasarlanmasıdır.
- 3. Haritaların bütünleştirilmesi:** Farklı uzmanlar tarafından geliştirilen kavram haritalarının bütünleştirilmesidir.
- 4. Görsel karmaşıklığın azaltılması:** Hazırlanan kavram haritasında önemli kavramlar ve aralarındaki ilişkilerin belirtilmesinin sade bir şekilde gerçekleştirilmesidir.

Kavram haritası oluşturmada bu zorlukları gidermek amacıyla kavram haritalarının otomatik oluşturulmasına ve oluşturulan kavram haritalarının verimliliğine yönelik araştırmalar bulunmaktadır (Tseng, Sue, Su, Weng, & Tsai, 2007; Bai & Chen, 2008; Chen, Wei, & Chen, 2008; Lee, Lee, & Leu, 2009; Lee & Segev, 2012; Watthananon & Mingkhwan, 2012; Arbizu, 2014; Lee, Park, & Yoon, 2015). Otomatik oluşturma araştırmaları incelendiğinde bu araştırmaların genellikle İngilizce metinlere yönelik gerçekleştirildiği görülmektedir. Özellikle, Türkçe metinlerin analizinin yapılması, belirli bir konu alanındaki önemli kavramların ve bu kavramlar arasındaki ilişkilerin ortaya konulması ve çevrimiçi öğrenme ortamlarında kavram haritalarının kullanımının yaygınlaşması açısından kavram haritalarının otomatik oluşturulmasına yönelik bir uygulamanın geliştirilmesi önem taşımaktadır. Bunun yanı sıra, geliştirilen uygulama ile üretilen çıktılar farklı bağlamlarda kavram haritalarının etkisinin incelenmesi açısından büyük kolaylık sağlayacağı düşünülmektedir. Bu noktadan hareketle bu araştırmada Türkçe metinlerde kavram haritalarının otomatik oluşturulması amacıyla bir uygulama geliştirilmiştir. Bu amaç çerçevesinde kavram haritalarının otomatik oluşturulmasına yönelik araştırmalar incelenmiş, Türkçe içeriklerin analiz edilmesine yönelik bir yöntem geliştirilmiş ve geliştirilen bu yöntemin uygulaması www.kavramharitasiolustur.com adresinden yayınlanmıştır.

Yöntem

Kavram Haritalarının Oluşturulması Süreci

Kavram haritasının otomatik oluşturulmasına yönelik uygulama geliştirilirken Chen vd. (2008) tarafından akademik makalelerden yararlanarak kavram haritası oluşturmak için belirlenen varsayımlar ve Lee ve Segev (2012) tarafından bilgi haritası oluşturma süreci incelenerek aşağıda belirtilen varsayımlar oluşturulmuştur:

1. Bir belge içerisinde sık kullanılan bir terim önemli bir kavramdır.

2. Bir belge içerisinde iki kavramın bulunması bu iki kavram arasında bir ilişki olduğunu belirtir.
3. Bir cümle içerisinde iki kavramın bulunma sıklığının yüksek olması, bu iki kavram arasındaki ilişkinin yüksek olduğunu belirtir.
4. Bir cümle içerisinde iki kavramın birbirine yakın kullanımı, bu iki kavram arasındaki ilişkinin yüksek olduğunu belirtir.

Yukarıda belirtilen varsayımlar göz önünde bulundurularak, kavram haritasının otomatik oluşturulması için aşağı belirtilen aşamalar izlenmiştir:

1. **Kelime köklerinin tespiti:** Bu aşamada, belge içerisinde kullanılan “araba, arabaya, arabanın, arabacı” gibi türetilmiş kelimelerin kökleri tespit edilmektedir. Bu sayede kavramların önem değerleri hesaplanırken türetilmiş kelimelerin tek bir kök altında hesaplanması sağlanmaktadır. Geliştirilen uygulamada kelime köklerinin tespit edilmesi amacıyla ITU Türkçe Doğal Dil İşleme Yazılım Zinciri (Şahin, Sulubacak, & Eryiğit, 2013; Eryiğit, 2014) uygulamasının morfolojik analiz bileşeninden yararlanılmıştır. Kullanıcı tarafından yüklenen dizin içerisinde dosyalarda yer alan metinler cümleler halinde uygulamaya gönderilmekte ve uygulamadan gelen cevaplar ayrıştırılarak kelimenin kökü belirlenmektedir. Uygulama tarafından birden fazla kelime kökü olabileceği tespit edilen kelimeler için harf uzunluğu en kısa metin esas alınmaktadır. Örneğin, “toplum” kelimesi için uygulama tarafından “toplum” ve “topl” kelimeleri kök olarak belirlenmektedir. Bu noktada “topl” kelimesinin harf sayısı “toplum” kelimesinin harf sayısına göre daha az olduğundan kelime kökü olarak “topl” ifadesi esas alınmaktadır. Kelime köklerinin belirlenmesinin ardından cümle temizleme işlemi gerçekleştirilmektedir.
2. **Cümle temizleme işlemi:** Bu aşamada, cümle içerisinde bulunan “ve, veya, ki, gibi, de, ol, bu, dolayı, birçok” gibi kelimeler cümle içerisinden temizlenmektedir. Bu sayede metin belgesi içerisinde sık kullanılan fakat konu açısından önemli olmayan kelimelerin eleme işlemi gerçekleştirilmektedir. Geliştirilen uygulamada kullanıcıların yüklemiş oldukları metinlerin içeriğine ilişkin herhangi bir kesinlik olmadığından dolayı uygulamaya kelime iptal etme modülü eklenmiştir. Bu modül ile cümle içerisinde yer alan kelimelerden analiz işlemine dahil edilmesi istenmeyen kelimeler belirlenebilmekte ve bu sayede daha anlamlı sonuçlar elde edilebilmektedir.
3. **Kavram ağırlıklarının belirlenmesi:** Kelime köklerinin belirlenmesinin ardından belirlenen kelimelerin önemini tespit edilmesi için kavram ağırlık değerleri belirlenmektedir. Kavram ağırlık değerlerinin belirlenmesi için Salton ve Buckley (1988) tarafından geliştirilen ve Pirrone, Cossentino, Pilato ve Rizzo (2003) tarafından önerilen terim frekans / devrik doküman frekansı (TF/IDF) algoritmasından yararlanılmıştır. TF/IDF algoritmasının kavram haritası oluşturmak için kullanılabileceğini önermişlerdir. Terim frekansı bir terimin veya kavramın belge içerisinde geçme yüzdesi olarak ifade edilebilir. Devrik doküman frekansı ise bir terimi veya kavramı içeren doküman sayısına bağlı bir değişkendir. Başka bir deyişle terim frekansı kavram sıklığını ifade ederken, devrik doküman frekansı kavramın ayırt ediciliğini ifade etmektedir. TF/IDF algoritmasına bağlı olarak terim ağırlıklandırma formül (1) kullanılarak yapılmaktadır.

$$a_{ik} = \frac{tf_{ik} \log(N/n_k)}{\sqrt{\sum_{k=1}^t (tf_{ik})^2 [\log(N/n_k)]^2}} \quad (1)$$

a_{ik} : k teriminin i dokümanındaki ağırlığı

tf_{ik} : k teriminin i dokümanındaki terim frekans değeri

N : toplam doküman sayısı

n_k : k terimini içeren doküman sayısı

t : terim sayısı

4. **Kavram için maksimum ağırlık değerinin belirlenmesi:** Hesaplanan her kavram farklı belgelerde farklı ağırlık değerlerine sahip olacağından bu aşamada kavram ağırlığının farklı belgelerde aldığı maksimum değer, kavramın ağırlık değeri olarak belirlenir.

$$A_k = \text{Max}(a_{ik}) \quad (2)$$

A_k : k teriminin ağırlık değeri

a_{ik} = k teriminin i dokümanındaki ağırlığı

Kavramların ağırlık değerlerinin hesaplanmasının ardından, hesaplanan değerlere göre kavramların büyükten küçüğe doğru sıralanması ile önem derecesine göre sıralama gerçekleştirilir. Sonraki aşamada başlangıçta belirlenen kavram sayısı miktarına göre kavramlar arasındaki ilişki değerleri hesaplanacaktır.

5. **Kavramlar arasındaki ilişki değerinin hesaplanması:** “Bir cümle içerisinde iki kavramın birbirine yakın kullanımı, bu iki kavram arasındaki ilişkinin yüksek olduğunu belirtir.” varsayımına dayalı olarak belirlenen önemli kavramlar arasındaki ilişki değerinin hesaplanması için aşağıda belirtilen yöntem kullanılmıştır (Lee & Segev, 2012). Bu yöntemde formül (3)’den de anlaşılacağı gibi kavram çiftlerinin kısa cümlelerde kullanımı, uzun cümlelere göre ilişkinin daha yüksek olmasını beraberinde getirmektedir.

$$R_{i,j} = \sum_{D_m} \sum_{S_n} \frac{2}{N_{D_m S_n}} \quad (3)$$

i,j: kavram çiftleri

$R_{i,j}$: i ve j kavramlarına ait ilişki değeri

m: belge sayısı

n: m nolu belgede yer alan cümle sayısı

S_n : n. cümle

D_m : m. belge

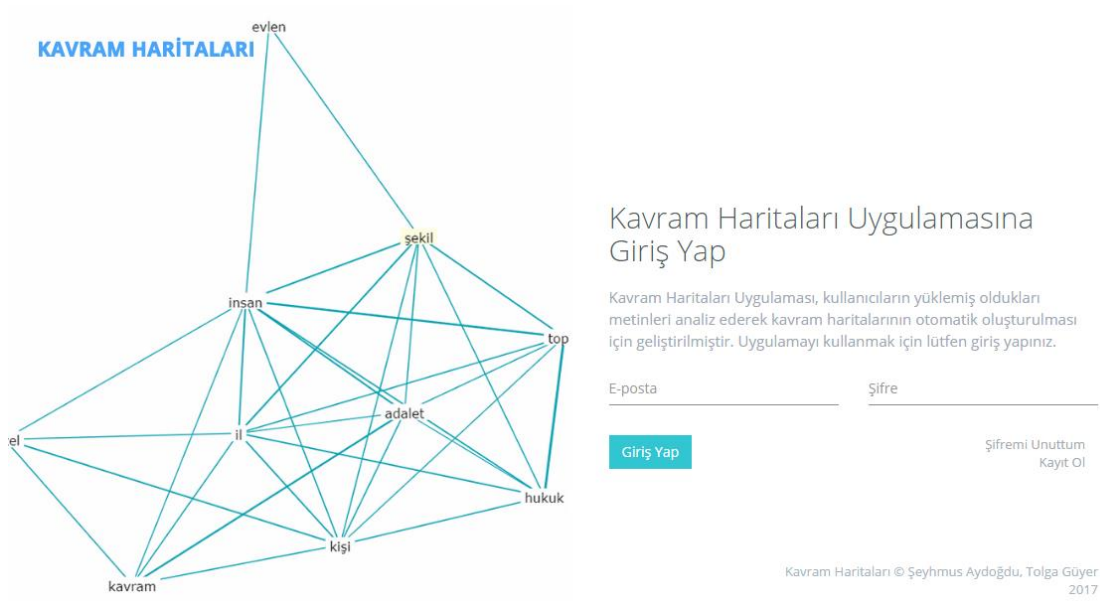
$N_{D_m S_n}$: m. belgede ve n. cümlede yer alan kelime sayısı

Araştırma kapsamında geliştirilen uygulamada anlamsal bir çıkarım yapılmadığından kavramlar arasındaki etiketlerde sadece ilişki değerleri belirtilmektedir. Ayrıca kavramlar arasındaki birleştirme çizgilerinin kalınlığı da hesaplanan ilişki değerine göre değişmektedir.

Geliştirilen Uygulamanın Yapısı

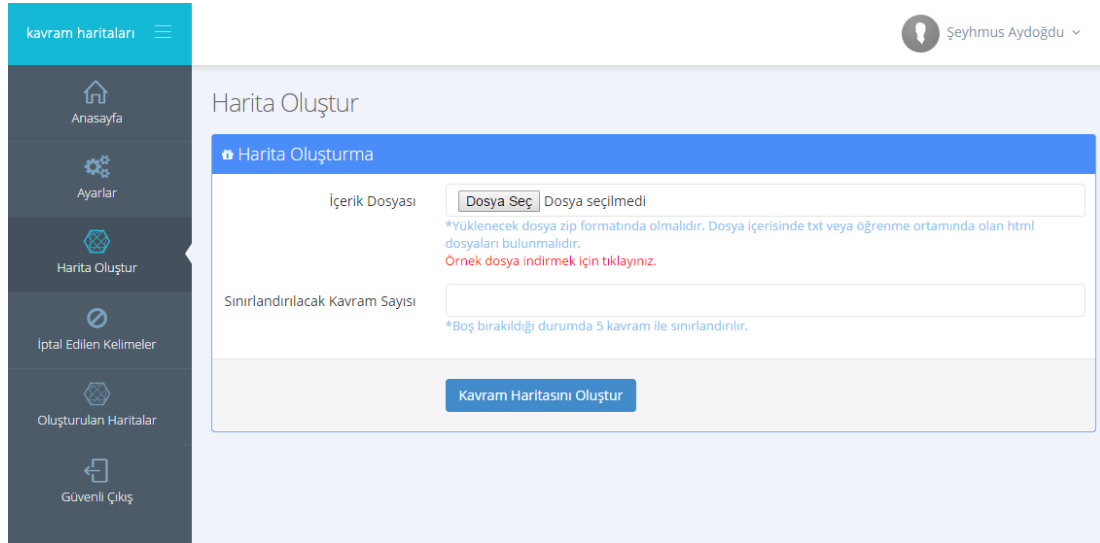
Araştırma kapsamında yukarıda belirtilen kavram haritalarının otomatik oluşturulmasına yönelik web tabanlı bir uygulama geliştirilmiştir. Geliştirilen uygulamada kullanıcıların sadece kavram haritasını otomatik oluşturmalarına değil aynı zamanda bu haritaları kaydetmelerine ve bilgisayarlarına indirebilmelerine olanak sağlanmıştır. Bu amaçlar doğrultusunda uygulama aşağıdaki modüllerden oluşmaktadır.

1. **Kullanıcı Yönetimi:** Kullanıcıların sisteme giriş yapmaları ve daha önce oluşturmuş oldukları kavram haritalarına erişim sağlamaları amacıyla bu modüle ihtiyaç duyulmaktadır. Bu modül kullanıcı girişi, kullanıcı kaydı ve şifre hatırlatma bileşenlerinden oluşmaktadır. Şekil 1 de uygulamanın kullanıcı girişi ekranına yer ver verilmiştir.



Şekil 1. Uygulama giriş ekranı

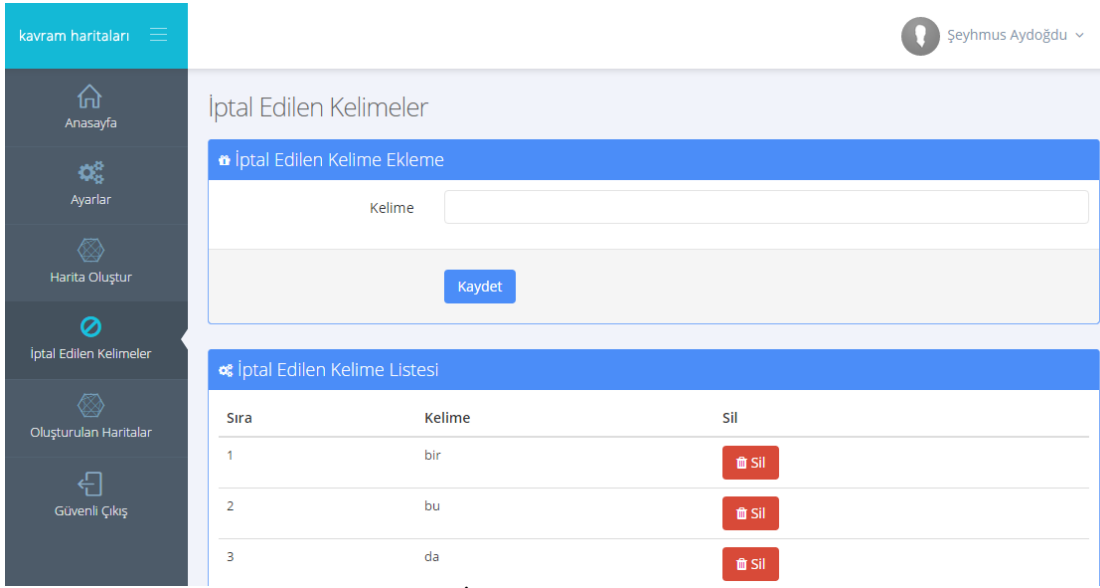
- Harita Oluşturma Modülü:** Bu modül ile kullanıcılar sisteme txt uzantılı dosyaları içeren zip formatındaki sıkıştırılmış dosyaları yükleyerek kavram haritasını oluşturabilmektedirler. Şekil 2 de harita oluşturma modülüne ilişkin ekran görüntüsü verilmiştir.



Şekil 2. Harita oluşturmak için içerik yükleme ekranı

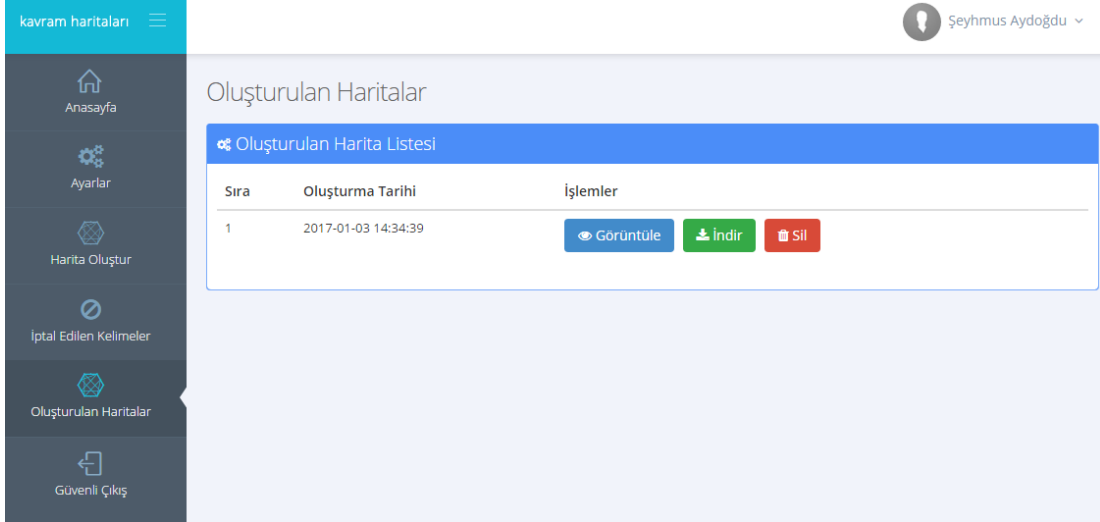
Bu modülde ayrıca sınırlandırılacak kavram sayısı ile kavram haritasında kaç tane kavramın yer alacağı da belirtilebilmektedir. Bu noktada kavram sayısı sınırlandırılırken terim ağırlıklandırma algoritmasına göre en büyük olan kavram ağırlığına göre sıralama yapılmaktadır.

- İptal Edilen Kelimeler Modülü:** Bu modül ile kavram haritası oluşturma sürecinde belirtilen cümle temizleme işlemindeki bağlaçlar, ekler veya cümlede analiz edilmesi istenmeyen diğer kelimeler tanımlanabilmektedir. Şekil 3 de modülün ekran görüntüsü verilmiştir. Burada iptal edilmek istenen kelimenin yalın halinin yazılması kelimenin analiz dışında tutulması için yeterlidir.



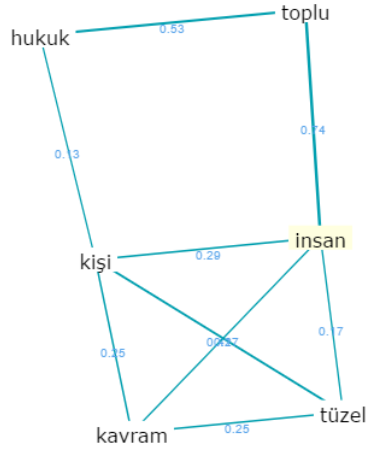
Şekil 3. İptal edilen kelimeler ekranı

4. **Oluşturulan Haritalar Modülü:** Bu modül ile kullanıcılar daha önce oluşturdukları kavram haritalarını görüntüleme, indirme ve silme seçeneklerine sahiptir. Şekil 4 de oluşturulan haritalar modülünün ekran görüntüsü verilmiştir.

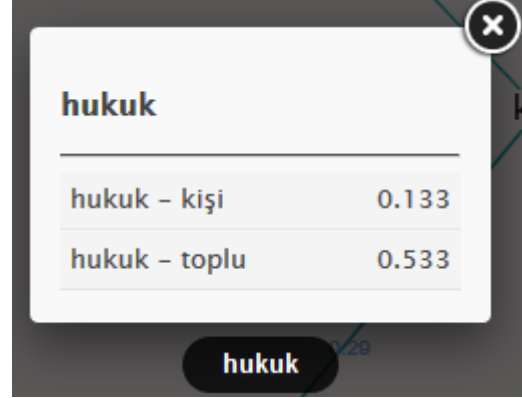


Şekil 4. Oluşturulan haritalar ekranı

Oluşturulan bir kavram haritasının görüntüsü Şekil 5’de görüldüğü gibidir. Bu örnekte wikipedia web sayfasından hukuk alanına yönelik içeriklerle oluşturulan ve 6 kavramla sınırlandırılan bir kavram haritası örneği verilmiştir.

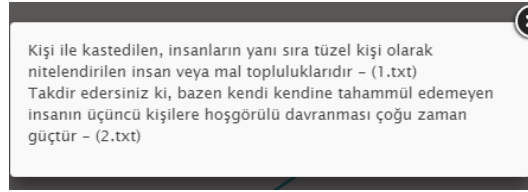


Şekil 5. ˆrnek bir kavram haritası grnm



Şekil 6. ˆrnek bir kavram haritası grnm

Kavram zerine tıkladıđında tıklanan kavramın diđer kavramlar ile olan listesi ekrana gelmektedir (Şekil 6). Daha sonra bu iliřkiler zerine tıkladıđında bu kavramları ieren metinler ve hangi dosya ierisinde yer aldıđı bilgisi ekrana gelmektedir (Şekil 7).



Şekil 7. Kavramları ieren metinlerin listesi

Bu sayede, bir konuya iliřkin temel kavramlar, bunlar arasındaki iliřkiler ve bu kavramların metin ierisinde kullanımı okuyucuya sunulabilmektedir.

Sonuç ve ˆneriler

Bu alıřmada, belirli bir konu alanındaki Trke metinler analiz edilerek kavram haritasının otomatik oluřturulmasına ynelik web tabanlı bir uygulama geliřtirilmiřtir. Geliřtirilen uygulamada kavram haritası oluřturma srecinde kelime kklerinin tespiti, cmle temizlenmesi, kavramların ađırlıklandırılması ve kavramlar arasındaki iliřki deđerlerinin belirlenmesi ařamaları izlenmiřtir.

Kelime kklerinin belirlenmesi amacıyla řahin, Sulubacak ve Eryiđit (2013) ve Eryiđit (2014) tarafından geliřtirilen ITU Trke Dođal Dil İřleme Yazılım Zinciri uygulamasının morfolojik analiz bileřeni kullanılmıřtır. Uygulamada birden fazla ˆnerilen kelime kkleri iin metin uzunluđu en kısa olan kelime, kelime kk olarak belirlenmektedir. Sonraki arařtırmalarda kelimenin cmle ierisinde anlamsal kullanımı dikkate alınarak kelime kkne karar verilmesi daha dođru kavram haritası oluřturulmasını sađlayabilir.

alıřma kapsamında geliřtirilen uygulamada izlenen yntemde kavramların sıklıđı esas alınmaktadır. Bu nedenle, cmle ierisinde kullanılmayan fakat herhangi bir kavramın belirtildiđi cmleler analiz dıřında kalmaktadır. Dil bilgisinde metnin btnlđu bađdařıklık kavramı ile ifade edilmektedir. Szckler arasındaki iliřkiler arasında btnlđu sađlanması szcksel bađdařıklık olarak tanımlanmaktadır. Szcksel bađdařıklıđı sađlama trleri ařađıda verilmiřtir (Keik & Uzun, 2003):

- Bir szcđn yinelenmesi
- Aynı kkten tremiř farklı szck trlerinin kullanımı

- Eő anlamlı veya yakın anlamlı sözcüklerin kullanımı
- Üst terim-alt anlamlılık iliřkisi içeren sözcüklerin kullanımı
- Aynı kavram alanından sözcüklerin kullanımı

Geliřtirilen uygulamada kavram kökleri dikkate alındığından sözcüksel bağdařıklık açısından yinelenen sözcükler ve aynı kökten türetilmiř farklı sözcükler analiz kapsamına dahil edilebilmektedir. Sonraki arařtırmalarda sözcüksel bağdařıklık türlerinde belirtilen eő anlamlılık, üst terim-alt anlamlılık ve aynı kavram alanındaki sözcüklerin kullanımı dikkate alınarak kavram haritalarının daha etkili bir şekilde oluřturulması sağlanabilir.

Geliřtirilen uygulamada, kullanıcıların yüklemiř oldukları txt uzantılı metin dosyaları analiz edilerek kavram haritası oluřturulmaktadır. Chen, Wei ve Chen (2008) tarafından gerçekteřirilen arařtırmada akademik makalelerde metin madencilięi yöntemleri ile analiz edilmiřtir. Sonraki arařtırmalarda elektronik ortamlarda genellikle kullanılan pdf ve doc uzantılı dosyaların iřlenmesi sağlanarak belli bir konu alanındaki Türkçe makaleler analiz edilebilir.

Teőekkür ve Bilgilendirme

Bu çalıřma “Çevrim içi öğrenme ortamlarında dijital kavram haritalarının öğrencilerin başarılarına ve kaybolmalarına etkisi” başlıklı doktora tezinden üretilmiřtir.

References

- Amadiou, F., & Salmern, L. (2014). Concept Maps for Comprehension and Navigation of Hypertexts. In D. Ifenthaler, & R. Hanewald, *Digital Knowledge Maps in Education: Technology-Enhanced Support for Teachers and Learners* (pp. 41-59). New York: Springer Science & Business Media. doi:http://dx.doi.org/10.1007/978-1-4614-3178-7_3
- Arbizu, A. V. (2014). Extracting knowledge from documents to construct concept maps. *Doctoral Thesis*. Indiana University.
- Bai, S., & Chen, S. (2008). Automatically constructing concept maps based on fuzzy rules for adapting learning systems. *Expert systems with Applications*, 35(1), 41-49. doi:https://doi.org/10.1016/j.eswa.2007.06.013
- Chen, N., Wei, C., & Chen, H. (2008). Mining e-learning domain concept map from academic articles. *Computers & Education*, 50(3), 1009-1021. doi:https://doi.org/10.1016/j.compedu.2006.10.001
- Eryiđit, G. (2014). ITU Turkish NLP Web Service. *Proceedings of the Demonstrations at the 14th Conference*. Gothenburg, Sweden.
- Holley, C. D., & Dansereau, D. F. (1984). *Spatial learning strategies: Techniques, applications, and related issues*. New York: NY: Academic.
- Horton, P. B., McConney, A. A., Gallo, M., Woods, A. L., Senn, G. J., & Hamelin, D. (1993). An investigation of the effectiveness of concept mapping as an instructional tool. *Science Education*, 77(1), 95-111. doi:http://dx.doi.org/10.1002/sce.3730770107
- Keik, İ., & Uzun, L. (2003). *Trke Yazılı ve Szl Anlatım. [Turkish Written and Oral Expression]* OrEskiřehir: Anadolu niversitesi Yayınları.
- Lee, C., Lee, G., & Leu, Y. (2009). Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning. *Expert Systems with Applications*, 36(2), 1675-1684. doi:https://doi.org/10.1016/j.eswa.2007.11.049
- Lee, J., & Segev, A. (2012). Knowledge maps for e-learning. *Computers & Education*, 59(2), 353-364. doi:https://doi.org/10.1016/j.compedu.2012.01.017
- Lee, S., Park, Y., & Yoon, W. C. (2015). Burst analysis for automatic concept map creation with a single document. *Expert Systems with Applications*, 42(22), 8817-8829. doi:https://doi.org/10.1016/j.eswa.2015.07.034
- McNeese, M. D., & Ayoub, P. J. (2011). Concept mapping in the analysis and design of cognitive systems: A historical review. In B. M. Moon, R. R. Hoffman, J. D. Novak, & A. J. Cañas, *Applied Concept Mapping: Capturing, Analyzing, and Organizing Knowledge* (pp. 47-65). CRC.
- Nesbit, J., & Adesope, O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, 413-448. doi:http://dx.doi.org/10.3102%2F00346543076003413
- Novak, J. D. (1990). Concept maps and vee diagrams: Two metacognitive tools to facilitate meaningful learning. *Instructional Science*, 19(1), 29-52. doi:http://dx.doi.org/10.1007/BF00377984
- Pirrone, R., Cossentino, M., Pilato, G., & Rizzo, R. (2003). Concept Maps and Course Ontology: a Multilevel Approach to E-learning. *II AI* IA Workshop on AI and E-learning*, (pp. 23-26).
- Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management*, 24(5), 513-523. doi:https://doi.org/10.1016/0306-4573(88)90021-0
- řahin, M., Sulubacak, U., & Eryiđit, G. (2013). Redefinition of Turkish Morphology. *Proceedings of the Tenth Symposium on Natural Language Processing (SNLP-2013)*. Phuket, Thailand.
- Tseng, S., Sue, P., Su, J., Weng, J., & Tsai, W. (2007). A new approach for constructing the concept map. *Computers & Education*, 49(3), 691-707. doi:https://doi.org/10.1016/j.compedu.2005.11.020
- Wathananon, J., & Mingkhwan, A. (2012). Optimizing knowledge management using knowledge map. *Procedia Engineering*, 32, 1169-1177. doi:https://doi.org/10.1016/j.proeng.2012.02.073