

First Stage of an Automated Content-Based Citation Analysis Study: Detection of Citation Sentences

Zehra Taşkın, Umut Al & Umut Sezen

{ztaskin, umutal, u.sezen}@hacettepe.edu.tr

Plan

- ❑ Need for “content-based citation analysis (CBCA)”
- ❑ An application of CBCA
- ❑ Methodology
- ❑ Findings
- ❑ Conclusion

Traditional Citation Analysis

- ❑ Research/er evaluations
 - ❑ Counting citations
 - ❑ Citation manipulations
 - ❑ Perfunctory citations
 - ❑ “Read before cite”



Citation Cartels: The Mafia

Citation Cartel Or Editor Gone Rogue?



THE WEEK



- Subscribe
- Give a gift
- Digital subscription

The science world is plagued by 'citation cartels'

FROM

STAT

Ivan Oransky and Adam Marcus

By IVAN ORANSKY @ivanoransky and ADAM MARCUS @armarcus / JANUARY 13, 2017



ded, and
makes
is one or
arly, the
unity; iii)
e these
ere
the time

ntly in

indulges
n into
stitutions
develop
onal peer
here is

One way to boost your uni's ranking: Ask faculty to cite each other

with 15 comments

Readers who follow scientific publishing will know the term “citation stacking” — as a profile-boosting technique, we’ve seen [journals ask authors to cite them](#), and individual scientists [work together to cite each other, forming “citation cartels.”](#) And now, we’ve seen a university do it.



“

All Academic Staff

Faculty of Engineering

Sir/Madam,

Key Performance Indicators (KPI) Confirmation for 2017

Please refer to the subject-matter stated above.

2. As had been informed earlier, the KPI Confirmation for all University of Malaya staff has been opened and the final date for the KPI Confirmation is 9 August 2017.
3. For the Academic Staff at the Faculty of Engineering, the First Appraisal Officer (PPP) has determined the KPI for each departmental staff (please refer to the KPI). Under Section 6, Faculty Specific Duties you are required to type:

(1) “Citation: To cite at least 3 relevant papers of colleagues in each of your publication”

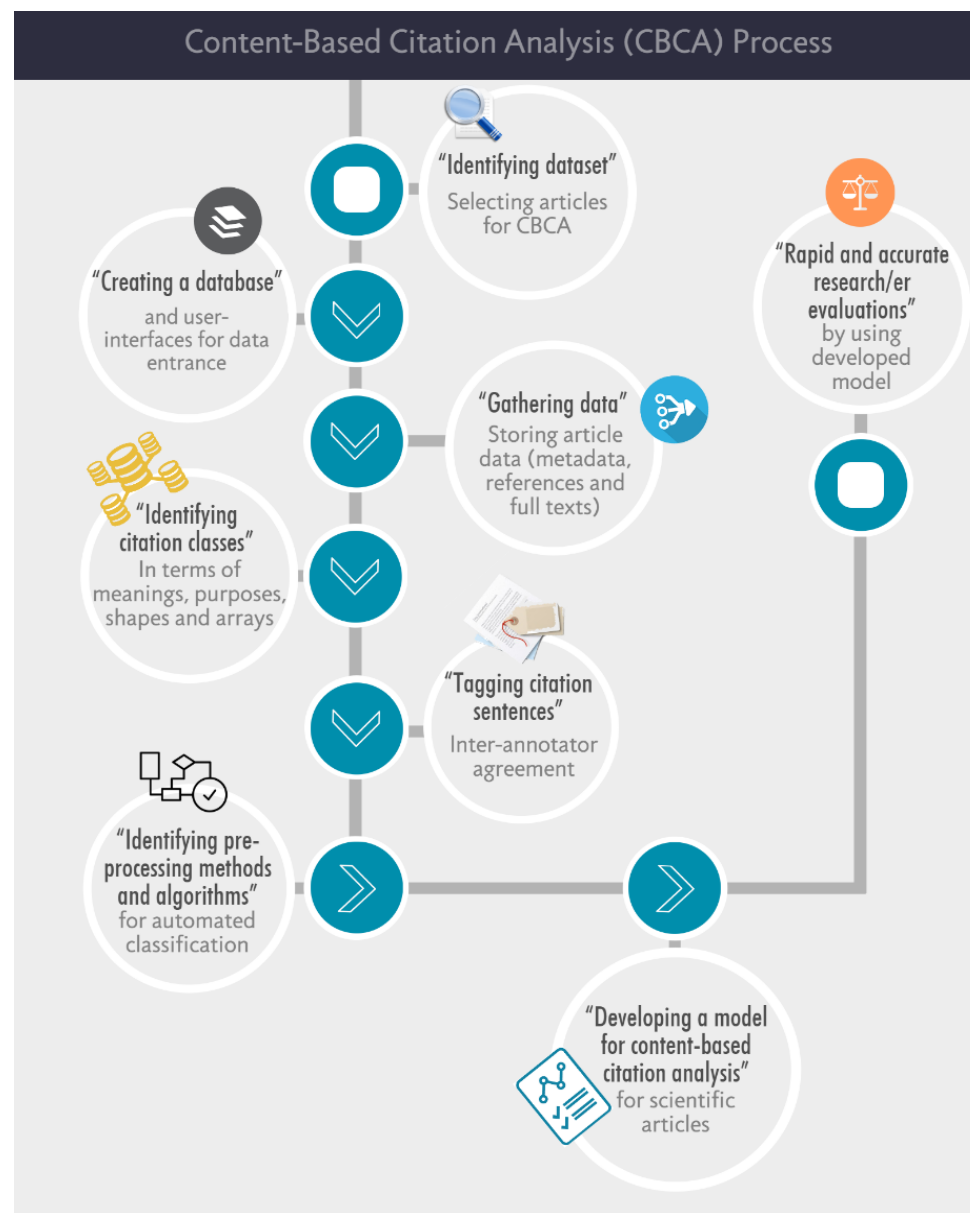
4. Other additional tasks are subject to the PPP of each PYD.

I would be pleased if you could take necessary action before **9 August 2017.**

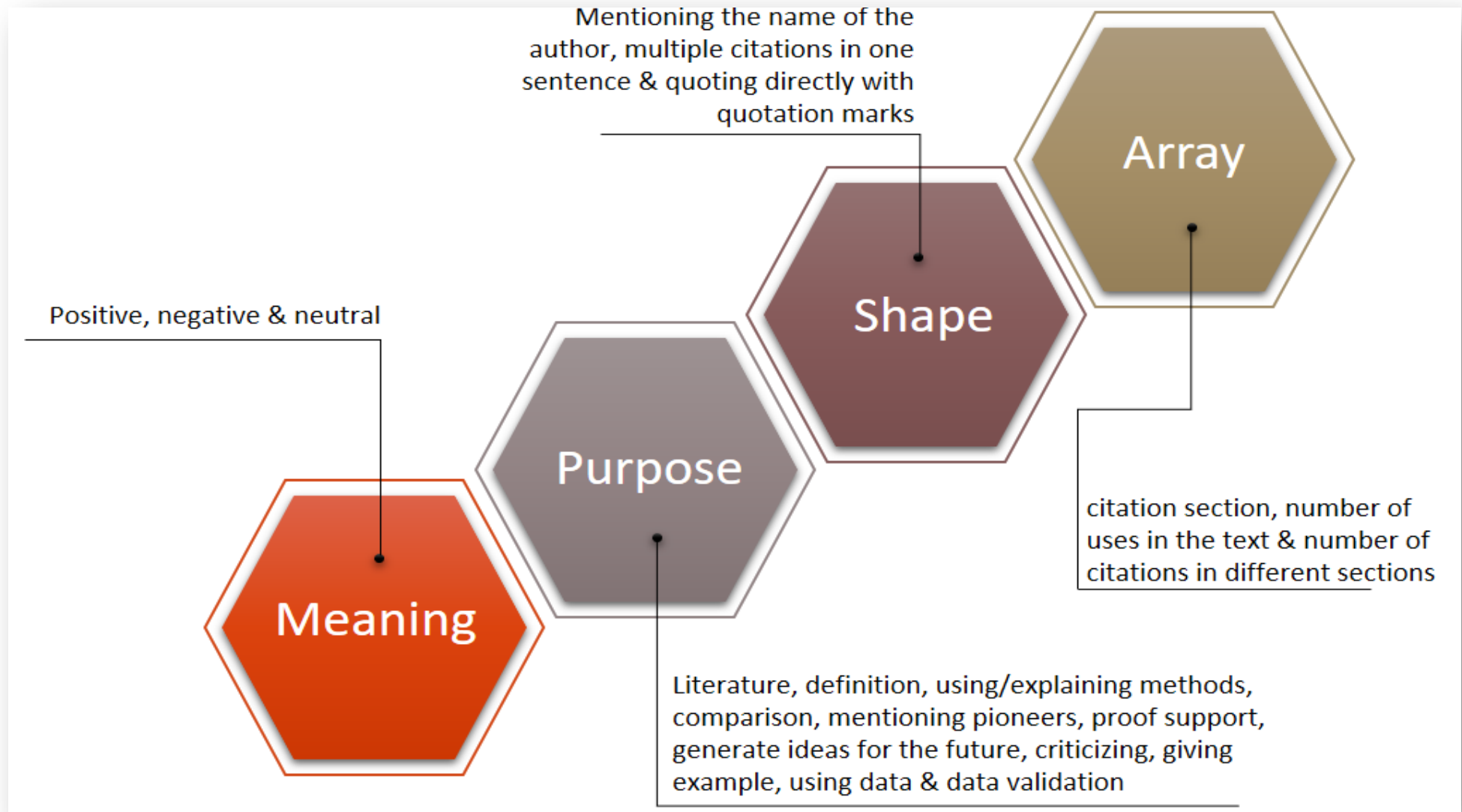
Content-Based Citation Analysis

- ❑ Evaluating citations by their contents
- ❑ Project supported by The Scientific and Technological Research Council of Turkey (115K440)
- ❑ Main aim:
 - ❑ Classifying citations by using machine learning algorithms

Content-Based Citation Analysis



Content-Based Citation Analysis



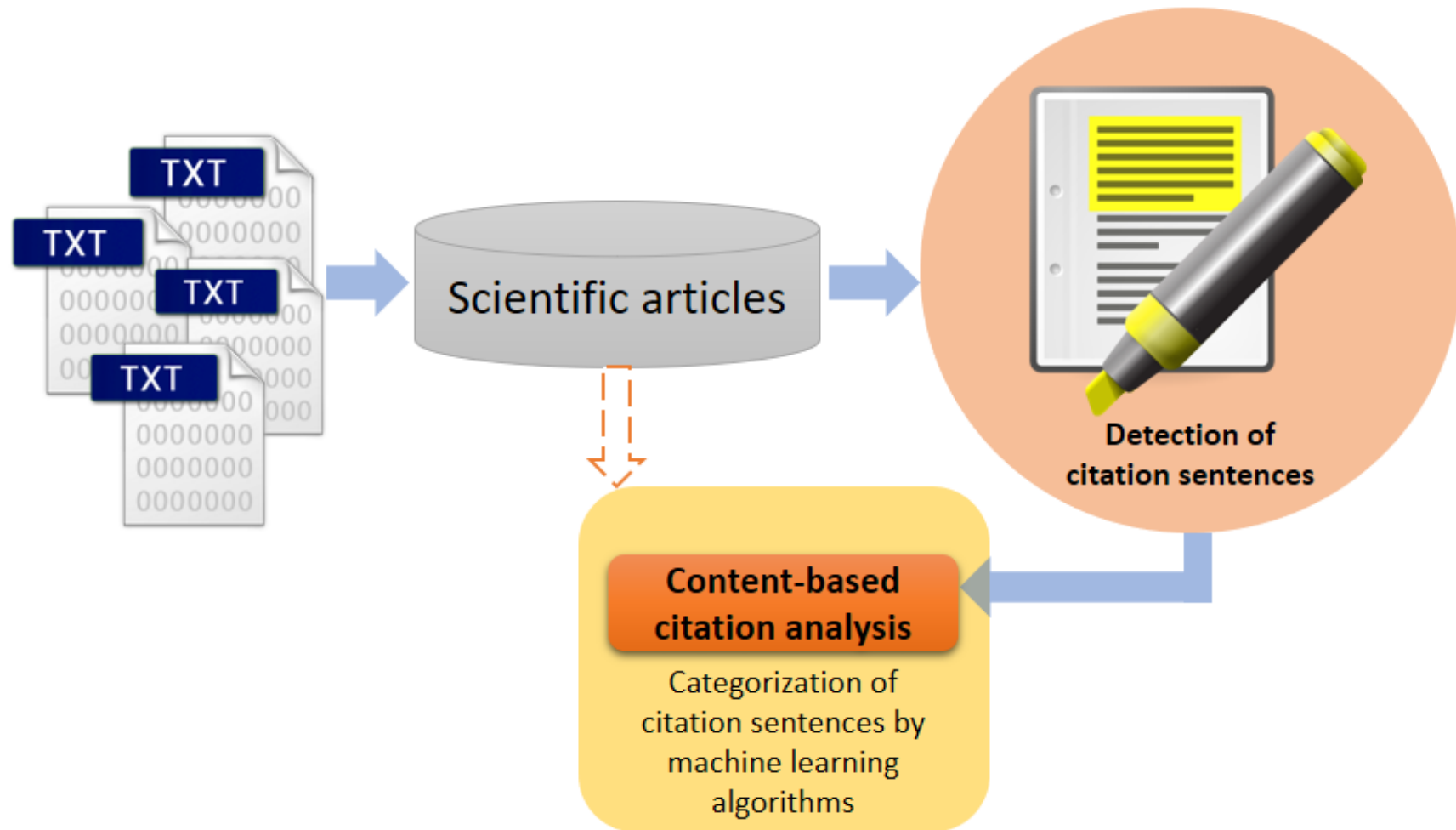
Content-Based Citation Analysis

- ❑ 423 peer-reviewed articles and their citation sentences are evaluated
- ❑ The automatic citation sentence classification achieved at least a **90%** success rate for all citation classes
- ❑ The need is:
 - ❑ Less human power
 - ❑ Automatic detection of citation sentences

Aim of Study

- ❑ Creating an automated citation extraction model by using finite state grammars to make content-based citation analysis easy and practical

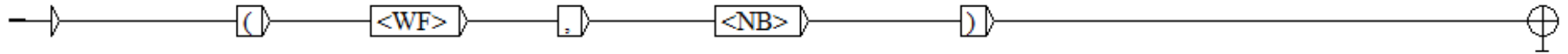
Aim of Study



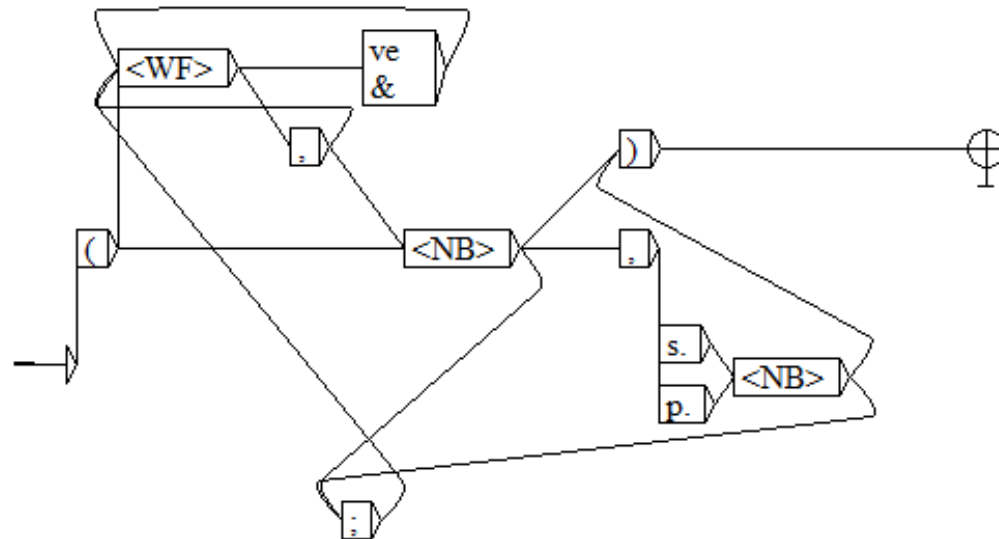
Methodology

- ❑ Creating rule list for citations given by APA Style
- ❑ Applying these rules to graphical syntax editor
- ❑ Software used: Nooj (<http://www.nooj-association.org/>)
- ❑ Detecting citation sentences
 - ❑ May include one or more sentences

Finite-State Grammar



(Taskin, 2017)



(Taskin, 2017)

(Taskin, 2017, s. 17)

(Taskin, 2016, s. 18; Taskin, 2017, s. 20)

Taskin (2017, s. 18)

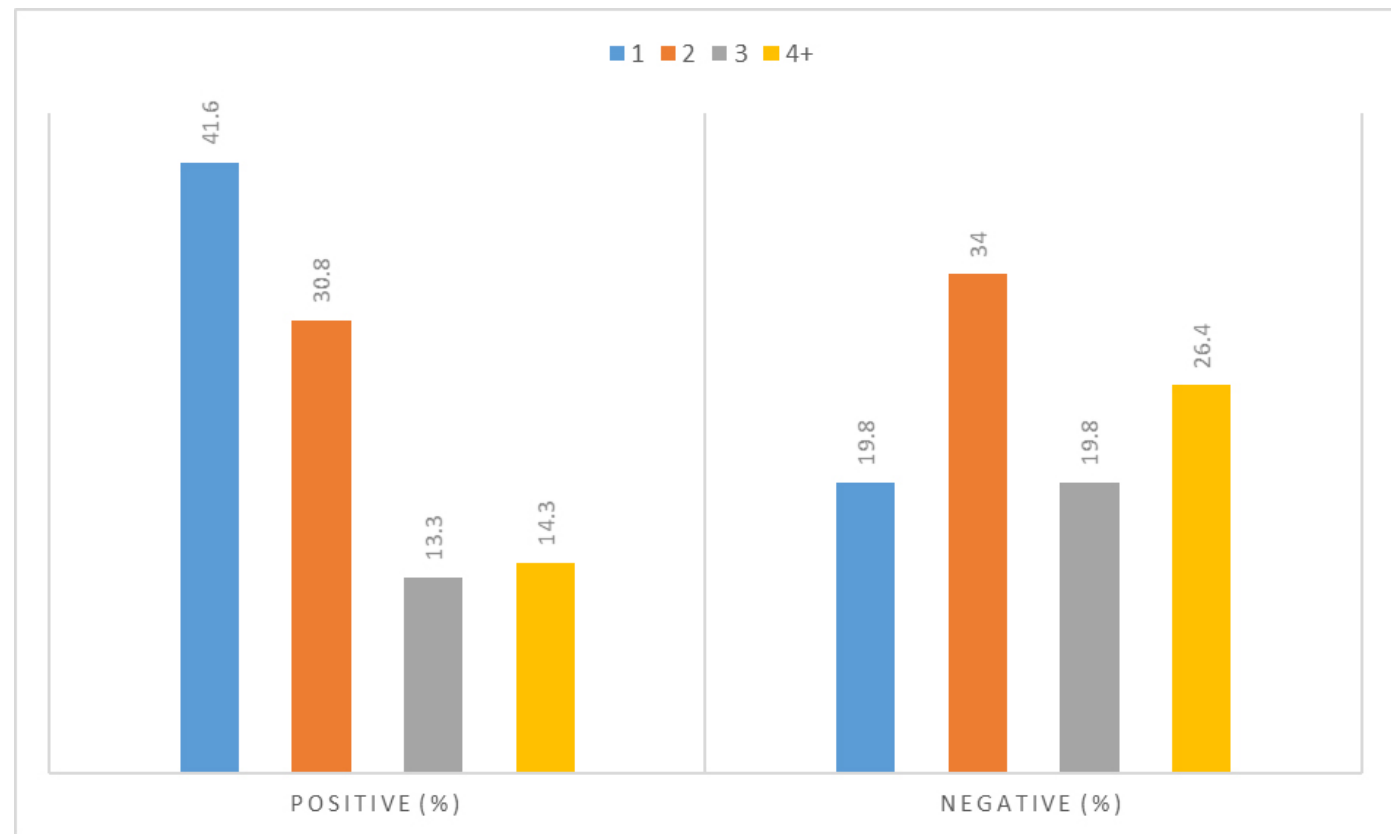
...

Finite-State Grammar

Display: 5 characters before, and 5 after. Display: ☒ Matches ☐ Outputs		
Before	Seq.	After
ve "açıklık" anlayışına bırakmasını sağlamıştır yönetiminin etkinliğinin artmasını da sağlayabilmektedir biri bilgiye özgürce erişebilmenin sağlanmasıdır ülkeler için temel haklardan biridir bir unsur olarak ortaya çıkmıştır edinme özgürlüğünün de temelini oluşturmaktadır anlayışının özümsemesine de katkı sağlamaktadır ve belgelere özgürce erişebilmeye bağlıdır gereğince uygulanmasına da katkı sağlamaktadır aykırı davranışların da önüne geçmektedir (Özdemirci, 2004, s. 66). Gülle'ye göre dışında devleti şeffaflaştırdığını söylemek mümkündür giren bir başka kanun izlemiştir dünya genelinde yasallaşmasına zemin hazırlamıştır bir bölümü olarak yer almıştır ve özgürlükler arasında yer verilmiştir hakkına sahip olduğuna dikkat çeker yılında İngiltere'de uygulamaya koymuştur gibi çeşitli etkinlikler de düzenlemektedir düşünceye/ bilgiye özgür erişimi savunmaktadırlar -i Vakayı" nin çıkarılmasıyla tamamlanmıştır yine bu hak çerçevesinde gözetilmektedir bir bölümü olarak yer almıştır kılan düzenlemeleri içerdiğine dikkat çekilmektedir doğru bir şekilde gerçekleştirmekte sorumludurlar değerlendirme Kurulu tarafından değerlendirilmesini içermektedir yer aldığı 29 ile 33. maddeleri kapsamaktadır edinme hakkı kapsamındaki işlemleri gerçekleştirmektedirler hatalı davranışları bu süreyi değiştirmez Edinme Değerlendirme Kurulu'na gönderirler incelenmesini kapsayan bir araştırma tekniğidir	(Köseoğlu, Selçuk ve Yurttaş, 2005, s.149; Çukurçayır, Özer ve Turgut, 2012, s. 9) (Uzun, 2005, s. 232) (Evren, 2010, s. 114) (İyimaya, 2003, s. 41; Canöz, 2008, s.144) (Uzun, 2005, s.231; Şavran, 2006; Yılmazoğlu, 2009, s. 1) (Tonta ve Çelik, 1996, s. 4) (Koçak, 2010, s.115) (İnaç ve Ünal, 2007, s. 16; Soykan, 2007, s. 65) (Uzun, 2005, s. 234) (Özdemirci, 2004, s. 66) (2006, s. 51) (Gülle, 2002, s. 89) (İyimaya, 2003, s. 42; Uzun, 2005, s. 237; Soykan, 2007, s.64; Canöz, 2008, s. 14... (Şavran, 2006, s. 47) (İyimaya, 2003, s. 42) (Cin, 2005, s.60) (İyimaya, 2003, s. 42) (Şavran, 2006, s. 48) (FAIFE, 2015) (Yılmaz, 2015, s. 139) (Demir, 2014, s. 59) (Cin, 2005, s. 42; Soykan, 2006, s. 27) (İyimaya, 2003, s. 42) (BEHK, 2003) (BEHK, 2003) (BEHK, 2003) (BEHK, 2003) (BEHK, 2003) (BEHK, 2003) (BEHK, 2003) (BEHK, 2003) (Yıldırım ve Şimşek, 1999)	. Demokratik olabilmeyen gereklerinden "açıklık" diğer . İdari işleyişin halkın karar ve . Bilgi edinme özgürlüğü, bireylerin idarenin . Yirminci yüzyılın ikinci yarısından itibaren . Bilgi Edinme Hakkı Kanunu, katılımcı . Bilgi Edinme Hakkı Kanunu ile . Yönetimin denetlenmesi yürütülen işler hakkında . Bu açıdan Bilgi Edinme Hakkı . Bilgi edinme hakkını gözetken yasalar . Gülle'ye göre (2006, s. 51) kamu . kamu kurumlarının ürettikleri resmi bilgilere . Bilgi edinme ya da bilgi . 1786 Virginia Bildirgesi, 1789 Fransa İnsan ve . Bilgi edinme hakkı, özerk bir . 1996 Güney Afrika Anayasası'nda bilgi . Bu düzenlemelerin yanı sıra, bilgi . Bilgi edinme hakkını başlı başına . Bilgiye erişimi kısıtlayan etkenlere tepki . IFLA, Amerikan Kütüphane Derneği (American . Bilgi edinme özgürlüğüyle ilgili bir . Takvim-i Vakayı halkın resmi . e-Devlet çalışmaları ve Avrupa . Türkiye'de 09.10.2003'te kabul edilen 4982 kanun . İkinci bölüm, "Bilgi Edinme Hakkı . Üçüncü bölüm ise "Bilgi Edinme . "Bilgi Edinme Hakkının Sınırları"nı . Buna göre, kanunun uygulanmasından sorumlu . Başvuru sahipleri gerçek ve tüzel . Başvurunun reddedilmesi halinde ret kararının . Türkiye'de Bilgi Edinme Hakkı . Bu kapsamda 2004-2014 yılları arasında yayımlanan

Citations Including Two or More Sentences

- ❑ Number of sentences are important to classify citations correctly
- ❑ Why?

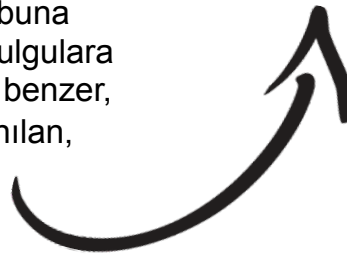


Detecting Important Conjunctions

- ❑ If detected citation sentence begin with conjunctions below, extract with previous.

örneğin, bu nedenle, bu bağlamda, bunun yanı sıra, buna göre, bu konuda, burada, yine, raporda, zira, zaten, yine, yine de, yazıda, yazara göre, yazar, yazarlar, yapılan bu, yani, üstelik, tüm bunlarla, tüm bu, tabi ki, tabii ki, sözü edilen, söz konusu, sonuçta, sonuç olarak, özellikle, oysa, oysa ki, oysaki, öyle ki, öte yandan, örneğin, örnek olarak, örnek vermek gerekirse, ona göre, o halde, nitekim, ne var ki, ne yazık ki, makalede, kuşkusuz, konuyla ilgili, kısaca, keza, kaldı ki, işte, incelenen, inceleme, ilki, ilkinde, ilk olarak, ikinci, her iki, her üç, her bir, hepsi de, hatta, görüldüğü üzere, gerçi, fakat, eğer, durum böyle iken, dolayısıyla, diğer taraftan, diğer bir deyişle, dahası, daha sonra, daha, çünkü, çalışmada, çalışma sonucunda, çalışma, buradan, burada, bununla birlikte, bunun, bununla, bunu, bunların, bunlardan, bunlara, bunlar, bundan dolayı, buna örnek olarak, buna rağmen, buna karşın, buna (göre, örnek olarak vs.), bulgulara göre, bu, böyle, böylelikle, böylece, bir başka deyişle, benzer, belirtilen, ayrıca, aynı, araştırmaya, araştırmının, anılan, ama, aksi, adı geçen

for this reason, in this context, in addition, in this regard, here, again, in this report, already, again, yet, according to article, writer, author, writers, in consequence, especially, whereas, on the other hand, for example, according to, all this, of course, however, indeed, unfortunately, undoubtedly, briefly, also, remaining, here, examined, both, in other words, furthermore, further, because, in work, this, that, these, those, therefore, according to the findings, this means that, thus, in other words, similarly, but, otherwise...



Detecting Important Conjunctions

- If following citation sentence begin with conjunctions below, extract with following.

ancak, oysa, oysa ki, oysaki, öte yandan, nitekim, ne var ki,
ne yazık ki, makalede, gerçi, fakat, eğer, durum böyle iken,
dolayısıyla, diğer taraftan, diğer bir deyişle, dahası, daha
sonra, daha, çünkü, çalışmada, çalışma sonucunda, çalışma,
burada, bunu, bununla, bunların, bunlardan, bunlara, bunlar,
bundan dolayı, buna örnek olarak, buna rağmen, buna
karşın, buna, bulgulara göre, bu, böyle, böylelikle, böylece,
bir başka deyişle, benzer, belirtilen, ayrıca, aynı, araştırmaya,
araştırmının, anılan, ama, aksi, adı geçen

however, whereas, on the other hand, unfortunately, though,
does not mean that, if the situation is so, then, from the other
side, in other words, because, here, therefore, these are,
therefore, for example, nevertheless, this, in other words,
similar, mentioned, also, same, but, otherwise, mentioned



Conclusion

- ❑ Creating an automated extraction tool for APA style may be used easily with the purpose of providing meaningful citation data to all content-based citation analysis studies
- ❑ It is possible to reduce the need for human power in such studies and to increase focus on the meanings of the citations.

First Stage of an Automated Content-Based Citation Analysis Study: Detection of Citation Sentences

Zehra Taşkın, Umut Al & Umut Sezen

{ztaskin, umutal, u.sezen}@hacettepe.edu.tr