

TÜRKÇE İÇİN METİN ÖZETLEME

YÜKSEK LİSANS TEZİ
Müh. Mesut TÜLEK
(504041524)

Tezin Enstitüye Verildiği Tarih : 7 Mayıs 2007
Tezin Savunulduğu Tarih : 14 Haziran 2007

Tez Danışmanı : Prof.Dr. Eşref ADALI
Diğer Jüri Üyeleri Prof.Dr. A.Coşkun SÖNMEZ (Yıldız Teknik Ü.)
Yrd.Doç.Dr. Zehre ÇATALTEPE (İ.T.Ü.)

HAZİRAN 2007

ÖNSÖZ

Tez çalışmam boyunca desteğini esirgemeyen değerli hocam ve danışmanım Prof. Dr. Eşref ADALI'ya teşekkürü bir borç bilirim.

Ayrıca, tüm hayatım boyunca her konuda destek olan, bir an olsun yalnız bırakmayan sevgili aileme ve bilgi erişimi ile ilgilenmeme vesile olan ve tez çalışmam boyunca destek olan değerli dostlarım M. Kıvanç TÜRKEEŞ ve Fatih KESGİN'e çok teşekkür ederim.

Mayıs, 2007

Mesut TÜLEK

İÇİNDEKİLER

KISALTMALAR	v
TABLO LİSTESİ	vi
ŞEKİL LİSTESİ	vii
ÖZET	viii
SUMMARY	ix
1. GİRİŞ	1
1.1 Doğal Dil İşleme ve Bilgi Erişimi	1
1.2 Metin Özetleme	2
1.3 Tezin Amacı	3
2. TÜRKÇENİN YAPISI	6
2.1 Türkçe	6
2.2 Kompozisyon	6
2.2.1 Konu - Anadüşünce	7
2.2.1.1 Konunun Maddesi	7
2.2.1.2 Yardımcı Düşünceler	7
2.2.1.3 Yardımcı Görüşler	7
2.2.2 Yazı Başlığı	8
2.2.3 Paragraf Planı	10
3. YAKIN ÇALIŞMALAR	12
3.1 Geleneksel Yaklaşımlar	12
3.1.1 Terim Sıklığı Yöntemi	13
3.1.2 Otomatik Çıkarımda Yeni Yöntemler	14
3.1.3 Kimya Özetçe Hizmeti'nde (KÖH) Otomatik Özetçe Araştırması	16
3.2 Derlem Tabanlı Yaklaşımlar	17
3.2.1 Eğitilebilir Belge Özetleyici	19
3.2.1.1 Özellikler	19
3.2.1.2 Derlem	21
3.2.1.3 Cümle Eşleştirme	21
3.2.1.4 Değerlendirme	22
3.2.2 İstatistiksel Tabanlı Belge Özetleme Sistemi	24
3.2.3 Güçlü DDİ Yöntemlerinden Bilgi Edinimi ile Eğitilebilir Özetleyici	25
3.2.4 SUMMARIST'te Otomatik Metin Özetleme	26
3.3 Hitap Yapısının Kullanılması	28
3.4 Zengin-Bilgi Yaklaşımları	28
3.5 Değerlendirme Yöntemleri	29
3.6 Yeni Özetleme Problem Alanları	30
4. ÖZETLEME	31
4.1 Gövdeleme	31
4.1.1 Gövdeleme ile İlgili Yapılmış Çalışmalar	32

4.1.2 Zemberek	33
4.1.2.1 Yapı	33
4.1.2.2 Biçimbirimsel Çözümleyici	36
4.1.2.3 Başarım	39
4.1.3 Biçimbirimsel Çözümleyici Kullanarak Gövdeleme	40
4.1.3.1 En Uzun Gövdenin Seçilmesi	40
4.1.3.2 İlk Beş Karakter Yöntemi	40
4.2 Özetleme Sistemi	40
4.2.1 Yapısal Çözümleme Birimi	42
4.2.2 Dilbilimsel Çözümleme Birimi	45
4.2.3 Özet Çıkarım Birimi	46
4.2.3.1 Başlık Yöntemi	46
4.2.3.2 Sözcük Sıklığı Yöntemi	46
4.2.3.3 İpucu Sözcük Öbekleri Yöntemi	48
4.2.3.4 Metnin İçindeki Konum Yöntemi	48
4.2.4 Yöntem Birleştirme Birimi	49
4.2.5 Başarımın Ölçülmesi	50
4.2.5.1 Keskinlik ve Anımsama	50
5. GERÇEKLENEN YAZILIM	53
5.1 Yazılımın Genel Yapısı	53
5.1.1 Biçimbirimsel Çözümleme	53
5.1.1.1 Eklerin Türünün Belirlenmesi	55
5.1.2 Gövdeleme	56
5.1.2.1 En Uzun Gövdenin Seçilmesi	56
5.1.2.2 İlk Beş Karakter Yöntemi	58
5.1.3 Özetleme	59
6. SONUÇLAR VE TARTIŞMA	64
KAYNAKLAR	72
EKLER	74
ÖZGEÇMİŞ	88

KISALTMALAR

DDİ	: Doğal Dil İşleme
BE	: Bilgi Erişimi
BÇ	: Biçimbirimsel Çözümleme
MÖ	: Metin Özetleme
DUC	: Document Understanding Conference
ARDA	: Advanced Research and Development Activity
GMAT	: Graduate Management Admission Test

TABLO LİSTESİ

	<u>Sayfa No</u>
Tablo 3.1: KPC derlemindeki dergiler	21
Tablo 3.2: KPC'nin eşleştirme dağılımı.....	22
Tablo 3.3: Özelliklerin başarımı	23
Tablo 3.4: Farklı özelliklerin ve filtreleme yöntemlerinin karşılaştırılması	25
Tablo 3.5: Çeşitli eğitim özellikleri kullanılmasının sonuçları	25
Tablo 4.1: Türkçe ve İngilizce derlemler için gövdeleme ile sıkıştırma oranı	31
Tablo 4.2: Gözlükçüdekilerden sözcüğü için Zemberek BÇ çıktısı.....	40
Tablo 4.3: Dilbilimsel çözümleme birimi çıktısı	46
Tablo 4.4: Örnek durma listesi.....	47
Tablo 4.5: Örnek bir sözcük sıklığı listesi	47
Tablo 4.6: İpucu sözcük öbekleri	48
Tablo 4.7: Yöntem birleştirme birimi çıktısı.....	50
Tablo 5.1: Gözlükçülük sözcüğü kök adayları.....	53
Tablo 5.2: Kalemin sözcüğü için gövdeleyici sonucu	58
Tablo 5.3: Gözlükçülüğün sözcüğü için BÇ çıktısı	59
Tablo 6.1: Gövdeleme yöntemlerinin sonuçları.....	64
Tablo 6.2: Gövdeleme yöntemlerinin sonuçları.....	65
Tablo 6.3: “Yaşam Kalitesi” belgesinin cümle puanları	66
Tablo 6.4: Yöntemlerin başarımları (K: Keskinlik, A: Anımsama).....	67
Tablo 6.5: Yöntemlerin tekil ve bağıl değerlendirilmesi	69

ŞEKİL LİSTESİ

	<u>Sayfa No</u>
Şekil 1.1: Metin özetleme sistemi mimarisi	5
Şekil 3.1: Luhn'un sözcük sıklık çizelgesi	13
Şekil 3.2: Luhn'un cümle önem derecesi hesaplama yöntemi	14
Şekil 3.3: Yöntemlerin ortalama seçim puanları	16
Şekil 3.4: Özet büyüklüğü – başarımlar	24
Şekil 3.5: SUMMARIST mimarisi	27
Şekil 4.1: Kök sözcük ağacı.....	35
Şekil 4.2: Yumuşama durumunda kök sözcük ağacının içeriği.....	36
Şekil 4.3: Çözümleyici blok çizelgesi	38
Şekil 4.4: Metin özetleme sistemi mimarisi	42
Şekil 4.5: Sözcük sonu belirleme algoritması.....	44
Şekil 4.6: Yapısal çözümleme birimi çıktısı.....	45
Şekil 4.7: Giriş ve sonuç paragrafı örneği	49
Şekil 4.8: Örnek bir bilgi talebi için keskinlik ve anımsama.....	51
Şekil 5.1: Kelime sınıfı yapısı	54
Şekil 5.2: Ek sınıfı özellikleri	55
Şekil 5.3: Ek XML dosyasındaki örnek bir ek düğümü	56
Şekil 5.4: En uzun gövde seçimi algoritması akış çizelgesi	57
Şekil 5.5: cBelge sınıfı arayüzü	60
Şekil 5.6: cParagraf sınıfı arayüzü.....	61
Şekil 5.7: cCumle sınıfı arayüzü.....	62
Şekil 5.8: cSozcuk sınıfı arayüzü.....	63
Şekil 6.1: Yöntemlerin keskinlik başarımları	68
Şekil 6.2: Yöntemlerin anımsama başarımları.....	68

TÜRKÇE İÇİN METİN ÖZETLEME

ÖZET

Günümüzde bilginin boyutu hızla artmaktadır. Bununla birlikte bilginin sayısal ortamda olma oranı da artmaktadır. 2003 yılında dünyadaki her insan için üretilen bilgi miktarı 800 MB, üretilen bilginin sayısal ortamda olma oranı da %90 olarak tahmin edilmiştir. Bu hızlı artışla birlikte aranan bir bilgiye makul sürede erişim sorunu da büyümektedir. Bu sorunu inceleyen bilgi erişimi genel şekliyle, depolanmış bilgi derleminden belirli bilgi gereksinimiyle ilgili bölümlere erişim yöntemine yönelik çalışma olarak tanımlanabilir.

Bilgi erişiminin altkümelerinden biri de metin özetlemedir. Metin özetleme, bir belgeyi girdi olarak alan ve çıktı olarak daha kısa, aslının yerine geçen ve onun en önemli içeriğini barındıran bir süreç olarak tanımlanabilir. Herkesin her şeyi okumaya vakti olmaması ihtiyacından yola çıkan metin özetleme şüphesiz birçok alanda ihtiyaç duyulabilecek bir işlemdir. Dergi ve gazeteler, bilimsel makaleler ve e-postalar özetlemenin verimliliği arttıracığı başlıca günlük hayat uygulamalarıdır.

Metin özetleme ile ilgili ilk çalışmalar bundan yaklaşık elli yıl önce İngilizce için yapılmıştır. Geçen elli yıllık süreçte probleme yönelik birçok yeni yöntem bulunup gerçekleşmesine rağmen, bugünkü araştırmalarda ve pratik uygulamalarda hala ilk çözüm yolları olan istatistiksel yöntemler kullanılmaktadır. Bu yaklaşımların verimliliğinin ve başarımının yüksek, uygulama maliyetinin de düşük olması günümüzde popülerliğini yitirmemesinin sebebidir.

Türkçe ise sondan eklemeli ve kurallı yapısı, çok az miktarda kurlsız sözcük içermesi nedeniyle bilgi erişimi araştırmacılarının ilgisini çekmiştir. Türkçenin bu sondan eklemeli yapısı sayesinde, aynı kökten farklı anlamlarda gövde adı verilen yeni sözcükler türemiştir. Bu özellik nedeniyle, gövdeleme yani bir sözcüğün eklenmiş çekim eklerinden arındırılarak gövde veya kökünün bulunması işlemi Türkçe için yapılan tüm bilgi erişimi sistemlerinde çok önemli bir yer edinmiştir.

Bu tezde, Türkçenin yapısı göz önüne alınarak, bir metnin özetlenmesi için farklı istatistiksel yöntemler tanımlanıp yazılımla gerçekleştirilmiş ve bu yöntemlerin Türkçeye uygunluğu tartışılmıştır. Diğer tüm Türkçe bilgi erişimi sistemlerinde de gerekli olduğu gibi, Türkçenin sondan eklemeli yapısının gözetilmesi amacıyla farklı gövdeleme algoritmalarının özetleme başarımına etkisi incelenmiştir. Başarımlarının daha yüksek olması amacıyla, gerçekleştirilen gövdeleme algoritmalarında sözcüklerin olası kök ve ek birleşimlerini üreten biçimbirimsel çözümleyici kullanılmıştır. Gövdelenmiş bu sözcükler farklı özetleme yöntemleri aracılığıyla incelenip her yöntem için özetle yer alacak cümleler belirlenmiştir. Daha sonra bu yöntemlerin ürettiği sonuçlar birleştirilerek son özet oluşturulmuştur.

TEXT SUMMARIZATION FOR TURKISH

SUMMARY

Today the size of information has been growing rapidly. Also the ratio of information in a digital form is also growing. It is estimated that in 2003 for each person on earth 800 MB of information was produced and the 90% of the currently produced information is in a digital form. This rapid growth brings the problem of retrieval of the searched information in a reasonable time. Information retrieval, which deals with this problem, can be broadly defined as the study of how to determine and retrieve the portions, which are relevant to particular information needs, from a corpus of stored information.

One of the subsets of information retrieval is text summarization. Text summarization can be defined as the process which takes a document as input and outputs a shorter document which is condensed and can be used instead of the original. Text summarization which starts from the fact that nobody has time to read everything is a process that can be needed at most areas certainly. Magazines and newspapers, scientific papers and e-mails are one of the most important daily life applications that summarization improves the efficiency.

The first works around text summarization were done about fifty years ago for English. Although many new methods have been found and implemented to solve the problem in the last fifty years, today's researches and practical applications still use the early statistical methods. High efficiency, high performance and low application cost of these approaches make them still popular today.

Turkish draws the attention of information retrieval researchers because of its agglutinative and rule based structure and of having a few number of irregular words. Due to its agglutinative structure, new words called stem with different meanings derived from the same root in Turkish. Because of this characteristic, stemming, the process of removing inflectional affixes to find the root or stem of the word, has a very important place at all Turkish information retrieval systems.

In this study, different statistical methods for text summarization are described and developed by taking into consideration the structure of Turkish. Also the validity of these methods for Turkish has been discussed. The effect of different stemming algorithms on summarization efficiency has been studied for the aim of taking into consideration the agglutinative structure of Turkish, as it is necessary in all other information retrieval systems for this language. Morphological analyzer, which outputs the root and affix combinations of the input word, has been used in stemming algorithms to increase the efficiency of the text summarization. These stemmed words have been studied by different summarization methods and sentences which will be included in the summary have been chosen. In the end, the final summary has been created by combining the results of these methods.

1. GİRİŞ

1.1 Doğal Dil İşleme ve Bilgi Erişimi

Günümüzde bilginin boyutu çok büyük hızla artmaktadır. Örneğin, 2003 yılında dünyadaki her insan için üretilen bilgi miktarı 800 MB olarak tahmin ediliyor. (Tabii bu rakamın içinde yazar kasa fişlerinin bile dâhil olduğunu unutmamak gerekir.) Ayrıca üretilen bilginin %90'ının sayısal ortamda olduğu tahmin edilmektedir. 10 yıl içinde de, en gerekli bilgilerin sayısal ortamda olacağı beklenmektedir [1]. Hızla artan bu bilgi miktarı, aranan bilgiye ulaşılması sorununu da beraberinde getirmektedir. Bu bilgi artışı ve bu bilgi yığnında istenen bilgiye ulaşma problemi araştırmacıların ilgisi çekmiştir ve bugün *bilgi erişimi* denen araştırma konusunu ortaya çıkarmıştır.

Bilgi erişimi genel şekliyle, depolanmış bilgi derleminden belirli bilgi gereksinimiyle ilgili bölümlere nasıl erişileceği çalışması olarak tanımlanabilir. Belirli bir alana ilişkin veya çeşitli konuların birleşiminden oluşan geniş bir bilgi derleminden oluşan bir dükkân varsayalım. Uygulamaya bağlı olarak bilgi yapısal bir şekilde veya dağınık olarak tutuluyor olsun. *Problem* çözmeyi bilmeyen dükkânın bir kullanıcısı belirli bir bilgiyi arıyor olsun. Dolayısıyla, bir şekilde kullanıcı *bilgi ihtiyacını* bir istek olarak dile getirmeli. Böylece BE; kullanıcının *talep* olarak ifade ettiği ve kullanılan belirli bir BE sistemine uyan *sorgu* şekline dönüştürülen ihtiyacına yönelik bilginin belirlenmesi ve erişimi ile ilgilenir. Bir BE sistemi normalde, belgeleri ve içindeki bilgileri temsil etmek için asıl belgenin yerine geçen bir temsil tutar [2].

Çoğu BE sisteminde belgelerin anlamı yalnızca belgelerin içindeki sözcüklerde yatar. Örneğin bu sistemlerde “Yediğimi görürüm” ile “Gördüğümü yerim” tam olarak aynı şeydir. Belgeleri oluşturan cümleleri oluşturan sözcüklerin sıralamasının ve seçiminin anlamı belirlemede etkisi yoktur. Sözbilimsel bilgiyi ihmal eden bu yaklaşımlar genelde *sözcük torbası* olarak adlandırılır [3]. Sözbilimsel ve dilbilimsel farklılıklar dört grup altında toplanabilir:

- Biçimbirimsel farklılıklar: Örneğin; *kitap*, *kitabım* ve *kitaplar* aynı sözcüğün farklı biçimleridir.
- Sözcüksel farklılıklar: Farklı sözcükler aynı anlamı temsil edebilir (eşanlam). Örnek olarak *hikâye* ve *öykü* verilebilir.
- Anlambilimsel farklılıklar: Bir sözcük farklı bağlamlarda farklı anlamlara gelebilir (eşseslilik). Örneğin, *site* hem Genel Ağ tasarımı hem de emlak bağlamında değerlendirilebilir.
- Sözbilimsel farklılıklar: *nehir yakınındaki hava kirliliği* ve *nehir kirliliği* sözcük öbeklerinin ortak iki sözcükleri olmasına rağmen farklı anlama gelirler [4].

Doğal Dil İşleme, doğal dillerin yani insanlar tarafından konuşma ve yazıda kullanılan dillerin kurallı yapısının çözümlenerek anlaşılması veya yeniden üretilmesi amacını taşır [5,6].

Bilgi erişimi araştırmaları, belge ambarlarından algoritma ve model geliştirme ile ilgilenir. BE'ye DDİ'nin bir alt alanı olarak bakılabilir. Çünkü BE, DDİ'nin özel bir uygulama alanı ile uğraşır. (Konuşma, görüntü ve video erişimi git gide yaygınlaşmasına rağmen geleneksel BE araştırması yazılı metinle uğraşır.) Ama aslında, biraz BE'nin kısmi ihtiyaçları DDİ alanında pek de ilgi çekici sorunlar olarak görülmediğinden biraz da BE'deki baskın yaklaşım olan istatistiksel yöntemlerin DDİ'de pek tercih edilmemesinden iki alan arasındaki etkileşim sınırlıdır.

DDİ'de nicel yöntemlerin yeniden güçlenmesi ile birlikte, iki alan arasındaki bağlantı artmıştır. Alanlar arasındaki etkileşimlere örnekler şöyledir: istatistiksel DDİ ve BE'de dikkat çekmeye başlayan bir sorun olan, belgelerdeki terim dağılımının olasılıklı modelleri; daha verimli belge erişimi için kullanılan bir DDİ yöntemi olan hitap bölümlenme; istatistiksel DDİ'de kullanılan iki BE yöntemi olan Vektör Uzayı Modeli ve Gizli Anlambilimsel Dizinleme [7].

1.2 Metin Özetleme

Metin özetleme, bir belgeyi girdi olarak alan ve çıktı olarak daha kısa, aslının yerine geçen ve onun en önemli içeriğini barındıran bir süreç olarak tanımlanabilir.

Buradaki “önem”, birçok farklı başvuru noktasına göre belirlenebilir. En yaygın olanı, bir konu ile ilişkili olma veya kullanıcıya belirli bir işi yapmada yardımcı olma gibi kullanıcı gereksinimlerini yansıtmaktır.

Metin özetlemenin gerek duyulabileceği veya verimliliği arttırabileceği alanlara; çeviri yapılması gereken bir metinde tüm metni çevirmek yerine çıkarılacak özetin çevrilmesi, BE sistemlerinde dizin terimlerinin çıkarımında özetin kullanılması örnek olarak gösterilebilir [8].

Bazıları bir özeti, aslında oldukça farklı anlamları olan *çıkarımsal özet* veya *özetçe* olarak düşünebilir. *Çıkarımsal özet*, konu ile en ilgili metin parçalarını, belki küçük değişikliklerle seçmeyle oluşturulur. *Özetçe*, gerekmedikçe içeriğe değinmeden belgenin içeriğini anlatan bir açıklama yazısıdır.

İki durumda da bazıları özetlemeyi, bir belgenin *sıkıştırılması* veya yoğunlaştırılması olarak düşünebilir. Çıkarımsal özet, konuyla daha az ilgili maddeleri çıkararak sıkıştırma işlemi yapar. Buna karşılık özetçe sıkıştırmayı, detayları gizleyip belirli bilgileri daha genel ifadelerle değiştirmek gibi daha karmaşık yöntemlerle yapar [9]. Bilgisayar tarafından üretilen özetçede amaç, insan tarafından üretilen özetçe kadar tutarlı özetçe üretmektir. Çıkarımsal özetde ise amaç, belgenin içeriğini yansıtan metin parçalarından oluşan bir küme çıkarmaktır [10].

Literatürde bulunabilecek bir başka ayırım *genel* ve *sorgu tabanlı* özetler arasındadır. Sorgu tabanlı özetler kendilerini sorguya bağlı içerikle sınırlandırırken genel amaçlı özetler belgenin içeriği hakkında genel bir kanı verirler. Sorgu tabanlı özetler, kullanma kılavuzu veya ders kitabı gibi büyük veya konu çeşitliliği olan belgelerle ilgilenilirken son derece faydalıdır [9].

Bir başka gruplama alanı belge kaynağına göre yapılan, *tek belge kaynaklı* ve *çoklu belge kaynaklı* özetlemedir. Tek belge kaynaklı özetleme, girdi olarak tek bir belge alınıp, yoğunlaştırılmış bir sonuç belgesi çıkarma işlemidir. Çoklu belge kaynaklı özetleme ise; birden fazla belgenin yorumlanıp, bu belgelerdeki benzer içeriklerin birleştirilerek çıktı olarak tek bir özet belgesi üretilmesidir.

1.3 Tezin Amacı

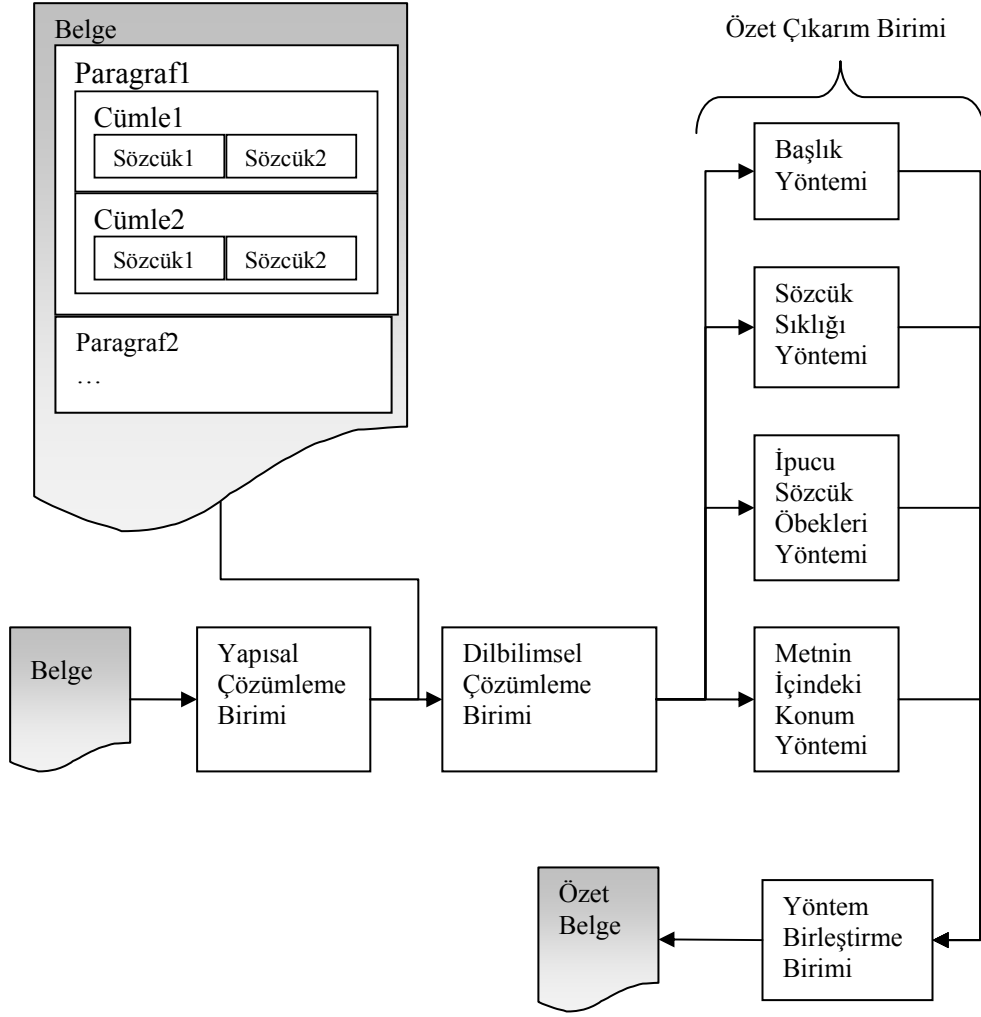
Bu tezde, Türkçenin Doğal Dil İşleme ile ilgili özellikleri göz önüne alınarak Türkçe metinler için bir özetleme sistemi tanıtılmıştır. DDİ araştırmacılarının dikkatini

çeken, Türkçenin kendine has kurallı yapısının ve sondan eklemeli bir dil oluşunun sonucu olan sözcüklerin biçimbirimsel çeşitliliğinden faydalanabilmek amacıyla çalışmada gövdeleyici de kullanılmıştır.

Metin özetleme yöntemleri, sözcüklerin, cümlelerin ve paragrafların metin içindeki yerlerine, diziliş ve görevlerine dayandığından anlatım yöntemleri ve bu yöntemlerin yazılı dildeki yapısı çalışmanın arka planını oluşturmaktadır. Bu nedenle çalışmanın ikinci bölümünde metin özetleme yöntemleri belirlenirken temel alınan Türkçenin yapısı ve özellikleri anlatılmıştır.

Bugüne kadar İngilizce için 50'li yıllardan başlayıp günümüze kadar birçok farklı yöntem tasarlanmış, mevcut yöntemlerin uygulama çeşitlilikleri denenmiştir. Ayrıca ortak bir başarı ölçütü belirleyebilmek için ortak bir veri kümesi ve derlem ile ilgili çalışmalar da yapılmıştır. Bu çalışmalardan tezin üçüncü bölümünde bahsedilmiştir. Maalesef Türkçe için bu denli zengin çalışmalar ve MÖ sistemlerinde kullanılacak ortak bir veri kümesi ve derlem çalışması yapılmamıştır.

Türkçenin sondan eklemeli-kurallı yapısı nedeniyle, diğer tüm BE sistemlerinde olduğu gibi, İngilizce için yıllardır geliştirilen bu MÖ yöntemlerinin de Türkçeye birebir uygulanması mümkün değildir. Gövdeleme, konuşma-parçası etiketleme gibi yöntemler Türkçe için yapılan bir MÖ sisteminin verimliliğini etkileyecektir. Tezin dördüncü bölümde kullanılan gövdeleyici ve MÖ yöntemleri anlatılmıştır. Beşinci bölümde gerçekleştirilen yazılımın detayları (Şekil 1.1'de mimarisi gösterilmiştir.) anlatılmıştır ve son bölümde de elde edilen sonuçlar yorumlanmıştır.



Şekil 1.1: Metin özetleme sistemi mimarisi

2. TÜRKÇENİN YAPISI

Türkçe için yapılan metin özetleme sistemi, Türkçe metinlerin içeriğini inceleyeceğinden Türkçenin yapısı, Türkçede kompozisyon çeşitleri, Türkçe kompozisyonların biçimi araştırma konusunun bir parçasını oluşturmaktadır. Bu bölümde, Türk dilinden, Türkçe kompozisyonların içeriğinden ve yazım biçimlerinden bahsedilmektedir.

2.1 Türkçe

Türkçe, dünya dilleri arasında kök bakımından yapılan sınıflandırmada, Ural-Altay dillerinin Altay kolunda yer almaktadır. Türk dilinin Altay kolunda bulunmasının bir özelliği, yapı bakımından yapılan sınıflandırmada bir yapı birliği meydana getirmesidir. Nitekim Ural-Altay dilleri eklemeli dillerdendir. Ünlü sayısı fazladır ve bu uyum hepsinde sağlamdır. Sözcük yapımı ve çekim şekli son eklerle sağlanır. Cümle yapısı Türkçede olduğu gibi yüklem sondadır. Bu benzerlik bilhassa Moğolca ile Türkçe arasında daha belirgindir.

Nasıl dil aileleri meydana getirirken, dünya dillerinin kök bakımından bir anadilden çıktığı düşünülürse (Hint-Avrupa, Hami-Sami, Fin-Ugor gibi), Türk dilinin de çok eskiden bir anadil olan Huncadan geldiği kabul edilir [11].

2.2 Kompozisyon

Kompozisyon, Türkçeye Fransızcadan geçmiş bir sözcük olup, “ayrı ayrı parçaları, unsurları bir araya getirerek uyumlu bir bütün oluşturma” demektir. Bir edebiyat terimi olarak kompozisyon, “okullarda yazı yazma alışkanlığını kazandırmak için öğrencilere verilen yazı ödevi, kalem alıştırmaları”, daha geniş anlamda ise “düşüncelerin, duyguların, olayların, tasarıların uyumlu bit bütün oluşturacak şekilde sözle veya yazı ile anlatılması”dır.

Fıkra, makale, sohbet, tenkit (eleştiri), deneme, roman, hikâye, masal, tiyatro, hatırat, seyahatname (gezi yazıları), mektup, dilekçe (eski: arzuhâl), şiir, nutuk, konferans,

röportaj, hayat hikâyesi (biyografi) edebiyatla ilgili birer kompozisyonudur [12]. Kompozisyon yazımındaki temel kavramlar şöyledir:

2.2.1 Konu - Anadüşünce

Eski dilde “mevzû” denilen konu, “kendisinden bahsolunan, kararlaştırılmış şey” demektir. Her yapının bir temeli, her fikrin bir dayanağı varsa, her yazının da mutlaka bir konusu vardır. Başboş, ne anlatmak istediği belli olmayan, konusuz yazı olmaz. Yazı yazmaya karar veren kişi için her varlık, her olay, her düşünce, her mesele, her şey bir kompozisyon konusu olabilir [12].

Bir konuyu oluşturan üç öge şöyledir:

2.2.1.1 Konunun Maddesi

Seçilen konuda “Açıklanacak olan nedir? Ne açıklanacaktır?” sorularının karşılığında “konunun maddesi” denir. Eğer konu, aynı zamanda bir *anadüşünce* olarak ele alınmışsa, konunun maddesi “düşüncenin özü” olur.

2.2.1.2 Yardımcı Düşünceler

Konunun maddesi ya da düşüncenin özü “tamlayan”dır; yardımcı düşünce/ler ya da görüş/ler ise “tamlanan”dır.

Yardımcı düşüncelere “açıklayıcı düşünceler” demek çok doğru olur. Bunlar, konunun maddesinin “ne ile, nelerle açıklanacağını” gösterir, sınırlar, belirtirler. Konunun maddesini ya da düşüncenin özünü “ne ile, nelerle açıklayacağım?” sorusuyla bulunurlar.

2.2.1.3 Yardımcı Görüşler

Yardımcı görüşler, görüşü açmaya, açıklamaya, daha geniş, ayrıntılı olarak açıklamaya yarayan görüşlerdir. Kimi durumlarda, düşüncenin özünü açıklayacağımız görüşü açıklamak yetmemektedir; görüşü de açıklamak gerekmektedir. İşte bu, yardımcı görüşlerle yapılır.

Yardımcı görüşler, hiçbir konuda verilmez, onları o konuyu açıklayacak olan, kendisi seçer, bulur. O halde yardımcı görüşler kişiseldir. Bir konuyu herkesin başka başka açıklamasının nedeni de işte budur. Kişinin kendisinin bulduğu görüşler ve yardımcı

görüşler, aynı konunun değişik kişiler tarafından çok değişik biçimde açıklamasına neden olmaktadır. Bu, bir bakıma konuya bakış açısıdır.

Açıklanmak üzere seçilen düşünceye *anadüşünce* denir. Buna göre anadüşünce, ya yazıda açıklanacak olan temel yargıdır ya da o yazıda vardırılabilecek olan temel yargıdır. Bu demektir ki anadüşünce, yazının giriş bölümünde de ortaya konabilir, sonuç bölümünde de sonuç olarak çıkarılabilir.

Konu “genel”, anadüşünce ise “özel, dar” anlamlıdır. Bu tanımdan şu anlaşılabilir: Anadüşünce, “konu”nun sınırlanmış biçimidir. Eğer konu “genel” anlamlı ise, konu ile anadüşünce birbirinden “ayrı”dır; konu, anadüşünce biçiminde sıralanmışsa, konu ile anadüşünce “aynıdır, birleştirilmiştir” [13].

2.2.2 Yazı Başlığı

Yazı başlığı, o yazıda işlenilmesi düşünülen ana fikrin aynasıdır. Canlı cansız her varlığın adı olduğu gibi, her yazının da bir başlığı vardır.

Yazı başlığı okuyucuların ilgisini çekmeli, onlara tesir etmelidir. Bu bakımdan başlık seçmek çok önemlidir. Okuyucu, başlığa bakınca, yazının nelerden söz edeceğini kestirebilmelidir.

Yazıların başlığı uzun veya kısa olabilir. Fakat, uzun başlıkların okuyucu üzerinde tesirli olmadığını unutmamak gerekir.

- Saatleri Ayarlama Enstitüsü (A. H. Tanpınar)
- Bursada Zaman (Ahmet Hamdi Tanpınar)
- Tarih İçinde Türk Edebiyatı (Faruk K. Timurtaş)

gibi kitap başlıkları, kısa, tesirli ve akılda kalabilecek başlıklardır.

Bir kitap başlığını, meydana getiren sözcüklerin hepsi büyük harfle yazılır. Bir dergide, gazetede veya antolojide yer alan herhangi bir edebî türden yazının hepsi büyük harfle olabileceği gibi, başlığı meydana getiren sözcüklerin yalnız ilk harfleri büyük olabilir [12].

Başlıklar, sayfada yer alışlarına göre iki türdür:

Büyük Başlıklar

Bunlar kitap, bölüm, konu, yazı... başlıklarıdır. Büyük başlıklara “genel başlık, ana başlık, asıl başlık” da denmektedir.

Büyük başlıklar, “kapsam” yönünden en geniş başlıklardır; başka deyişle, başlıkların başlığına büyük başlık denir. Büyük başlık altında yer alan bütün küçük başlıkları içine alan başlık, büyük başlıktır.

Küçük Başlıklar

Büyük başlık altında sıralanan bölümlerin başlıklarına “küçük başlık” denir. Küçük başlığa “ara başlık, bölüm başlığı” da denmektedir.

Neler Başlık Olur

Adlar: Türkiye, Atasözleri, Önder, Bilim ve Teknik, Sanat ve Kitap, Nem ve Gam, Meşe ile Kamış, Kurt ile Kuzu

Tek addan ya da bir bağlaçla bağlı iki addan başlık yapılabilir.

Belirtisiz Ad Takımları: Yurt Sevgisi, Gece Çalışması, Mohaç Türküsü, Türk Gençliği

Belirtili Ad Takımları: Atatürk’ün Onuncu Yıl Söylevi, Tiyatronun Değeri, Robenson’un Kulübesi, Tom Savyer’in Maceraları...

Sıfat Takımları: Sivil Mustafa Kemal, Dört Kanatlı Kuş, Bencil Yolcu, En İyi Dost, Al Başlıklı Kız, Yatılı Öğrenci...

Adeylemler: Görmek, Bakmak; Görme, Anlama Merakı; Okumak Sanatı, Gülmek, Cumhuriyeti Sevmek...

Ortaçlar: Verilen Sözü Tutmak, Gündüzünü Kaybeden Kuş, Yükselen İnsanlık...

Zamirler: Bizim Akdeniz, Bizimkiler, Biz ve Onlar, Bizim Köy...

Kısa Cümleler: Mustafa Kemal’ler Tükenmez, Tabur Karanlıkta And İçti, Bizim Köy Balıkçı Köyüdür, İstanbul’u Dinliyorum...

Kısa Soru Cümleleri: Nasıl Yazılmalı? Fen Adamları Nasıl Çalışır? Bu Vatan Kimin? ...

2.2.3 Paragraf Planı

Paragraf, “bir yazının iki satır başı arasındaki kısmına”, daha geniş anlamda bir yardımcı düşünceyi işleyip, geliştiren cümleye veya cümle grubuna denir. Bir nesir, bir zincirin halkaları gibi, paragrafların birbirine bağlanmalarıyla meydana gelir [12].

Paragrafta birlik her cümlenin anadüşünce cümlesine sıkı sıkıya bağlanmasıyla sağlanır. Anadüşünce cümlesi, ya doğrudan doğruya belirtilir ya da dolaylı olarak. Paragrafta yer alan bütün cümleler, anadüşünce cümlesine göre gelişmeleri kontrol edilerek arasındaki bağlantı sağlanır, birlik ve bütünlük gerçekleşir.

Anadüşünce cümlesi her zaman paragrafın başında bulunmaz. Bu, yazarın tutumuna bağlıdır. Yazar, önce okuyucunun ilgisini uyandırmak, onu konu üzerinde düşündürmek istiyorsa anadüşünceyi paragrafın ortasında ya da sonunda belirtir. Bunu, anlatımı bir örneklikten kurtarma, ona bir çeşitlilik kazandırma amacıyla da yaparlar.

Paragraflar, bir ağacın ana kolları, cümleler ise bu kollara bağlı dallar gibidir. Bu kollar ve dallar birleşerek koca bir gövdeyi meydana getirirler. Onun için paragrafların birbirlerine manaca bağlı olmaları, bir düşünce birliği içinde bulunmaları gerekir. Bu düşünceler de plânlı olarak sıralanmalıdır. Sonucun ne olacağı daha ilk paragrafta belli oluyorsa, okuyucu yazının bütününe okumaya ihtiyaç duymaz.

Her paragraf, genellikle hüküm bildiren bir cümle ile başlar. Fikirler değiştikçe paragraflar da değiştirilir. Bu bakımdan, yazı yazarken, paragraf yapmanın türlü faydaları görülür: yazının okunması kolaylaşır; çabuk kavranması sağlanır; okuyucuya bıkkınlık verilmez ve dikkati dağıtılmaz.

Paragraf, bir yardımcı düşünceyi işleyip geliştiren cümleler grubu olduğuna göre, paragraf meydana getiren her cümle, kendinden önceki cümleye dil ve düşünce yönünden bağlı olmalı; ilgili yardımcı düşünceyi açıklayıcı, geliştirici, sonuca götürücü bir özellik taşımalıdır.

Paragraf bir düşünce birimi olduğuna göre, her cümlenin birbirine iyice zincirlenmesi, bir anlam bütünlüğü yaratması gerekir. Bu da cümle düzeniyle ilgilidir. Her cümle, kendinden öncekine hem dil hem de düşünüş yönünden iyice bağlanmalıdır. Bir cümleden diğerine geçerken doğal geçişler sağlanmalı; boşluklar bırakılmamalıdır.

Paragraf bir düşünce birimidir ama, o düşünceyi destekleyecek, açıklayacak, örnekleyecek başka düşünceler de gerekir. İşte bunların hepsi bir anadüşünceyle ilgili olmakla birlikte aralarında birtakım boşluklar olabilir. İyi düzenlenmiş bir paragrafta bu boşluklar arasında köprüler kurulur. Bu köprü işini gören sözcüklere “bağlayıcı öğeler” denir. Bağlayıcı öğeler, sözcük, sözcük öbeği, cümlecikler, sözcük ve düşünce yinelemesidir.

Paragrafın uzunluğu için kesin bir şey söylenemez. Bunun için bir ölçü koymak gerekirse paragraf, bir düşünceyi tam açıklayacak kadar uzun, okuyucunun ilgisini canlı tutacak kadar da kısa olmalıdır. Bir başka deyişle, paragrafın uzunluğu şu üç etmene bağlıdır: Konu, okuyucunun durumu, yazarın tutumu. Çünkü bir yazar, paragraf kurmanın bütün kurallarını bilir de, gene yapıca sağlam bir paragraf kuramaz. Bu, kural bellemeden, onlara bağlı kalmadan ziyade, konunun eksiksiz anlaşılmasına, okuyucunun tam tanınmasına, amacın açıkça bilinmesine, söyleneceklerin iyi seçilmesine bağlıdır [14].

Bir yazar, yazısını yazarken, yazının düzeni için üç tür paragraf uygular:

- Giriş paragrafı: Okuyucuya konuyu tanıtmak, onu konuya yöneltmek gayesini taşır. Bu sebepten giriş paragrafının kuvvetli ve tesirli bir ifadesi olması gerekir. Fıkra, röportaj, makale gibi kısa yazılarda “giriş paragrafı” bir tanedir. Roman, hikâye, tiyatro oyunu gibi uzun türlerde ise olay, çevre ve olayı yaratan kişiler tanıtılacağı için, “giriş paragrafı” iki, üç veya daha fazla olabilir.
- Gelişme paragrafı: Giriş paragrafını takip eden ve birkaç paragraftan meydana gelen bölüm olup, sonuç bölümünden önce bulunur. Bu sebepten gelişme paragrafına giriş ile sonuç arasındaki paragraftır diyebiliriz. Bu paragrafta konu, türlü yönleriyle açıklanır, tartışılır, olgunlaştırılır; tasvirler yapılır, karakterler tanıtılır, adım adım sonuca yaklaşılır.
- Sonuç paragrafı: Yazıyı sona erdiren paragraf olup, bu paragrafta konu derli toplu birkaç cümle ile özetlenir. Konuyla ilgili bir hüküm verilmek isteniyorsa, bu paragrafta ifade edilir [12].

3. YAKIN ÇALIŞMALAR

Tez çalışmasının bu bölümünde metin özetleme ile ilgili yapılmış çalışmalar altı başlık altında anlatılmıştır:

1. Geleneksel yaklaşımlar
2. Derlem tabanlı yaklaşımlar
3. Hitap yapısının kullanılması
4. Zengin-bilgi yaklaşımları
5. Değerlendirme yöntemleri
6. Yeni özetleme problem alanları [8]

3.1 Geleneksel Yaklaşımlar

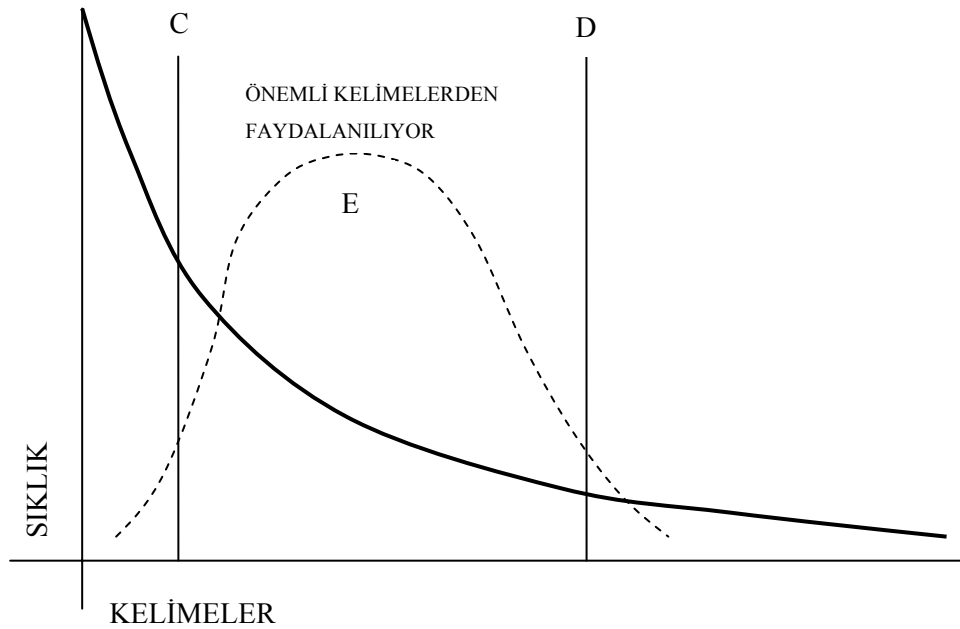
Geleneksel yaklaşımlar, 40 yıldan beri uygulanırlar. Bu yaklaşımları geleneksel yapan; sonradan ortaya çıkan araştırmaları teşvik ettiği gibi modern pratik uygulamalara da temel oluşturmasıdır. Bu yaklaşımdaki makaleler yüzey seviyesinde yaklaşım kullanırlar. Luhn makalesinde; önemli bilginin *terim sıklığına* dayanarak seçilmesine dayanan istatistiksel yöntemi açıklar. Sonraki makalede, Edmundson, *terim sıklığı* yöntemini, *ipucu sözcük öbekleri*, *başlık sözcükleri* ve *cümle konumu* yöntemleri ile karşılaştırır. Pollock ve Zamora tarafından yazılmış makale ise; kimya alt alanlarına has *ipucu sözcük öbekleri* yöntemi kullanımına dayanan, Kimya Özetçe Hizmeti'nde yapılmış bir özetçe programıdır [8].

Mani ve Maybury'ye göre metin özetlemenin temelini atan bu sistemler metin özetleme disiplini de başlatır. Özellikle bugünün ticari uygulamalarında, sözcük sıklığı, ipucu sözcük öbekleri, başlık yöntemi ve cümle konumu yöntemleri hala önemli yöntemler olarak kalmaktadır. Bu yöntemlerde sistemlerin başarımı genel olarak insan ve makine performansı karşılaştırılarak ölçülür. Bugün hala üzerinde kaygı bulunan konular konu ilk ortaya atıldığından beri varlığını sürdürmektedir. Bunlardan bazıları: Özetleme için en güçlü ve en genel özellikler nelerdir? Bu

özellikler nasıl birleştirilip birlikte kullanılırlar? Sistemin başarımı nasıl ölçülmeli [8]? Otomatik metin özetleme alanındaki ilk çalışmalar olmasına rağmen hala en çok kullanılan, başarımı ve uygulaması en yüksek olan yöntemler bu geleneksel yaklaşım yöntemleridir.

3.1.1 Terim Sıklığı Yöntemi

Luhn'un [15] bulduğu bu yöntemin temel ilkesi bir yazarın bir tezi savunurken veya bir konu hakkındaki düşüncelerini aktarırken belli sözcükleri tekrar etmesidir. Buna göre vurgu, önemliliğin bir göstergesidir. Fakat en sık geçen sözcükler günlük hayatta sık kullanılan sözcüklerdir. Bu sözcükler her tip belgede, her zaman en sık geçen sözcüklerdir, ama bunlar içerik belirtmez. Bu sözcüklerden oluşan sabit bir liste oluşturulup, özetleme sisteminde bunlar ihmal edilmelidir. Luhn, yöntem olarak da, yüksek sıklık için bir kesme değeri belirlenip, bu değerin üzerindeki alınılmasını önerir. Sözcük sıklığı, yöntemin temel ölçütü olduğundan, benzer şekilde bir alt kesme değeri de belirlenmelidir. Bu iki kesme noktasının en iyi değerleri, yüksek örnek sayılarına ulaşılarak bulunmalıdır [15]. (bkz. Şekil 3.1)

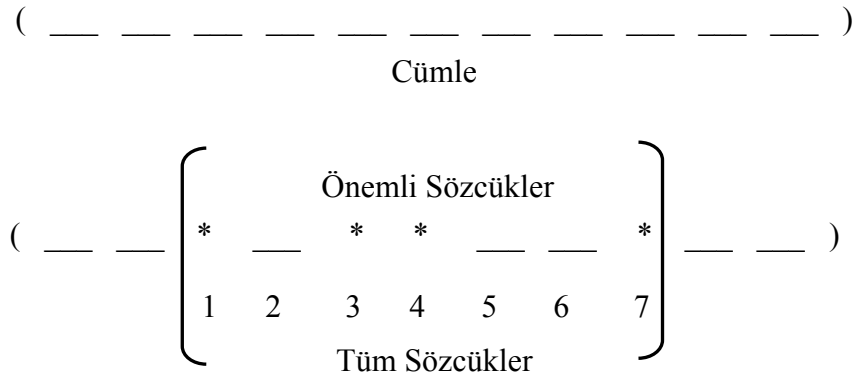


Şekil 3.1: Luhn'un sözcük sıklık çizelgesi

Biçimbirimsel farklılıkları ortadan kaldırmak için sözcükler normalize edilir. Bu normalizasyon, eklemeli dillerde kullanılan karmaşık gövdeleme algoritmaları gibi değil, sözcüklerdeki benzer harf sayılarına dayanarak yapılır. Buna göre; iki sözcük

her karakter konumunda eşleştirilir ve eşleşmeyen bir karakter bulunduğunda kalan karakter sayısı hesaplanır ve eğer 6 veya 6'dan az ise aynı sözcük oldukları varsayılır (Örneğin; “similar”, “similarity”) [8].

Her cümleye sözcük sıklıklarına göre puan verme işlemi, her cümlenin en az bir önemli terim ve dörtten fazla önemsiz terim içermeyecek şekilde parçalara bölünmesi ile devam eder. Her parçadaki önemli sözcük sayısının karesi, parçadaki toplam sözcük sayısına bölünür. En yüksek puana sahip parçanın puanı, cümlenin puanı olarak seçilir. Daha sonra en yüksek puanlara sahip cümleler, Şekil 3.2’de görüldüğü gibi özet cümleleri olarak seçilir [15].



Şekil 3.2: Luhn’un cümle önem derecesi hesaplama yöntemi

Luhn ayrıca alana bağlı olarak seçilmiş sözcüklere ek prim getirecek yöntemden bahseder. Luhn’ın öngördüğü özetleme kullanım alanları; çeviri yapılması gereken makalelerde, tüm makale yerine özeti çevirisinin yapılması, BE sistemlerinde bu yöntem kullanılarak dizin terimlerinin çıkarılmasıdır [8].

3.1.2 Otomatik Çıkarımda Yeni Yöntemler

Edmundson 1969 yılında, sözcük sıklığına ek olarak, üç yeni özellikten bahseder. Bunlar; *ipucu sözcük öbekleri*, *başlık sözcükleri* ve *cümle konumu*. Edmundson yönteminde bir derlem kullanır. Derlemin bir kısmını test amacıyla, diğer bir kısmını da yöntemlerindeki parametreleri bulmak için kullanır.

İpucu sözcük öbekleri yöntemi; bir cümlenin konu ile ilgili olma ihtimalinin, “önem”, “imkansız”, “ancak” gibi yararlı sözcüklerin varlığından etkilenme hipotezine dayanır. Yöntem daha önceden derlemden seçilmiş terimleri ipucu sözcük öbekleri

olarak kullanır. İpucu sözcük öbeklerinden oluşan bu sözlük, üç alt sözlükten ibarettir: Olumlu şekilde konu ile ilgili *olumlu sözcükler*, olumsuz şekilde konu ile ilgili *olumsuz sözcükler* ve konu ile ilgisiz *ilişkisiz sözcüklerdir*. Her cümle için son *ipucu ağırlık*, o cümlenin sözcüklerinin anahtar ağırlıkları toplamıdır [16].

Anahtar yöntemi temelde Luhn'un yöntemine [15] dayanmaktadır. Yöntemin dayandığı hipoteze göre, içerikte yüksek sıklıkta bulunan sözcükler olumlu şekilde konu ile ilgilidir. Yöntem her belge için, idealde belgenin gövdesinden istatistiksel olarak seçilen konu sözcüklerinden oluşan sözlük oluşturur. *İpucu sözlüğünde* olmayan anahtar yöntemine göre bulunan sözcükler azalan sıklığa göre sıralanır. Belli bir eşik değerine kadar olan ve ipucu sözlüğünde olmayan sözcükler, sıklıkları kadar olumlu ağırlıklandırılırlar. Her cümle için son *anahtar ağırlık*, cümlenin sözcüklerinin anahtar ağırlıkları toplamına eşittir [16].

Başlık yöntemi, bir yazarın başlığının yazısında bahsettiklerini içine alacak şekilde yazacağı hipotezine dayanır. Ayrıca bir yazar yazdığı metni, bölümlere ayırırken ve bu bölümlere uygun başlık seçerken ilgili bölümü özetlemiş olur. Başlık sözcüklerinin olumlu şekilde konu ile ilgili olması, % 99 güven aralığı ile istatistiksel olarak kabul edilir. Başlık yöntemi, tüm belgelerin başlıklarındaki ve alt başlıklarındaki anlamsız olmayan her sözcükten oluşan bir sözlük derler. Başlık sözlüğündeki sözcüklere olumlu ağırlıklar verilir. Her cümle için son *başlık ağırlık*, cümlenin sözcüklerinin başlık ağırlıkları toplamına eşittir [16].

Edmundson'un ortaya koyduğu *cümle konumu yönteminin* dayandığı iki hipotez vardır. Birincisi; belirli başlık altındaki cümleler olumlu şekilde konu ile ilgilidir. İkincisi ise konu belirten cümleler ve onların paragrafları bir belgede çok önce veya çok geç görünme eğilimindedirler [16].

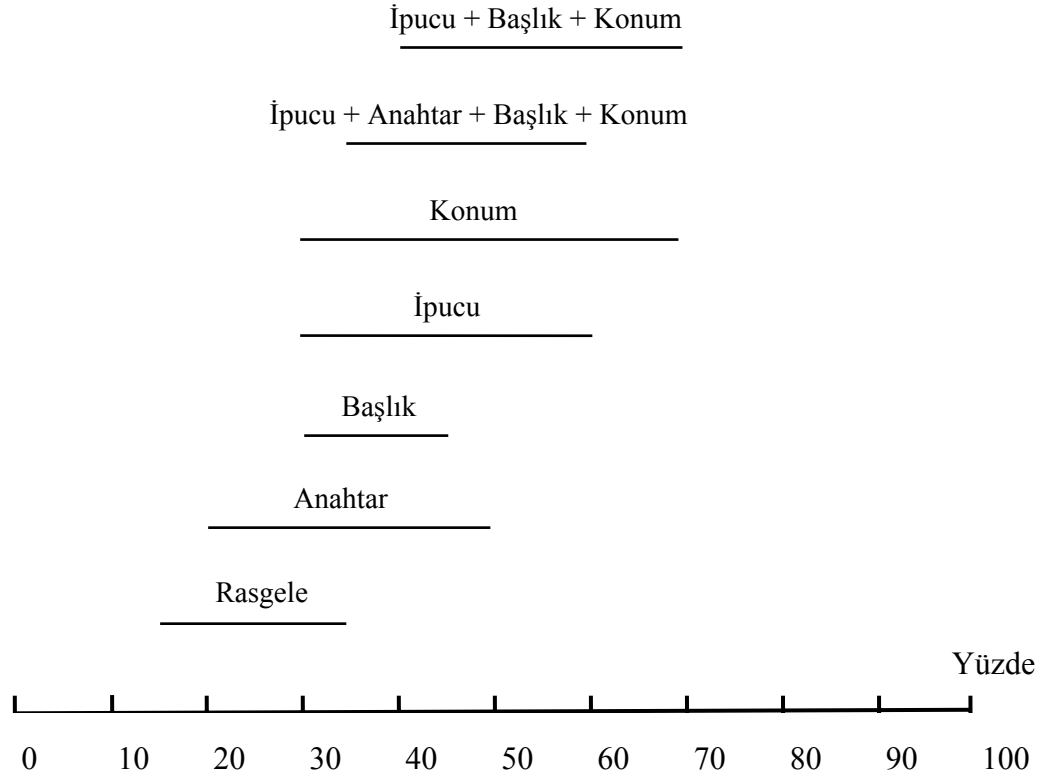
Sistemin sonunda, yukarıdaki dört yöntem sonucu hesaplanmış göreceli ağırlıklar lineer bir fonksiyonda parametrik hale dönüştürülür (Denklem (3.1)):

$$a_1I + a_2A + a_3B + a_4K \quad (3.1)$$

Burada a_1 , a_2 , a_3 ve a_4 İpucu, Anahtar, Başlık ve Konum ağırlıkları için pozitif tamsayı parametrelerdir.

Edmundson, her çıkarım yöntemi için hem ayrı ayrı hem de birleşimleri şeklinde, otomatik ve hedef çıkarımlardaki seçilen cümle sayısının yüzdeleri hesaplar. Buna

göre önemli ortalama yüzde değerler, örnek ortalaması artı eksi bir örnek standart sapma değerini çevreleyen aralıklarla birlikte Şekil 3.3'te gösterilmiştir. Rasgele çıkarım (cümlelerin %25'inin rasgele seçimi) ve otomatik çıkarımın sonuçları karşılaştırma amacı ile verilmiştir. Tek başına Anahtar yöntemi en düşük otomatik yöntem puanına sahipken, İpucu-Başlık-Konum yöntemlerinin birleşimi en yüksek ortalama seçim puanına sahip gözükmemektedir [16].



Şekil 3.3: Yöntemlerin ortalama seçim puanları

Bu sonuca göre Edmundson, tasarladığı çıkarım sisteminde Anahtar yöntemini çıkarır. Edmundson, her ne kadar dizinleme için önemli de olsa, Anahtar sözcüklerin çıkarım için önemli olmadığı sonucuna varmaktadır.

3.1.3 Kimya Özetçe Hizmeti'nde (KÖH) Otomatik Özetçe Araştırması

Pollock ve Zamora tarafından yapılan araştırma, belirli bir alana has otomatik çıkarımın başarımını ölçmeye yöneliktir. Yaptıkları sistem, kimya makalelerini %20 oranında sıkıştırarak çıkarım yapmayı amaçlar. Sistemleri temelde, kimya alt alanlarına ait ipucu sözcükler kullanılmasına dayanır. İpucu sözcük öbekleri, cümlenin özetinde yer alması açısından olumlu veya olumsuz olarak kullanılırlar.

Terim sıklıkları, daha çok ipucu sözcük öbeklerinin etkisini ayarlamak için kullanılır. Yönteme göre, çok sayıda olumlu sözcük öbeğinden oluşan metinler için daha uzun bir özet çıkarılmasını dengelemek için metinde sıklıkla yer alan olumlu sözcüklerin ağırlıkları düşürülür. Benzer şekilde sıklıkla yer alan olumsuz sözcüklerin de ağırlıkları düşürülür [8].

Diğer iki yöntem ile karşılaştırıldığında, Pollock ve Zamora eleme işlemini cümle sıkıştırma ile çözer. Sözcük listesindeki terimlerin, metindeki virgülleri sınıflandırmak için daha çok bağlamsal testlerde, daha sonra ise takip eden cümlecğin sınıflarının belirlenmesinde kullanılan, konuşma-parçası (POS) bilgisi de (isim, fiil, vb.) vardır. Giriş niteliğinde ve parantez içine alınmış, “that” ile biten öbekler (muhtemelen ardından bir sonucun geleceği) veya “in” ile başlayan (“in conclusion” gibi) cümlecikler son özette silinir. Bu çalışmadaki ilgi çekici başka bir özellik ise, çıktındaki sözcük dağarcığının normalizasyonu. Normalizasyon ile çıktının, sözcüklerin veya öbek kısaltmalarının ABD dışındaki yazımlarının standartlaştırılmasından ve kimyasal bileşik isimlerinin formülleriyle değiştirilmesinden oluşan KÖH standartlarını karşılaması sağlanır. Yazarlar yöntem sonucunda, otomatik oluşturulan özette kalitesinin elle oluşturulandan daha düşük olduğu sonucuna varmakla birlikte, otomatik oluşturulanların da fonksiyonel olarak yeterli olduğuna inanırlar [8].

3.2 Derlem Tabanlı Yaklaşımlar

Edmundson’un çalışmasını, terim istatistiğinin elde edilebileceği ve özelliklerin uygun birleşimine karar verilebileceği derlem tabanlı yaklaşımlar takip eder. Bu bölüm altında değerlendirilen makaleler genel olarak yüzey seviyesinde yaklaşımları kapsar. Kupiec, Pederson ve Chen (KPC) makalelerinde cümle çıkarımı için Bayes sınıflandırıcısı kullanımını açıklarlar. Sınıflandırıcı; özetçe ile karşılaştırıldığında çıkarım için değerli olarak etiketlenmiş tam metin cümlelerden oluşan *özellik vektörleri* kullanılarak eğitilir. Edmundson’un makalesindeki gibi, KPC’nin kendi verileri üzerinde bulduğuna göre; *cümle konumu* yöntemi tek başına en iyi özellik, özellikler birleştirildiğinde ise *cümle konumu*, *ipucu sözcük öbekleri* ve *cümle uzunluğu* yöntemlerinin birleşimi en iyi özellik birleşimini oluşturmaktadır. Myaeng ve Jang tarafından yazılmış sonraki makale, Korece yazılmış teknik makalelere uygulanan, KPC yönteminin değişik bir biçimini açıklar. Sonucunda, *ipucu sözcük*

öbekleri, cümle konumu ve başlık sözcükleri yöntemlerinin birleşiminin en iyi sonucu verdiğini bulurlar. Aone, Okurowski, Gorlinsky ve Larsen tarafından kaleme alınan bir diğer makale, KPC yaklaşımı ile birlikte terim tabanlı istatistiklerin kullanımında, biçimbirimsel, eşanlamsal ve tam isim çeşitliliklerini kullanmanın özetleme başarımını etkilediğini açığa çıkarır. Bu bölümdeki Hovy ve Lin tarafından yazılmış diğer bir makale üç fazlı yaklaşımın daha uzak bir örneğini tartışır ve konuma bağlı bilginin tanımlanması için yeni bir yöntem tanıtır. Ek olarak, makale, metindeki kavramları genelleştirmek ve toplamak için bir eşanlamlılar tabanlı yöntemden bahseder [8].

Derlem tabanlı yaklaşımlarda çok yaygın bir kullanım ağırlıkların terim sıklığına göre hesaplanmasıdır. Metin özetleme ile birlikte bilgi erişimi'nde de sıklıkla kullanılan, *tf.idf* ölçüsü bir belgeyi derlemdeki diğer belgelerden ayırt etmede kullanılır. Bu ölçütün birçok çeşidi olmakla beraber, bunlardan biri Denklem (3.2)'de gösterilmiştir:

$$w_{ij} = tf_{ij} * \log_2 \frac{N}{n} \quad (3.2)$$

Burada w_{ij} , d_i belgesindeki t_i teriminin ağırlığıdır. tf_{ij} ise d_j belgesindeki t_i teriminin sıklığı, N derlemdeki belge sayısı, n , derlemdeki t_i teriminin bulunduğu belge sayısını temsil eder.

Derlem tabanlı yaklaşımlarda karşılaşılan birtakım zorlukları Mani ve Maybury [8] şöyle özetlemiştir:

- Uygun metinlerden oluşan bir derlem oluşturmak. Özetlemede kullanılabilir, temsili örneklerden oluşan, kullanıp dağıtma izinleri alınmış, gerekli açıklayıcı notlar eklenmiş metinlerden oluşan bir derlem oluşturmak gereklidir.
- Uygun özet kümesinin olduğu sağlanmalı. Özetleri olmayan belgelerden oluşan bir derlem MÖ için pek yararlı olmayacaktır. Bilimsel makalelerin özetçeleri mevcuttur, fakat diğer türlerdeki metinler için bu geçerli olmayabilir. Özetler bir yazar tarafından, profesyonel özet çıkaran kişiler tarafından veya bilirkişilerden özet cümlelerinin çıkarımını istemek yöntemleriyle temin edilebilir. Özeti yazar tarafından sağlanmadığı

durumlarda, bu özetlerin oluşturulmasındaki ölçütlere dikkat edilmelidir. Ayrıca, yazar tarafından sağlanan özetçeler, belirli kuralları uygulamak üzere eğitilmiş profesyonel özet çıkaran kişilerin hazırladığı özetler kadar sistematik olmayacaktır.

- Görünmeyen test verisi bakımından, yöntemlerin değerlendirilmesi, gerekli eğitim malzemesinin miktarı, elle yapılan mühendislik birikimine göreli olarak maliyet ve yararı ve bazı kullanıcı işlerine etkisi
- Bu yöntemler genişletilerek, tutarlı özet ve özetçelerin üretimi
- Özetle ilişkili bilginin ayrımının sağlanmasına yönelik, farklı türler için yeni özellikler bulunması
- Kavram seviyesinde toplanmış terim istatistiklerinin kullanılması
- Özetlemede kullanılan genellemeleri çeşitlendirecek, derlem kullanımının kavram hiyerarşisi ile birleştirilmesi

3.2.1 Eğitilebilir Belge Özetleyici

Kupiec, Pederson ve Chen (KPC) tarafından ileri sürülen bu yöntem, özellik kümelerinin birlikte kullanımı ve bu birlikte kullanım için derlem ile eğitimi önerir. KPC'nin kullandığı özellikler şöyledir:

3.2.1.1 Özellikler

Cümle Uzunluğu Kesme Özelliği'nin temeli, kısa cümlelerin özette yer almama eğiliminde olmasıdır. Verilen eşik değerine göre (örneğin 5), değerden uzun olan cümleler için özellik doğru, aksi durumda da yanlış olacaktır.

Belirli Sözcük Öbeği Özelliği'ne göre, belirli sözcük öbekleri listesindeki genellikle iki sözcük uzunluğundaki öbeklerden (örneğin; “bu mektup...”, “sonuç olarak...”) herhangi birini barındıran cümleler veya “özetle”, “sonuçlar”, vb. anahtar sözcüklerin birini barındıran cümlelerden sonra gelen cümlelerin özette olma ihtimali vardır. KPC'ye göre bu özellik, belirlenen 26 sözcük öbeğinden birini içeren cümleler veya özel anahtar sözcüklerden birini içeren cümlelerden sonra gelen cümleler için doğrudur.

Paragraf Özelliği belgedeki ilk on ve son beş paragrafı tutar. Paragraftaki cümleler, paragrafın ilk cümlesi olmasına, paragrafın son cümlesi olmasına (bir cümleden uzun

paragraflar için) ve paragrafın ortasında olmasına (paragraf uzunluğu iki cümleden büyük olan paragraflar için) göre ayrılır.

Konuya Has Sözcük Özelliği'ne göre en sık gözlenen içerik sözcükleri konuya has sözcükler olarak belirlenirler. Bu sözcüklerden bir kısmı seçilir ve bunların geçtiği cümleler, sözcüklerin sıklıklarına bağlı bir fonksiyonla puanlandırılırlar. Bu özellik ikilidir. Yani bir cümlenin en yüksek puanlı cümleler kümesinde olup olmamasına bağlıdır. Ölçeklenmiş cümle puanları sözde olasılıklar olarak kullanılarak deneyler yapıldığında ikinci derecede başarımlar sağlanmıştır.

Büyük Harfli Sözcük Özelliği temelde tam isimlerin ve kısaltmaların genelde önemli olduğu ilkesine dayanır. Bir büyük harfli sözcüğün, cümle başı olmayıp büyük harfle başlama kısıdı ile bir önceki özelliğe benzer şekilde hesaplanır. Ayrıca, sözcük birkaç kere tekrarlanmalıdır ve kısaltma ölçü birimi olmamalıdır (örneğin F, C, Kg). Bu tür sözcükleri içeren cümleler, bu sözcükler ilk kez karşılaşıldığında sonraki rastlanmalara göre iki kat puanlandırılmalıdır [17].

KPC tanıttığı özelliklere ek olarak bu özelliklerin birleştirilmesi için Bayes sınıflandırıcısı kullanır. Buna göre:

Verilen k tane özellik için, her s cümlesi için S özetinde olma olasılığı Bayes kuralı kullanılarak Denklem (3.3)'teki gibi ifade edilebilir:

$$P(s \in S | F_1, \dots, F_k) = \frac{P(F_1, \dots, F_k | s \in S)P(s \in S)}{P(F_1, \dots, F_k)} \quad (3.3)$$

Özelliklerin istatistiksel olarak bağımsız olduğu varsayımı yapılırsa:

$$P(s \in S | F_1, \dots, F_k) = \frac{\prod_{j=1}^{j=k} P(F_j | s \in S)P(s \in S)}{\prod_{j=1}^{j=k} P(F_j)} \quad (3.4)$$

$P(s \in S)$ sabittir ve $P(F_j | s \in S)$ ve $P(F_j)$ gözlemlenme sayısı sayılarak eğitim kümesinden direkt olarak bulunabilir. Tüm özellikler ayrık olduğundan, formül olabirlik yerine olasılık cinsinden yazılabilir. Bu, üretilmiş özetinde yer almak üzere cümle seçiminde kullanılabilir, her s cümlesine puan atayan basit bir Bayes sınıflandırıcı fonksiyonuna dönüşür [17].

3.2.1.2 Derlem

KPC derlemi hazırlarken profesyonel özet çıkaran kişilerden destek alır. Söz konusu derlemde, Tablo 3.1’de görülen, bilimsel/teknik alanda 21 yayından örneklenmiş 188 belge/özet çifti yer almaktadır. Bu özetler genel olarak belirticidir ve ortalama uzunluğu üç cümledir. KPC hazırlanan derlemdeki belge başına ortalama cümle sayısını 86 olarak hesaplamıştır [17].

Tablo 3.1: KPC derlemindeki dergiler

Amerika Uzay ve Havacılık	Üretim Mühendisliği
Amerika Laboratuvarı	Metal Bitirme
İnşaat Mühendisliği	Modern Plastikler
Kimya Mühendisliği Eğitimi	Yağ ve Petrol Dergisi
Uluslararası Beton	Uluslararası Hamur ve Kâğıt
IEEE İletişim Dergisi	Robotbilim Dünyası
IEEE Kontrol Sistemi	Metalürji ve Malzeme Mühendisliği
Hücrel Plastik Dergisi	Duyargalar
Malzeme Bilimi Harfleri Dergisi	Su Mühendisliği ve Yönetimi
Japon Demiryolu Mühendisliği	Uluslararası Kablo Birliği ‘93
Makine Tasarımı	

Yöntem kullanacağı derlemde belge/çıkarm çiftlerine ihtiyaç duymaktadır. Fakat KPC’nin elinde, ana metinden esinlenilerek elle hazırlanmış özetler bulunmaktadır. Dolayısıyla, özetleme işi, bir uzmanın özet metni hazırlarken, içeriği koruyarak kelimesi kelimesine veya küçük değişikliklerle kullanabileceği, bir belgeden cümleleri çıkarımdır [17].

3.2.1.3 Cümle Eşleştirme

Eğitimi sağlayabilmek için elle hazırlanmış özet cümleleri ile asıl metindeki cümlelerin eşleştirilmesi gerekmektedir. Bu eşleştirme işlemi sonuçlarını KPC şöyle özetler: Elle çıkarılmış özet cümlesi, birebir veya küçük değişikliklerle asıl belgeden çıkarılabilmesi durumunda *direkt cümle eşleştirme* gözlenen durumdur. Elle çıkarılmış özet cümlesini oluşturabilmek için asıl belgeden iki veya daha fazla cümlelerin kullanıldığı durumu *direkt birleşim* olarak adlandırır. Eğer yazarın özet cümlesini genel okuma sonucunda (asıl belgeden belirli bir cümle kullanmadan) çıkardığı açıkça gözleniliyorsa veya böyle bir ihtimal söz konusu ise cümle *eşleştirilememiş* olarak etiketlenir. Şu iki durumda da özet cümleleri *tamamlanmamış* olarak etiketlenir: Birincisi, özet cümle ile asıl cümle arasında bir kesişimin söz konusu olduğu fakat asıl cümlelerin içeriğinin özette korunmadığı durum. İkincisi ise,

özet cümlesinin asıl belgeden bir cümleyi içerdiği fakat buna ek olarak direkt birleşim ile eşleştirilmemiş başka bilgiyi de içerdiği durumdur [17]. Buna göre KPC'nin derlemi üzerinde yaptığı özet ile asıl metin eşleştirme çalışmasının sonuçları Tablo 3.2'deki gibidir.

Tablo 3.2: KPC'nin eşleştirme dağılımı

Direkt Cümle Eşleştirme	451	%79
Direkt Birleşim	19	%3
Eşleştirilememiş Cümleler	50	%3
Tamamlanmamış Tek Cümleler	21	%4
Tamamlanmamış Birleşikler	27	%5
Toplam Elle Çıkarılmış Özet Cümleleri	568	

3.2.1.4 Değerlendirme

Derlemdeki belge yetersizliğinden değerlendirme kısmında KPC, bir dergideki belgeleri test için kullanılırken diğer tüm belge ve özetleri eğitim için kullanır. Eşleştirilemeyen veya tamamlanmamış olarak etiketlenmiş cümleler eğitimden ve testten çıkarılır. Bunlar çıkarıldığında toplam 498 benzersiz cümle sayısına ulaşırlar. Başarımı ölçmek için iki yöntem kullanılır:

Birinci yöntem, elle çıkarılmış özet cümlelerinin özetleyici program tarafından tam olarak tekrar üretilmiş cümlelere oranıdır. Dolayısıyla başarımlar, metin alıntıları ile limitlidir ve erişilebilecek en yüksek başarımlar tüm direkt cümle eşleştirmeleri ve direkt birleşimlerin toplamıdır. Tablo 3.2 referans alınarak:

$$\frac{451+19}{568} = \%83 \quad (3.5)$$

Özetleyici tarafından üretilen bir cümle şu koşulları sağladığı takdirde doğrudur:

1. Direkt cümle eşleştirmesi vardır ve elle çıkarılmış özetinde mevcuttur.
2. Direkt birleşimin bir parçası olarak elle çıkarılmış özetinde vardır ve birleşimin diğer tüm üyeleri üretilebilmiştir (böylece birleşimdeki tüm bilgi korunmuş olur) [17].

Her test belgesi için, eğitilmiş özetleyici elle çıkarılmış özetindeki cümle sayısı ile aynı sayıda cümle üretir. 568 cümle için, 201 tane doğru belirlenmiş özet cümlesi için, 195 cümle, direkt cümle eşleştirme ve 6 cümle de direkt birleşim olarak belirlenir.

Dolayısıyla özetleyici elle çıkarılmış özetin %35'ini çıkarır. Fakat bu sonuç bir belge için sadece bir tane “doğru” özet olduğu varsayımı üzerine çıkarılmıştır ki bu durum pek de muhtemel değildir. Gerçekten de yapılan çalışmada özet cümlelerin çıkarılması istendiğinde çok çeşitlilik gözlenmektedir. KPC bu durumu Rath'e referans vererek şöyle der: “Rath, dört farklı bilirkişi tarafından yapılan çıkarım seçiminde sadece %25 kesişim olduğunu gözlemlemiştir. Bir bilirkişinin belli zaman sonra yaptığı çıkarımda da sadece %55 kesişim olduğu gözlemlenmiştir.” [17]

İkinci değerlendirme yöntemi ise özetleyici tarafından doğru bir şekilde tanımlanabilen 498 eşleştirilebilir cümleye oranıdır (dolayısıyla teorik olarak %100 başarımla elde etmek mümkün). Özetleyiciye, elle çıkarılmış özetdeki cümle sayısı ile aynı miktarda cümle için özet çıkarttırıldığında, 498 cümleden 211 tanesi (%42) doğru şekilde bulunur [17].

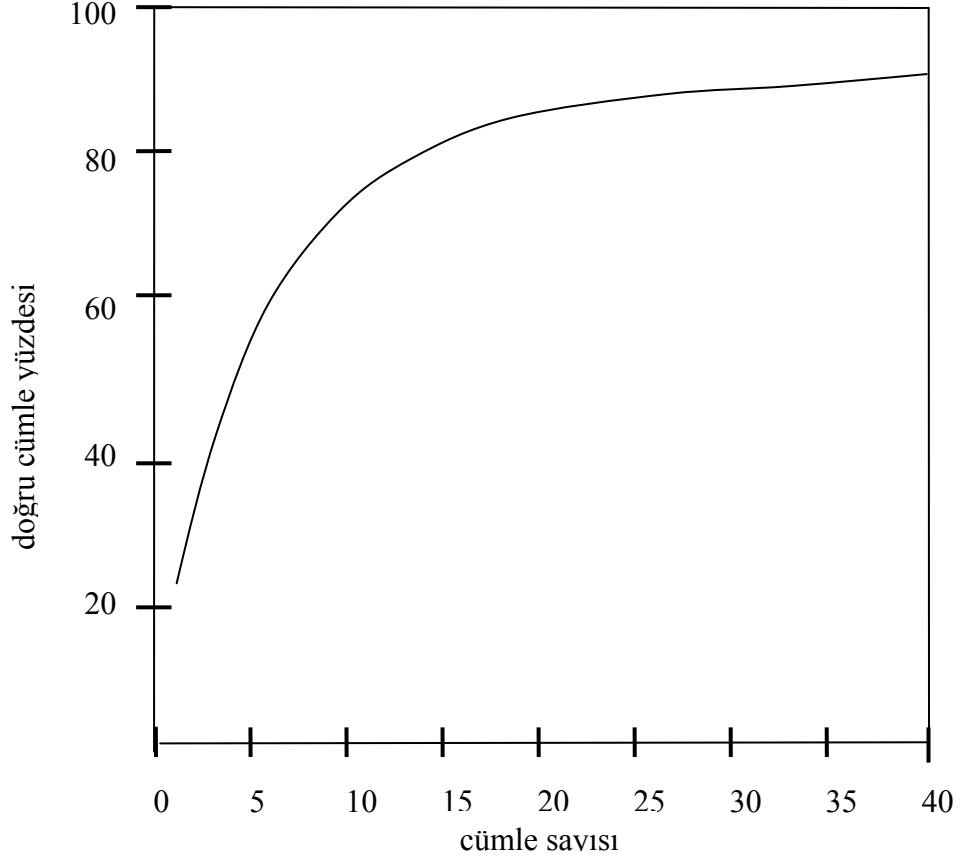
Tablo 3.3'ün ikinci sütunu özelliklerin tek başına cümle bazında başarımlarını gösterir. Cümlelerin eşit olasılıklara sahip olduğu durumlarda belge sırasına göre sıralanırlar. Dolayısıyla, tek başına kullandığında, cümle uzunluğu kesme özelliği, belgenin başındaki başlıklar dışında kalan metnini döndürür [17].

Tablo 3.3: Özelliklerin başarımları

Özellik	Tek Başına Cümle Doğruluğu	Birikimli Cümle Doğruluğu
Paragraf	163 (%33)	163 (%33)
Belirli Sözcük Öbeği	145 (%29)	209 (%42)
Cümle Uzunluğu	121 (%24)	217 (%44)
Konuya Has Sözcük	101 (%20)	209 (%42)
Büyük Harfli Sözcük	100 (%20)	211 (%42)

Tablo 3.3'teki üçüncü sütun, özellikler başarılı şekilde birleştirilip bir araya getirildiğinde başarımların nasıl değiştiğini gösterir. Özelliklerin en iyi birleşim şekli paragraf+belirli sözcük öbeği+cümle uzunluğu'dur. Sık geçen anahtar sözcükler özelliğinin eklenmesi (konuya has sözcük ve büyük harfli sözcük) genel başarımda küçük bir düşüşe neden olur.

Tüm özellikler kullanılarak özetleyicinin başarımları Şekil 3.4'te gösterilmiştir. Asıl belgedeki cümlelerin %25'ini otomatik olarak seçen özet oluşturulurken, Edmundson %44 cümle seviyesinde başarımdan bahseder. Benzer şekilde, KPC'nin derlemindeki ortalama belge uzunluğunun (86 cümle) %25'i yaklaşık 20 cümledir. Şekil 3.4'e göre bu da %84 başarıma karşılık gelir.



Şekil 3.4: Özet büyüklüğü – başarıım

3.2.2 İstatistiksel Tabanlı Belge Özetleme Sistemi

Myaeng ve Jangbu makalelerinde, KPC'nin kullandığı yöntemin bir çeşidini Korece teknik metinlere uyguladılar. Giriş ve gelişme kısmındaki malzemeleri tek başına değerlendirirler ve buradaki her cümleyi, arka planı temsil etmesine, ana konu ile ilgili olmasına, belge yapısını açıklamasına veya geleceğe dönük yapılabilecek çalışmaları anlatmasına göre el ile etiketlerler. Özet cümlelerinin %96'dan fazlasının ana konu cümleleri olduğu sonucuna varırlar. Cümleler ayrıca insan tarafından elle oluşturulan bir özetle olup olmasına göre de ayrıca etiketlenirler. Eğitim yöntemleri öncelikle Bayes sınıflandırıcısı kullanarak bir cümlenin ana konuya bağlı olup olmadığını bulur. Ardından çoklu Bayes sınıflandırıcılarından gelen delilleri kullanarak bir cümlenin özete ait olup olmadığını belirler. Son olarak gereksiz cümleler için bir filtre uygulanır. Tablo 3.4'te de [18] görüldüğü gibi, ipucu sözcük, cümle konumu ve başlık yöntemlerinin birlikte kullanımının kendi veri kümeleri üzerinde en iyi sonucu verdiğini bulurlar [8].

Tablo 3.4: Farklı özelliklerin ve filtreleme yöntemlerinin karşılaştırılması

Kullanılan özellikler	Filtreleme yöntemleri			
	Filtresiz	Metin bileşenleri ile filtreleme	Anahtar sözcükler ile filtreleme	Metin bileşenleri ve anahtar sözcükler ile filtreleme
Tüm altı özellik	0.3865 (+5)	0.3949(+7)	0.3913(+6)	0.3996(+9)
İpucu sözcük olmadan	0.3116	0.3203	0.3172	0.3259
Konum özelliği olmadan	0.3544	0.3572	0.3594	0.3594
Metin bileşeni özelliği olmadan	0.4007(+9)	0.4167(+13)	0.4228(+15)	0.4228(+15)
Başlık benzerliği yöntemi olmadan	0.3439	0.3591	0.3487	0.3639
İpucu sözcük özelliği olmadan	0.3976(+8)	0.4033(+10)	0.4022(+9)	0.4075(+11)
Merkeziyet özelliği olmadan	0.4318(+17)	0.4337(+18)	0.4365(+19)	0.4389(+19)
Başlık + ipucu sözcük + konum özelliği	0.4274(+16)	0.4363(+18)	0.4420(+20)	0.4436(+20)

3.2.3 Güçlü DDİ Yöntemlerinden Bilgi Edinimi ile Eğitilebilir Özetleyici

Aone, Okurowski, Gorfinsky ve Larsen, KPC'nin kullandığı yöntemle benzer bir yöntem kullanırlar. Yazarların ana düşüncesi, terim tabanlı istatistiklerin kullanımında terimlerin bir araya getirilmesinde farklı yöntemler kullanımının özetleme başarımını etkilemesi ilkesidir. Örneğin, bir sözcüğün biçimbirimsel yönden farklı şekillerini aynı kök terim olarak saymaya ek olarak eşanlamlı çeşitliliklerini de aynı kök terim olarak saymak. Ayrıca, tam isimlerin ve bunların diğer isimlerinin de terimler olarak sayılması önerilebilir (örneğin, “International Business Machines” için “IBM”). Yazarlar, bu farklı tiplerdeki terimlerin ağırlıklandırılmasına dayanan bir cümle çıkarım algoritması ile elle çıkarılmış özetleri karşılaştırırlar. Yer isimleri ve organizasyon isimleri terim olarak nitelendirildiğinde ve insan isimleri filtrelendiğinde başarımın artabileceğini gösterirler. İnsan isimlerinin belirlenmesi ve gizlenmesi gerektiğine sebep olarak eğitim ve test için kullandıkları TREC derleminin (Harman ve Voorhees 1996) kişi odaklı olmamasını gösterirler [8]. Bu yöntemle, eğitim için farklı özellikler kullanılarak elde edilen sonuçlar Tablo 3.5'te [19] görülmektedir.

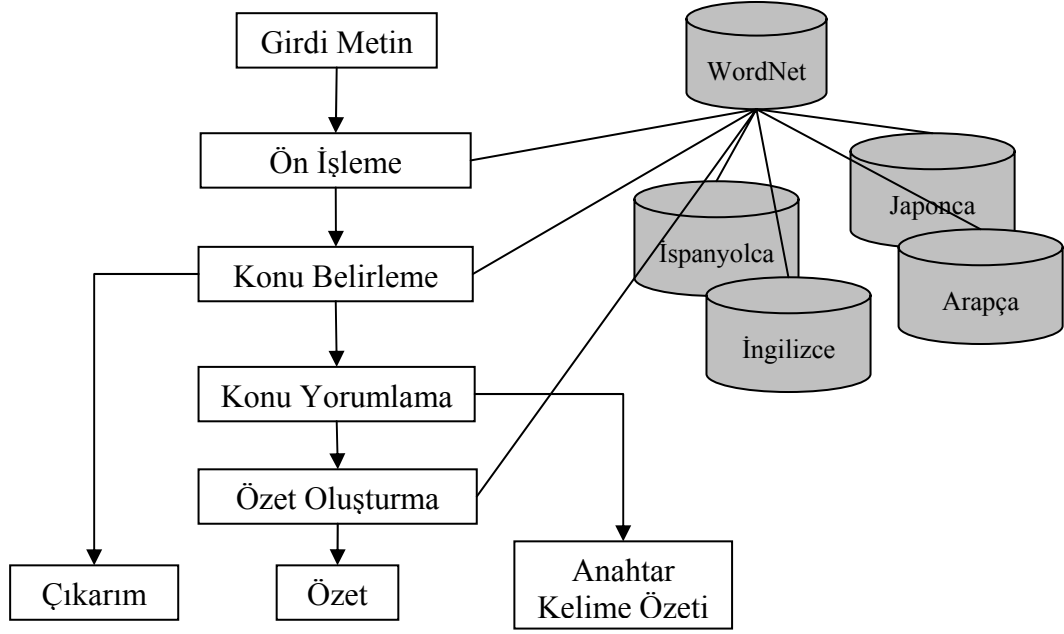
Tablo 3.5: Çeşitli eğitim özellikleri kullanılmasının sonuçları

F-Ölçüsü	Keskinlik	Anımsama	Uzunluk	Yüksek Puan	Konum	Paragraf
24.6	22.6	27.1	-	-	-	+
24.6	22.6	27.1	+	-	-	+
39.2	36.0	43.1	+	-	-	-
39.7	36.4	43.6	-	-	-	-
39.7	36.4	43.6	-	+	-	-
39.7	36.4	43.6	+	+	-	-

39.7	36.4	43.6	-	+	-	+
39.7	36.4	43.6	+	+	-	+
43.8	40.2	48.2	-	-	+	-
45.1	41.4	49.5	-	-	+	+
45.5	41.8	50.0	+	-	+	+
45.7	42.0	50.2	+	-	+	-
46.6	42.7	51.1	-	+	+	-
46.6	42.7	51.1	+	+	+	-
48.4	44.4	53.2	-	+	+	+
49.9	45.8	54.8	+	+	+	+

3.2.4 SUMMARIST'te Otomatik Metin Özetleme

Hovy ve Lin tarafından kaleme alınan bu makalede, özetlemede kullanılan birkaç derlem tabanlı yöntemden bahsedilir. Şekil 3.5'te [20] görüldüğü gibi, üç aşamalı bir süreçten bahsederler: konu belirleme, kavram yorumlama ve özet oluşturma. Konu belirleme, bir belgedeki daha sonra çıkarım için cümleleri ağırlıklandırmada kullanılacak olan belirgin kavramların çıkarılmasını amaçlar. Konuma bağlı bilginin bulunması için yazarlar yeni bir yöntem önerirler. Yöntemleri, anahtar sözcükler olarak konuların belirtildiği belgelerden oluşan bir derlemde eğitim yaparak, konu ile en ilişkili olan anahtar sözcükleri içerme eğiliminde olan, puanlamaya göre sıralanmış cümle konumları listesi oluşturur. Sistem, metinler, özetçeler ve anahtar sözcüklerden oluşan, 13,000 makalelik bir derlem (Ziff-Davis) üzerinde eğitilmiştir. Sonuçta bulunan En Uygun Konum Yöntemi önemli sonuçlara imza atar. Yazarlar ayrıca çeşitli özellik birleşim algoritmalarının kullanımını açıklarlar [8].



Şekil 3.5: SUMMARIST mimarisi

Hovy ve Lin'in kavram yorumlama (biraz başlangıç niteliğindeki) üzerine çalışması, daha genel konulara varmak için metindeki ilişkili kavramları bir araya getirmeyi dener. Uygun bir genelleştirmeye varmak için kavram hiyerarşisini kullanarak, WordNet kavramlar dizininden (Miller 1995) kavramların metinlerindeki sıklıkları sayarlar. Bu bir araya getirme ve genelleştirme işlemleri sadece çıkarımsal özetler yerine özetçe oluşturulmasında da kullanılabilir. Birçok beklenen ilişkinin bulunması WordNet'te zor olduğundan, metin sınıflandırmayı (metinleri mevcut sınıflara yerleştirme) ve metin kümelemeyi (metinleri kendi içinde gruplama) kapsayan diğer derlem tabanlı yöntemleri kullanırlar. Yazarlar metin sınıflandırma ile ilgili, 30,000 metinden oluşan, her makalenin 32 olası konu etiketi ile etiketlendirildiği Wall Street Journal derlemi üzerindeki tecrübelerini de aktarırlar. Bu her konu için, bir terim ağırlıklandırma metriğine göre puanlama sonucu en yüksek puana sahip 300 terime konu imzası gibi davranılır. Belgenin konusunu bulmak için test belgelerindeki terimler bu imzalarla eşleştirilir. Özet oluşturmada bu yöntemlerin başarımını ölçmek için ileriye dönük çalışmalar yapılması gerekmektedir [8].

3.3 Hitap Yapısının Kullanılması

Daha önce bahsedilen çalışmalar, yüzey seviyesindeki yaklaşımlara odaklanırken, bu alandaki çalışmalar diğer yaklaşımlara odaklanırlar. Bu alanda yazılmış makaleler hitap yapısını kullanırlar.

Boguraev ve Kennedy'nin ve Barzilay ve Elhadad'ın kaleme aldığı makaleler, metin bağlılığına dayanan ilişkileri kullanan, varlık seviyesinde yaklaşım kullanırlar. Boguraev ve Kennedy, sağlam sözdizimsel ayrıştırma ve terimler arasındaki anaforsal ilişki çözünürlüğüne dayanan tamlama terimlerini bir araya getirip kullanırlar. Barzilay ve Elhadad ise; eşanlam ve geniş anlam gibi ilişkilere dayanarak terimleri sözcüksel zincirler şekline getirirler ve bu zincirleri cümle seçiminde kullanırlar.

Bu alandaki diğer çalışmalar, metindeki makro seviyedeki ilişkilerin geniş çaplı hitap seviyesinde modelini oluştururlar. Marcu, metindeki cümleciklerden sözbilimsel yapı teorisi (SYT) ağaçları inşa etmek için ipucu sözcük öbeklerini kullanan sınırlandırılmamış metinler için sözbilimsel ayrıştırıcı kullanır. Bu ağaç yapısı daha sonra, cümlecik seçimi için önem fonksiyonu inşa etmek amacıyla kullanılır. Strzalkowski ve diğ. makalesi özetlemeye, verilen bir sorgu için parçaları puana göre sıralama sorunu gibi davranır. Buradaki sorgu kullanıcının girdiği sorgunun terimlerinin başlıkta olup olmamasından ve metinde sık tekrarlanan terimler olup olmamasından oluşturulur. Bağlam dışındaki çıkarımları yapabilmek için gereken bağlamsal bilgiyle sorguya bağlı bilgiyi birleştirmek için haber hikâyesi özetlerinin hitap yapısının basit bir hitap modelini kullanırlar. Teufel ve Moens ise yazdıkları makalede, özetleme işini ikiye böler: cümlelerin çıkarımı ve çıkarılan her cümle için hitap rolünün (yedi rolden oluşan kümeden) belirlenmesi. İki aşamada da KPC'nin yaklaşımında modellediği Bayes sınıflandırıcısı kullanılır [8].

3.4 Zengin-Bilgi Yaklaşımları

Özel alanlara ilişkin zengin özetleme gereksinimlerini modelleyen yaklaşımlardır. Bu alandaki çalışmaların hepsi görel olarak, özetleme sürecinin girdi olarak kullandığı zengin yapısal temsillerin oluşturulduğunu varsayarlar. Hitap yapısının kullanıldığı yaklaşımlar daha çok özetlemenin analiz sürecine yoğunlaşırken, bu alandaki çalışmalar dönüşüm ve sentez evresine yoğunlaşır.

Lehnert çalışmasında, hikâye özetlemek için, “grafik birimleri” etrafında kurulmuş hikâyelerdeki yapısal ilişkilere dayanan çıkarım tabanlı bir yöntem (uygulamaya geçirilmemiş) anlatır. Hahn ve Raimer yaptıkları çalışmada özetleme için üç aşamalı bir süreçten bahsederler: Önce, önem belirteçlerinden oluşan dağarcık paragraflara uygulanır. Ardından, konu açıklamaları belirlenir ve uygun biçimde paragraflar bazında bir araya getirilir. Son olarak da, hiyerarşik metin grafi oluşturmak için, konu açıklamalarına genelleme operatörleri uygulanır. Sonucunda da hitap seviyesi ve varlık seviyesi yaklaşımları birleştirilmiş olur.

McKeown, Robin ve Kukich, iki farklı özetleyici ile fırsatçı bir şekilde bilgiyi paketleyip cümleler oluşturmak için yöntem anlatırlar: revizyon yaklaşımı ile basketbol maç özetleri hazırlayan STREAK ve ağ planlama aktivitelerini hitap planlama ile özetleyen PLANDOC. Özetleyiciler tarafından uygulanan dilbilimsel operasyonlar, bu iki alandan insan özetlerinin derlem analizi tarafından harekete geçirilirler. Konu ile ilgili, Maybury tarafından kaleme alınan bir diğer makalede SumGen açıklanır. Olay seçimi, yoğunlaşma ve sunum için yöntemler kullanan SumGen askeri benzetimlere ek olarak iş haberleri için de kullanılmaktadır.

3.5 Değerlendirme Yöntemleri

Değerlendirme yöntemleri dışsal (extrinsic) değerlendirme ve içsel (intrinsic) değerlendirme olarak ikiye ayrılır. Dışsal değerlendirme yöntemlerinde, bir özetin kalitesi başka işlerin tamamlanmasını nasıl etkilediğine bakılarak ölçülür. İçsel değerlendirmede ise, insanlar özetin kalitesini doğrudan, elle oluşturulmuş ideal bir özetle özetin akıcılığı, kapsamı ve benzerliği yönünden karşılaştırırlar.

Rath, farklı zamanlarda deneklere bir makaledeki, makaleyi temsil edebilecek cümleleri seçmelerini isteyerek, ideal özetin doğasında ne gibi zorlukları olduğunu, deneysel deliller ile sunar. Bu bağlamda, Brandow, Mitze ve Rau terim sıklığının konuya has özelliklerine karşılık sadece metnin başlangıç kısmını seçen bir sistemi, karşılaştıracak içsel bir değerlendirme yöntemi anlatır.

Morris, Kasper ve Adams, GMAT (Graduate Management Admission Test) okuduğunu anlama egzersizleri bölümündeki soruları cevaplamayı da kapsayan bir değerlendirme yöntemi önerir. Firmin ve Chrzanowski tarafından kaleme alınan bir diğer makale, metin özetleme sistemlerinin TIPSTER değerlendirmesinin ilk

provasını açıklar. Bu makaleler, bu süreçlerde oluşan özetlerin tam metinler kadar etkili olabileceğini gösterir.

3.6 Yeni Özetleme Problem Alanları

Bu kısımda, önemi ve bunlara duyulan ilgisi artan iki yeni alandan bahsedilebilir: çoklu belge özetleme ve çoklu ortam özetleme

Salton, Singhal, Mitra, Buckley'nin ve Mani, Bloedorn'un makalelerinde, bir metin biriminin diğerlerine ne kadar kuvvetle bağlı ise o kadar önemli olduğu metinler için bağlanabilirlik modeli kullanır. Salton ve diğ. yazdığı çoklu belge özetlemede, bir veya daha fazla belgedeki paragraflar sözcük dağarcığındaki benzerliğe göre karşılaştırılır. Belirli bir benzerlik eşik değerinin üstünde olan paragraflar, daha sonra özet çıkarmak için çeşitli yöntemlerle üzerinden geçilebilecek “metin ilişki haritası” grafi çıkarmak üzere birleştirilir. Mani ve Bloedorn, düğümleri metindeki terim görünme sayıları (konumları) olan, kenarları görünmeler arasındaki bağlılık ilişkisi (yakınlık, yinleme, eşanlam, geniş anlam) olan graf temsili kullanarak kullanıcı odaklı çoklu belge özetleme yapar. Bu temsil, sorgu ile ilişkili her belgedeki bilgiyi seçmek için kullanılır. Sonra da bu bilgi benzerlik ve farklılıkları bulmak için kullanılır. McKeown ve Radev tarafından yazılan bir diğer makalede, terörist kazaları hakkındaki farklı haber hikâyeleri arasındaki ilişki, kabul etmek ve inkâr etmek gibi mantıksal ilişkiye dayanan benzer çıkarılmış şablonlar bir araya getirilerek operatörler tarafından onaylanır. Birleşim sürecinde, bir araya getirilmiş bu şablonlar şablonların açıklamalarını ve ilişkilerini üreten bir cümle üreticine verilir.

Diğer ortaya çıkan bir alan da girdinin ve/veya çıktının metin olmasının gerekmediği çoklu ortam özetlemedir. Merlino ve Maybury, TV haber yayınında arama, tarama ve özetleme için bir araçları olan Haber Yayın Kılavuzu'nun bağlamındaki haber yayın videosunun özeti için tek-ortam ve karışık-ortam sunum yöntemleri için değerlendirme yöntemi anlatırlar. Futrelle, bir belgeden bir veya daha fazla şekli seçmek, bir şekli basitleştirmek için özünü çıkarmak (metin yoğunlaştırarak elemeye benzer), çoklu şekilleri birleştirmek (metin temsillerinin bir araya getirilmesi ve birleştirilmesine benzer) ve son olarak da sentez aşamasındaki özet çizelgesinin grafiksel şeklini üretmek için çeşitli yöntemleri anlatır.

4. ÖZETLEME

Tez çalışmasının bu bölümünde gerçekleştirilen metin özetleme sisteminin tasarım detayları anlatılmıştır. Özetleme sistemi başarımına etkisinin öneminden ötürü, gövdeleme bu bölümde ayrı bir altbölüm olarak ele alınmıştır.

4.1 Gövdeleme

Türkçede kök sözcüğe yapım ekleri getirilerek türetilen yeni sözcüklere gövde denir. Bir sözcüğün eklenmiş çekim eklerinden arındırılarak gövde veya kökün bulunması işlemine *gövdeleme* adı verilir [6].

İngilizce gibi analitik dillerde, gövdeleme görece olarak karmaşık değildir. Çünkü Tablo 4.1’de [21] de görüldüğü gibi bir sözcüğün biçimbirimsel çeşitlilikleri sınırlıdır. Diğer taraftan, Türkçe gibi eklemeli dillerde, teorik olarak sonsuz sayıda olası sözcük biçimleri üretme kapasitesi olduğundan gövdeleme hala zor bir problem olarak durmaktadır. Eklemeli dillerde, Biçimbirimsel çözümleyiciler BE sistemleri için tek doğru araçlardır. Fakat bu da düşük depolama ve yüksek performans gibi iki gereksinimi karşı karşıya getirir. Gövdeleme BE sistemleri tarafından, eklemeli diller için depolama karmaşıklığını yenmek ve keskinlik/anımsama metriklerini geliştirmek amacıyla talep edilir. Aslında, biçimbirimsel çözümleyicilerin gövdeleyici olarak kullanılmasından ortaya çıkan yüksek seviye hesaplama karmaşıklığı nedeniyle, BE sistemlerinin toplam performansı düşer [21]. Tez çalışmasında performans yüksek öncelikli bir kıstas olarak ele alınmadığından, gövdeleyici için biçimbirimsel çözümleyici kullanmakta bir mahsur görülmemiştir.

Tablo 4.1: Türkçe ve İngilizce derlemler için gövdeleme ile sıkıştırma oranı

Derlem	Sözcük belirtileri	Farklı terimler	Farklı gövdeler	Sıkıştırma (%)
Türkçe	376.187	41.370	6.363	84,6
İngilizce	567.574	18.384	11.671	36,4

4.1.1 Gövdeleme ile İlgili Yapılmış Çalışmalar

En eski yayınlanmış Türkçe BE çalışması Köksal tarafından yapılan, yirmi sorgu ile birlikte bilgisayar bilimleri hakkında 570 belge kullanır (başlık, anahtar sözcükler, bölüm başlıkları ve özetçe). Çeşitli dizinleme belge-sorgu eşleşme yaklaşımlarının verimliliğini anımsama-keskinlik grafiği ve çıkarım işleminde kullanmamak üzere Türkçede sık geçen sözcüklerden (“bir”, “ve” gibi) oluşan, 274 uzunluklu durma listesi kullanarak ölçer. Köksal, gövdeleme amacı ile sözcüklerin ilk beş karakterini (5-önek) kullanır. Bu seçim çeşitli önek boyutları ile denenerek elde edilmiştir.

Solak ve Can 533 haber makalesinden ve 71 sorgudan oluşan bir derlem kullanır. Gövdeleme için, bir biçimbirimsel çözümleyici kullanılır ve çalışma çeşitli sorgu-belge eşleştirme fonksiyonları kullanır. Çalışma gövdelemeyle yapılan çalışmanın gövdelemesiz olana göre daha verimli olduğunu gösterir. Bahsedilen deneylerde yedi değişik terim ağırlıklandırma yöntemi anlatılır.

Sever ve Bitirim 2468 yasa belgesine on beş sorguya dayanan sistemin gerçekleşmesini anlatırlar. İlk olarak, önceki iki gövdeleyeciye kıyasla yeni gövdeleyicinin başarımının daha iyi olduğundan bahsederler. Sonra da, çekimsel ve türetimsel gövdeleyicilerinin gövdeleyici kullanmamaya göre %25 keskinlik artımı sağladığını gösterirler.

Pembe ve Say Türkçe için BE problemini Türkçenin, biçimbirimsel, sözcüksel-anlambilimsel ve sözbilimsel seviyelerinde çalışırlar. Gövdelemenin etkisini bazı sorgu zenginleştirme yöntemleri ile değerlendirirler. Deneylerinde, Genel Ağ'dan çeşitli konular hakkında 615 Türkçe belge ve beş uzun doğal dil sorgusu kullanırlar. Yedi değişik dizinleme ve erişim birleşimi kullanırlar ve başarım sonuçlarını ölçerler [1].

Alpkoçak ve diğ. yaptığı çalışmada, sözcük bitene kadar veya gövde bulunana kadar sözcüğün sonundan karakter silinerek, sözlükte sözcük aranır.

Ekmekçioğlu ve diğ. bir BE sisteminin parçası olarak, bir dizi gövdeleme araştırması yaparlar. Terim birleştirmesi için N-gram eşleştirme kullanırlar ve gövdeleme kullanıldığında n-gram eşleştirme ile çıkarım başarımının arttığını bulurlar.

Duran tarafından yapılan çalışma ise, dilin sözcüksel kurallarına dayanan temelde Türkçenin biçimbirimsel çözümlemesi olan bir algoritma verir. Verilen sözcüğün öneki sözlükte aranır ve kalan karakterlerin ek olduğu varsayılır. Algoritma kök

alternatiflerinde seçim için karmaşık bir yöntem uygulamaz. Daha çok türetilmiş biçimbirimlere kıyasla, kullanıma göre değişen biçimbirimlerle çözümleme ve daha az türetilmiş biçimbirimlere öncelik verilir [22].

Oflazer tarafından yapılan biçimbirimsel çözümleyici her ne kadar gövdeleyici olmasa da biçimbirimsel çözümleyici kullanan gövdeleyicilerde başarımlı etkileyecektir. Biçimbirimsel çözümleme için sonlu durumlu makinelerin kullanıldığı ve Türkçeye has tüm özel durumların hesaba katıldığı bu çalışmada yaklaşık 22.000 kök sözcükten oluşan bir kök sözcüğü kullanılmıştır [23].

4.1.2 Zemberek

Zemberek, Türkçe ve diğer Türkî diller için yazılmış, biçimbirimsel çözümleme, yazım denetimi, sözcük üretme gibi temel DDİ işlemlerini yapabilen açık kaynak kodlu DDİ kütüphanesidir. Geliştirmeye açık yapısı, açık kaynak kodlu olması ve BÇ'nin tatmin edici doğruluk seviyesinde çalışmasından dolayı çalışmada gövdeleyici olarak Zemberek tercih edilmiştir.

Türkî diller Avrupa'dan Sibiryaya 140 milyon kişi tarafından konuşuluyor olmasına rağmen, açık hesaplama kütüphanelerinin eksikliğinden dolayı yazım denetimi gibi en basit bir DDİ işleminin yapılması bile zahmetli olabilmektedir. Zemberek altyapısının desteklediği DDİ işlemleri şöyledir:

- yazım denetimi
- biçimbirimsel çözümleme
- gövdeleme
- sözcük üretimi
- sözcük önerme
- sadece ASCII karakterle yazılmış sözcükleri Türkçe karakterli hale çevirme
- heceleme [24]

4.1.2.1 Yapı

Kütüphane iki ana bölümden oluşur; dil yapı bilgisi ve DDİ işlemleri. Çekirdek kütüphane, DDİ'ye has algoritmaları içerir ve dillerin gerçekleştirilmesi için gerekli araçları sağlar. Çekirdek kütüphane Türkî diller için tasarlanmış olmasına rağmen,

herhangi bir dile özel uygulama içermemektedir. Bu esnekliđi sađlayabilmek için çeřitli yardımcı mekanizmalar ve soyutlamalar kullanılmıřtır. Her dil gereklenmesi önceden belirlenmiř dilbilgisi gereksinimlerine uymakla ve gerekli dil verisini sađlamakla sorumludur. Bir dil gereklendikten sonra, ekirdek DDİ fonksiyonları bu bilgileri genel olarak kullanıp son kullanıcıya kullanımı kolay yazılım eriřim mekanizması üzerinden hizmetler sađlar.

Bir Türkİ dilin gereklenmesi görel olarak kolaydır. Zemberek'te dil geliřtiricilerin iřini kolaylařtırmak için bazı dil verileri metin tabanlı kurulum dosyalarına haricileřtirilmiřtir. Ama bu durum esnekliđi ve bařarımı olumsuz yönde etkiler. Bu yüzden, özel durumlar ve ek üretimi gibi bazı bilgiler kodun içinde tutulmuřtur [24].

Harfler ve Alfabe

Zemberek kullanılacak her dil için harf ve alfabe bilgisine ihtiya duyar. Bunlar bir sözcüğün istenen DDİ iřlemleri yürütülmeden önce geçerli karakterler içerip içermediđini belirlemede kullanılır. Bu bilgi basite bir metin dosyasında tutulur [24].

Ek Bilgisi

Tüm Türkİ dillerin temel yapıtařı eklerdir. Zemberek'te ekler bir XML dosyasında tutulur. Ekler daha rahat kullanımı için kümelendi (ođul ekleri, sahiplik ekleri gibi).

Biimbirimsel özümlemede kullanılması amacıyla bu dosyada; ekleri takiben hangi ekin veya ek kümelerinin geleceđi, ekin nasıl ekleneceđi, ek ađacının bařlangı noktasını belirlemede kullanılacak ön ekler, ek özel duruma (yumuřama, sertleřme, ünlü düřmesi gibi) sebep oluyorsa özel durum üretim yöntemi tutulur.

Bařlangıta tüm ek dosyası ek nesnelere transfer edilir ve uygulama alıřması sırasında bu nesnelere kullanılır [24].

Kök Sözcük Sözlüğü

Kök sözcük, ek içermeyen anlamlı sözcük anlamına gelir. Her dil için kök sözcükler bir metin dosyasında tutulur. Sözcüğün kendisine ek olarak, sözcüğün tipi ve özel durum(lar) da tutulur.

Türkİ dillerde genellikle yabancı dillerden gemiř sözcüklerde görülmekle birlikte özel durumlar da görülür. Örneđin;

saat → *saatler*, *saatlar* değil (Arapça kökenli)

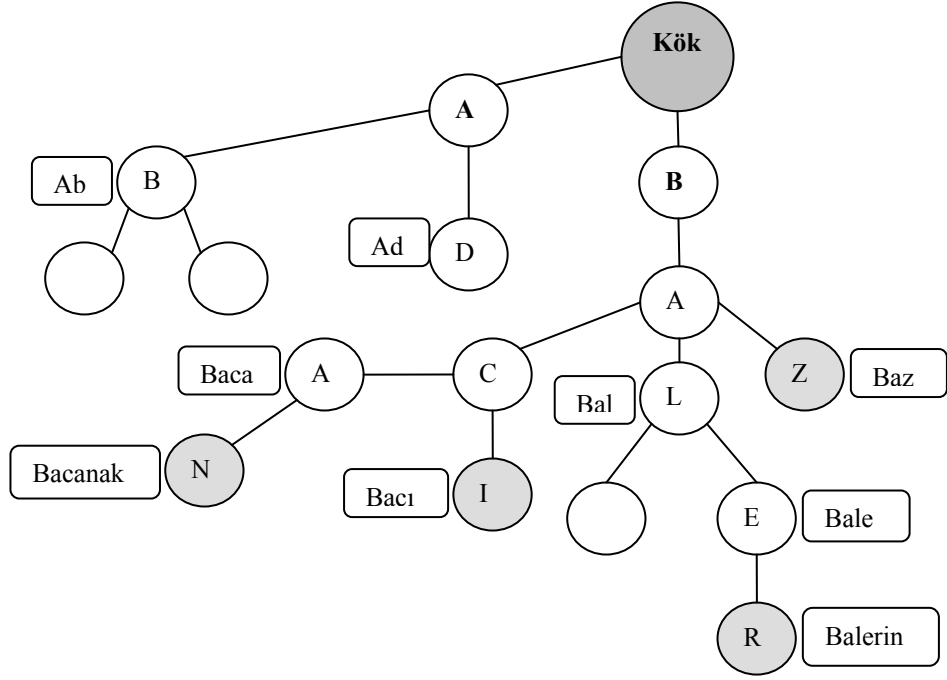
red → *reddi*, *redi* değil (Arapça kökenli)

burun → *burnu*, *burunu* değil (Ünlü düşmesi, Türkçe kuralı)

su → *suyu*, *sunu* değil (Sadece su sözcüğünde görülen bir durum.)

ben → *bana*, *bene* değil (Sadece birinci ve ikinci tekil şahıslar olan “ben” ve “sen”de görülür.)

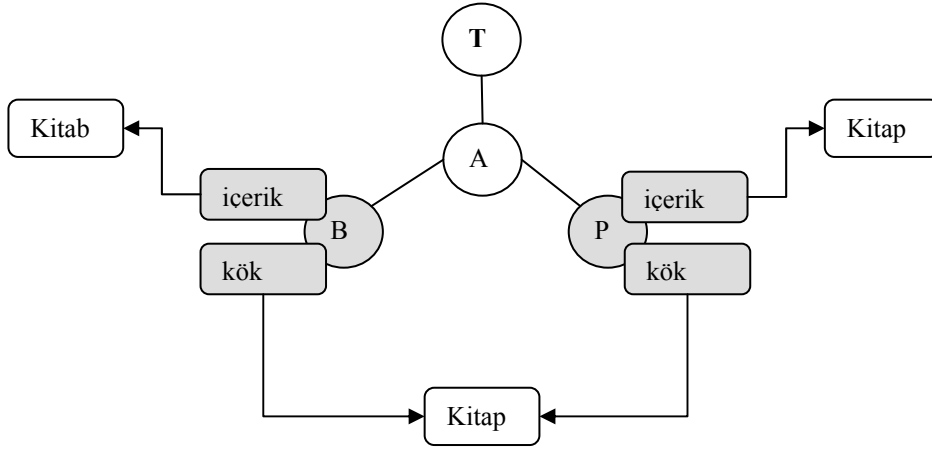
Kütüphane başlangıç durumuna getirilirken, Zemberek önce ikili kök dosyasını bir nesneye yükler. Bir kök sözcük okunduktan sonra, ilgili özel durumlar da kök nesnesine eklenir ve sonuç nesnesi, hızlı erişim ve büyütme kolaylığı sağlamak için Şekil 4.1’te görülen özel bir Direkt Düz Sözcük Graf’ı ağacında saklanır.



Şekil 4.1: Kök sözcük ağacı

Türkçedeki benzeşme, yumuşama gibi kurallardan dolayı çözümlenecek sözcük kök sözcüğü her zaman birebir barındırmayabilir. Bu nedenle, kök sözcük ağaca eklendikten sonra sistem, özel bir durum uygulandıysa ortaya çıkabilecek kök sözcüğün bozulmuş şekillerini de hesaplar. Örneğin, Türkçede “kitaba” sözcüğü, “kitap” sözcüğünü birebir barındırmaz. Çünkü sert sessizle biten bir sözcüğe sesli harf ile başlayan bir ek getirildiğinde sondaki sert sessiz yumuşara. Dolayısıyla, Şekil

4.2’de görüldüğü gibi, hesaplanan bozulmuş “kitab” asıl kök nesnesi referans gösterilerek ağaca eklenir.



Şekil 4.2: Yumuşama durumunda kök sözcük ağacının içeriği

Zemberek’te bu ağaç sözlük olarak kullanılır ve bu ağaç sayesinde çeşitli kök seçiciler kolaylıkla geliştirilebilir. Zemberek’te şu anda üç tane kök seçici mevcuttur. Birincisi; verilen bir sözcük için normal tam bir kök seçimi için kullanılır. Örneğin; “elmaslar” sözcüğü için “el”, “elma”, “elmas” köklerini bulur. İkinci kök seçici bir karakter katarı benzerlik algoritması kullanarak kök adaylarını seçerken belirli bir hoşgörü seviyesi uygular. Doğal olarak ilk yöntemle göre daha fazla kök bulur. Üçüncü seçicinin ASCII kodlamasında yer almayan harfler için hoşgörüsü vardır [24].

4.1.2.2 Biçimbirimsel Çözümleyici

Biçimbirimsel çözümleyici basitçe bir sözcüğün tüm olası kök ve eklerini bulur. Yapısı basit sözlük tabanlı yukarı-aşağı çözümleyici olarak tanımlanabilir.

Bir girdi sözcüğün biçimbirimsel çözümlenmesinde dört ana adım vardır:

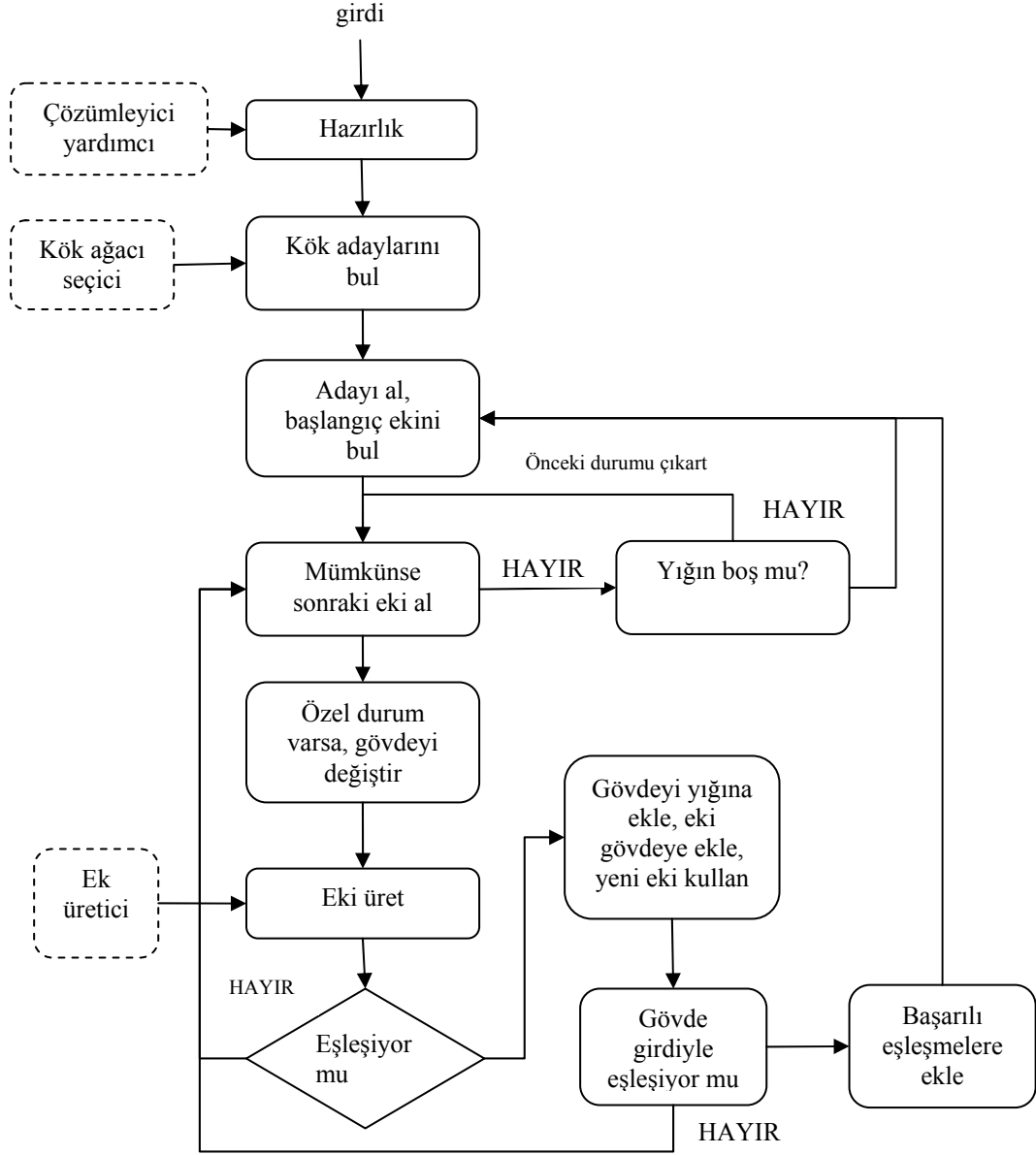
- Sözcük önışlemeden geçirilir.
- Uygun kök seçici kullanılarak girdi sözcük için kök adayları bulunur.
- Her kök adayı için, girdi sözcük inşa edilene kadar veya geriye hiçbir ek seçeneği kalmayana kadar olası eklerin köke eklenmesine devam edilir, gerekliyse özel durumlar uygulanır.
- Çözümleyici sonuçları son işlemeden geçirilir.

Önişlem, sözcükten “tire” karakterlerinin, aksanların atılması ve küçük harfli yazıma çevrilmesi gibi sözcüğün çözümlene işlemine hazır hale getirilmesi adımlarından oluşur. Sözcükte dilin alfabesinde olmayan harf bulunursa işlem durdurulur. Kök adayları, girdi sözcüğün başlangıç kısmı olabilecek kök sözcüklerdir. İşleme bağlı olarak, kök adaylarını bulmak için uygun kök seçiciler kullanılır.

Kök adayları bulunduktan sonra, sözcüğün tipine bağlı olarak sistem ilk *kök ekini* bulur. *Başlangıç eki*, neredeyse diğer tüm eklerde olduğu gibi, onu takip edebilecek eklerin listesini içerir.

Dolayısıyla ilk ekten başlanır ve ekin özel duruma yol açıp açmadığına bakılır. Eğer ek özel bir durum oluşturuyorsa, değişiklik kök sözcüğe uygulanır. Örneğin, kök sözcük *burun* ise, takip eden ek sesli harfle başlıyorsa, ekin bir parçası olarak tanımlanmış sözcük değiştirici uygulanarak, çözümleyici kökteki son sesli harfi düşürür (kök *burn* haline dönüşür). Ardından, girdi sözcük ve değiştirilen sözcük kullanılarak bir ek fonksiyonu çağrılır. Fonksiyon parametreleri alır ve kendi ek üretim elemanlarını kullanarak, *ek üretici* denen dile bağlı bir yöntemi kullanarak değiştirilmiş sözcük için ek üretir. Her ek üretici dile bağımlıdır. Dile bağlı yöntemler uygulayarak somut ekler üretmekten sorumludurlar.

Bir ek üretildikten sonra, doğru şekilde eşleşip eşleşmediğini belirlemek için girdi sözcük ile karşılaştırılır. İşleme bağlı olarak *tam*, *ASCII toleranslı*, *hata toleranslı* gibi çeşitli sözcük karşılaştırıcılar kullanılabilir. Başarılı bir ek üretildiyse, sözcüğün yeni şekli ve ek durumu bir yığında tutulur ve işlem takip eden ek bilgileri ile devam eder. Girdi sözcük ve başarılı ekler eklenmiş kök düzgün şekilde eşleştirildiğinde sistem sonucu döner (örneğin sistem bir yazım denetimi ise) veya eşleşen ek kalmayana kadar diğer olası çözümleri denemekle devam eder. Çözümleyicinin basitleştirilmiş blok çizelgesi Şekil 4.3’de gösterilmiştir.



Şekil 4.3: Çözümleyici blok çizelgesi

Genel olarak, çözümlene işlemi dilden bağımsızdır. Ama doğru olası çözümler bulunduğundan sonra, sembollerin ve büyük harflerin girdi sözcükte doğru kullanıldığı kontrolü için bir son işleme işlemi yapılır. Çünkü önışlem kısmında bunlar çıkarılmıştır. Bu durum, kısaltmaların ve “Ahmet’in” veya “prof.” gibi özel karakterleri içeren sözcüklerin çözümlenmesinde önemlidir. Bu son işleme adımı dile bağlıdır ve tüm dil gerçeklemelerinde tanımlanan bir yardımcı sınıf ile yapılır.

Türkî diller önek almazlar. Bu kuralı bozan, Türkçedeki “mas+mavi”, “bem+beyaz” gibi yaygın olmayan örnekler vardır. Ama bu önekler belli bir kalıba uymazlar.

Yabancı dillerden giren bazı sözcükler de önek alırlar. Zemberek önekleri çözümlememektedir.

Ayrıca Zemberek, *-it*; yak→yakıt, kes→kesit gibi bazı sözcük üretim eklerini de çözümlememektedir. Bunun yerine sentezlenmiş sözcüğü sözlükte tutar. Bununla birlikte Zemberek'te, kök sözcük olarak işaretlenerek bazı sözcük üretim ekleri uygulanır [24].

4.1.2.3 Başarım

Zemberek tasarlanırken, gerçek dünya uygulamalarında kullanılacağı düşünülerek tasarlandığından sistem performansının önceliği yüksektir. Özellikle en çok görülecek fonksiyon olan yazım denetiminin oldukça hızlı olması gerekir. Uygulamada başarımı arttırmak için özellikle yapılan bazı geliştirmeler şöyledir:

- İkili kök dosyası kullanımı kök ağacının başlangıç durumuna getirilmesi zamanını düşürür.
- Karakterlerin direkt eşleştirilmesi için alfabe büyük karakter dizileri kullanır. Türkçe için harf nesnelere ile bu durum giderildi.
- Sıradan “hash table” veya liste yapılarına göre daha karmaşık bir yapıya sahip olmasına rağmen, özel kök ağacı kullanımı sıradan “hash table” tabanlı yaklaşımlara göre çok daha iyi başarım sergiler.
- Standart çözümleme esnasında, eklerin olası ilk harfleri ve bu verinin gerçekten yaratılmadan ve bir ekle karşılaştırılmadan kontrolü başarımı artırır.
- Bazı ekler diğer bazı ekleri takip etmede istatistiksel olarak daha yüksek şansa sahiptir. Dolayısıyla ek dosyasında “öncelikli-ekler” diye bir bilgi mevcuttur ve bazı ekler çözümleme esnasında önceliğe sahiptir. Bu ek bilgisi istatistikleri Zemberek'in istatistiksel bilgi toplama aracı ile elde edilmiştir.

Yazım denetimi başarımı 71,295 sözcükten oluşan Türkçe bir roman ile A. A. Akın ve M. D. Akın tarafından test edilmiştir ve ortalama test sonucu saniyede 75,000 sözcük olarak ölçülmüştür. Yazım denetimi biçimbirimsel çözümlemeye göre daha yavaş bir işlemdir. AMD Athlon 64 3000+ mikroişlemcili, 512 MB belleğe sahip bir bilgisayarda Windows XP işletim sistemi ve Java SE 6 Beta-2 test koşullarında, Zemberek'in saniyede 12,000 sözcüğü çözümleyebildiği saptanmıştır [24].

4.1.3 Biçimbirimsel Çözümleyici Kullanarak Gövdeleme

Biçimbirimsel çözümleyiciler verilen bir sözcüğün olası tüm kök-ek birleşimlerini getirirler. Bu seçenekler içinden uygun gövdenin seçilmesi gerekmektedir. Gövdelemenin, Türkçe için yapılmış diğer tüm BE sistemlerinin verimliliğinde olduğu gibi metin özetleme sisteminde de önemi büyüktür. Bu nedenle sistemin verimliliğini ve başarımını ölçmek için iki farklı gövdeleyici kullanılmıştır.

4.1.3.1 En Uzun Gövdenin Seçilmesi

Gövdeleme işlemi, bir sözcüğün çekim eklerinden arındırılıp kök+yapım ekleri alınmış hali olarak tanımlandığından ve Türkçede istisnalar dışında çekim eklerinden sonra yapım ekleri gelmeyeceğinden en uzun gövde, sözcüğün çekim eklerinden arındırılmış gövdesini verecektir.

Tablo 4.2: Gözlükçüdekilerden sözcüğü için Zemberek BÇ çıktısı

Kök	Kök Tipi	Ekler
<i>Gözlük</i>	ISIM	ISIM_KOK + ISIM_ILGI_CI + ISIM_KALMA_DE + ISIM_BULUNMA_KI + ISIM_COGUL_LER + ISIM_CIKMA_DEN
<i>Göz</i>	ISIM	ISIM_KOK + ISIM_BULUNMA_LIK + ISIM_ILGI_CI + ISIM_KALMA_DE + ISIM_BULUNMA_KI + ISIM_COGUL_LER + ISIM_CIKMA_DEN

Yöntem çözümlenecek sözcüğe anlamsal olarak en yakın gövdenin (çekim eklerinden arındırılmış hali) seçilmesi esasına dayanır. Dolayısıyla Tablo 4.2’de görülen *gözlükçüdekilerden* sözcüğü için BÇ çıktılarından gözlükçü gövdesi seçilir.

4.1.3.2 İlk Beş Karakter Yöntemi

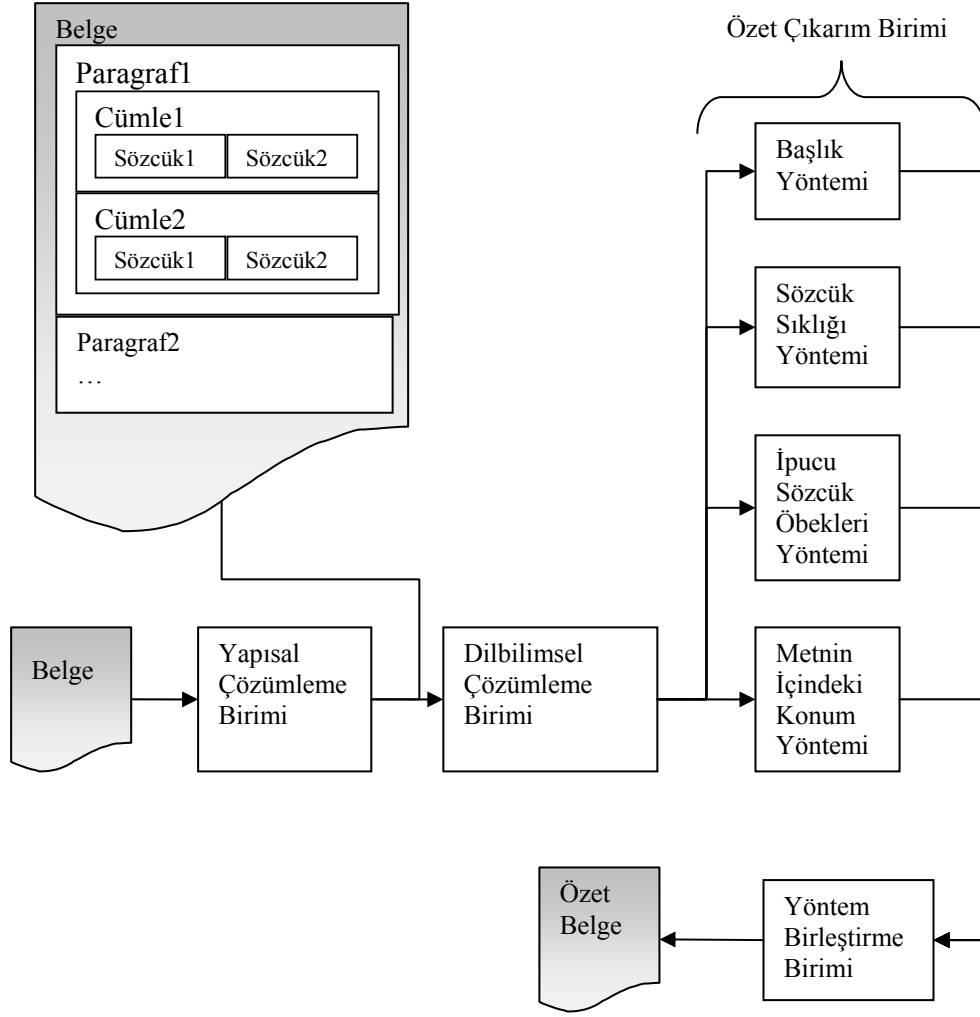
Bu yöntem, Köksal [1] tarafından istatistiksel deneyler sonucunda bulunan, Türkçe sözcüklerin ortalama 5 gövde uzunluğunda olduğu ve çoğu sözcüğün bu kurala uyduğu hipotezine dayanır. Bu yöntemle göre, Tablo 4.2’de BÇ çıktısı gösterilen *gözlükçüdekilerden* sözcüğünün gövdelerinden *gözlük* gövdesi seçilmelidir.

4.2 Özetleme Sistemi

Bir metin özetleme sisteminin ana amacı, sisteme yollanan belgenin veya belge kümesinin kısa ve özlü bir temsilini oluşturmaktır. Özetleri aşağıdaki gibi kategorize etmek mümkündür:

- *Tipe göre*: arama sorgusuna veya kullanıcı sorusuna en yakın içeriği çıkaran *sorgu-tabanlı* özetlemeye karşılık belgenin içeriği hakkında genel bilgi veren *genel amaçlı* özetleme
- *Amaca göre*: bir belgenin veya belge kümesinin içeriği hakkında genel bilgi veren *belirtici* özete karşılık en önemli bilginin çıkarımına izin veren konuyla en ilgili bilginin sunumunu amaçlayan *bilgi verici* özet
- *Özetlenecek belge sayısına göre*: *tek belge* özeti tek belgenin özetine izin verirken *çoklu belge* özeti birden çok sayıda belgenin özetine izin verir
- *Tipe göre*: farklı tipteki asıl belgeler farklı yapı ve uzunlukta olmakla birlikte farklı sözcük dağarcığı kullanırlar. Farklı tipler; haber, bilimsel makaleler, e-postalar, Genel Ağ sayfaları, gazete-dergi makaleleri olabilir.
- *Asıl metne benzerliğine göre*: asıl metindeki cümlelerden alıntılardan oluşan *çıkarımsal özet*, asıl metnin ana içeriğini yansıtan aslını yansıtan yeni cümlelerden oluşan *özetçe*

Yapılan tez çalışmasında *sorgu tabanlı olmayan, bilgi verici, tek belge özetlemeye dayanan, tipten bağımsız çıkarımsal özet* oluşturulması amaçlanmıştır. Şekil 4.4'te metin özetleme sisteminin genel yapısı görünmektedir.



Şekil 4.4: Metin özetleme sistemi mimarisi

4.2.1 Yapısal Çözümleme Birimi

Bu birimde yapılan ana işlem; girdi olarak verilen bir metin belgesinin, diğer birimlerde rahat ve hızlıca kullanılmasını sağlayacak belge nesnesine doldurulmasıdır. Belge nesnesi, belgenin içindeki paragrafların uygulamadaki temsili olan paragraf nesnelere oluşan bir liste ve belgenin başlığını tutar. Benzer şekilde paragraf nesnelere, paragrafı oluşturan cümlelerden ibaret olan cümle nesnelere oluşan cümle listesi ve cümle nesnelere de sözcüklerden oluşan sözcük listesi yer almaktadır.

Bu işlemin yapıldıktan sonra da dilbilimsel çözümleme yapılarak özetleme yöntemlerinin uygulanabilmesi için girdi belge, Bölüm 2.2.1'de belirtilen Türkçe metin standartlarına uygun olmalıdır. Belge bir başlıktan, satır başı ile biten paragraflardan ve paragraflar da gerekli noktalama işaretleri ile birbirinden ayrılmış

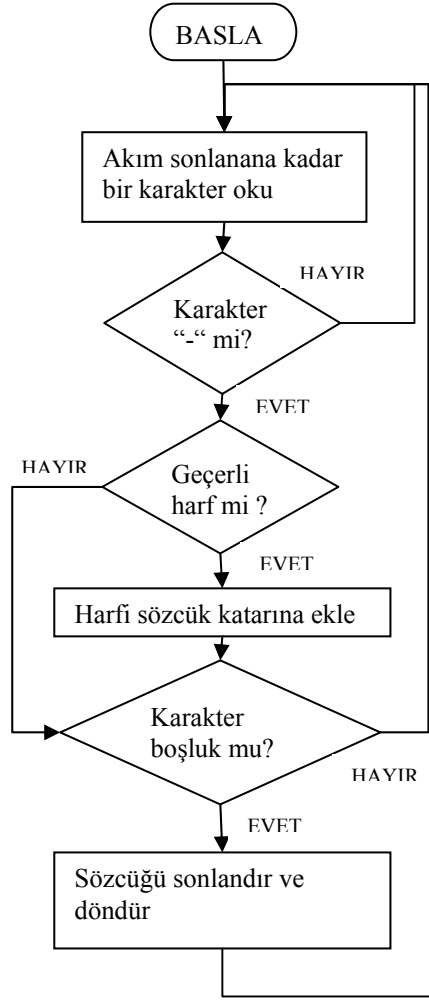
cümlelerden oluşmalıdır. Aksi durumda, yapısal çözümleme birimi sistemin genelinde ihtiyaç duyulan paragraf-cümle-sözcük hiyerarşisini sağlayamayacaktır.

Belge metninden, belge-paragraf-cümle-sözcük nesnelерinin oluşturulabilmesi için paragraf sonu, cümle sonu ve sözcük sonu belirlenmelidir.

Paragraf sonu belirlemede basitçe satır sonu karakterine ulaşılması koşulu kabul edilmiştir. Buna göre metin çözümlenirken satır sonu karakterine gelindiğinde, bir paragrafın sonlandığı varsayımı yapılmıştır.

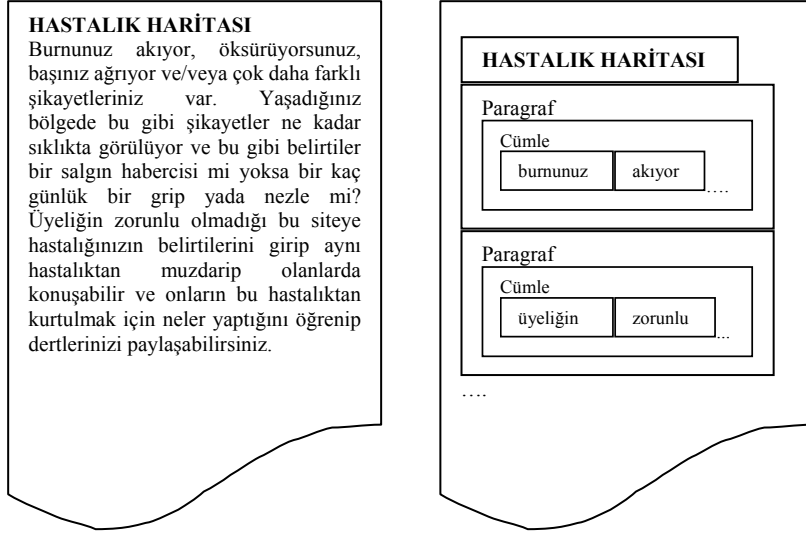
Cümle sonu belirlemede Zemberek kütüphanesindeki mevcut cümle sonu belirleme yöntemi kullanılmıştır. Buna göre; bir karakter katarının cümle olarak değerlendirilebilmesi için ilk karakterin geçerli bir harf, “” (kesme işareti), “-” (tire) noktalama işaretlerinden biriyle başlaması ve “.”, “:”, “!”, “?” noktalama işaretlerinden biriyle bitmesi gerekir.

Zemberek kütüphanesindeki, sözcük sonu bulma algoritması ise biraz daha karmaşıktır. Buna göre Şekil 4.5’te görüldüğü gibi, bir karakter katarı karakter bazında işlenir. Katarda sözcük başının belirlenmesi için geçerli bir harfle başlaması gerekmektedir. Daha sonra sözcük yine harflerle veya “”, “-” işareti ile devam eder. (“-” işareti satır sonu ayraçı olarak kullanılabileceğinden) Herhangi bir anda “boşluk” karakteri ile karşılaşıldığında sözcüğün sonlandığı varsayılır. Bu işlem sırasında tüm harfler küçük harflere dönüştürülür.



Şekil 4.5: Sözcük sonu belirleme algoritması

Yapısal çözümleme birimi, sonuçta Şekil 4.6'da görülen çıktıyı elde eder.



Şekil 4.6: Yapısal çözümleme birimi çıktısı

4.2.2 Dilbilimsel Çözümleme Birimi

Dilbilimsel çözümleme biriminin temel işlevi paragraf-cümle-sözcük hiyerarşisindeki sözcüklerin Zemberek kütüphanesindeki biçimbirimsel çözümleyici aracılığıyla çözümlenip, çözümleme çıktıları içinden uygun gövdenin seçilmesinden oluşur. Uygun gövdenin seçilmesinde bölüm 4.1.3'te belirtilen yöntemler kullanılmıştır. Buna göre sözcük sınıfında, her sözcüğün metin olarak içeriğine ek olarak sözcüğün BÇ çıktıları ve sözcük için seçilen gövde de tutulur.

Türkçe eklemeli bir dil olduğundan, gövdelemenin Türkçe BE sistemleri başarımına etkisi çok önemlidir. Örneğin, *burun* sözcüğünün biçimbirimsel çeşitlilikleri olan *burnu*, *burnum*, *burnun*, *burnundaki*, *burnunun*, *burnunları* sözcükleri Türkçe bir belgede geçiyor olsun. Gövdeleme kullanılmadan sözcük sıklığı yöntemine göre terim sıklıkları hesaplanıyor olsun. *Burun* sözcüğünün yukarıda sayılan altı adet biçimbirimsel çeşitliliğinin hepsi ayrı birer terim olarak değerlendirilecektir. Dolayısıyla aslında *burundan* bahseden 6 değişik yer söz konusu iken, bunlar birikimli olarak değerlendirilememiş olacaktır. Gövdelemenin bu denli önemli olması ve başarımı etkiliyor olmasından ötürü, BÇ çıktısı kullanan iki farklı gövdeleyici gerçekleştirilmiş ve başarımına etkisi test edilmiştir.

Tablo 4.3: Dilbilimsel çözümleme birimi çıktısı

Asıl Sözcük	BÇ Çıktısı	Seçilen Gövde
<i>burnunuz</i>	Kok: <i>burun</i> Ekler: ISIM_KOK + ISIM_SAHİPLİK_SEN_IN + ISIM_KISI_BIZ_IZ Kok: <i>burun</i> Ekler: ISIM_KOK + ISIM_SAHİPLİK_SIZ_INIZ	<i>burun</i>
<i>akıyor</i>	Kok: <i>ak</i> Ekler: FIIL_KOK + FIIL_SIMDİKIZAMAN_IYOR	<i>ak</i>
<i>öksürüyorsunuz</i>	Kok: <i>öksür</i> Ekler: FIIL_KOK + FIIL_SIMDİKIZAMAN_IYOR + FIIL_KISI_SIZ	<i>öksür</i>

4.2.3 Özet Çıkarım Birimi

Özetleme, istatistiksel olarak, her cümlemin puanlandırılması ve daha sonra bu cümlelerden ek yüksek puana sahip belli miktardakilerin seçilip özetin oluşturulmasına dayanır. Özet çıkarım birimi, dört farklı yöntem kullanır. Her cümleye her yöntem sonucunda bir puan verilir. Kullanılan yöntemler şöyledir:

4.2.3.1 Başlık Yöntemi

Edmundson tarafından ortaya atılan bu yöntemin dayandığı temel hipotez, bir metnin başlığının konu hakkında genel, özet bir bilgi verdiğidir. İstatistiksel olarak %99 doğrulukla başlık sözcüklerinin konu ile ilgili olduğu kabul edilir [16]. Özellikle haberler, bilimsel makaleler gibi kurallı yapıya sahip belgelerde doğruluğu yüksektir. Sohbet kayıtları, forum başlıkları gibi Genel Ağ'da her kullanıcının girişine açık alanlarda doğruluk düşebilecektir.

Belgenin başlığındaki tüm sözcüklerden bir başlık sözcük listesi oluşturulur. Bu sözcükler de dilbilimsel çözümleme birimi tarafından gövdelenir. Belgedeki sözcükler gövdelenindikten sonra, başlık listesinde olan sözcüklere pozitif puan verilir. Ardından her cümlemin başlık yöntemi puanı, cümledeki tüm sözcüklerin başlık puanları toplanarak elde edilir.

4.2.3.2 Sözcük Sıklığı Yöntemi

İlk defa Luhn [15] tarafından ortaya atılan yöntemin dayandığı temel ilke, bir metinde yüksek sıklıkta görülen sözcüklerin konuyla ilişkili olduğudur. Dolayısıyla

bu sözcükleri barındıran cümleler de belgenin konusuyla ilişkilidir ve çıkarımsal özetinde yer almalıdır. Buna göre vurgu, önemliliğin bir göstergesidir.

Her belgenin içinde en sık geçen sözcükler sabittir. Bunlar günlük hayatta sıklıkla kullanılan sözcüklerdir. Çoğunlukla bağlaçlardan (“ve”, “ile”, “veya”) ve bazı sayı sıfatlarından (“bir”, “iki”) oluşurlar ve belge ile ilgili konu belirtmezler. Bu sözcüklerden sabit bir *durma listesi* oluşturulup, sözcük sıklığı hesaplanırken bu sözcükler ihmal edilmelidir. Luhn her ne kadar sıklığı ifade etmek için alt limit alındığı gibi bir üst limit de alınarak bu sözcüklerin elenebileceğini vurgulasa da bu sözcükler bir dil için her türlü belgede aynı sözcükler olacağından böyle bir kullanıma gidilmemiştir.

Tablo 4.4: Örnek durma listesi

Sözcük	Sözcük
bir	ancak
ve	eğer
de	da
bu	ya
şu	veya
ile	mi

Yönteme göre, belgedeki sözcüklerin biçimbirimsel çözümlenip gövdelenmiş hallerinden bir liste oluşturulur. Liste oluşturulurken *durma listesinde* yer alan sözcükler ihmal edilir. Bu listedeki gövdelenmiş sözcüklerin metin içinde tekrarlanma sayısı hesaplanır ve tekrarlanma sayısına göre azalan sırada liste sıralanır. Belge içinde bu anahtar sözcük öbeklerinin yer aldığı sözcüklere sözcüklerin sıklık miktarı ile orantılı olumlu puan verilir. Son olarak her cümlenin sözcük sıklığı yöntemi puanı, cümledeki tüm sözcüklerin sözcük sıklığı yöntemi puanları toplanarak hesaplanır. Örnek bir sözcük sıklık listesi Tablo 4.5’de görülebilir.

Tablo 4.5: Örnek bir sözcük sıklığı listesi

Sözcük	Sözcük Sıklığı
ve	58
bilgisayar	42
gelecek	42
tasarım	37
teknoloji	36
commodore	33

4.2.3.3 İpucu Sözcük Öbekleri Yöntemi

Yöntem, Edmundson [16] tarafından tanıtılmıştır. Buna göre; bir cümle belli bazı sözcük veya sözcük öbeklerini taşıyorsa, o cümle konu belirten dolayısıyla özetle yer alabilecek bir cümledir. Test derleminden deneysel gözlemler sonucu belirlenen bu sözcük öbekleri iki grupta toplanır: Bulunduğu cümleye konu belirtme özelliği katan *olumlu sözcük/sözcük öbekleri*, bulunduğu cümleye konu belirtmeme özelliği katan *olumsuz sözcük/sözcük öbekleri*.

Tablo 4.6: İpucu sözcük öbekleri

Olumlu Sözcükler	Olumsuz Sözcükler
özetle	çünkü
sonuçta	öyleyse
neticede	ancak
	ama
	ayrıca
	artık
	halbuki
	sonra
	bu
	şu

İpucu sözcükler, deneysel yöntemler ile çıkarılmıştır. Buna göre, test aşamasında kullanılan belgelerin çıkarımsal özetlerindeki durma listesinde yer almayan sözcükler, sözcük sıklığı yönteminde olduğu gibi sıklıklarına göre ters sıralı bir şekilde oluşturulmuştur. Ardından da en çok belgede ortak olarak geçen sözcükler belirlenip, insan tarafından incelenip olumlu sözcük listesi oluşturulmuştur. Olumsuz sözcük listesi de benzer işlemin, asıl belgede yer alıp çıkarımsal özetle yer almayan cümleler için yapılması ile elde edilmiştir.

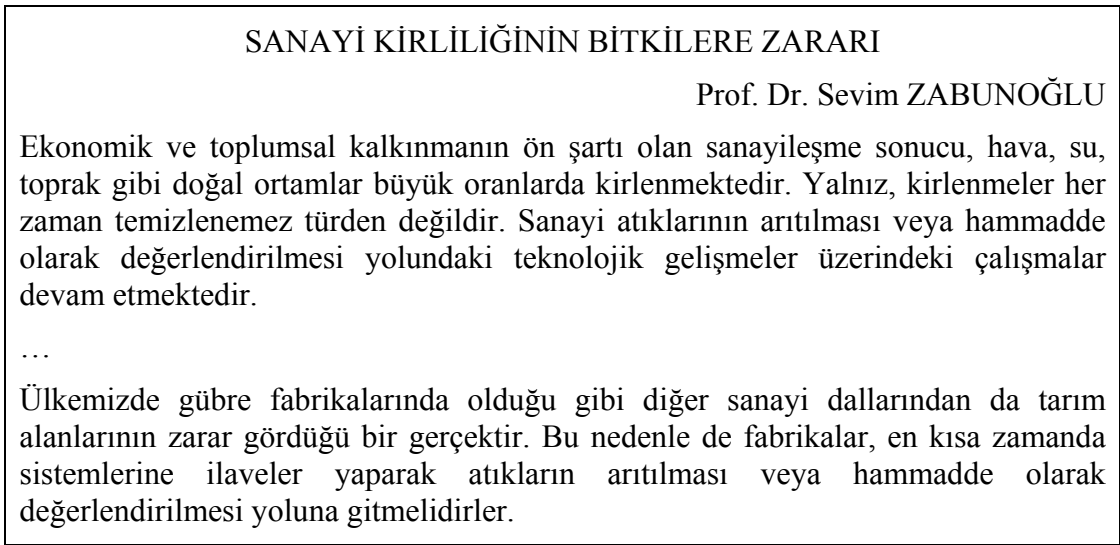
Özetlenecek belgede her cümle ipucu sözcükleri barındırmasına göre olumlu-olumsuz puanlandırılırlar.

4.2.3.4 Metnin İçindeki Konum Yöntemi

Edmundson [16] tarafından ilk kez ortaya atılmıştır. Bu yöntem, belge içindeki belli paragrafların, paragraflardaki belli cümlelerin konu belirtme olasılıklarının yüksek olması temeline dayanılarak uygulanmıştır. Her belgenin ilk ve son paragrafı, her paragrafın ilk ve son cümlesi bu yönteme göre olumlu olarak ağırlıklandırılır.

Bölüm 2.2.3'te belirtildiği gibi, Türkçe belgelerde de paragraf genellikle hüküm belirten bir cümle ile başlar. Genelde belirtmek istenen fikir paragrafın ilk cümlesinde veya son cümlesinde verilir. Aradaki cümlelerde bu fikri destekleyen düşüncelere yer verilir.

Yapısal çözümleme birimi asıl belgeyi paragraf-cümle-sözcük hiyerarşisi oluşturacak şekilde ayrıştırır. Belgenin ayrıştırılmış bu hali üzerinden belgenin ilk ve son paragraflarına ve her paragrafın ilk ve son cümlelerine kolaylıkla ulaşılır ve olumlu olarak ağırlıklandırılır. Şekil 4.7'de örnek bir belgenin giriş ve sonuç paragrafları görülmektedir [12].



Şekil 4.7: Giriş ve sonuç paragrafi örneği

4.2.4 Yöntem Birleştirme Birimi

Özet çıkarım birimi, asıl belgenin paragraf-cümle-sözcük hiyerarşisi kazandırılmış şekli üzerinden bir takım istatistiksel yöntemler uygulayarak metindeki her cümleye her farklı yöntem için matematiksel puan hesaplar. Yöntemlerin birbirine baskınlığına göre; her yöntem için hesaplanan puanlar ağırlıklandırılarak her cümle için son puan değeri elde edilir. Her cümle için hesaplanan cümle puanı, Denklem (4.1)'e göre hesaplanır. Buradaki a_1 , a_2 , a_3 ve a_4 katsayıları deneysel yöntemlerle hesaplanmıştır.

$$Puan(S_i, B_j) = a_1 I_{ij} + a_2 A_{ij} + a_3 B_{ij} + a_4 K_{ij} \quad (4.1)$$

Yöntem birleştirme birimi Tablo 4.7'da görülen puanlarına göre azalan sırada sıralanmış cümleler oluşturur.

Tablo 4.7: Yöntem birleştirme birimi çıktısı

Cümle	Puan
Ekonomik ve toplumsal kalkınmanın ön şartı olan sanayileşme sonucu, hava, su, toprak gibi doğal ortamlar büyük oranlarda kirlenmektedir.	17
Ülkemizde gübre fabrikalarında olduğu gibi diğer sanayi dallarından da tarım alanlarının zarar gördüğü bir gerçektir.	15
Bu nedenle de fabrikalar, en kısa zamanda sistemlerine ilaveler yaparak atıkların arıtılması veya hammadde olarak değerlendirilmesi yoluna gitmelidirler.	13
Sanayi atıklarının arıtılması veya hammadde olarak değerlendirilmesi yolundaki teknolojik gelişmeler üzerindeki çalışmalar devam etmektedir.	11

Puana göre ters azalan sırada sıralanmış cümlelerden, daha önce kullanıcı tarafından belirlenen özetleme yüzdesi miktarına karşılık gelen sayıda cümle listenin başından seçilir. Son olarak seçilen bu cümleler, belgede geçme sırasına göre birleştirilerek sonuç özeti oluşturulur.

4.2.5 Başarımın Ölçülmesi

4.2.5.1 Keskinlik ve Anımsama

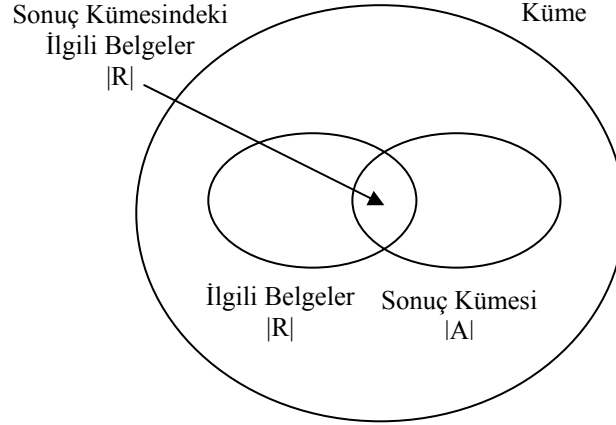
Genel olarak BE sistemlerinde başarımlar ölçütleri olarak kullanılan *keskinlik* ve *anımsama* Şekil 4.8’de de şematik olarak gösterildiği üzere şöyle tanımlanır:

- *Anımsama* konu ile ilgili getirilen belgelerin konu ile ilgili toplam belge sayısına oranı:

$$Anımsama = \frac{|R_a|}{|R|} \quad (4.2)$$

- *Keskinlik* de konu ile ilgili getirilen belgelerin getirilen toplam belge sayısına oranı:

$$Keskinlik = \frac{|R_a|}{|A|} \quad (4.3)$$



Şekil 4.8: Örnek bir bilgi talebi için keskinlik ve anımsama

Verilen bir sorgu için bir belge kümesinden ilgili belgeleri döndüren bir BE sisteminin başarımının ölçümünde sabit bir anımsama-keskinlik değer çifti elde edilmez. Bunun yerine erişim sistemi algoritmasında sorgu ile ilişkililiğe göre sıralanmış olarak dönen sonuç kümesindeki belgeler için baştan başlanarak birikimli olarak ilerlenir ve her adımda anımsama-keskinlik değer çifti hesaplanır ve sonuç olarak sorgu için bir anımsama-keskinlik grafiği elde edilir. Bir sorgu için elde edilen bu grafik değerleri, farklı anımsama-keskinlik grafikleri için şu şekilde hesaplanır:

$$\bar{P}(r) = \sum_{i=1}^{N_q} \frac{P_i(r)}{N_q} \quad (4.4)$$

Burada $\bar{P}(r)$ r anımsama seviyesindeki ortalama keskinlik, N_q kullanılan sorgu sayısı ve $P_i(r)$ de i 'inci sorgu için r anımsama seviyesindeki keskinliktir [25].

Tüm BE sistemlerinde olduğu gibi, metin özetlemede de düşük keskinlik oranı gereksiz kayıtların getirilmesine, düşük anımsama oranı ise sonuç kümesinde yer almayan yararlı belgeler olmasına yol açacaktır ki bunların ikisi de istenmez [4].

Tez çalışması kapsamında tüm yöntemlerin ayrı ayrı ve tüm birleşimlerinin keskinlik ve anımsama değerleri hesaplanmış ve sonuçları çalışmanın son bölümünde tartışılmıştır. Ayrıca bu başarımlara çeşitli gövdeleme algoritmalarının etkisini irdelemek amacıyla, farklı gövdeleyiciler kullanılarak özetleme başarımının değişimi gözlemlenmiştir. Sistemin başarımını ölçmek için 10 belgeden oluşan bir test derlemi kullanılmıştır. Derlemede farklı alanlardan (gazete köşe yazıları, tez çalışmaları, ders kitabı), çeşitli uzunluklarda belgeler ve bu belgelerin insan tarafından elle çıkarılmış

özetleri yer almaktadır. Bu derlemde bazı belgeler, bunların elle çıkarılmış özetleri ve özetleme uygulaması tarafından çıkarılmış sonuç özetleri ekler bölümünde yer almaktadır.

5. GERÇEKLENEN YAZILIM

5.1 Yazılımın Genel Yapısı

Tez kapsamında, metin özetleme algoritmasının uygulanıp, sonuçlarının değerlendirilebilmesi için yazılım gerçekleştirilmiştir. Yazılım, Zemberek doğal dil işleme kütüphanesinin yazıldığı, geniş dokümantasyona sahip ve tam nesneye dayalı programlama dili olan Java’da gerçekleştirilmiştir. Yazılımın Java dilinde geliştirilmesi için, bütünleşik geliştirme ortamı (IDE) olarak NetBeans tercih edilmiştir. Yazılımın temel parçaları; biçimbirimsel çözümleme, gövdeleme ve özetlemedir.

5.1.1 Biçimbirimsel Çözümleme

Biçimbirimsel çözümleme işlemi, girdi olarak verilen sözcüğün küçük harflere çevrilip içindeki uyumsuz karakterlerin ayıklanmasıyla başlar. Ardından çözümleme işlemi, sözcük için olası köklerin bulunmasıyla devam eder. *gözlükçülük* sözcüğü için kök bulucunun bulduğu olası kökler, Tablo 5.1’de gösterilmiştir. Şekil 4.3’te detaylı akış çizelgesi verilen çözümleme işlemi, üretilmiş köklere ek üretici vasıtasıyla ekler eklenip girdi sözcüğün oluşturulması ile devam eder. Girdi sözcüğü oluşturabilen tüm kök + ek birleşimleri seçilerek işlem sonlanır.

Tablo 5.1: Gözlükçülük sözcüğü kök adayları

Kök adayları
<i>G</i>
<i>Göz</i>
<i>Göze</i>
<i>Gözle</i>
<i>Gözlük</i>

Zemberek biçimbirimsel çözümleyicisi çıktı olarak *Kelime* nesnesi dizisi döndürür. Bir sözcük için biçimbirimsel çözümleyici birden fazla kök + ek birleşimi döndürebileceğinden, bir tane *Kelime* nesnesi yerine *Kelime* nesnesi dizisi döndürmektedir. Kelime sınıfının özellik ve metotları Şekil 5.1’de görülebilir.

```
private static final HarfDizisi BOS_ICERIK = new HarfDizisi(0);
private HarfDizisi icerik=BOS_ICERIK;
private Kok kok;
private List<Ek> ekler = new ArrayList<Ek>(3);
private KelimeTipi tip;

public Kelime();
public Ek[] ekDizisi();
public Kelime clone();
public List<Ek> ekler();
public String ekZinciriStr();
public Kelime(Kok kok, Alfabe alfabe);
public Kelime(Kok kok, HarfDizisi dizi);
public void setIcerik(HarfDizisi icerik);
public int ekSayisi();
public TurkceHarf sonHarf();
public HarfDizisi icerik();
public int boy();
public Ek sonEk();
public String icerikStr();
public void ekEkle(Ek ek);
public Kok kok();
public String toString();
public void icerikEkle(HarfDizisi eklenecek);
public boolean gercekEkYok();
```

Şekil 5.1: Kelime sınıfı yapısı

Kelime sınıfının içinde çözümlenmiş sözcüğe gelen ekleri tutan, *Ek* sınıftan türetilmiş nesnelere tutan *ekler* özelliği vardır. *Ek* sınıfının yapısı Şekil 5.2’de gösterilmiştir.

```
//bu ekten sonra gelebilecek eklerin listesi.  
private List<Ek> ardisilEkler = new ArrayList<Ek>();  
private String ad;  
//ekin sesli ile baslayip baslayamayacagini belirler. bu bilgi otomatik olarak ek  
olusum  
//kurallarına göre baslangicta belirlenir.  
private boolean sesliIleBaslayabilir = false;  
//kurallara göre ek olusumunu belirler. dile göre farklı gerçeklemleri olabilir.  
private EkUretici ekUretici;  
//bu ekin üretim kullarının listesi.  
private List<EkUretimBileseni> uretimBilesenleri;  
//bu eke ilişkin özel durumlar.  
private List<EkOzelDurumu> ozelDurumlar = new  
ArrayList<EkOzelDurumu>(1);  
private boolean sonEkOlamaz = false;  
private boolean halEki = false;  
private boolean iyelikEki = false;  
private Set<TurkceHarf> baslangicHarfleri;  
private String ekTipi;
```

Şekil 5.2: Ek sınıfı özellikleri

5.1.1.1 Eklerin Türünün Belirlenmesi

Zemberek'teki biçimbirimsel çözümleyici verilen sözcüğün tüm kök ve ek birleşimlerini eklerin adını belirterek döndürür. Gerçeklenen gövdeleyicilerde eklerin yapım veya çekim eki olmasına göre işlem yapılacağından eklerin yapım/çekim olarak nitelendirilmesi gerekmektedir. Bunun için BÇ'nin kullandığı ek XML dosyasındaki *ek* düğümlerine, Şekil 5.3'de görülen, *cekim*, *yapim* veya *kok* değerlerini alan *tip* isimli özellik eklenmiştir. Her ekin tipi tek Türk dili kitapları aracılığıyla belirlenmiş ve XML dosyası güncellenmiştir.

Zemberek kaynak kodlarında ek XML'inin diskten okunup kök ağacının oluşturulduğu yere bu özelliğin değerinin okunması ve okunan değer *Ek* sınıfına eklenen *ekTipi* özelliğine atanması işlemleri eklenmiştir. Böylece BÇ'nin ürettiği *Kelime* nesne listesindeki birleşimlerin içindeki *Ek* sınıfı listesi aracılığıyla bir ekin yapım eki mi çeki mi olduğuna kolayca erişilebilmiştir.


```
<ek ad="FIIL_DONUSUM_EN" uretim="+yAn" tip="yapim">
  <ardisil-ekler>
    <kume>ISIM_HAL</kume>
    <kume>ISIM_SAHİPLİK</kume>
    <kume>ISIM_KISI</kume>
    <kume>İMEK_ZAMAN</kume>
    <aek>ISIM_COĞUL_LER</aek>
    <aek>ISIM_YOKLUK_SIZ</aek>
    <aek>ISIM_BULUNMA_LİK</aek>
    <aek>ISIM_DURUM_LİK</aek>
    <aek>ISIM_BULUNMA_Lİ</aek>
    <aek>ISIM_BİRLİKTELİK_LE</aek>
    <aek>ISIM_DONUSUM_LES</aek>
  </ardisil-ekler>
  <oncelikli-ekler>
    <oek>ISIM_KISI_ONLAR_LER</oek>
  </oncelikli-ekler>
</ek>
```

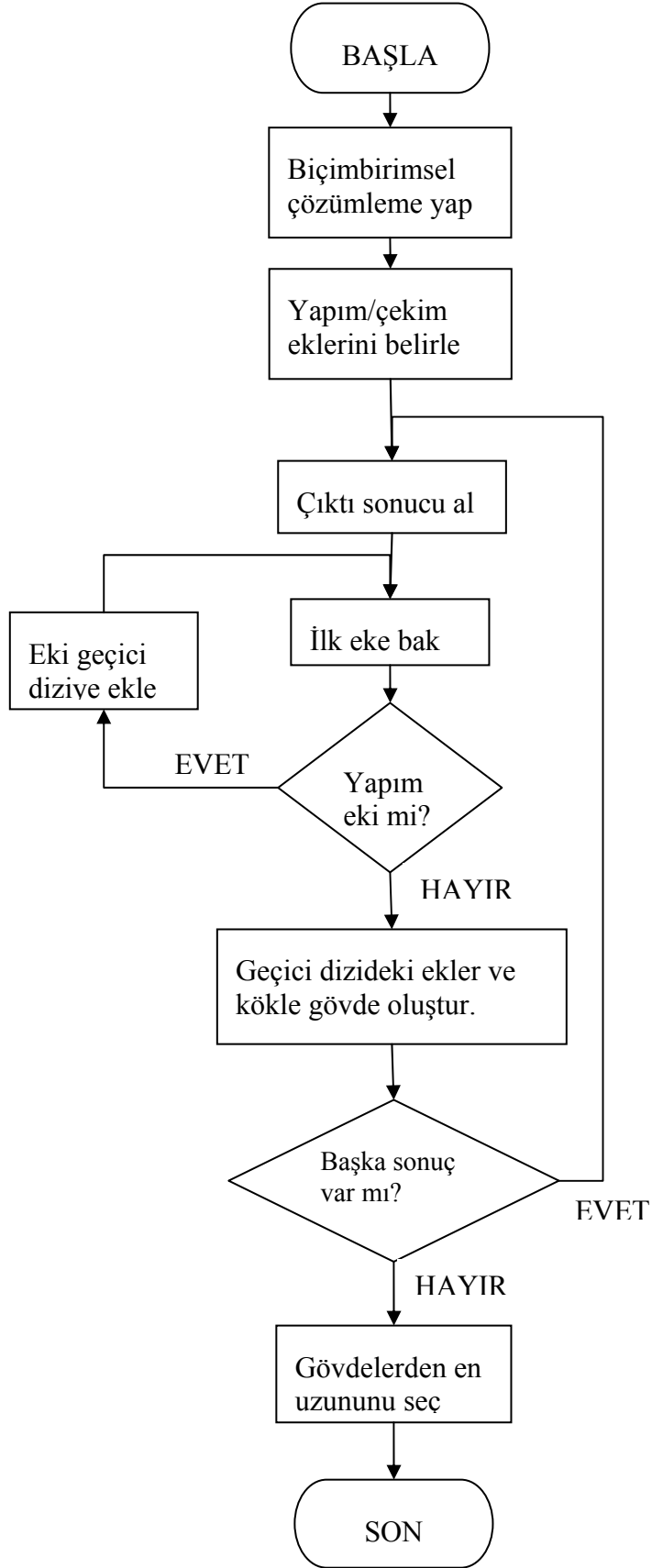
Şekil 5.3: Ek XML dosyasındaki örnek bir ek düğümü

5.1.2 Gövdeleme

Türkçe için yapılan özetleme sisteminde gövdeleme başarımı etkileyen önemli unsurlardandır. Bu nedenle tez kapsamında iki farklı gövdeleme algoritması gerçekleştirilmiş ve bu algoritmaların özetleme başarımına etkisi incelenmiştir. Bu bölümde gerçekleştirilen iki gövdeleyici anlatılmıştır.

5.1.2.1 En Uzun Gövdenin Seçilmesi

Türkçede istisnalar dışında çekim eklerinden sonra yapım eklerinin gelmemesi, yani kökten sonra yapım ekleri ardından da çekim eklerinin gelmesi kuralını esas olan bu yöntem biçimbirimsel çözümleme çıktısını girdi olarak alır. Şekil 5.4’de akış çizelgesi görülen en uzun gövdenin seçilmesi yöntemi, BÇ sonucu üretilen olası kök ve ek birleşimlerinden en uzun gövdeyi seçer.



Şekil 5.4: En uzun gövde seçimi algoritması akış çizelgesi

Algoritma biçimbirimsel çözümleme çıktılarından bir sözcüğün seçilmesi ile başlar. BÇ'nin ürettiği *Kelime* sınıfına ait nesnenin *ekler* özelliği kök sözcüğe eklenen ekleri tipleriyle birlikte tutar. BÇ biriminde yapılan güncellemeler sayesinde, çözümleme yapılırken, eklerin türleri (yapım/çekim) belirlenir. Ardından kök-ek hiyerarşisine sahip gövde adayının ekleri ilk ekten başlanarak incelenir. Sözcüğün sonundan çekim ekleri temizlenerek kök ve yapım eklerinden oluşan hali seçilir. Bu işlem tüm biçimbirimsel çözümleme çıktıları için tekrarlanır. Sonuçta elde edilen gövdelerden en uzun olanı sözcüğün gövdelenmiş hali olarak kabul edilir. Tablo 5.1'de *kalemin* sözcüğü için gövdeleme algoritmasının sonuçları görülmektedir. Buna göre, çekim eklerinden arındırılmış gövde adaylarından en uzun olanı (*kalem*) gövde olarak seçilir.

Tablo 5.2: Kalemın sözcüğü için gövdeleyici sonucu

Asıl Sözcük	BÇ Çıktısı	Gövde
<i>kalemin</i>	Kök: <i>kalem</i> Ekler: ISIM_KOK + ISIM_TAMLAMA_IN	<i>kalem</i>
	Kök: <i>kalem</i> Ekler: ISIM_KOK + ISIM_SAHİPLİK_SEN_IN	<i>kalem</i>
	Kök: <i>kale</i> Ekler: ISIM_KOK + ISIM_SAHİPLİK_BEN_IM + ISIM_TAMLAMA_IN	<i>kale</i>

5.1.2.2 İlk Beş Karakter Yöntemi

Türkçe tüm kök sözcüklerin beş harfli olma hipotezinin farklı bir şekli Köksal'ın [1] çalışmasında ele alınmıştır. Köksal yaptığı gövdeleme çalışmasında, elde ettiği deneysel sonuçlara göre sözcüklerin ilk beş harfini gövde olarak kabul etmiştir. Bu yöntemin metin özetleme verimliliğine etkisinin incelenmesi amacıyla, ilk beş karakter yöntemini kullanan gövdeleyici yazılımı geliştirilmiştir.

Aslen Köksal [1] tarafından biçimbirimsel çözümleyici kullanmadan sözcüğün ilk beş karakterinin gövde olmasına dayanan yöntem, tez çalışması kapsamında biçimbirimsel çözümleyici kullanacak şekilde yeniden yorumlanmıştır. Buna göre biçimbirimsel çözümleyicinin ürettiği kök-ek birleşimlerinden 5 karaktere en yakın olan çıktı gövde olarak seçilir.

Biçimbirimsel çözümleyicinin döndürdüğü *Kelime* sınıfı türünden nesne dizisi teker teker incelenir. Buna göre her çıktı için, öncelikle *kök*'ün beş karakterden uzun olup olmamasına bakılır. Eğer beş karakterden kısa ise, *Kelime* sınıfının bir üyesi olan

ekler dizisindeki ekler sıra ile köke eklenir. Oluşan sözcük beş karakter veya beş karaktere en yakın değere ulaşıncaya kadar bu işlem devam eder. İşlem tüm biçimbirimsel çözümleneçliktarı için tekrarlanır ve sonuç listesinden beş karaktere en yakın sözcük gövde olarak seçilir.

Yönteme göre biçimbirimsel çözümleneçliktarı Tablo 5.3'te görölen *gözlükçülüğü*n sözcüğü'nün gövdesi *gözlük* olarak belirlenecektir.

Tablo 5.3: Gözlükçülüğü'nün sözcüğü için BÇ çliktısı

Asıl Sözcük	BÇ Çliktısı
<i>gözlükçülüğü</i> n	Kok: <i>göz</i> Ekler: ISIM_KOK + ISIM_BULUNMA_LIK + ISIM_ILGI_CI + ISIM_DURUM_LIK + ISIM_SAHİPLİK_SEN_IN
	Kok: <i>gözlük</i> Ekler: ISIM_KOK + ISIM_ILGI_CI + ISIM_DURUM_LIK + ISIM_SAHİPLİK_SEN_IN
	Kok: <i>gözlük</i> Ekler: ISIM_KOK + ISIM_ILGI_CI + ISIM_DURUM_LIK + ISIM_TAMLAMA_IN
	Kok: <i>göz</i> Ekler: ISIM_KOK + ISIM_BULUNMA_LIK + ISIM_ILGI_CI + ISIM_DURUM_LIK + ISIM_TAMLAMA_IN

5.1.3 Özetleme

Girdi olarak bir belge alan ve çliktı olarak daha kısa, aslı'nın yerine geçene ve onun en önemli içeriğini barındıran bir süreç olarak tanımlanan özetleme sisteminin gerçekleşmesi için temelde dört adet sınıf kullanılmıştır:

1. cBelge
2. cParagraf
3. cCumle
4. cSozcuk

Bölüm 4.2.1'de detayları anlatılan Yapısal Çözümleneçliktarı Birimi tarafından, işlenmemiş metin alınarak, özetleme sisteminin kullanabileceği nesnelere oluşturulur. Bu nesnelere türetildiği sınıflardan ilki Şekil 5.5'te arayüzü görölen *cBelge* sınıfıdır. Bu sınıfta ham belgenin içeriğinin ayrıştırılması işlemi yapılır. Başlık ve paragraflardan oluşan düzenli bir belge için *cBelge* sınıfından oluşturulan nesnede;

- *baslik* özelliği belgenin başlık sözcüklerinin ayrıştırılmış hallerini *cSozcuk* sınıfından türetilmiş nesnelere dizisi olarak
- *paragrafDizisi* özelliği belgenin olduğu paragrafların ayrıştırılmış hallerini *cParagraf* sınıfından türetilmiş nesnelere dizisi olarak

tutarlar.

```

private String dosyaAdi = "";
private String encoding = "";
private List<cParagraf> paragrafDizisi = null;
private List<cAnahtarSozcuk> anahtarSozcukListesi = null;
private List<String> durmaListesi = null;
private List<String> olumluSozcukListesi = null;
private List<String> olumsuzSozcukListesi = null;

private Hashtable anahtarSozcukDizisi = null;
private List<cOzetCumle> ozetCumleListesi = null;
private List<cSozcuk> baslikSozcukListesi = null;

public cBelge(String pDosyaAdi, String pEncoding, List<String>
    pDurmaListesi, List<String> pOlumluSozcukListesi, List<String>
    pOlumsuzSozcukListesi)
public void oku();
public String toString();
public void icerikYazdir() throws IOException;
public void cozumle(Zemberek zemberek);
public void enUzunGovdeSec(Zemberek zemberek);
public void nHarfGovdeSec(Zemberek zemberek, int n);
public void dizinOlustur();
public void dizinYazdir(String pDosyaAdi) throws IOException;
public void hesaplaAnahtarSozcukPuan();
public void hesaplaBaslikPuan();
public void hesaplaIpucuPuan();
public void hesaplaKonumPuan();
public void hesaplaSonPuan();
public String ozetOlustur(int pOran);
public void ozetYazdir(String pDosyaAdi) throws IOException;
public void cumlePuanlariYazdir(String pDosyaAdi) throws IOException;
public void yazdirBaslikPuan(String pDosyaAdi) throws IOException;
public void yazdirIpucuPuan(String pDosyaAdi) throws IOException;

```

Şekil 5.5: cBelge sınıfı arayüzü

cBelge sınıfının *cozumle* metodu biçimbirimsel çözümleyici aracılığıyla belgedeki tüm paragrafların içindeki cümlelerin içindeki sözcüklerin biçimbirimsel olarak

çözümlemesini sağlar. *enUzunGovdeSec* ve *nHarfGovdeSec* metotları da benzer mantıkla gövdeleme algoritmalarını kullanarak gövdelenmesini sağlar. *hesaplaAnahtarSozcukPuan*, *hesaplaBaslikPuan*, *hesaplaIpucuPuan* ve *hesaplaKonumPuan* metotları ise ilgili yöntemler aracılığıyla cümlelerin puanlandırılması işlemini gerçekleştirir. *hesaplaSonPuan* metodu farklı yöntemlere göre cümlelere verilmiş puanları birleştirerek her cümle için son puan oluşturur. Son olarak da *ozetOlustur* metodu özetle yer alacak cümleleri seçip birleştirir ve sınıfın *ozetCumleListesi* özelliğini doldurur.

cBelge sınıfı içinde dizi nesnesi tutulan, Şekil 5.5'te arayüzü görülen *cParagraf* sınıfı benzer metot ve özelliklere ek olarak *paragrafKonumPuan* özelliği ile paragrafların Başlık 4.2.3.4'te açıklanan metnin içindeki konum yöntemine göre hesaplanmış paragraf puanları bilgisi tutulur.

```
private List<cCumle> cumleDizisi = null;
private String icerik = "";
private String encoding = "";
private float paragrafKonumPuan = 0.0F;

public cParagraf(String pParagraf, String pEncoding);
private void oku();
public String toString();
public void cozumle(Zemberek zemberek);
public void enUzunGovdeSec(Zemberek zemberek);
public void nHarfGovdeSec(Zemberek zemberek, int n);
```

Şekil 5.6: *cParagraf* sınıfı arayüzü

cParagraf sınıfında *cCumle* dizisi tipinde bir özellik olarak tutulan Şekil 5.7'te arayüzü görülen *cCumle* sınıfı, yöntemin cümle tabanlı puanlamaya dayanmasından dolayı önemli bir yere sahiptir. Sınıfta *cumleBaslikPuan*, *anahtarSozcukPuan*, *cumleIpucuPuan* ve *cumleKonumSonPuan* olarak isimlendirilen dört adet özellik mevcuttur. Bu özellikler, cümlelerin ilgili yöntemler uygulanarak hesaplanan puanlarını daha sonra yöntem birleştirme biriminde kullanılmak üzere tutarlar.

```

private List<cSozcuk> sozcukDizisi = null;
private String icerik = "";
private String encoding = "";

private int belgedekiSira = 0;

private float anahtarSozcukPuan = 0.0F;
private float cumleKonumPuan = 0.0F;
private float cumleKonumSonPuan = 0.0F;
private float cumleBaslikPuan = 0.0F;
private float cumleIpucuPuan = 0.0F;
private float cumleSonPuan = 0.0F;

public cCumle(String pCumle, String pEncoding);
private void oku();
public String toString();
public void cozumle(Zemberek zemberek);
public void enUzunGovdeSec(Zemberek zemberek);
public void nHarfGovdeSec(Zemberek zemberek, int n);

```

Şekil 5.7: cCumle sınıfı arayüzü

Metin özetleme sistemi mimarisinin yapıtaşını oluşturan *cSozcuk* sınıfından türetilmiş bir nesne dizisi benzer şekilde *cCumle* sınıfında özellik olarak yer almaktadır. Şekil 5.8’te arayüzü gösterilen bu sınıfta biçimbirimsel çözümleyicinin ürettiği sonuçlar *cozumlemeListesi* dizisinde, bu çözümleme sonuç listesi içinden seçilen çözümleme ise *secilenCozumleme* özelliğinde tutulur. Özetleme yöntemlerinde sıklıkla kullanılan sözcüğün gövdelenmiş hali ise *secilenGovde* isimli özellikte ve sözcüğün durma sözcüğü olup olmadığını gösteren bayrak ise sınıfın *durmaSozcuk* özelliğinde saklanır. Sözcük seviyesinde hesaplama yapan yöntemler olan başlık, sözcük sıklığı ve ipucu sözcük öbekleri yöntemleri sonucu hesaplanan puanlar *baslikPuan*, *anahtarSozcukPuan* ve *ipucuPuan* özelliklerinde tutulur.

```
private String icerik = "";
private Kelime[] cozumlemeListesi = null;
private Kelime secilenCozumleme = null;
private String secilenGovde = "";
private boolean durmaSozcuk = false;
private float anahtarSozcukPuan = 0.0F;
private float baslikPuan = 0.0F;
private float ipucuPuan = 0.0F;

public cSozcuk(String pIcerik);
public void cozumle(Zemberek zemberek);
public void enUzunGovdeSec(Zemberek zemberek);
public void nHarfGovdeSec(Zemberek zemberek, int n);
public String toString();
```

Şekil 5.8: cSozcuk sınıfı arayüzü

6. SONUÇLAR VE TARTIŞMA

Günümüzde hızla artan bilgi miktarı, aranan bilgiye ulaşmayı hedefleyen bilgi erişimi araştırmalarını doğurmuştur. Bu tez kapsamında bilgi erişimi araştırması konularından olan metin özetleme, Türkçe dili için ele alınmıştır.

Türkçe gibi sondan eklemeli ve kurallı bir dil için tasarlanan diğer tüm bilgi erişimi sistemlerinde olduğu gibi, metin özetleme sisteminde de kullanılan yöntemlerin doğası gereği gövdelemenin sistem başarımını etkileyen bir ölçüt olabileceği öngörülmüştür. Bu amaçla, sözcüklerin olası kök ve ek birleşimlerini veren biçimbirimsel çözümleyici kullanılarak iki gövdeleyici gerçekleştirilmiştir ve bu gerçekleştirilen gövdeleyicilerin metin özetleme sisteminin başarımına etkisi gövdeleme kullanmama ile karşılaştırılmıştır. Ayrıca özetlemede kullanılan yöntemlerin birbirine baskınlığı, tek başına veya birlikte kullanımda verdikleri sonuçlar irdelenmiştir.

Sonuçların irdelenmesinde 10 adet belge ve bu belgelerin insan tarafından elle çıkarılmış özetlerinden oluşan test derlemi kullanılmıştır.

Öncelikle iki gövdeleme yönteminin belgelerdeki bazı sözcükler için sergilediği farklı davranışlara örnek olarak Tablo 6.1’de görülen sözcükler gösterilebilir:

Tablo 6.1: Gövdeleme yöntemlerinin sonuçları

Sözcük	En Uzun Gövde Seçimi	5 Harf Gövde Seçimi	Açıklama
takımlar	takımla	takım	BÇ hatası
yaklaşan	yaklaşan	yaklaş	Yöntem farkı
seçimler	seçimle	seçim	BÇ hatası
düzenlediği	düzenledik	düzen	Yöntem farkı
düşüncesinin	düşünce	düşün	Yöntem farkı

Bu tabloda da görüldüğü gibi biçimbirimsel çözümleyicinin ürettiği yanlış çıktılar gövdeleme dolayısıyla da özetleme sistemini olumsuz yönde etkilemektedir. Örneğin, çoğul eki (-lar) biçimbirimsel çözümleyici tarafından -la + -r (yapım eki + zaman eki) şeklinde de yorumlanabilmektedir. Yöntemlerin doğası gereği farklı davrandıkları sözcüklere örnek olarak da yine Tablo 6.1’de görülen “yaklaşan” sözcüğü

gösterilebilir. En uzun gövde seçim algoritması sözcüğün gövdesi olarak “yaklaşan” sözcüğünü bulurken, beş harf gövde seçim algoritması “yaklaş” sözcüğünü bulmaktadır.

Bu farklı yöntemlerin belgedeki dizinleme sonuçlarını nasıl etkilediği ise Tablo 6.2’de görülebilir. Bu tabloda derlemdeki “Yaşam Kalitesi” başlıklı belgenin, en uzun gövdenin seçilmesi, beş harfli gövdenin seçilmesi ve gövdeleme uygulanmadan sözcüğün metinde geçtiği şeklinin kullanılması ile oluşturulan belge dizin sonuçlarının bir kısmı görülmektedir. Görüldüğü üzere belgede “yaşam”, “yaşamdan”, “yaşama”, “yaşaması”, “yaşamdaki”, “yaşamı” şeklinde geçen fakat aslında aynı kavramı (“yaşam”) ifade eden sözcük için beş harfli gövde algoritması bu ilişkiyi başarılı şekilde kurmakta ve “yaşam” kavramının 28 kez tekrar edildiğini çıkarmaktadır. En uzun gövde seçim algoritması ise “yaşam” kavramına ek olarak “yaşama” kavramının “yaşam” kökünden yapım eki ile türetilmiş bir gövde olduğundan ayrı bir kavram olarak değerlendirilmesi gerektiği sonucun üretmektedir. Benzer durum daha iyi bir şekilde “değer” kavramı için de görülmektedir.

Tablo 6.2: Gövdeleme yöntemlerinin sonuçları

En Uzun Gövde		Beş Harf Gövde		Gövdelemesiz	
Gövde	Sıklık	Gövde	Sıklık	Sözcük	Sıklık
yaşam yaşama	27 1	yaşam	28	yaşam yaşamdan yaşama yaşaması yaşamdaki yaşamı	22 2 1 1 1 1
kalite	21	kalite	21	kalitesi kalitesini kalitesinin	16 4 1
hasta hastalık hastala	10 7 2	hasta	19	hastanın hastalığın hastalar hastaların hasta hastalarda hastalıklarda hastalıklarıyla hastalığı	6 4 2 2 1 1 1 1 1
değerlendirme değer değerlendirilme değerlendirmek	4 2 2 2	değer	12	değerlendirme değerlendirmek değer değerlendirilirken	2 2 1 1

değerlendiril değerlendirilmek	1 1			değerlendirilmektedir değerlendirilmesini değerlendirilmesinin değerlendirmeden değerlendirmesi değerlere	1 1 1 1 1
sağlık	9	sağlık	9	sağlık sağlığa sağlığı	4 3 2

Tablo 6.3’de metin özetleme sisteminin “Yaşam Kalitesi” belgesi için oluşturduğu cümle puan tablosu görülmektedir. Tablodaki sonuçlar en uzun gövde algoritmasının seçimi ve asıl belgeden %25 oran ile özet belgesi oluşturulması parametresi ile elde edilmiştir. Cümle son puanları, her yöntem için cümleye -10 ile 10 arasında verilen puanların katkısı eşit verilerek elde edilmiştir.

Tablo 6.3: “Yaşam Kalitesi” belgesinin cümle puanları

Cümle Sıra No	İçerik	Anahtar Sözcük Puan	Başlık Puan	İpucu Puan	Konum Puan	Cümle Son Puan
16	MS'li hastalarda yapılan ...	10,00	10	0	10	7,50
0	Hastanın yaşamdan aldığı ...	3,94	3,75	0	10	4,42
19	Bazı çalışmalarda yaşam kalitesi ...	2,61	2,5	0	10	3,78
8	Özellikle "şifa"nın mümkün ...	6,14	3,75	0	5	3,72
1	"Bireyin içinde yaşadığı ...	5,54	3,75	0	5	3,57
9	Yaşam kalitesi, gerek ...	4,74	3,75	0	5	3,37
7	Yaklaşık olarak son 20 yıldır ...	5,82	2,5	0	5	3,33
17	Yaşam kalitesi, hastanın ...	5,14	2,5	0	5	3,16
3	Çok çeşitli tanımları ...	2,33	0	0	10	3,08
10	Subjektif olan yaşam kalitesi ...	4,02	2,5	0	5	2,88
2	Sağlığa ilişkin yaşam ...	3,90	2,5	0	5	2,85
18	Bunun yanısıra gerek ...	3,21	2,5	0	5	2,68
12	MS'li hastanın bakımında da ...	2,97	2,5	0	5	2,62
11	Bireysel yaşam kalitesi ...	2,65	2,5	0	5	2,54
14	MS'in yarattığı çeşitli ...	3,98	3,75	0	0	1,93
5	Yaygın olarak kullanılan ...	3,86	2,5	0	0	1,59
15	Üstelik etkilenen yaşam ...	5,30	3,75	-10	5	1,01
6	Bir başka deyişle, ...	2,25	0	0	0	0,56
4	Yaşam kalitesi kavramı, ...	4,02	2,5	-10	5	0,38
13	MS'te eskiden hastalığın ...	3,65	2,5	-10	0	-0,96

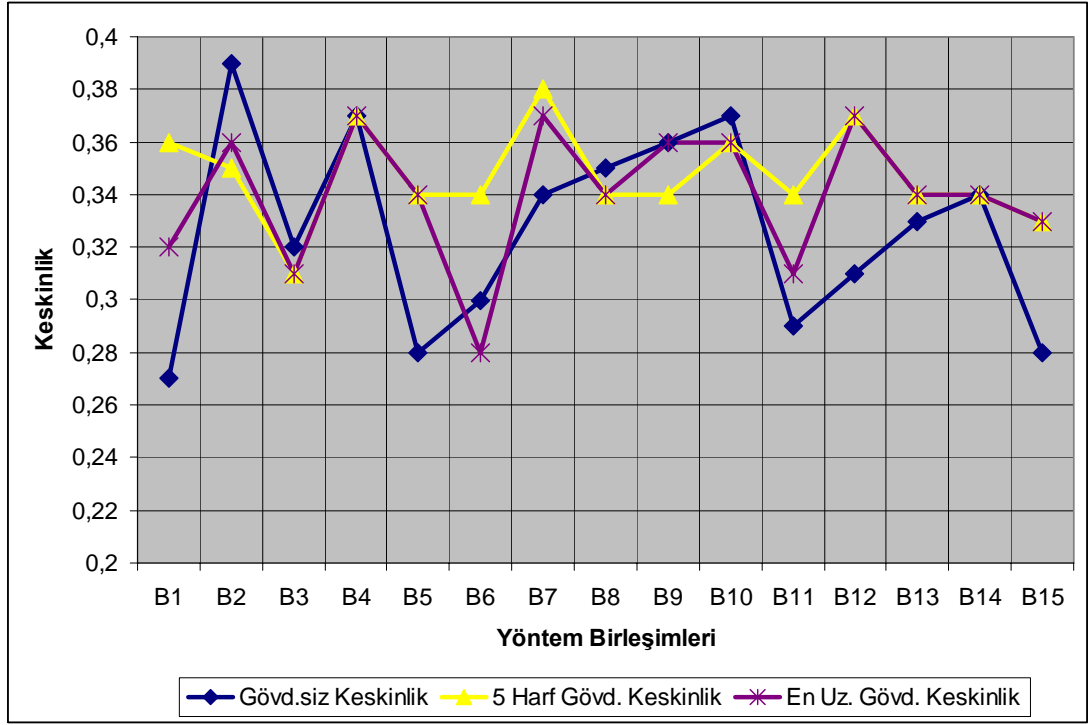
Derlemdeki belgelerden ve belgelerin elle çıkarılmış özetleri üzerinde uygulamanın %25 özet boyutu ile farklı yöntem birleşimleri için gövdeleme yapmadan, beş harf

gövde seçim algoritması ve en uzun gövdeleme algoritması kullanılarak elde edilen sonuçlar Tablo 6.4'teki gibidir:

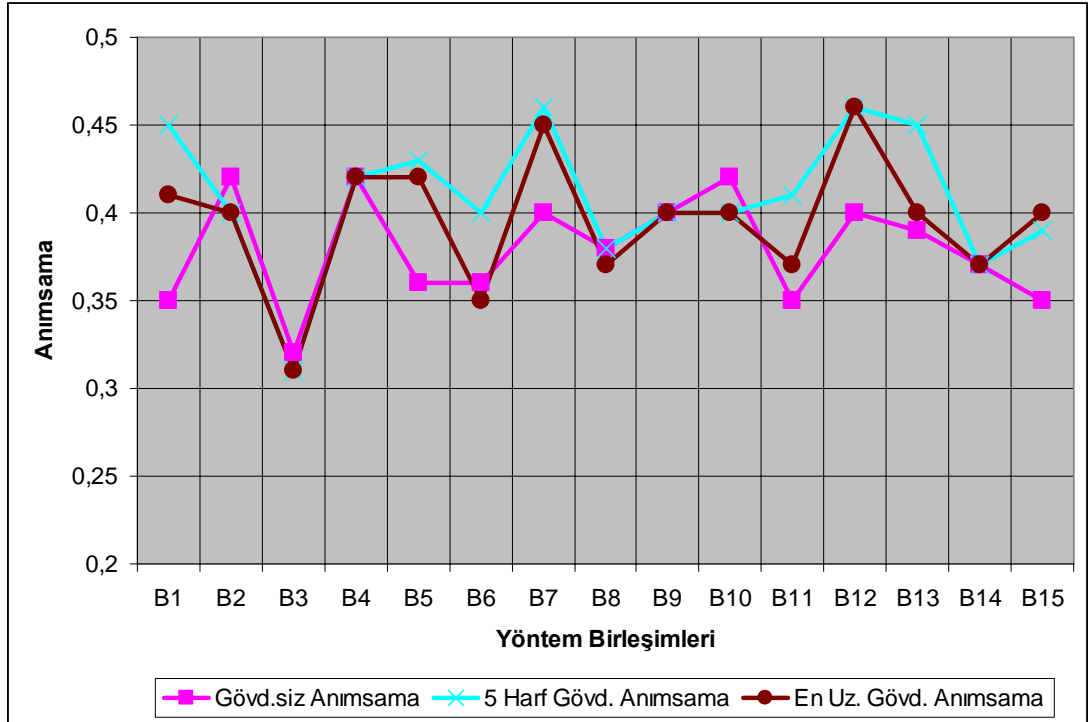
Tablo 6.4: Yöntemlerin başarımları (K: Keskinlik, A: Anımsama)

Gövde.siz		5 Harf Gövd.		En Uz. Gövd.		Anahtar Sözcük	Başlık	İpucu	Konum	Bir. No
K	A	K	A	K	A					
0,27	0,35	0,36	0,45	0,32	0,41	+	-	-	-	B1
0,39	0,42	0,35	0,40	0,36	0,40	-	+	-	-	B2
0,32	0,32	0,31	0,31	0,31	0,31	-	-	+	-	B3
0,37	0,42	0,37	0,42	0,37	0,42	-	-	-	+	B4
0,28	0,36	0,34	0,43	0,34	0,42	+	+	-	-	B5
0,30	0,36	0,34	0,40	0,28	0,35	+	-	+	-	B6
0,34	0,40	0,38	0,46	0,37	0,45	+	-	-	+	B7
0,35	0,38	0,34	0,38	0,34	0,37	-	+	+	-	B8
0,36	0,40	0,34	0,40	0,36	0,40	-	+	-	+	B9
0,37	0,42	0,36	0,40	0,36	0,40	-	-	+	+	B10
0,29	0,35	0,34	0,41	0,31	0,37	+	+	+	-	B11
0,31	0,40	0,37	0,46	0,37	0,46	+	+	-	+	B12
0,33	0,39	0,34	0,45	0,34	0,40	+	-	+	+	B13
0,34	0,37	0,34	0,37	0,34	0,37	-	+	+	+	B14
0,28	0,35	0,33	0,39	0,33	0,40	+	+	+	+	B15

Tablo 6.4'ün ilk dört satırında özetlemede kullanılan yöntemlerin tekil olarak farklı birleşimleri ile keskinlik ve anımsama değerleri görülmektedir. Ayrıca farklı gövdeleme algoritmalarının bu birleşim kümelerinden (B1, B2, ... B14) elde edilen sonuçlara etkisi de görülmektedir. Bu değerler neticesinde, yöntem birleşimlerinin farklı gövdeleme algoritmaları kullanıldığında ürettiği sonuçların keskinlik grafiği Şekil 6.1'deki, anımsama grafiği de Şekil 6.2'deki gibidir.



Şekil 6.1: Yöntemlerin keskinlik başarımı



Şekil 6.2: Yöntemlerin anımsama başarımı

Şekil 6.1 ve Şekil 6.2'den görüldüğü üzere, gövdeleme kullanıldığında neredeyse tüm yöntem birleşimleri daha yüksek keskinlik ve anımsama değerleri üretme eğilimindedir. Gövdeleme yöntemleri içinden ise beş harf gövdeleme yönteminin

diğer gövdeleme yöntemi olan en uzun gövdenin seçilmesine göre başarımının daha yüksek olduğu söylenebilir.

Tablo 6.4'e bakılarak, tek başına kullanıldığında konum yönteminin en iyi yöntem olduğu görülmektedir. Tek başına en düşük keskinlik-anımsama değerlerine sahip yöntem olarak ipucu sözcük öbekleri yöntemi görülmektedir. Tüm yöntem birleşimleri içinde ise anahtar sözcük, başlık ve konum yöntemlerinin birleşiminin (B12) en yüksek başarıyı verdiği, anahtar sözcük ve ipucu sözcük öbekleri yöntemlerinin birleşiminin (B6) de en düşük başarıya sahip olduğu gözlenebilir.

Tablo 6.5'te görüldüğü gibi, anahtar sözcük ve konum yöntemleri diğer yöntemler ile birlikte kullanıldığında en yüksek bağıl keskinlik ve anımsama artışına sahip yöntemlerdir.

Tablo 6.5: Yöntemlerin tekil ve bağıl değerlendirilmesi

Yöntem	Tekil Keskinlik	Tekil Anımsama	Bağıl Keskinlik Artışı	Bağıl Anımsama Artışı
Anahtar Sözcük	%36	%45	%9 (Başlık ve konum yöntemi ile birlikte kullanıldığında)	%15 (Başlık ve konum yöntemi ile birlikte kullanıldığında)
Başlık	%35	%40	-%2,6 (Anahtar sözcük ve konum yöntemi ile birlikte kullanıldığında)	%0 (Anahtar sözcük ve konum yöntemi ile birlikte kullanıldığında)
İpucu	%31	%31	-%2,7 (Konum yöntemi ile birlikte kullanıldığında)	-%4,8 (Konum yöntemi ile birlikte kullanıldığında)
Konum	%37	%42	%8,8 (Anahtar sözcük ve başlık yöntemi ile birlikte kullanıldığında)	%7 (Anahtar sözcük ve başlık yöntemi ile birlikte kullanıldığında)

Yöntemlerin başarımını arttırmak amacıyla, deneyler sonucunda en düşük başarıyı veren ipucu sözcük öbekleri yönteminin kullandığı olumlu ve olumsuz sözcükler listesi, daha geniş bir derlemde yapılacak çalışmalar sonucu genişletilebilir. Ayrıca sözcük öbeklerinin cümlenin içinde bulunmasına değil de cümlenin başında bulunmasına bakılması keskinliği arttırabilecektir. Ana başlığa ek olarak içerisinde alt başlıklar da olan metinlerde başarımın arttırılması için başlık yöntemi alt başlıkları da kapsayacak şekilde genişletilebilir.

Tasarlanan metin özetleme sistemi, girdi olarak verilen bir başlığa sahip, e-posta, bilimsel makale, gazete makaleleri gibi Türkçe metinlerin özetini farklı istatistiksel yöntemleri birleştirerek çıkarabilmektedir. Sohbet kayıtları, reklâm ve ilanlar gibi daha düzensiz metinlerde çok olumlu sonuç vermeyecektir.

Gövdelemenin gerçekleşen sistemin başarımına etkisi bir önceki bölümde anlatılmıştı. Metin içindeki sözcüklerin biçimbirimsel çeşitliliklerinin göz önüne alınmasının özetleme başarımını arttırdığı sonucundan yola çıkarak, sözcüksel çeşitliliklerin (eşanlamlı sözcükler), anlambilimsel (eşsesli sözcükler) ve sözbilimsel çeşitliliklerin (tamlama analizi) hesaba katılmasının özetleme başarımına olumlu etkide bulunacağı öngörülebilir. Ayrıca farklı biçimbirimsel çözümleyici ve gövdeleyiciler kullanılarak sistem başarımına etkisi incelenebilir.

Gerçeklenen metin özetleme sistemi herhangi bir alana bağlı değil de genel amaçlı tasarlandığından herhangi bir sözcük işleme programı veya diğer biçimler yerine salt metin girdilerle çalışmaktadır. Fakat özetleme hedef alanı biraz daha daraltılarak yeni yöntemler geliştirilebilir. Örneğin Genel Ağ sayfalarının özetlenmesi hedefleniyorsa, koyu veya yatık yazılmış sözcükleri barındıran cümlelerin özetle olma olasılığının bulundurmayanlara göre daha yüksek olduğundan söz edilebilir.

Genel amaçlı olmayan fakat bilgili, ne aradığını bilen kullanıcıya hitaben yapılabilecek metin özetleme sisteminde kullanıcıdan anahtar sözcüklerin alınarak bu sözcüklerin de olumlu sözcükler gibi değerlendirilmesi başarımı arttıracaktır. Benzer şekilde belirli bir alana has özetleme yapılmak istendiğinden alana ilişkin terimlerin sisteme girdi olarak verilmesi özetleme başarımını arttıracaktır.

Girdi olarak tek belge alan özetleme sistemi kullanılarak, çoklu belge özetleyen sistem geliştirilerek sözcük sıklığı yöntemi, ters belge sıklığı ile birlikte daha verimli kullanılabilir.

Türkçe bilgi erişimi sistemlerinde kullanılacak genel amaçlı kavramsal derlem kullanılarak, Türkçenin hitap yapısından anlamsal zincirler şeklinde temsil edilerek faydalanılabilir. Ek olarak belgeler ve bunların özetlerinden oluşan genel bir eğitim derlemi hazırlanarak, yöntemlerin birleşmesinde kullanılan katsayıların ve metnin içindeki konum yöntemindeki konum katsayılarının derlemden eğitim yoluyla hesaplanmasının başarımı arttıracığı öngörülmektedir.

İngilizce için, Gelişmiş Araştırma ve Geliştirme Hareketi (ARDA) destekleyiciliğinde düzenlenen amacı araştırmacılara yöntemlerini ve uygulamalarını geniş çaplı veri kümesi üzerinde test edip sonuçlarını paylaşabilecekleri ortam sağlamak olan Belge Anlama Konferansı (DUC) benzeri konferanslar düzenlenerek gerçekleştirilen özetleme sistemlerinin başarımının ölçülmesine yönelik evrensel kıstaslar belirlenerek yapılan çalışmaların bu koşullar altında test edilmesi geleceğe yönelik çalışmalara daha doğru yön verecektir.

KAYNAKLAR

- [1] **Can, F.**, 2006. Turkish Information Retrieval: Past Changes Future, in *Lecture Notes in Computer Science*, Vol. 4243, pp. 13-22, Springer Verlag, Berlin Heidelberg.
- [2] **Sembok, T.M.T.**, 2003. Character Strings to Natural Language Processing in Information Retrieval, in *Lecture Notes in Computer Science*, Vol. 2911, pp. 26-33, Springer Verlag, Berlin Heidelberg.
- [3] **Jurafsky, D. and Martin, J.H.**, 2000. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice Hall, N.J.
- [4] **Pembe, F.C. and Say, A.C.C.**, 2004. A Linguistically Motivated Information Retrieval System for Turkish, in *Lecture Notes in Computer Science*, Vol. 3280, pp. 741-750, Springer Verlag, Berlin Heidelberg.
- [5] Dođal Dil İşleme, Wikipedia: Özgür Ansiklopedi, 15 Nisan 2007 <http://tr.wikipedia.org/wiki/Do%C4%9Fal_dil_i%C5%9Flleme>
- [6] **Kesgin, F.**, 2007. Türkçe metinler için konu belirleme sistemi, *Yüksek Lisans Tezi*, İ.T.Ü. Fen Bilimleri Enstitüsü, İstanbul.
- [7] **Manning, C.D. and Schütze, H.**, 1999. Foundations of Statistical Natural Language Processing, The MIT Press, Massachusetts
- [8] **Mani, I. and Maybury, M.T.**, 1999. Advances in automatic text summarization, MIT Press, London.
- [9] **Jackson, P. and Moulinier, I.**, 2002. Natural language processing for online applications: text retrieval, extraction, and categorization, Amsterdam.
- [10] **Jones, S., Lundy, S., Paynter, G.W.**, 2002. Interactive Document Summarization Using Automatically Extracted Keyphrases, *35th Annual Hawaii International Conference on System Sciences (HICSS'02)-Volume 4*, Hawaii.
- [11] **Çatıkkaş, M.A.**, 1996. Üniversiteler için uygulamalı Türk dili ve kompozisyon kılavuzu, Alfa, İstanbul.
- [12] **Korkmaz, Z. ve diğ.**, 2001. Yüksek öğretim öğrencileri için Türk dili ve kompozisyon bilgileri, Yargı, İstanbul.
- [13] **Sarıca, S., Gündüz, M.**, 1992. Güzel konuşma yazma: kompozisyon, Fil, İstanbul.
- [14] **Özdemir, E.**, 1998. Anlatım sanatı (kompozisyon), Remzi Kitabevi, İstanbul.
- [15] **Luhn, H.P.**, 1958. The automatic creation of literature abstracts, in *IRE National Convention*. Reprinted in: Mani, I., Maybury, M.T. (eds.) *Advances in automatic text summarization*, pp. 15-21, MIT Press, London.

- [16] **Edmundson, H.P.**, 1969. New methods in automatic abstracting, in *Journal of the Association for Computing Machinery* 16 (2). 264-285. Reprinted in: Mani, I., Maybury, M.T. (eds.) *Advances in Automatic Text Summarization*, pp. 23-42, MIT Press, London.
- [17] **Kupiec, J., Pedersen, J. and Chen, F.**, 1995. A trainable document summarizer, in *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 68-73, Washington. Reprinted in: Mani, I., Maybury, M.T. (eds.) *Advances in Automatic Text Summarization*, pp. 55-60, MIT Press, London.
- [18] **Myaeng, S.H. and Jang, D.H.**. A trainable document summarizer. Reprinted in: Mani, I., Maybury, M.T. (eds.) *Advances in Automatic Text Summarization*, pp. 61-70, MIT Press, London.
- [19] **Aone, C., Okurowski, M.E., Gorfinsky, J. and Larsen, B.**. A trainable summarizer with knowledge acquired from robust NLP techniques. Reprinted in: Mani, I., Maybury, M.T. (eds.) *Advances in Automatic Text Summarization*, pp. 71-80, MIT Press, London.
- [20] **Hovy, E. and Lin, C.Y.**, 1997. Automated text summarization in SUMMARIST, in *Proceedings of the Workshop on Intelligent Scalable Text Summarization*, pp. 18-24, Madrid. Reprinted in: Mani, I., Maybury, M.T. (eds.) *Advances in Automatic Text Summarization*, pp. 81-94, MIT Press, London.
- [21] **Dinçer, B.T., Karaođlan, B.**, 2003. Stemming in agglutinative languages: A probabilistic stemmer for Turkish, in *Lecture Notes in Computer Science*, Vol. 2869, pp. 244-251, Springer Verlag, Berlin Heidelberg.
- [22] **Altıntaş, K. and Can, F.**, 2002. Stemming for Turkish A comparative evaluation, in *Proceedings of the 11th Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN)*, pp. 181-188, Istanbul University Press, Istanbul.
- [23] **Oflazer, K.**, 1994. Two-level Description of Turkish Morphology, *Literary and Linguistic Computing*, Vol. 9, No:2.
- [24] **Akın, A.A. and Akın, M.D.**, 2007. Zemberek, an open source NLP framework for Turkic languages.
- [25] **Baeza-Yates, R. and Ribeiro-Neto, B.**, 1999. Modern Information Retrieval, Addison-Wesley, England.

EKLER

EK A

Bu bölümde test derleminde yer alan belge, elle çıkarılmış özet ve uygulama çıktısı özetlerden bazıları yer almaktadır:

Belge I

YAŞAM KALİTESİ

Hastanın yaşamdan aldığı doyumun ve tedavi/bakımdan gördüğü yararın değerlendirilmesinin en iyi yöntemlerinden biri yaşam kalitesini ölçmektir. "Bireyin içinde yaşadığı kültür, değer sistemi, kişisel hedefleri, normları ve ilgileri doğrultusunda yaşamdaki yerini algılayış biçimi" şeklinde tanımlanan yaşam kalitesi, geniş ve kompleks bir kavram olup kişinin fiziksel sağlığı, psikolojik durumu, bağımsızlık düzeyi, sosyal ilişkileri ve çevresiyle olan bağlantısından etkilenir (43, 145). Sağlığa ilişkin yaşam kalitesi kavramı ise kişinin fiziksel, psikolojik ve sosyal işlevsellik boyutlarını değerlendirme durumudur. Çok çeşitli tanımları yapılmış ve her bir tanımda kapsamı farklı tanımlanmış olan bu kavram Vickrey tarafından fiziksel, mental ve sosyal sağlığı kapsayan çok boyutlu yapı olarak tanımlanmıştır (150).

Yaşam kalitesi kavramı, algısal olması ve ancak kişi tarafından rapor edilebilmesi itibarıyla "dışarıdan gözlenebilen" ve "direkt olarak fiziksel ölçüm yapılabilen" objektif değerlendirmeden farklılık göstermektedir. Yaygın olarak kullanılan objektif ölçüm araçlarından farklı olarak, yaşam kalitesi ölçümü ölçülen fizyolojik değerlere değil hastanın algılamasına odaklanmaktadır. Bir başka deyişle, sağlık ekibinin hastanın durumuna ilişkin görüşleri kadar kişinin kendini nasıl hissettiğini de önemseyen ve sorunlara hastanın perspektifinden bakan bir yaklaşım sunmaktadır. Yaklaşık olarak son 20 yıldır sağlık bakımının hastalığı tedavi etmenin ötesinde, sağlık ve esenliği koruma ve geliştirme kavramlarına odaklanması da sağlığa ilişkin yaşam kalitesi kavramının geniş perspektifiyle örtüşmekte ve araştırma değişkeni olarak kabul görmesini kolaylaştırmaktadır (118).

Özellikle "şifa"nın mümkün olmadığı, hastalığın yarattığı değişikliklerin tüm yaşamı etkilediği ve bakımın hastaların hastalıklarıyla optimal şekilde yaşaması üzerine odaklandığı kronik hastalıklarda, sağlığa ilişkin yaşam kalitesi kavramının önem kazanması çok doğaldır. Yaşam kalitesi, gerek hastaların profilini belirlemek ve yaşamdan memnuniyet derecelerine ayna tutmak gerek hastalar için uygulanan tedavi /bakım girişimlerinin etkinliğini ölçmek için değerlendirilmektedir (94).

Subjektif olan yaşam kalitesi ölçümü yukarıda sözü edilen özellikleri ve çok boyutlu bir değerlendirme sağlayarak hemşirelikte önemli olan holistik yaklaşıma hizmet etmesinin yanısıra, bireysellik kavramına da atıfta bulunması nedeniyle hemşirelikte kabul görmektedir. Bireysel yaşam kalitesi anlayışları içinden evrensele ulaşma amacı güden bu çalışmaların hemşirelik literatüründe anlamlı yeri vardır.

MS'li hastanın bakımında da yaşam kalitesi çalışmalarının önemi giderek artmaktadır. MS'te eskiden hastalığın sadece görünür etkileri yani yetiyitimi değerlendirilirken, günümüzde hastalığın etkilerinin daha geniş bir perspektifle değerlendirilmesini sağlayan yaşam kalitesini ölçme yaklaşımı benimsenmektedir (85, 91, 123, 177). MS'in yarattığı çeşitli etkilerle yaşamın tüm boyutlarını etkilemesi çok boyutlu yaşam kalitesi değerlendirmesi için gerekçe oluşturmaktadır. Üstelik etkilenen yaşam boyutları hastalar tarafından genel sağlık durumunun en iyi belirleyicisi olarak rapor edildiğinden, hastalığın yaşama etkisini değerlendirmek için yaşam kalitesini değerlendirmek kaçınılmaz olmaktadır (118, 204).

MS'li hastalarda yapılan yaşam kalitesi çalışmaları incelendiğinde bunların bir bölümünün yaşam kalitesi ölçüm araçlarının geçerlik ve güvenilirliği çalışmaları olduğu, bir bölümünün yaşam kalitesinin ve yaşam kalitesini etkileyen faktörlerin belirlenmesi çalışmaları olduğu görülmektedir. Yaşam kalitesi, hastanın başka tedavi veya yöntemlere de gereksinimi olup olmadığını ve uygulanan girişimlerin hasta tarafından da etkili bulunup bulunmadığını anlamak, tedaviden elde edilen yarar ile yan etkilerin neden olduğu zarar ve tedavinin güçlükleri konusunda fikir edinmek için kullanılmaktadır. Bunun yanısıra gerek fiziksel gerek emosyonel boyutlarıyla yaşam kalitesi, fiziksel sonuçların önemli bir prediktörü de olabilmektedir. Bazı çalışmalarda yaşam kalitesi ölçeklerinin, yetiyitimi değişikliğinin göstergesi olduğu belirtilmektedir (118, 178, 204).

YAŞAM KALİTESİ – ÖZET

Hastanın yaşamdan aldığı doyumun ve tedavi/bakımdan gördüğü yararın değerlendirilmesinin en iyi yöntemlerinden biri yaşam kalitesini ölçmektir. "Bireyin içinde yaşadığı kültür, değer sistemi, kişisel hedefleri, normları ve ilgileri doğrultusunda yaşamdaki yerini algılayış biçimi" şeklinde tanımlanan yaşam kalitesi, geniş ve kompleks bir kavram olup kişinin fiziksel sağlığı, psikolojik durumu, bağımsızlık düzeyi, sosyal ilişkileri ve çevresiyle olan bağlantısından etkilenir (43, 145). Sağlığa ilişkin yaşam kalitesi kavramı ise kişinin fiziksel, psikolojik ve sosyal işlevsellik boyutlarını değerlendirme durumudur. Çok çeşitli tanımları yapılmış ve her bir tanımda kapsamı farklı tanımlanmış olan bu kavram Vickrey tarafından fiziksel, mental ve sosyal sağlığı kapsayan çok boyutlu yapı olarak tanımlanmıştır (150). Yaşam kalitesi kavramı, algısal olması ve ancak kişi tarafından rapor edilebilmesi itibarıyla "dışarıdan gözlenebilen" ve "direkt olarak fiziksel ölçüm yapılabilen" objektif değerlendirmeden farklılık göstermektedir. Yaygın olarak kullanılan objektif ölçüm araçlarından farklı olarak, yaşam kalitesi ölçümü ölçülen fizyolojik değerlere değil hastanın algılamasına odaklanmaktadır. Yaşam kalitesi, gerek hastaların profilini belirlemek ve yaşamdan memnuniyet derecelerine ayna tutmak gerek hastalar için uygulanan tedavi /bakım girişimlerinin etkinliğini ölçmek için değerlendirilmektedir (94). Subjektif olan yaşam kalitesi ölçümü yukarıda sözü edilen özellikleri ve çok boyutlu bir değerlendirme sağlayarak hemşirelikte önemli olan holistik yaklaşıma hizmet etmesinin yanısıra, bireysellik kavramına da atıfta bulunması nedeniyle hemşirelikte kabul görmektedir. MS'li hastanın bakımında da yaşam kalitesi çalışmalarının önemi giderek artmaktadır. MS'te eskiden hastalığın sadece görünür etkileri yani yetiyitimi değerlendirilirken, günümüzde hastalığın etkilerinin daha geniş bir perspektifle değerlendirilmesini sağlayan yaşam kalitesini ölçme yaklaşımı benimsenmektedir (85, 91, 123, 177). MS'li hastalarda yapılan yaşam kalitesi çalışmaları incelendiğinde bunların bir bölümünün yaşam kalitesi ölçüm araçlarının geçerlik ve güvenilirliği çalışmaları olduğu, bir bölümünün yaşam kalitesinin ve yaşam kalitesini etkileyen faktörlerin belirlenmesi çalışmaları olduğu görülmektedir.

YAŞAM KALİTESİ – OTOMATİK ÖZET

Hastanın yaşamdan aldığı doyumun ve tedavi/bakımdan gördüğü yararın değerlendirilmesinin en iyi yöntemlerinden biri yaşam kalitesini ölçmektir. Yaklaşık

olarak son 20 yıldır sağlık bakımının hastalığı tedavi etmenin ötesinde, sağlık ve esenliği koruma ve geliştirme kavramlarına odaklanması da sağlığa ilişkin yaşam kalitesi kavramının geniş perspektifiyle örtüşmekte ve araştırma değişkeni olarak kabul görmesini kolaylaştırmaktadır (118).MS'li hastalarda yapılan yaşam kalitesi çalışmaları incelendiğinde bunların bir bölümünün yaşam kalitesi ölçüm araçlarının geçerlik ve güvenilirliği çalışmaları olduğu, bir bölümünün yaşam kalitesinin ve yaşam kalitesini etkileyen faktörlerin belirlenmesi çalışmaları olduğu görülmektedir. Yaşam kalitesi, hastanın başka tedavi veya yöntemlere de gereksinimi olup olmadığını ve uygulanan girişimlerin hasta tarafından da etkili bulunup bulunmadığını anlamak, tedaviden elde edilen yarar ile yan etkilerin neden olduğu zarar ve tedavinin güçlükleri konusunda fikir edinmek için kullanılmaktadır. Bazı çalışmalarda yaşam kalitesi ölçeklerinin, yetiyitimi değişikliğinin göstergesi olduğu belirtilmektedir (118, 178, 204).

Belge II

LİBERALİZM

Bilindiği gibi, Orta Çağlarda Rönesans hareketi sanat alanında ve Reformasyon hareketi de din alanında insan düşüncesine bir serbesti, bir çeşit hürriyet getirme amacını gütmüştür. Rönesans ile birlikte sanatkar tabiata daha geniş bir serbestlik ve hürriyet içinde bakmaya, ilhamlarını daha geniş sınırlar içinde işlemeye başlamıştır. Reformasyon hareketi ise, o zamana kadar egemen din olarak katolikliğin sert katı ve hoşgörüsüz din kalıbını kırarak, Tanrı ile insan arasındaki münasebetlere bir hürriyet getirmek amacını gütmüştür. Bunun sonucu olarak; Protestanlık denen yeni bir din şekli ve onun da çeşitli mezhepleri ortaya çıkmıştır. Protestanlık ve onun çeşitli mezhepleri, Tanrı-İnsan münasebetlerine, katoliklikten çok farklı yeni yeni şekiller getirmiştir.

Fakat şu var ki, ne sanat alanındaki hürriyet gelişimi ve ne de din alanındaki hürriyet gelişimi, insanın toplum içindeki siyasal yaşayışına herhangi bir serbestlik veya hürriyet getirmekten çok uzak kalmıştır. Sanat ve din alanında insan düşüncesine bir dereceye kadar hürriyet gelmiş, lakin hürriyet insanın toplum içindeki siyasal durumunu değiştirememiştir. İnsanlar yine, kudretini ve yetkisini Tanrıdan aldığını iddia eden, kral, imparator, prens veya hükümdar denen bir tek insanın sert, mutlak ve sınırsız otoritesine tabi olarak ve onun keyfi idaresi altında yaşamaya devam

etmişlerdir. Daha açık bir deyişle, toplum içinde kişinin insan olarak hiçbir değeri tanınmamıştır.

İşte 1789 İhtilali ile ki, toplumların bu siyasal düzeni yıkılmaya başlamıştır. Şimdi hükümdarın sınırsız otoritesine karşı, kişinin varlığı ve bu varlığın, insan olmak haysiyeti dolayısıyla sahip bulunduğu temel hak ve hürriyetler, sınırlayıcı bir unsur olarak çıkıyordu. Siyasal düzen, hükümdarın otoritesi ile, kişinin insanca yaşama ilkesi arasında kurulan bir dengeye dayandırılmak isteniyordu. Fransız İhtilalcileri daha ihtilal kaynaşmalarının ilk aylarında, 28 Ağustos 1789 da, yayınladıkları "İnsan ve Vatandaş Hakları Demeci" ile bu dengeyi açıkça ilan ettiler. Bu demecin esasları şöyle idi: İnsanlar hakları bakımından hür ve eşit doğarlar ve öyle kalırlar. Bu haklar hürriyet, mülkiyet, güvenlik ve zulme karşı direnme haklarıdır. Her türlü egemenlik esas olarak millettendir. Kanun millet egemenliğinin ifadesidir. Her vatandaş hür bir şekilde konuşabilir, yazabilir ve yayında bulunabilir. Kamu düzenine dokunmadıkça, kimse dini ve siyasi inançlarından dolayı kınanamaz.

Görülüyor ki bu demeç, insanın insan olmak dolayısıyla, daha doğduğu andan itibaren, bir takım temel hak ve hürriyetlere sahip bulunduğunu ortaya koyuyor ve hükümdarın otoritesini de bu temel hak ve hürriyetlerle sınırlıyordu. Bu, Avrupa'da insanların ilk defa gördükleri ve işittikleri bir şeydi. (İnsan ve Vatandaş Hakları Demeci'nin ortaya attığı bu hürriyet ilkeleri, daha önce 4 Temmuz 1776 tarihli Amerikan Bağımsızlık Demeci'nde de ortaya atılmıştı. Fakat, o zamanın ulaşım imkanlarının yetersizliği ve iki kıta, arasında geniş bir deniz parçasının bulunması dolayısıyla, Amerika'daki bu hürriyetçi hareket Avrupa'ya fazla tesir edememiş, ancak birkaç Avrupalı aydının dikkatini çekmiştir. Mamafih, Fransız İhtilali liderlerinin Amerikan Bağımsızlık Demecinden örnek aldıkları da bir gerçektir.)

Fransız İhtilali ile ortaya çıkan bu yeni siyasal düzen anlayışına Liberalizm veya Hürriyetçilik hareketi denmektedir. Fakat kişinin bu temel hak ve hürriyetlerinin gerek hükümdar, gerek insanlar tarafından kabul edilmesi yeterli değildi. Bu haklar ve hürriyetler bir anayasada açıkça belirtilmedikçe ve yine bu anayasada hükümdarın otorite ve yetkilerinin ne şekilde ve nasıl kullanılacağı belirlenmedikçe, kişinin siyasal varlığı yeterli bir teminata sahip olamazdı. Onun içindir ki, Liberalizm hareketinin en mühim unsuru anayasacılık olmuştur. Yani, Liberalizm anayasalı bir hürriyet düzeni kurma amacını gütmüştür. Bu anayasalı düzende hükümdar yine hükümdar olarak kalmaktadır. Lakin yetkilerinin sınırı ve kullanılma şekli bir

anayasa ile çizilecektir. Hemen ilave edelim ki, 19'uncu yüzyılın Liberalizm hareketleri içinde Cumhuriyetçilik eğilimi çok az görülmüştür. Fransa'da meydana gelen bu hürriyetçilik hareketini Avrupa'nın diğer mutlak hükümdarlarının hemen kabul etmesi beklenemezdi. Çünkü Fransa'da kralın otoritesini yıkan bu hareket, kendilerine de bulaşırsa, bunlar da otoritelerinden yoksun kalabilirler ve hatta tahtlarını kaybedebilirlerdi. Bu sebeple, Avusturya, Prusya, Rusya ve İngiltere gibi büyük devletlerle Avrupa'nın küçük krallıkları daha ilk günden itibaren Fransız İhtilaline cephe aldılar ve bu da İhtilal Fransa'sı ile bu devletler arasında, 1792'de başlayıp, 1815'e kadar devam edecek uzun savaşların patlamasına sebep oldu.

1792-1815 arasındaki Fransız İhtilali savaşları, Napolyon'un elinde, bütün Avrupayı Fransız hegemonyası altına sokmak isteyen bir kuvvet emperyalizmi niteliğini kazanmışsa da, şurası da bir gerçektir ki, bu savaşlar hürriyet kavramının bütün Avrupaya ve özellikle kitlelere yayılmasını da kolaylaştırmıştır. 1815'in Avrupasında, siyasal düzen konusundaki insan düşüncesi 25 yıl öncesinden artık çok farklıydı.

Bununla beraber, mutlak hükümdarlar bu önemli değişikliğin gerçek mahiyetini anlamamış görünüyorlar. 1815'de Napolyon Fransa'sını yenerek Fransa'yı ihtilalden önceki sınırları içine sokan Avrupanın büyük devletleri (İngiltere, Avusturya, Rusya ve Prusya) hürriyetçilik fikirlerini de yenilgiye uğrattıklarını sanmışlardır. Hatta küçük krallıklar bile aynı düşüncüyü taşıyorlardı. Mamafih Hürriyetçilik (Liberalizm) akımının tehlike ve korkusunu da içlerinde hissetmiş olmalı ki, 1815 Viyana Kongresi ile Avrupanın toprak ve sınır düzenlemelerini kendi politik çıkarlarına göre yaparlarken, aynı zamanda, bundan sonra patlak verebilecek herhangi bir hürriyetçilik hareketini beraberce bastırmak hususunda da anlaşmışlardır. Bu işbirliğinin rahatlığı içinde Avrupanın mutlakiyetçi hükümdarları 1815'den sonra toplumlarını yine eski düzen üzerinden idare etmeye başlamışlardır. Bu ise toplumların değişen düşüncesi ile tam bir çelişme haline gelmiş ve bunun neticesi olarak, Avrupa, 1818-1822, 1830 ve 1848'de olmak üzere, üç devre halinde bir dizi ayaklanma ve ihtilallere sahne olmuştur. Biraz aşağıda da değineceğimiz gibi, bu üç devreli ayaklanmalar sadece liberalist mahiyette değil, aynı zamanda Nasyonalist mahiyette de kendisini göstermiştir.

1818-22 devresinde Liberalist ayaklanmalara Almanya'da, İtalya'da ve İspanya'da rastlamaktayız. Bilhassa Alman üniversiteleri Hürriyetçi ayaklanmaların bir beşiği

haline gelmiştir. Fakat büyük devletlerin işbirliği bu ayaklanmaların kısa bir sürede bastırılmasını sağlamıştır.

1822'den sonra Avrupa kısa bir süre içinde bir sükunete girmişse de, 1830'da bir çok ülkelerde yeniden patlamalar meydana gelmiştir. Fransa'da basın hürriyetinin kısıtlanmasından patlak veren ve birkaç gün süre ile Paris'i kanlı çarpışmalara sahne yapan ayaklanma hemen başka ülkelere de sıçramıştır. Belçikalılar bir yandan Hollandalıların egemenliğinden kendilerini kurtarıırken, aynı zamanda, zamanın en ileri hürriyetçi anayasasını da kabul ediyorlardı.

1830'un hürriyetçi ayaklanmaları Almanya ve İtalya'nın küçük krallıklarının birçoğunda da görüldü. Bir çok Alman devletlerinde hürriyetçi anayasalar kabul edildi. İtalyan devletlerindeki hürriyetçi hareketleri ise Avusturya sert bir şekilde bastırdı. Ne olursa olsun, Liberalizm hareketi 1830 İhtilallerinde önemli bir adım atmıştır. 1848-49'larda Avrupa yeniden bir dizi ayaklanmalar içine girdi. Nasyonalist hareketin ağır bastığı bu ayaklanmalarda, liberalizm artık önemli bir zafer sağlamıştır. 19'uncu yüzyılın ortalarında anayasalı rejimler artık normal bir siyasal düzen haline gelmektedir. Fransa'da işçi hakları ve toplantı hürriyeti meselesinden doğan 1848 ayaklanması, Cumhuriyet rejiminin kurulması ile sonuçlanmış, lakin bu cumhuriyet ancak dört yıl kadar sürerek, 1852 de imparatorluğa dönüşmüştür. Fakat ne var ki, 1848'de cumhurbaşkanı seçilen Louis Napoleon (Napoleon Bonaparte'ın yeğeni), 1852'de imparatorluk rejimini ancak bir halk oylaması ve halkın tasvibi ile kurabilmiştir. Avusturya'da ise, 1848 ihtilali, 1815'den beri mutlakiyetçiliğin bayraktarlığını ve liberalizm düşmanlığının öncülüğünü yapan Başbakan Metternich'in ülkesinden kaçması ile neticelenmiştir. Mutlakiyetçi Prusya ise, 1850'de bir anayasa kabul etmek zorunda kalmıştır. Keza Hollanda, İsviçre ve Danimarka'da da gayet liberal anayasalar kabul edilmiştir. Hasılı, 19'uncu yüzyılın ortalarında artık hürriyetçilik ve anayasalı rejim, Avrupa ülkelerinin ciddi bir meselesi olmaktan çıkmıştır.

LİBERALİZM – ÖZET

Fransız İhtilali ile ortaya çıkan bu yeni siyasal düzen anlayışına Liberalizm veya Hürriyetçilik hareketi denmektedir. Fakat kişinin bu temel hak ve hürriyetlerinin gerek hükümdar, gerek insanlar tarafından kabul edilmesi yeterli değildi. Bu haklar ve hürriyetler bir anayasada açıkça belirtilmedikçe ve yine bu anayasada hükümdarın

otorite ve yetkilerinin ne şekilde ve nasıl kullanılacağı belirlenmedikçe, kişinin siyasal varlığı yeterli bir teminata sahip olamazdı. Onun içindir ki, Liberalizm hareketinin en mühim unsuru anayasacılık olmuştur. Yani, Liberalizm anayasalı bir hürriyet düzeni kurma amacını gütmüştür. Bu anayasalı düzende hükümdar yine hükümdar olarak kalmaktadır. Lakin yetkilerinin sınırı ve kullanılma şekli bir anayasa ile çizilecektir. Hemen ilave edelim ki, 19'uncu yüzyılın Liberalizm hareketleri içinde Cumhuriyetçilik eğilimi çok az görülmüştür.

Ne olursa olsun, Liberalizm hareketi 1830 İhtilallerinde önemli bir adım atmıştır. 19'uncu yüzyılın ortalarında anayasalı rejimler artık normal bir siyasal düzen haline gelmektedir.

LİBERALİZM – OTOMATİK ÖZET

Bilindiği gibi, Orta Çağlarda Rönesans hareketi sanat alanında ve Reformasyon hareketi de din alanında insan düşüncesine bir serbesti, bir çeşit hürriyet getirme amacını gütmüştür. Reformasyon hareketi ise, o zamana kadar egemen din olarak katolikliğin sert katı ve hoşgörüsüz din kalıbını kırarak, Tanrı ile insan arasındaki münasebetlere bir hürriyet getirmek amacını gütmüştür. Bunun sonucu olarak; Protestanlık denen yeni bir din şekli ve onun da çeşitli mezhepleri ortaya çıkmıştır. Protestanlık ve onun çeşitli mezhepleri, Tanrı-İnsan münasebetlerine, katoliklikten çok farklı yeni yeni şekiller getirmiştir. Fakat şu var ki, ne sanat alanındaki hürriyet gelişimi ve ne de din alanındaki hürriyet gelişimi, insanın toplum içindeki siyasal yaşayışına herhangi bir serbestlik veya hürriyet getirmekten çok uzak kalmıştır. Görülüyor ki bu demeç, insanın insan olmak dolayısıyla, daha doğduğu andan itibaren, bir takım temel hak ve hürriyetlere sahip bulunduğunu ortaya koyuyor ve hükümdarın otoritesini de bu temel hak ve hürriyetlerle sınırlıyordu. Fransız İhtilali ile ortaya çıkan bu yeni siyasal düzen anlayışına Liberalizm veya Hürriyetçilik hareketi denmektedir. Onun içindir ki, Liberalizm hareketinin en mühim unsuru anayasacılık olmuştur. Hemen ilave edelim ki, 19'uncu yüzyılın Liberalizm hareketleri içinde Cumhuriyetçilik eğilimi çok az görülmüştür. Bu sebeple, Avusturya, Prusya, Rusya ve İngiltere gibi büyük devletlerle Avrupa'nın küçük krallıkları daha ilk günden itibaren Fransız İhtilaline cephe aldılar ve bu da İhtilal Fransa'sı ile bu devletler arasında, 1792'de başlayıp, 1815'e kadar devam edecek uzun savaşların patlamasına sebep oldu. 1792-1815 arasındaki Fransız İhtilali savaşları, Napolyon'un elinde, bütün Avrupayı Fransız hegemonyası altına sokmak

isteyen bir kuvvet emperyalizmi niteliğini kazanmışsa da, şurası da bir gerçektir ki, bu savaşlar hürriyet kavramının bütün Avrupaya ve özellikle kitlelere yayılmasını da kolaylaştırmıştır.1830'un hürriyetçi ayaklanmaları Almanya ve İtalya'nın küçük krallıklarının birçoğunda da görüldü. Ne olursa olsun, Liberalizm hareketi 1830 İhtilallerinde önemli bir adım atmıştır. Nasyonalist hareketin ağır bastığı bu ayaklanmalarda, liberalizm artık önemli bir zafer sağlamıştır. Avusturya'da ise, 1848 ihtilali, 1815'den beri mutlakiyetçiliğin bayraktarlığını ve liberalizm düşmanlığının öncülüğünü yapan Başbakan Metternich'in ülkesinden kaçması ile neticelenmiştir. Hasılı, 19'uncu yüzyılın ortalarında artık hürriyetçilik ve anayasalı rejim, Avrupa ülkelerinin ciddi bir meselesi olmaktan çıkmıştır.

Belge III

NASYONALİZM

Nasyonalizm veya milliyetçilik akımının esası milli bağımsızlıktır. Başka devletlerin hegemonyası altında yaşayan milletlerin milli bağımsızlıklarını kazanmaları ve kendi bağımsız devletlerini kurmaları hareketidir. Bu hareket de kaynağını Fransız İhtilalinden almaktadır. Bir bakıma kişi hürriyeti kavramının milletlere de tatbididir. Nasıl bir insan, insan olması dolayısıyla bir takım temel hak ve hürriyetlere sahip bulunuyorsa, bir millet'de, bir bütün olarak, hürriyetine yani bağımsızlığına sahip olma hakkına sahiptir.

Öte yandan, milliyetçilik hareketinin ortaya çıkmasında, İhtilal Fransa'sı müessir bir rol oynamıştır. Bilhassa İhtilal Fransa'sının kaderine 1797'den itibaren hakim olan Napolyon Bonapart, Avrupa'nın büyük devletleri ile savaşırken ve bu devletlerin topraklarına girerken, bu topraklardaki milletleri bağlı oldukları devletlere karşı ayaklandırmış ve Fransa'nın bu milletlere hürriyet getirdiğini söylemiştir. Bu kışkırtmaların bu milletler üzerinde tesirsiz kaldığını söylemeye imkan yoktur.

Fakat 1815 Viyana Kongresinde, büyük devletler Avrupa haritasını kendi çıkarlarına göre düzenlerken, nasyonalizm bakımından, liberalizm konusunda yaptıkları aynı hatayı işlemişlerdir. Milletler ya parçalanmış ve bu parçalar başka devletlerin sınırları içine sokulmuş veya çizilen sınırlar içinde çeşitli milletler bulunmuştur. Mesela, Polonya toprakları bir kere daha Avusturya, Rusya ve Prusya arasında paylaşılmış ve Polonyalıların büyük bir kısmı Rusya'nın hegemonyası altına girmiştir. Halbuki Napolyon 1807 yılında, Prusya ile Rusya arasında bir tampon olmak üzere,

Varşova Büyük Dükalığı adı ile bağımsız bir Polonya devleti kurmuştu. Polonyalılar bağımsızlığın tadını almışlardı. 1815 Viyana Kongresinde ise, bu bağımsız Polonya bir kere daha ortadan kaldırılarak toprakları üç devlet arasında bölüşülmüştür. Öte yandan, Avusturya İmparatorluğunun sınırları içinde, başta Macarlar olmak üzere, Polonyalılar, Çekler, Hırvatlar ve Romenler gibi birçok millet bulunmaktaydı. Hatta Osmanlı İmparatorluğunun Avrupa toprakları için de aynı şey söylenebilir.

Bu topraklar içinde pek çok yabancı milletler yaşamaktaydı. Nasyonalizm hareketi bunlara da tesir etmekten geri kalmamış, 1829'da Yunanistan, 1878'de Sırbistan, Karadağ ve Romanya, 1908'de Bulgaristan ve 1913 yılında da Arnavutluk, birer bağımsız devlet olarak Osmanlı İmparatorluğundan kopmuştur. Bu sebepten ötürü 1830-1848 ihtilallerinde milli bağımsızlık teşebbüslerine de rastlamaktayız. Mesela, 1830 yılında Rusya sınırları içinde yaşayan Polonyalılar bağımsızlıklarını almak için ayaklanmışlar ve tam bir yıl süre ile Rus ordularına karşı savaşmışlardır. Fakat kanlı bir yenilgiye uğramışlardır. 1848-49 da ise Macarlar bağımsızlıklarını almak için Avusturya'ya karşı savaşmışlardır. Avusturya Macarlarla başa çıkamayınca Rusya'dan yardım istemiş ve Macaristan'a giren 150000 kişilik bir Rus ordusu 1849 yazında Macar milli bağımsızlık hareketini çok kanlı bir şekilde bastırmıştır. 1815'den sonra milliyetçilik akımının en önemli meselesi Alman ve İtalyan milli birlikleri olmuştur. Bu mesele de kaynağını, 1815 Viyana Kongresi kararlarından almaktaydı ve burada da baş rolü Avusturya oynuyordu. Avusturya sınırları içinde çeşitli milletler bulunduğu için, Fransız İhtilalinin milliyetçilik konusundaki tesirlerini gözönünde tutan bu devlet, milliyetçi akımlara karşı da kuvvetle cephe almıştır. Çünkü, Avusturya sınırları içindeki milletler bağımsızlıklarını alacak olurlarsa, Avusturya İmparatorluğunun dağılması işten bile değildi. Avusturya bu noktayı çok iyi görmüştür. Nitekim 1914-18 Birinci Dünya Savaşının sonunda Avusturya yenilip de bütün bu milletler bağımsızlıklarını alınca, geriye bugünkü küçük Avusturya devleti kalmıştır. Aynı şey Osmanlı İmparatorluğu için de olmuştur. Birinci Dünya Savaşının hemen öncesinde İmparatorluk sınırlarının Meriç nehrine kadar gerilemesinde, Balkan ülkelerinin bağımsızlıklarını kazanmaları baş rolü oynamıştır. Alman ve İtalyan milli birliklerinin kurulması ihtimali ise, Avusturya için en fazla korkutucu olmuştur. Bu iki milli birliğin kurulması ve birer devlet olarak ortaya çıkmasının Avusturya sınırları içindeki çeşitli milletlere yapacağı tesir bir yana, Avusturya'nın, biri batısında, diğeri de güneyindeki iki güçlü

devletin kurulması, bu devletin Avrupa'daki durumunu zayıflatırdı. Avusturya böyle bir ihtimalin gerçekleşmesini önlemek için, 1815 Viyana Kongresi kararlarıyla, hem Almanya'yı ve hemde İtalya'yı dağınık tutmaya muvaffak olmuştu. Kutsal Germen İmparatorluğunun 360 devleti, Viyana Kongresinde ancak 36'ya indirildi. Yani Almanya'da 36 ayrı devlet bulunuyordu. İtalya'da da aynı durum vardı. İtalya denen topraklar üzerinde bir çok krallık bulunduğu gibi, Avusturya, bazı İtalyan krallıklarının başına Avusturya İmparator ailesinden prensleri getirdi. Bu suretle bazı İtalyan Krallıkları vasıtasıyla İtalya üzerinde doğrudan doğruya kontrol tesis etmiş olmaktaydı. Avusturya, Almanya için de aynı şeyi yaptı. Almanya'daki 36 devlet Germen Konfederasyonunu teşkil ediyordu ve bu konfederasyonun başına da Avusturya Prensleri getirilmişti. Yani, Avusturya, Alman devletleri üzerinde de doğrudan doğruya kontrole sahip bulunuyordu.

Fakat her iki ülkede de Avusturyanın önemli bir problemi vardı. Almanya'da Prusya ve İtalya'da Piyemonte devletleri, biri Alman, diğeri de İtalyan milli birliğin kurulmasını istiyorlardı. Onun içindir ki, Avusturya 1815 den itibaren her iki devletle de sinsi bir mücadelenin içine girmek zorunda kaldı.

1848 Martında Viyana sokaklarında halkın hürriyet için de Macarların da bağımsızlık için ayaklanmaları Avusturya'yı güç duruma sokunca Almanya'daki milliyetçiler Alman milli birliğini kurmak için harekete geçtiler ve Prusya'yı da lider olarak seçtiler. Hatta bir Alman İmparatorluğu Anayasası yapılarak Prusya Kralı Alman İmparatoru ilan edilmek istendi. Lakin Avusturya'nın Prusya'ya yönelttiği tehdit o kadar şiddetli oldu ki, Prusya gerilemek zorunda kaldı. O günün şartları içinde Prusya Avusturya ile bir savaşı göze alamadı. Prusya geri çekilince Alman milli birliği için yapılan bu teşebbüs de neticesiz kaldı.

İtalya'da ise, Piyemonte harekete geçti. Fakat Piyemonte'nin teşebbüsü de sonuçsuz kaldı. Piyemonte 1848 ve 1849 yıllarında iki defa Avusturya ile savaşa girişti, fakat her ikisinde de yenildi. Bu yenilgiler diğeri İtalyan devletlerinin de cesaretini kırdı. Fakat bu iki teşebbüs Piyemonte'ye bir şeyi öğretmişti; İtalyan milli birliği ancak Avusturyanın savaş meydanında yenilmesiyle gerçekleştirilebilir ve böyle bir savaş da küçük Piyemonte tek başına yapamazdı. Avusturya'yı yenmek için Piyemontenin bir büyük Avrupa devletini yanına alması gereklidir. Gerçekten, 1854-56 Kırım savaşına Fransa ve İngiltere'nin yanında katılarak bu iki büyük devletin sempatisini kazanan Piyemonte, milliyetçi hareketleri destekleyen 3'üncü Napolyon (1848

seçimlerinde cumhurbaşkanı seçilen Louis Napolyon, 1852'de İmparator olunca 3'üncü Napolyon adını almıştır.) Fransa'sı ile 1858 yılında bir ittifak yapmaya muvaffak oldu. Fransa'yı ittifakına alan Piyemonte 1859 yılında Avusturya'ya savaş açtı ve Avusturya'yı, Fransa ile birlikte, iki defa savaş meydanında yenilgiye uğrattı. Piyemonte'nin bu zaferi üzerine, diğer İtalyan devletleri Piyemonte'ye katılarak İtalyan Milli birliğini kurdular. 1861 Şubatında Torino'da ilk İtalyan parlamentosu açıldı ve İtalya Krallığı ilan edildi.

İtalyan milli birliğinin kuruluşunu Alman birliğinin kuruluşu takip etmiştir. Bu birliğin kuruluşu da Prusya'nın ve onun başbakanı Bismarck'ın eseri olmuştur. Alman birliği üç safhada gerçekleşmiş olup, bunların herbiri bir savaştır. 1864 Prusya-Danimarka savaşı ile Prusya, Danimarka'nın elindeki bazı Alman topraklarını geri alıp Germen Konfederasyonuna katmıştır. 1866 Prusya-Avusturya savaşı ise, Avusya'nın Prusya karşısında yenilgisi üzerine, bu devleti Germen Konfederasyonunun dışında bırakmak suretiyle, Almanya üzerindeki Avusturya kontrolünü sona erdirmiştir. Fakat buna rağmen Bismarck için Alman milli birliğini kurmak hemen mümkün olamadı. Çünkü, Fransa'nın nüfuzu altında bulunan Katolik Güney Alman Devletleri birliğe yanaşmadıkları gibi, şimdi kuzeyinde kuvvetli bir Almanya'nın ortaya çıkmasından endişe etmeye başlayan Fransa'nın Prusya'ya karşı durumu sertleşmiştir. Alman birliği yolundaki Fransa engelini bertaraf etmek, Bismarck için, ancak, Fransa'yı 1870-71 savaşında ağır bir yenilgiye uğratmakla mümkün oldu. Bismarck'ın dediği gibi, Alman milletinin milli bütünlüğü ancak "kan ve demirle" gerçekleştirilebilmiştir.

İtalya ve Almanya'nın bağımsız devletler olarak Avrupa'daki milletlerarası münasebetlere girmesiyle, Nasyonalizm akımı çok büyük bir adım atmış olmaktadır. Fakat milliyetler meselesi de tamamen çözülmüş değildi. Milliyetler meselesinin milletlerarası münasebetleri bulandırması, İkinci Dünya Savaşından sonra bile devam edecektir. Yalnız şu varki, bilhassa Alman milli birliğinin kuruluşu ve Alman İmparatorluğunun Avrupa'daki kuvvet münasebetleri içinde 1871'den itibaren birdenbire sivrilmesi, kıtada milletlerarası münasebetlere yepyeni bir yapı ve gelişme seyri vermiş ve bundan sonra da Almanya etrafında şekillenmeye başlayan diplomatik gelişmeler, fikir akımlarının etkisini geri plana itmeye başlamıştır. 1871'in Avrupa'sı 1815'in Avrupa'sından artık çok farklıdır.

NASYONALİZM – ÖZET

Nasyonalizm veya milliyetçilik akımının esası milli bağımsızlıktır. Başka devletlerin hegemonyası altında yaşayan milletlerin milli

bağımsızlıklarını kazanmaları ve kendi bağımsız devletlerini kurmaları

hareketidir. Bu hareket de kaynağını Fransız İhtilalinden almaktadır. Bir bakıma kişi hürriyeti kavramının milletlere de tatbikidir. Nasıl bir insan, insan olması dolayısıyla bir takım temel hak ve hürriyetlere sahip bulunuyorsa, bir millet'de, bir bütün olarak, hürriyetine yani bağımsızlığına sahip olma hakkına sahiptir. Öte yandan, milliyetçilik hareketinin ortaya çıkmasında, İhtilal Fransa'sı müessir bir rol oynamıştır. Bilhassa İhtilal Fransa'sının kaderine 1797'den itibaren hakim olan Napolyon Bonapart, Avrupa'nın büyük devletleri ile savaşırken ve bu devletlerin topraklarına girerken, bu topraklardaki milletleri bağlı oldukları devletlere karşı ayaklandırmış ve Fransa'nın bu milletlere hürriyet getirdiğini söylemiştir. Bu kışkırtmaların bu milletler üzerinde tesirsiz kaldığını söylemeye imkan yoktur.

Fakat 1815 Viyana Kongresinde, büyük devletler Avrupa haritasını kendi çıkarlarına göre düzenlerken, nasyonalizm bakımından, liberalizm konusunda yaptıkları aynı hatayı işlemişlerdir. Milletler ya parçalanmış ve bu parçalar başka devletlerin sınırları içine sokulmuş veya çizilen sınırlar içinde çeşitli milletler bulunmuştur.

İtalya ve Almanya'nın bağımsız devletler olarak Avrupa'daki milletlerarası münasebetlere girmesiyle, Nasyonalizm akımı çok büyük bir adım atmış olmaktadır. Fakat milliyetler meselesi de tamamen çözülmüş değildi. Milliyetler meselesinin milletlerarası münasebetleri bulandırması, İkinci Dünya Savaşından sonra bile devam edecektir. Yalnız şu varki, bilhassa Alman milli birliğinin kuruluşu ve Alman İmparatorluğunun Avrupa'daki kuvvet münasebetleri içinde 1871'den itibaren birdenbire sivrilmesi, kıtada milletlerarası münasebetlere yepyeni bir yapı ve gelişme seyri vermiş ve bundan sonra da Almanya etrafında şekillenmeye başlayan diplomatik gelişmeler, fikir akımlarının etkisini geri plana itmeye başlamıştır. 1871'in Avrupa'sı 1815'in Avrupa'sından artık çok farklıdır.

NASYONALİZM – OTOMATİK ÖZET

Nasyonalizm veya milliyetçilik akımının esası milli bağımsızlıktır. Başka devletlerin hegemonyası altında yaşayan milletlerin milli bağımsızlıklarını kazanmaları ve kendi bağımsız devletlerini kurmaları hareketidir. Nasıl bir insan, insan olması dolayısıyla

bir takım temel hak ve hürriyetlere sahip bulunuyorsa, bir millet'de, bir bütün olarak, hürriyetine yani bağımsızlığına sahip olma hakkına sahiptir.Öte yandan, milliyetçilik hareketinin ortaya çıkmasında, İhtilal Fransa'sı müessir bir rol oynamıştır.Fakat 1815 Viyana Kongresinde, büyük devletler Avrupa haritasını kendi çıkarlarına göre düzenlerken, nasyonalizm bakımından, liberalizm konusunda yaptıkları aynı hatayı işlemişlerdir. Hatta Osmanlı İmparatorluğunun Avrupa toprakları için de aynı şey söylenebilir. Nasyonalizm hareketi bunlara da tesir etmekten geri kalmamış, 1829'da Yunanistan, 1878'de Sırbistan, Karadağ ve Romanya, 1908'de Bulgaristan ve 1913 yılında da Arnavutluk, birer bağımsız devlet olarak Osmanlı İmparatorluğundan kopmuştur. Alman ve İtalyan milli birliklerinin kurulması ihtimali ise, Avusturya için en fazla korkutucu olmuştur. Bu iki milli birliğin kurulması ve birer devlet olarak ortaya çıkmasının Avusturya sınırları içindeki çeşitli milletlere yapacağı tesir bir yana, Avusturya'nın, biri batısında, diğeri de güneyindeki iki güçlü devletin kurulması, bu devletin Avrupa'daki durumunu zayıflatırdı. İtalya denen topraklar üzerinde bir çok krallık bulunduğu gibi, Avusturya, bazı İtalyan krallıklarının başına Avusturya İmparator ailesinden prensleri getirdi.Fakat her iki ülkede de Avusturyanın önemli bir problemi vardı. Almanya'da Prusya ve İtalya'da Piyemonte devletleri, biri Alman, diğeri de İtalyan milli birliğinin kurulmasını istiyorlardı. Onun içindir ki, Avusturya 1815 den itibaren her iki devletle de sinsi bir mücadelenin içine girmek zorunda kaldı.1848 Martında Viyana sokaklarında halkın hürriyet için de Macarların da bağımsızlık için ayaklanmaları Avusturya'yı güç duruma sokunca Almanya'daki milliyetçiler Alman milli birliğini kurmak için harekete geçtiler ve Prusya'yı da lider olarak seçtiler. Prusya geri çekilince Alman milli birliği için yapılan bu teşebbüs de neticesiz kaldı. 1861 Şubatında Torino'da ilk İtalyan parlamentosu açıldı ve İtalya Krallığı ilan edildi.İtalyan milli birliğinin kuruluşunu Alman birliğinin kuruluşu takip etmiştir.İtalya ve Almanya'nın bağımsız devletler olarak Avrupa'daki milletlerarası münasebetlere girmesiyle, Nasyonalizm akımı çok büyük bir adım atmış olmaktaydı. 1871'in Avrupa'sı 1815'in Avrupa'sından artık çok farklıdır.

ÖZGEÇMİŞ

1983 yılında İstanbul'da doğan Mesut TÜLEK, 2000 yılında Pertevniyal Lisesi, 2004 yılında da İstanbul Teknik Üniversitesi Elektrik-Elektronik Fakültesi Bilgisayar Mühendisliği bölümünden mezun olmuştur. 2004 yılından beri, Bizitek Yazılım ve İnternet Teknolojileri'nde Yazılım Mühendisi olarak çalışmaktadır.