

# ÖZGEÇMİŞLERDEN VE İŞ İLANLARINDAN DOĞAL DİL İŞLEME YÖNTEMLERİ İLE BİLGİ ÇIKARIMI

Gülşen Bal

Kariyer.net

[gulsenb@kariyer.net](mailto:gulsenb@kariyer.net)

Cem Gültunca

Kariyer.net

[cgultunca@kariyer.net](mailto:cgultunca@kariyer.net)

Aşkın Karakaş

Kariyer.net

[askin@kariyer.net](mailto:askin@kariyer.net)

Tunga Güngör

Boğaziçi Üniversitesi

[gungort@boun.edu.tr](mailto:gungort@boun.edu.tr)

## ÖZET

İnternet günlük hayatımızın her alanında kullandığımız iletişim ağı olmasıyla birlikte, sadece alışveriş yaparken veya banka işlemlerimizi gerçekleştirirken değil aynı zamanda iş ararken de kullandığımız bir iletişim yolu haline gelmiştir. Aynı şekilde işverenler/firmalar da İnsan Kaynakları sitelerini kullanarak hızlı ve kolay bir şekilde iş ilanlarını yayınlayıp birçok adaya ulaşabilmektedirler. İnsan Kaynakları siteleri içerisinde kullanıcılar özgeçmiş oluşturarak, yetkinliklerini, kişisel bilgilerini, iş tecrübelerini, eğitim bilgilerini, projelerini, sertifikalarını, referanslarını vb. bilgilerini paylaşırlar. Ancak iş arayanların kendi iş tecrübelerine, yetkinliklerine, eğitim bilgilerine vb. özelliklerine tam anlamıyla uygun olan iş ilanlarını bulmaları zor bir işlemdir. Aynı zamanda işverenler için de, iş ilanlarına uygun özellikteki adayları bulmak yine zaman alan ve zor bir işlemdir. Bu çalışmada iş ilanlarından ve özgeçmişlerden doğal dil işleme yöntemleri ile bilgi çıkarımı gerçekleştirilip, elde edinilen bilgilerin eşleştirilmesi yapılarak en uygun adayın ve en uygun iş ilanının adaya/işverene sunulmasına yönelik bir sistemin geliştirilmesi amaçlanmıştır.

## Anahtar Kelimeler

**Bilgi Çıkarımı, Anlamsal Web, Doğal Dil İşleme, İş İlanı, Özgeçmiş, Ontoloji, Tanıma-Tabanlı Eşleştirme**

## SUMMARY

Besides internet is a network that we use in our daily lives, it has become a connection tool not only for shopping and banking operations but also for job seeking. Likewise, employers/companies use Human Resources Websites to publish job advertisements and reach job applicants expeditiously and easily. Human resources websites allow users to represent themselves with a resume which is a brief written formal structure that has basic parts for defining job seeker's information such as qualifications, personal information, work experience, education information,

projects, certificates, references and so on. However, it is a difficult process for a job applicant to find the best matched jobs with his/her work experiences, qualifications, education experiences, projects and so on. At the same time, it is also a difficult and time-consuming task for a company to find the best-qualified candidates among applicants. This project aims to present a system that extracts information from job advertisements and resumes with natural language processing (NLP) methods and, using the extracted information, perform matching to find the most suitable job advertisements for job seekers and to find the most suitable job applicants for employers.

## Keywords

**Information Extraction, Semantic Web, Natural Language Processing, Job Advertisement, Resume, Ontology, Recognition-Based Matching**

## GİRİŞ

İnternetin günümüzde yaygın bir biçimde kullanılmasıyla birlikte internet aracılığı ile çevrimiçi insan kaynakları siteleri üzerinden iş aramak veya işveren için aday aramak yaygın bir hale gelmiştir. Çevrimiçi insan kaynakları siteleri kullanıcılara özgeçmiş oluşturmalarına imkan verip, işverenler tarafından yayınlanan iş ilanlarına başvurma olanağı sağlar. Kullanıcılar da özgeçmişlerine eğitim bilgileri, kişisel bilgiler, genel bilgiler, iş tecrübeleri, yetkinlikler, projeler, sertifikalar, vb. bilgilerini ekleyerek kendilerine uygun gördükleri iş ilanlarına başvuruda bulunurlar. Fakat işveren için de iş ilanına uygun olan adayın bulunması ve yüzlerce/binlerce başvuru arasından aranan iş tecrübesine ve yetkinliklere uygun adayların elde edilebilmesi çok fazla zaman gerektiren ve zahmetli bir iş haline gelmektedir. Bu ihtiyaca bir çözüm bulmak amacıyla Kariyer.net<sup>1</sup> ve TÜBİTAK desteğiyle yürütülmekte olan bu proje, özgeçmişlerde ve iş ilanlarında bulunan bilgilerin doğal dil işleme yöntemleri ile elde

<sup>1</sup> <http://www.kariyer.net/>

edilip, eşleştirme yöntemleri ile adayların özelliklerine uygun iş ilanlarının ve işverenlerin iş ilanlarına uygun özellikteki adayların bulunmasına olanak sağlayacaktır. Bu çalışmada, projenin ilk kısmını oluşturan özgeçmiş ve iş ilanlarından bilgi çıkarım işlemi detaylı bir şekilde anlatılacaktır.

Bilgi çıkarımı(*information extraction*) uzun yıllar boyunca araştırmacıların yoğun bir şekilde üzerinde çalıştığı bir alan olmuştur. Bilgi çıkarımı, yapısal olmayan veya yarı-yapısal formatlara ait yapılardan bilgiyi elde etme işlemine denir[1], [2], [3], [4], [5], [6]. Bilgi çıkarım yöntemleri olarak son yıllarda kural bazlı bilgi çıkarımı, istatistiksel bilgi çıkarımı ve Sarıcı tümevarım(*Wrapper Induction*) yöntemleri kullanılmıştır. Kural bazlı bilgi çıkarım yönteminde oluşturulan kurallar insanlar tarafından elle girilen veya örneklerden öğrenme yöntemleri ile oluşturulmuş kurallardır.

İstatistiksel öğrenme metodlarında gizli Markov modeller (*hidden Markov models*) [7], [8], [9], [10] ve hiper düzlem ayırıcıları (*hyperplane seperators*) yöntemleri kullanılır. Gizli Markov Modelleri doğal dil işleme alanı da dahil bir çok alanda kullanılmaktadır. Hiper düzlem ayırıcılar ise öğrenilecek kavram için, özellik uzayında pozitif örnekleri negatiflerden ayıran hiper düzlemleri öğrenirler. Hiper düzlem ayırıcılarının büyük ölçekteki özelliklerle çalışabiliyor olabilmesi, Hiper düzlem ayırıcılarını özellikle doğal dil işleme ve bilgi çıkarımı konularında uygun hale getirmektedir.

Sarıcı tümevarım metodunda ise amaç; belirli bir kaynaktan bulunan içeriğin elde edilmesi ve öntanımlı temsil edilmiş olan uygun içeriğin teslim edilmesini sağlamaktır[11]. Genellikle web ortamında uygulanan Sarıcı tümevarım; HTML ve XML formatında kaydedilmiş bilgilerin veri yapılarına uygun hale çevrilip ileride kullanılması için dolaylı yoldan saklanmasını sağlayan bir işlemci olarak tanımlanabilir.

Bu çalışmada, özgeçmişlerden bilgi çıkarım işlemi semantik web kullanılarak ve kural tabanlı bilgi çıkarım yöntemi uygulanarak gerçekleştirilmiştir[12]. Yapısal olmayan özgeçmişler içerisinde genel bilgiler, kişisel bilgiler, eğitim bilgileri, yetkinlikler, iş tecrübesi bilgisi, yabancı dil bilgisi, sertifikalar, projeler, referanslar vb. bilgilerin çıkarımı sağlanmıştır. Bu bilgi çıkarım işleminde önceden tanımlanmış olan kurallar çerçevesinde oluşturduğumuz bilgi tabanı kullanılarak bilgi çıkarım işlemi uygulanmıştır. İş ilanları içerisinde ise eğitim bilgisi, iş tecrübesi bilgisi, yetkinlikler, sertifikalar ve yabancı dil bilgisi çıkarımı gerçekleştirilmiştir.

## METODOLOJİ

### ÖZGEÇMİŞ ANALİZ

Çalışmanın bu kısmında türkçe yazılmış yapısal olmayan özgeçmişlerin metin kısmında bulunan kişisel bilgiler, genel bilgiler, eğitim bilgileri (lise, üniversite adı, fakülte ve bölümü), yetkinlikler, sertifikalar, projeler, yabancı dil bilgisi, iş tecrübesi bilgisi (şirket adı, pozisyon, sektör ve iş açıklaması), referanslar ve ek bilgiler bilgilerinin çıkarımı gerçekleştirilmiştir. Özgeçmiş içerisinde bilgi çıkarım işlemi, Word dökümanlarının iç yapısının anlaşılması zor karakterler içermesinden ve etiket yapısı olmamasından dolayı, özgeçmişler anlaşılması ve incelenmesi daha iyi olan ve belirli bir etiket yapısına sahip olan HTML yapısına çevrilir.

HTML yapısı içerisindeki etiketler sayesinde metinlerin anlaşılabilirliği ve ayrıştırılması sağlanmıştır. HTML yapısına çevrilmiş düz metin içerisinde paragraf ve cümleler oluşturularak, tanımlanan kurallar ile bilgi çıkarımı gerçekleştirilir. Çıkarılan bilgiler Kariyer.net ve genel özgeçmiş yapısına uygun olarak oluşturulan Özgeçmiş Ontolojisine kayıt edilir. Özgeçmiş Ontolojisindeki bilgileri işlemek için OWL API<sup>2</sup> (Ontoloji derleyici) kullanılmıştır. Çıkarım yapılan bilgilerden yeni kavramlar elde etmek ve çıkarım yapabilmek için Semantik Web Kural Dili (SWRL)<sup>3</sup> kullanılarak kurallar yazılmıştır. Yapısal olmayan özgeçmiş içerisindeki bilgilerin anlaşılabilirliği için, bilgi tabanındaki bilgilerle karşılaştırma işlemi, örüntü eşleşmesi ve özgeçmiş içerisindeki başlıklardan yararlanılmıştır.

Özgeçmiş içerisindeki anlaşılması en zor olan bölüm iş tecrübesi açıklaması bölümüdür. Serbest metin formatında yazılmış olan iş açıklaması bölümünde, iş tecrübesi açıklamasının işlenmesi ve anlamlandırılabilirliği sadece bilgi tabanlı bir sistemle mümkün değildir. Bu nedenle metin ilk olarak morfolojik analiz (*morphological analysis*) işleminden geçirilmiştir. Morfolojik analiz işlemi, bir kelimenin yapısını (kelimenin kökü ve aldığı ekler) ortaya çıkarmak amacıyla uygulanır. Bu işlem, aynı kökten türeyen kelimelerin ilişkili kelimeler olduğunun anlaşılmasına yöneliktir. Türkçe gibi eklemeli dillerde, genellikle bir kelimeyi yazıldığı gibi almak yerine kelimenin kökünü tespit ederek kökü (ve gerekiyorsa ekleri) kullanmak daha başarılı sonuçlar vermektedir. Morfolojik analiz işleminden sonra ise morfolojik muğlaklık giderme (*morphological disambiguation*) işlemi

<sup>2</sup> <http://owlapi.sourceforge.net/>

<sup>3</sup> <http://www.w3.org/Submission/SWRL/>

gerçekleştirilmiştir. Morfolojik muğlaklık giderme işlemi ise, birden fazla anlamı bulunan bir kelimenin doğru anlamının bulunması amacıyla uygulanır. Türkçe kelimelerin önemli bir bölümü birden çok anlama sahiptir. Bir dokümanda yer alan bir kelimenin, kullanıldığı bağlama (*context*) bakılarak doğru anlamının tespit edilmesi gerekmektedir. Bu çalışmamızda Boğaziçi Üniversitesi tarafından geliştirilen morfolojik analiz ve muğlaklık giderici programları kullanılmıştır[13]. Örnek bir iş tecrübesi açıklaması Tablo 1’de ve morfoloji işlemleri sonucu elde edilen çıktı Tablo 2’de gösterilmiştir.

*Cari hesap hareketlerinin kontrolü*  
*Günlük gelen tahsilatların vadelerinin kontrol edilmesi*  
*Bankalar ve kredi kartı hesaplarının kontrolü*  
*Fiyat farkı ve vade farkı tutarlarının hesaplanması*  
*Satış elemanlarının performansların çıkarılması*  
*Avukatlık müşterinin takibi*  
*Dönemsel olarak alıcı ve satıcı hesapların mutabakatının yapılması*  
*Yeni cari muhasebe stok kartlarının açılması*

**Tablo 1. İş Tecrübesi Açıklaması**

**Cari cari[Adj]**  
 hesap **hesap[Noun]+[A3sg]+[Pnon]+[Nom]**  
 hareketlerinin  
**hareket[Noun]+lAr[A3pl]+SH[P3sg]+NHn[Gen]**  
 hareket[Noun]+[A3sg]+lArH[P3pl]+NHn[Gen]  
 hareket[Noun]+lAr[A3pl]+SH[P3pl]+NHn[Gen]  
 hareket[Noun]+lAr[A3pl]+Hn[P2sg]+NHn[Gen]  
 kontrolü **kontrol[Noun]+[A3sg]+SH[P3sg]+[Nom]**  
 kontrol[Noun]+[A3sg]+[Pnon]+YH[Acc]  
 Günlük **günlük[Adj]**  
 Gün[Noun]+[Prop]+[A3sg]+[Pnon]+[Nom]-  
 lHk[Adj+FitFor]  
 günlük[Noun]+[A3sg]+[Pnon]+[Nom]  
 gün[Noun]+[A3sg]+[Pnon]+[Nom]-  
 lHk[Noun+Ness]+[A3sg]+[Pnon]+[Nom]  
 gün[Noun]+[A3sg]+[Pnon]+[Nom]-  
 lHk[Adj+FitFor]  
 gelen **gel[Verb]+[Pos]-YAn[Adj+PresPart]**  
 gelen[Adj] gele[Noun]+[A3sg]+Hn[P2sg]+[Nom]  
 tahsilatların  
**tahsilat[Noun]+lAr[A3pl]+[Pnon]+NHn[Gen]**  
 tahsilat[Noun]+lAr[A3pl]+Hn[P2sg]+[Nom]  
 vadelerinin  
**vade[Noun]+lAr[A3pl]+SH[P3sg]+NHn[Gen]**  
 vade[Noun]+[A3sg]+lArH[P3pl]+NHn[Gen]  
 vade[Noun]+lAr[A3pl]+SH[P3pl]+NHn[Gen]  
 vade[Noun]+lAr[A3pl]+Hn[P2sg]+NHn[Gen]

kontrol **kontrol[Noun]+[A3sg]+[Pnon]+[Nom]**  
 edilmesi **et[Verb]-HI[Verb+Pass]+[Pos]-  
 mA[Noun+Inf2]+[A3sg]+SH[P3sg]+[Nom]**  
 Bankalar **banka[Noun]+lAr[A3pl]+[Pnon]+[Nom]**  
 ve **ve[Conj]**  
 kredi **kredi[Noun]+[A3sg]+[Pnon]+[Nom]**  
 kartı **kart[Noun]+[A3sg]+SH[P3sg]+[Nom]**  
 kart[Noun]+[A3sg]+[Pnon]+YH[Acc] kart[Adj]-  
 [Noun]+[A3sg]+[Pnon]+YH[Acc] kart[Adj]-  
 [Noun]+[A3sg]+SH[P3sg]+[Nom]  
 Hesaplarının  
**hesap[Noun]+lAr[A3pl]+SH[P3sg]+NHn[Gen]**  
 hesap[Noun]+[A3sg]+lArH[P3pl]+NHn[Gen]  
 hesap[Noun]+lAr[A3pl]+SH[P3pl]+NHn[Gen]  
 hesap[Noun]+lAr[A3pl]+Hn[P2sg]+NHn[Gen]  
 kontrolü **kontrol[Noun]+[A3sg]+SH[P3sg]+[Nom]**  
 kontrol[Noun]+[A3sg]+[Pnon]+YH[Acc]

**Tablo 2. Morfoloji işlemleri sonucu elde edilen çıktı**

İlk olarak iş tecrübesi açıklamasına morfolojik analiz işlemi uygulanır. Morfolojik analiz programı sonucunda verilen cümle boşluklara ayrılarak kelimelere bölünür. Her bir kelime de kökleri ve aldığı eklere göre ayrılır[14]. Morfolojik analiz çıktısı olarak her bir kelimenin kullanıldığı tüm durumlar gösterilir. Bu yüzden bir kelime için birden fazla sonuç elde edilir. Muğlaklık giderici ise, morfolojik analiz çıktısını işleyerek kelimenin birden fazla olan kullanımları arasından en çok kullanılan durumunu seçer. Kelimenin en çok kullanılan ve doğru olan kullanımı muğlaklık gidericisi çıktısında sonuçlar arasında en başta yer alır. Bu sonuç Tablo 1’de kalın font ile belirtilmiştir.

### İLAN ANALİZİ

İş ilanları içerisinde bilgi çıkarım işleminde eğitim bilgisi, yabancı dil, iş tecrübesi ve yetkinlikler bilgilerinin çıkarımı hedeflenmiştir. Bu bilgileri elde edebilmek için Kariyer.net iş ilanları arasında incelemeler yapılmış ve iş ilanlarının birçoğunun ortak bir yapıda olduğu belirlenmiştir. Bilgi çıkarım için ortak yapılara uygun şablonlar oluşturulmuştur.

Eğitim kalıbı için:

*eğitim(<seviye>,<branş>,<branş,...>,<tercihen>)*  
 şablonu kullanılmıştır. Eğitim bilgisini bulmak için cümle içerisinde eğitim derecesini belirten ‘seviye’ özelliği (lisans, yüksek lisans, doktora), branş özelliği (üniversite, bölüm, fakülte) kullanılmıştır. Tercihen özelliği ise cümlede ‘tercihen’ kalıbının geçip geçmediğini belirtir. Tercihen özelliği tutularak ilerde iş ilanı ve özgeçmiş eşleştirmesinde tercihen kalıbı kullanılmış ise, o kriterlere sahip özgeçmişin daha

yüksek puan alması hedeflenmiştir. Bilgi çıkarımı işlemi yapılırken birden fazla kelimenin birleşmesiyle oluşan kelime gruplarının bulunabilmesi için n-gram Algoritması<sup>4</sup> uygulanır. Bu algoritma sayesinde örneğin ‘Bilgisayar Mühendisliği’ kelime grubu anlaşılıp bilgi çıkarımı gerçekleştirilir.

N-gram algoritması ile bilgi çıkarımı uygulanmış eğitim cümlesi ve sonucu aşağıda belirtilmiştir.

‘*Tercihen üniversitelerin işletme, endüstri mühendisliği, ekonomi bölümlerinden mezun*’

Sonuç:

eğitim(lisans,brans(işletme,endüstri mühendisliği ,ekonomi), doğru)

Elde edilen eğitim şablonunda, özellikle eğitim derecesi belirtilmediği için ön tanımlı olarak lisans derecesi kayıt edilmiştir. N-gram algoritması sonucunda oluşan kelime ve kelime grupları ile eğitim ontolojisi değerleri karşılaştırılarak üniversite bölümleri elde edilip şablona kayıt edilmiştir. Cümlede tercihen kelimesi geçtiği için şablon içerisindeki değerine ‘doğru’ ataması yapılmıştır.

Yabancı dil bilgisi için şablon:

*yabancıDil(<dil>,<dil>,...),<seviye>,<tercihen>)*

Yabancı dil bilgisi bulunurken bilinen yabancı dile onun seviyesine ve tercihen kalıbına bakılır.

İş tecrübesi ve yetkinlik şablonu:

*işYetkinlik(<sektör>,<sektör>,...),(<pozisyon>,<pozisyon>,...),(<yetkinlik>,<yetkinlik>,...),<yıl>,<seviye>,<tercihen>)*

İş tecrübesi ve yetkinlik bilgi çıkarımı yapılırken cümle içerisinde sektör bilgisi, pozisyon bilgisi, yetkinlik, tecrübe yılı, tecrübe seviyesi ve tercihen kalıbı bilgileri elde edilir. Bu bilgiler elde edilirken yetkinlik, sektör ve pozisyon ontolojileri kullanılır.

Örneğin;

‘*En az 3 yıl proje müdürü olarak çalışmış ve çok iyi derecede Java bilen*’ cümlesinin bilgi çıkarımından elde edilen sonuç:

işYetkinlik(pozisyon(proje müdürü), yetkinlik(java), 3, çok iyi, yanlış)’dır.

Oluşturulan şablonlar sayesinde bilgi çıkarımı sonucunda elde edilen bilgiler yapısal formatta kaydedilmiş olup ilan ve özgeçmiş eşleşmesi için kullanılacaktır.

## UYGULAMA

Çalışmamız .NET ortamında geliştirilmekte olup C# dili ile uygulanmıştır. Bilgi tabanımızı oluştururken

Graph veritabanlarından olan AllegroGraph<sup>5</sup> kullanılmıştır. AllegroGraph bir kaynak tanımlama çerçevesi (*resource definition framework*) üçlü deposudur. AllegroGraph içerisinde bilgiyi sorgulamak ve elde etmek için SPARQL(*protocol and RDF query language*) dili kullanılmıştır. Kaynak tanımlama çerçevesi bilgiyi sunmak ve web’de paylaşımını sağlamak için kullanılır. AllegroGraph içerisindeki her bir üçlü özne, eylem ve nesne <subject,predicate,object> birleşiminden oluşur. AllegroGraph içerisinde yabancı dil, üniversite isimleri, üniversite bölümleri, fakülte isimleri, yetkinlikler, sektörler, pozisyonlar ve seviyeler üçlü yapıda tutulmaktadır. AllegroGraph C# istemcisi kullanılarak ilanlar için tanımlanmış olan şablonlarla bu üçlü yapılar karşılaştırılarak bilgi çıkarım işlemi gerçekleştirilir.

---

1	İlanların HTML etiketlerini temizle
2	Elde edilen temizlenmiş ilan metnini paragraf ve
3	cümlelere ayırarak ilgili listesine kaydet
4	WHILE cümleler var ise cümleler listesinde
5	Cümle’deki kısaltmaları aç
6	WHILE cümleyi kelimelere böl
7	FOR a=1’den a<8’e kadar
8	N-gram algoritmasını uygula ve o cümle
9	için n-gram listesi oluştur.
10	Bu listeler içerisinde sırayla arama yap
11	Eğitim ontolojisi ile karşılaştırarak eğitim
12	şablonunu çıkar
13	İş ve yetkinlik ontolojisi ile karşılaştırarak
14	işYetkinlik şablonunu çıkar
15	Yabancı dil ontolojisi ile karşılaştırarak
16	yabancıDil şablonunu çıkar
17	bitiş
18	bitiş

---

**Tablo 3. İlan bilgi çıkarım işlemi sözde kodu**

Tablo 3’de ilanlardan bilgi çıkarım işleminin sözde kodunda belirttiği üzere; bilgi çıkarım işleminde ilk önce HTML yapısı halinde bulunan iş ilanı metnini elde edebilmek için <img>, <script>, <table>, <tfoot>, <thead>, <u> vb. HTML etiketleri temizlenir. Elde edilen metin içerisinde paragraflar ve cümleler elde edilerek listelere kayıt edilir. Bu işlemlerden sonra elde edilen cümleler içerisindeki kelimeler taranarak kısaltmalar var ise bu kısaltmaların genişletilmiş hali kullanılır. Daha sonra her bir cümlede her bir kelimesi ve kelime grupları elde edilip mevcut ontolojiler ile karşılaştırılarak elde edilen bilgiler ilgili şablona kaydedilir.

<sup>4</sup> <http://en.wikipedia.org/wiki/N-gram>

<sup>5</sup> <http://franz.com/agraph/allegrograph/>

## TEST

Özgeçmiş içerisinde bilgi çıkarım işlemi 30.000 Word formatında yazılmış türkçe özgeçmişler üzerinde test edilmiştir. 30.000 özgeçmiş içerisinde ortalama 343,20 kelime bulunmaktadır. Tablo 4'te özgeçmiş bilgi çıkarım işlemi sonucu belirtilmektedir. Bilgi çıkarım işleminde %71 doğru bilgi çıkarımı elde edilmiş olup, çalışma süresi ortalama 10 saniye sürmektedir

Özgeçmiş Sayısı	Doğruluk	Çalışma Süresi
30.000	%71	10 saniye

**Tablo 4. Özgeçmiş Bilgi Çıkarım İşlem Sonuçları**

Özgeçmiş içerisindeki iş tecrübeleri açıklaması bölümünde morfolojik analiz yöntemi ile 8.000 adet muhasebe uzmanı özgeçmişini üzerinde analizler yapılmış ve toplamda 934.508 kelime incelenmiştir. İş ilanından bilgi çıkarım işlemi 200 iş ilanı üzerinde test edilmiştir. 200 iş ilanı içerisinde ortalama 132 kelime bulunmaktadır. Bilgi çıkarımı uygulanmış olan örnek iş ilanı aşağıda, işlem sonucu Tablo 5'te belirtilmiştir.

İş İlanı :

*İşletme, Endüstri Mühendisliği, Bilgisayar Mühendisliği, Matematik ve benzer bölümlerinden mezun.*

*Ekip çalışmasına yatkın.*

*Şirketimizin dinamik yapısına uyum sağlayabilecek.*

*Analitik düşünme yeteneğine sahip.*

*İyi derecede İngilizce bilen.*

*Yurtdışı ve şehirdışı projelerde görev alabilecek.*

*Erkek adaylar için askerliğini tamamlamış*

*Esnek çalışma saatlerine uyabilecek.*

*Tercihen ERP SAP proje deneyimi olan.*

*MS Office uygulamalarına hakim.*

Eğitim Şablonu :

Üniversite Bölüm: İşletme, Matematik, Endüstri Mühendisliği, Bilgisayar Mühendisliği

Derece: Lisans

Tercihen : Yanlış

Yabancı Dil Şablon :

Yabancı Dil : İngilizce

Derece : İyi

Tercihen : Yanlış

İş Tecrübesi/Yetkinlik Şablonu :

Yetkinlik : ERP, SAP, MS Office

Tercihen : Doğru

**Tablo 5. Bilgi Çıkarım İşlemi Sonucu**

200 adet iş ilanı içerisinde yetkinliklerin %56'sı, üniversite isimlerinin %66.50'si ve yabancı dil bilgisinin tamamı bulunabilmiştir.

## SONUÇ

İş arayan adayların binlerce iş ilanı arasından kendi yetkinlik, eğitim ve iş tecrübelerine uygun iş ilanlarını bulması çok fazla zaman alan ve efor sarf ettiren bir işidir. Aynı şekilde eleman arayan firmalar için de açık pozisyonları için yüzlerce/binlerce başvuru arasından pozisyon niteliklerine uygun adayların bulunması işe alım sürecini yavaşlatan bir işlemdir. Türkiye'de ilk defa yapılmakta olan bu çalışmada, iş ilanlarının ve özgeçmişlerin karşılaştırılması ile, iş arayan adaylara uygun olan iş ilanlarının bulunması ve işverenlerin iş ilanlarına uygun olan adayların bulunması amaçlanmıştır. Çalışmamızın ilk aşaması olan iş ilanlarından ve özgeçmişlerden doğal dil işleme yöntemleri ile bilgi çıkarım işlemi tamamlanmış olup, iş ilanı ve özgeçmişlerin karşılaştırılması için çalışmalarımız devam etmektedir.

## TEŞEKKÜR

Bu çalışma TÜBİTAK TEYDEP programı tarafından 3130841 proje numarası ile desteklenmektedir.

## KAYNAKÇA

- [1] E. Riloff, "Automatically constructing a dictionary for information extraction tasks," in Proceedings of the National Conference on Artificial Intelligence. JOHN WILEY & SONS LTD, 1993, pp. 811–811.
- [2] "Automatically generating extraction patterns from untagged text," in Proceedings of the national conference on artificial intelligence, 1996, pp. 1044–1049.
- [3] J.-T. Kim and D. I. Moldovan, "Acquisition of linguistic patterns for knowledge-based information extraction," Knowledge and Data Engineering, IEEE Transactions on, vol. 7, no. 5, pp. 713–724, 1995.
- [4] J. Y. Chai and A. Biermann, "The use of lexical semantics in information extraction," in Proceedings of the ACL Workshop on Natural Language Learning, 1997.
- [5] J. Y. Chai, A. W. Biermann, and C. I. Guinn, "Two dimensional generalization in information extraction," in Proceedings of the Sixteenth National Conference on Artificial Intelligence, 1999, pp. 431–438.
- [6] R. Basili, M. T. Pazienza, M. Vindigni, P. Bank et al., "Corpusdriven learning of event recognition rules," in In Proceedings of Machine Learning for Information. Citeseer, 2000.
- [7] E. Agichtein and V. Ganti, "Mining reference tables for automatic text segmentation," in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004, pp. 20–29.
- [8] D. M. Bikel, S. Miller, R. Schwartz, and R. Weischedel, "Nymble: a high-performance learning name-finder," in Proceedings of the fifth conference on Applied natural language processing. Association for Computational Linguistics, 1997, pp. 194–201.
- [9] V. Borkar, K. Deshmukh, and S. Sarawagi, "Automatic segmentation of text into structured records," in ACM SIGMOD Record, vol. 30, no. 2. ACM, 2001, pp. 175–186.
- [10] K. Seymore, A. McCallum, and R. Rosenfeld, "Learning hidden markov model structure for information extraction," in AAAI- 99 Workshop on Machine Learning for Information Extraction, 1999, pp. 37–42.
- [11] Tang. H, Ye. J. "A Survey for Information Extraction Method" CS411
- [12] Celik. D, Karakas. A, Bal. G, Gultunca. C, Elci. A, Buluz. B, Alevli. M.C." Towards an Information Extraction System Based on Ontology to Match Resumes and Jobs" Computer Software and Applications Conference Workshops (COMPSACW), 2013 IEEE.
- [13] Sak, H., Güngör, T. ve Saraçlar, M. "Resources for Turkish Morphological Processing", Journal

of Language Resources and Evaluation, Cilt 45, s. 249-261, 2011.

- [14] Oflazer, K., Say, B., Hakkani Tür, D.Z. ve Tür, G. "Building a Turkish Treebank", Treebanks Building and Using Parsed Corpora, 15. Bölüm, S. 261-277, Kluwer, 2003.

## ÖZGEÇMİŞLER

**Gülşen Bal** Yeditepe Üniversitesi Bilgisayar Mühendisliği bölümünden lisans derecesini almıştır. Şu anda Yeditepe Üniversitesi MBA programında yüksek lisans eğitimine devam etmektedir. 2 yıldır Kariyer.net Araştırma Geliştirme merkezinde yazılım uzmanı olarak çalışmaktadır.



**Cem Gültunca** İstanbul Ticaret Üniversitesi Bilgisayar Mühendisliği bölümünden lisans derecesini almıştır. Şu anda İstanbul Ticaret Üniversitesi Bilgisayar Mühendisliği programında yüksek lisans eğitimine devam etmektedir. 2 yıldır Kariyer.net Araştırma Geliştirme merkezinde yazılım uzmanı olarak çalışmaktadır.



## Aşkın Karakaş

Aşkın Karakaş doktora derecesini 2000 yılında İstanbul Teknik Üniversitesi Makine Mühendisliği bölümünden almıştır. Halen Kariyer.Net in Ar-Ge Merkezi koordinatörlüğünü yürütmektedir. Yüksek performanslı veritabanlarının dizayn edilmesi ve yönetilmesi konusundaki tecrübesinin yanında, Ontoloji oluşturma ve eşleştirme, bilgi çıkarımı, doğal dil işleme ve kompleks ağlar ile ilgili projelerde görev almıştır.



## Tunga Güngör

Doktora derecesini 1995 yılında Boğaziçi Üniversitesi, Bilgisayar Mühendisliği bölümünden almıştır. Şu anda aynı bölümde öğretim üyesidir. Araştırma alanları doğal dil işleme, makine öğrenmesi, makine tercümesi ve örüntü tanımadır. Yazarın 70 civarında yayınlanmış bilimsel makalesi vardır; birçok araştırma projelerine ve konferans düzenlemelerine katılmıştır.

