

# TÜRKÇE METİNLERDE ŞARTLI RASTGELE ALANLARLA VARLIK İSMİ TANIMA

## NAMED ENTITY RECOGNITION BY CONDITIONAL RANDOM FIELDS FROM TURKISH INFORMAL TEXTS

Serap Özkaya<sup>1</sup>, Banu Diri<sup>1</sup>

1. Bilgisayar Mühendisliği Bölümü, Yıldız Teknik Üniversitesi  
serap.ozkaya@gmail.com, banu@ce.yildiz.edu.tr

### ÖZETÇE

*Varlık İsmi Tanıma, Doğal Dil İşleme'nin alanlarından biri olup, resmi ve resmi olmayan dokümanlarda alana bağlı veya alandan bağımsız olarak kişi, yer, kurum, tarih, formül ve para gibi farklı varlık isimlerini bulmayı hedeflemektedir. Kural tabanlı veya makine öğrenmesi teknikleri veya bunların birlikte kullanıldığı sistemler ile tanıma yapılabilmektedir. Bu çalışmada, konudan bağımsız ancak e-posta gibi resmi olmayan dokümanlardan Şartlı Rastgele Alanlar kullanılarak kişi, kurum ve yer isimlerinin etiketlenmesi gerçekleştirilmiştir.*

### ABSTRACT

*Named Entity Recognition (NER) being one of the areas of Natural Language processing can be domain dependent or independent for formal and informal texts aims to extract information about name entity such as person, location, organization, dates, formula and money. Rule Based methods and machine learning methods can be implemented in the system. In this study, Conditional Random Fields has been used to extract name entities which are person, location and organization names from informal texts.*

### 1. GİRİŞ

Varlık İsmi Tanıma-VİT olarak Türkçe'ye çevirdiğimiz "Name Entity Recognition-NER", veri ve doküman madenciliği, doğal dil işleme, bilgi çıkarımı ve bilgiye erişim gibi birçok disiplinle ilişkilidir. VİT'in amacı resmi olan ve olmayan bir dilde yazılmış, belli bir çalışma alanına bağlı veya bağımsız olan tüm dokümanlar içerisinde dile bağımlı veya bağımsız olarak varlık isimlerini bulmak ve bunları kişi, kurum ve yer ismi, zaman, tarih ve parasal ifadeler olarak sınıflandırmaktır.

Varlık İsmi Tanıma ile ilgili çalışmalar 1900'lü yıllarda başlamış olup, "Message Understanding Conference-MUC" platformunda farklı diller için yapılan çalışmalar yer almıştır. Çalışmalarda özellikle yaygın olarak kullanılan dil İngilizce olup, Varlık İsmi Tanıma da başarımlar %90'larda olmaktadır. İngilizce'nin dışında Arapça, Çince, Japonca gibi farklı dillerde de çok sayıda çalışma yapılmıştır. Bu alandaki ilk çalışmalardan biri Rau [1] tarafından yapılan doküman içerisinde şirket isimlerini çıkarmak ve tanımdır. VİT için kullanılan ilk sistemler kural tabanlı algoritmalar iken, son zamanlarda makine öğrenmesi tekniklerini kullanan modern sistemlerde geliştirilmiştir. Hem kural tabanlı hem de makine öğrenmesini birlikte kullanan hibrit sistemler de mevcuttur. Yapılan çalışmalar, kullanılan dile (tek dil veya çok dilli),

dokümanın türüne veya alanına, ve varlık isminin çeşidine göre değerlendirilmektedir [2].

Varlık İsmi Tanıma'da kullanılan öğrenme yöntemleri eğitici, yarı-eğitici ve eğitici olmak üzere üç grupta ele alınmaktadır. Eğitici sistemlerde Saklı Markov Modelleri [3], Karar Ağaçları [4], Maksimum Entropy Modeli [5], Destek Vektör Makineleri [6], ve Şartlı Rastgele Alanların [7,8] kullanımı örnek olarak verilebilir. Bu sistemlerde işaretlenmiş geniş bir derlem, ezberlenecek varlıklar listesi ve belirsizliği giderecek kurallar kullanılır. Yarı eğitici sistemlerde ise, başlangıçta öğrenmek için küçük bir küme kullanılarak eğitim yapılır, sonrasında sistem kendi kendine öğrenir [9]. Eğitici yöntemlerde, ise grup içi benzerlikler minimum, gruplar arası benzerlikler maksimum tutularak varlık isimleri tanınmaya çalışılır. Bu yöntemlerde bazen WordNet gibi sözlüksel kaynaklarda kullanılır [10].

Türkçe için yapılan VİT çalışmaları ise, sınırlıdır. Bunlardan ilki [11] tarafından gerçekleştirilen dilden bağımsız olan varlık tanıma çalışmasıdır. 2003 yılında [12] tarafından Türkçe için, istatistiksel isim etiketleme, finans haber metinlerinden kişi isimlerinin çıkarılması [13], Türkçe haber metinleri için varlık isimlerini etiketleme [14] ve yine alandan bağımsız dört farklı varlık ismi tanıma üzerine [15] çalışmalar yapılmıştır.

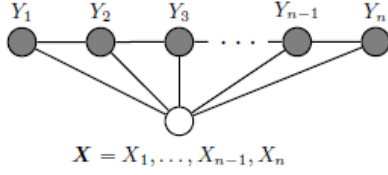
Bu çalışmada resmi olmayan bir dilde yazılmış olan Türkçe e-postalar içerisinde Şartlı Rastgele Alanlar kullanılarak yer, kurum-kuruluş ve özel isimlerin etiketlenerek çıkarılması için bir sistem geliştirilmiştir. Makalenin ilerleyen bölümlerinde sırasıyla kullanılan yöntem, geliştirilen sistem, özellik çıkarımı ve alınan deneysel sonuçlardan bahsedilmektedir.

### 2. KULLANILAN YÖNTEM

Geliştirilen uygulamada Şartlı Rastgele Alan yöntemi kullanılarak, kural tabanlı varlık ismi tanıma gerçekleştirilmiştir. Doğal Dil işlemede sözdizimsel analiz ile sıralı veri etiketlenebilmektedir [16]. Örneğin;

<İsim> Ahmet <İsim> <TamlayanEki> ın <TamlayanEki> <İsim> bab a <İsim> <TamlayanEki> sı <TamlayanEki>, <Sıfat> beyaz <Sıfat> <İsim> koyun <İsim> <NesneEki> u <NesneEki> <İsim> araba <İsim> <DiğerZarflar> ile <DiğerZarflar> <İsim> köy <İsim> <DolaylıTümleçEki> e <DolaylıTümleçEki> <Yüklem> getir <Yüklem> <ZamanEki> di <ZamanEki>

Şartlı Rastgele Alan ise sıralı veriyi işaretlemek ve bölümlere ayırmak için kullanılan, Maksimum Entropi Markov Model ve Gizli Markov Modelin genel halini yansıtan bir olasılık sunar. Şartlı Rastgele Alanda Şekil 1'de görüldüğü üzere,  $y$  gibi belirli bir işaret dizisinin  $x$  değeriyle şartlı olasılığını hesaplamak için, yönsüz çizge modelini kullanmaktadır [16].



Şekil 1. Yönsüz Çizge Model

Eşitlik 1'de  $f_k(y_{t-1}, y_t, x, t)$  fonksiyonu, tüm gözlem dizisinin ve  $t$ ,  $t-1$  pozisyonundaki işaretlerin geçiş özelliğini ifade etmektedir.  $\lambda_k$  ise öğrenilen veri sonucunda optimize edilen parametreyi ifade etmektedir.  $Z_k$  da normalizasyon faktörüdür. Geçiş özelliği fonksiyonun her biri, gerçek gözlem özellik değerlerinin  $f(x, i)$  birini alır. Dolayısıyla tüm özellik fonksiyonları gerçek değerlere sahiptirler.

$$P_\lambda(y|x) = \frac{1}{Z_x} \exp\left(\sum_{t=1}^T \sum_k \lambda_k f_k(y_{t-1}, y_t, x, t)\right) \quad (1)$$

Örneğin,

$$f_k(y_{t-1}, y_t, x, t) = \begin{cases} 1 & y_{t-1} \text{ sıfat ise, } y_t \text{ isimdir} \\ 0 & \text{Diğer Durum} \end{cases}$$

Öncelikli amaç, işlenecek verinin benzerliğini maksimuma çıkarmak için tahminlerde bulunmaktır. İfadenin logaritması alındıktan sonra (Eşitlik 2) elde ettiğimiz formülde benzerlik-log (log-likelihood) fonksiyonu konkavdır ve global bir optimuma sahiptir, o yüzden bir optimizasyon problemidir.

$$\log P_\lambda(y|x) = \left(\sum_{t=1}^T \sum_k \lambda_k f_k(y_{t-1}, y_t, x, t)\right) - \log Z_x \quad (2)$$

Türevini alıp 0 değerine eşitledikten sonra, maksimum entropi sınırını elde etmiş oluruz. Parametreler ard arda yapılacak ölçeklendirme algoritmasıyla bulunabilir (Eşitlik 3).

$$\text{maximize} \left(\sum_{t=1}^T \sum_k \lambda_k f_k(y_{t-1}, y_t, x, t)\right) - \log Z_x \quad (3)$$

İkinci amacımız ise, verilen gözlem dizisine en benzer işaret dizisini bulmaktır (Eşitlik 4).

$$s = \arg \max p_\lambda(s|o) \quad (4)$$

Sınırlı Viterbi algoritmasını, optimal  $s$  çözümünü bulmak için kullanabiliriz [17].

Bu çalışmada, Şartlı Rastgele Alan için Java programlama dilinde geliştirilmiş kütüphane kullanılmıştır[18].

### 3. GELİŞTİRİLEN SİSTEM

Varlık İsmi Tanıma sistemi bir java uygulaması olarak yazılmış ve e-posta içerisindeki her kelime bir java nesnesi olarak düşünülmüştür. Oluşturulan bu nesnelere kelime ve kelimenin uzunluğu, adedi, kısaltma olup olmadığı, tipi (isim, sıfat, zamir, vd.), cümlelerin ilk kelimesi mi gibi özellik değişkenleri tutulmaktadır. Uygulama çalıştırıldığında, veri setindeki her e-posta ve içerisindeki her farklı kelime için bir nesne oluşturulup:

- Özellik çıkarımında kullanılmak üzere kelime sıklıkları ve karakter tabanlı 1-gram, 2-gram, 3-gram'lar çıkarılır.
- Çıkarılan her kelimenin daha önceden oluşturulan kaynaklar listesinde yer almadığı kontrol edilir.

- Kelimenin büyük harfle başlaması, noktalama işaretlerini içermesi, başlık bilgisinde yer alması gibi özelliklere bakılarak, her özellik için kelime nesnesindeki özellik etiket değerine 1 atanır.
- Etiketleme işlemi bittikten sonra veri setinden çıkartılan özellikler kullanılarak sınıflandırma için optimizasyon parametreleri belirlenir.

Son aşamada Şartlı Rastgele Alanlar yöntemiyle etiket sınıfı belirlenir. Ancak, sınıflandırma işleminden önce uygulamaya belirlenen kural tanımları verilmelidir. Örneğin, bir kelimenin kişi ismi olabilmesi için çıkarılan kurallardan bir kaçısı aşağıda verilmektedir.

<kisi\_ismi>From/BüyükHarfleBaşlama/<kisi\_ismi>

Örneğin: From: Serap Özkaya [serap.ozkaya@gmail.com](mailto:serap.ozkaya@gmail.com)

<kisi\_ismi>ForwardedBy/BüyükHarfleBaşlama/BüyükHarfleBaşlama-<kisi\_ismi>

Örneğin: Forwarded By Banu Diri

<kisi\_ismi>Ünvan/BüyükHarfleBaşlama/<kisi\_ismi>

Örneğin: Dr. Haluk Bey

<kisi\_ismi>BüyükHarfleBaşlama/Ünvan-<kisi\_ismi>

Örneğin: Merhaba Ahmet Bey

<kisi\_ismi>DiğerÖzelKelime/BüyükHarfleBaşlama-<kisi\_ismi

Örneğin: Sayın Banu Hanım

<kisi\_ismi>DiğerÖzelKelime/Ünvan/BüyükHarfleBaşlama-<kisi\_ismi>

Örneğin: Saygılarımla Av. Ayşen

### 4. VERİ SETİ VE KULLANILAN ÖZELLİKLER

Resmi olmayan Türkçe metinlerden kişi, kurum ve yer varlık isimlerini çıkarmak için kullanılan veri seti akademik, kurumsal ve kişisel e-postalardan eşit sayıda toplanmış 150 adet e-postadan oluşmaktadır. E-postalar, maksimum 792 kelime, minimum da 98 kelime olmak üzere ortalama 158 kelime uzunluğundadır. E-postalar yazım kurallarına dikkat edilerek yazılmış dokümanlar olmasa da kendilerine özgü bir formata sahiptirler. Bir e-posta iletilişinin özelliklerinden biri daha az söz dizimine sahip iyi tanımlanmış bir başlık bilgisi ve iletilişin sonunda yer alan ve genellikle kişi ismi olan imza bilgisidir. Ancak e-postalarda metin bilgisi yapısal ve gramer olarak bir çok hatayı içerisinde bulunduran, alıcı ve gönderici tarafından anlaşılabilen özel kısaltmalar ve jargonları da içermektedir (Şekil 2).

Message-ID: <7789106.1072119990022.JavaMail.evans@thyme>

Date: Mon, 6 SEP 2010 09:27:45 -0800 (PST)

From: serap.ozkaya@gmail.com

To: banu@ce.yildiz.edu.tr

Subject: Updated: Ilgaz Turu

Cc: vildan@ce.yildiz.edu.tr

Mime-Version: 1.0

Content-Type: text/plain; charset=us-ascii

Content-Transfer-Encoding: 7bit

Merhaba Banu,

Ilgaz turmuzun ayrıntılı programı ektedir.

10 kişi ve üzeri sayıda gruplar için tur fiyatı 15 YTL indirimlidir.

keyifli gezilerde bulusmak üzere.

İyi çalışmalar

Serap Ozkaya

Şekil 2: Örnek E-Posta

#### 4.1 Kullanılan Dış Kaynaklar

Temel özelliklerin çıkarılmasında yardımcı olan kısaltmalar, ünvanlar ve bazı özel kelimelerin tutulduğu sözlüklere de ihtiyaç duyulmaktadır. Bu çalışmada 175 adet *kısaltma* listesi (İng., Fr., Tşk, Sok., Mah., TD, Sn, A.Ş., Ltd,...), 35 adet *ünvan* listesi (Prof., Dr., Av., Gen.,...), ve 32 adet *özel*

*kelimeler* listesi (Sayın, Hanım, Bey, Hocam, Üniversite, Bakanlık, Hastane, Dağ, Tepe,...) kullanılmıştır.

#### 4.2 Temel Özellikler

Temel özellikler, herhangi bir kaynak kullanılmadan sadece veri setinden çıkartılmaktadır. E-posta için çıkartılan bu özellikler; e-posta başlık bilgisi özellikleri, etiketlenebilen özellikler ve kural tabanlı özellikler şeklinde 3 farklı grupta değerlendirilmektedir.

##### Başlık Bilgisi Özellikleri

Her e-posta bir başlık bilgisiyle başlamaktadır. İyi tanımlanmış bir başlık bilgisinde varlık isimleri yer alabileceğinden uygulamada aşağıdaki başlık bilgisi alanlarına bakılmaktadır.

- From, To, Cc ,Bcc, Send, Sender, In-Reply-to, Followup-to, Reply-to, Forwarded by
- Konu bilgisi

##### Etiketlenebilen Özellikler

Şartlı Rastgele Alan yönteminde, veri setindeki kelimelerin özellikleri etiket bilgisi ile gösterilebilmektedir. Örneğin, “*ilk kelime mi=1*” ataması kelimenin bir cümlede ilk kelime olduğunu göstermektedir. Eğer kelime cümledeki ilk kelime değilse bu değer 0 olacaktır [16].

E-postada geçen bir kelime aşağıdaki özelliklere sahip olduğu durumda etiket değerlerine 1 atanır. Böylece kelimenin sahip olduğu özellik bilgisi tutulur.

- Büyük harfle başlama : Kelime büyük harf ile başlıyor ise,
- Noktalama işaretleri : Kelime (-; ,” “) gibi noktalama işaretleriyle birlikte yazılmış ise,
- İlk kelime : Cümlenin ilk kelimesi ise,
- Başlık bilgisi : Kelime başlık bilgisinde yer alıyorsa,
- Ünvanla başlama : Kelimeden önce gelen kelime ünvan listesinde var ise,
- Özel kelimelerle başlama : Kelimeden önce gelen kelime özel kelimeler listesinde var ise,

özellik etiket değerlerine 1 ataması yapılır.

##### Kural Tabanlı Özellikler

Kural tabanlı özellikler bir varlık isminin kişi, yer ve kurum isimlerinden biri olup olmadığının tahmininde kullanmak için e-postalardan çıkartılan özelliklerdir.

- Ard arda gelen kelimeler : Veri setinde ele alınan her kelime için kelimeden önce gelen ve kelimeden sonra gelen kelimeler belirlenir. Bu özellik daha çok isim-soy isim eşleştirmesinde kullanılmaktadır.
- Karakter tabanlı 1-gram, 2-gram, 3-gram : Bazı son ekleri belirleyebilmek için bu özellik kullanılmaktadır.
- Kelime uzunluğu : Kelimenin kısaltma, ünvan, vd. olup olamayacağı tahmin edilebilmektedir.
- İsim, Soy İsim : Ard arda gelen kelimelerden kişi ismi olabilecek kelimelerin hangisinin isim, hangisinin soy isim olduğunun belirlenmesidir.

### 5. DENEYSEL SONUÇLAR

Geliştirilen sistemde test için akademik, kurumsal ve kişisel olmak üzere üç farklı alanın herbiri için 50 farklı e-posta seçilmiştir. Bu e-postalar içerisinde bulunan kişi, yer ve kurum isimleri etiketlenmeye çalışılmıştır. Sistemin başarısını ölçmek için eşitlik-5’deki kesinlik (precision), (6)’deki çağrı

(recall), (7)’deki kesinlik ve ölçüm değerlerinin harmonik ortalaması olan F-ölçüm (F-measure) kullanılmıştır.

$$\text{Kesinlik}(P)=\text{Doğru tespit edilen varlık sayısı}/\text{Tespit edilen varlık sayısı} \quad (5)$$

$$\text{Çağrı}(R)=\text{Doğru tespit edilen varlık sayısı}/\text{Test kümesindeki toplam varlık sayısı} \quad (6)$$

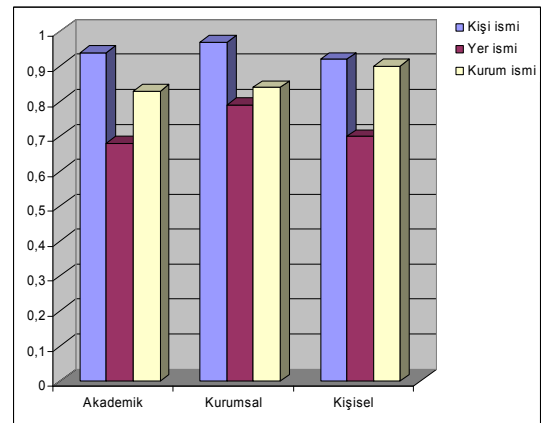
$$\text{F-Ölçüm}=(2*P*R) / (P+R) \quad (7)$$

Tablo-1’de üç farklı doküman türünde yer alan toplam metin sayıları, her alan içerisindeki yer, kurum ve isim varlık sayıları, sistem tarafından tespit edilen varlık ismi sayısı, sistem tarafından doğru tespit edilen varlık ismi sayısı ve sistemden bağımsız elle yapılan varlık ismi tespitleri gösterilmektedir. Uygulamanın doğru tespit sayısı, uygulama tarafından tespit edilen varlık ismi sayılarından hatalı tespitler çıkartılarak oluşturulmuştur.

Tablo 1. Tespit edilen varlık sayıları (UT, Uygulamanın Tespiti, DT, Uygulamanın Doğru Tespiti, GT, Gerçek Tespiti)

e-Posta Türü	Akademik UT/DT/GT	Kurumsal UT/DT/GT	Kişisel UT/DT/GT
Kişi İsmi	159/146/150	175/172/180	160/141/145
F-Ölçüm	0.94	0.97	0.92
Yer İsmi	32 / 17 /18	54 / 35 /35	36 /22 /26
F-Ölçüm	0.68	0.79	0.70
Kurum İsmi	37 / 27 /28	47 / 37 /41	17 / 14 /14
F-Ölçüm	0.83	0.84	0.90

Şekil-3’te üç farklı e-posta türünde kişi, yer ve kurum varlık isimlerinin tanınmadaki başarısı f-ölçüm cinsinden verilmektedir. Geliştirilen kurallar ile en başarılı şekilde bulunan varlık *kişi ismi* olmuştur. İkinci sırada ise *kurum ismi* gelmektedir. En düşük performans *yer isimleri* olmuştur.



Şekil 3. Varlık isimlerinin farklı alanlarda ki F- Ölçüm değerleri

Kurum varlık isimlerinde başarının kişi isimlerine göre daha düşük olmasının sebebi öncelikli olarak yazım kurallarına

uyulmamış olmasıdır. Örneğin, e-postanın imza bölümünde “Tesekkürler” kelimesinden sonra gelen

“E.Yıldız

Mediation Solutions”

bilgisi olsun. *E.Yıldız* kelimesi yazılırken ilk isim ile soyadı arasında boşluk olmaması ve arada nokta işaretinin yer almasından dolayı ifadenin tamamı kurum ismi olarak etiketlenmektedir.

Tanımda başarının en düşük olduğu varlık, yer isimleridir. E-postalar içerisinde yer isimlerinin yazılmasında, yer isminin sonra gelen iyilik, durum ve bildirme eklerinde kesme işaretinin kullanılmadığı gözlemlenmiştir. Bu durumda yer isimlerinin hatalı etiketlenmesine sebep olmaktadır.

En başarılı e-postalar kurumsal içerikli olup, ortalama %87 doğrulukla tanıma yaparken, en az hata ile tanınan varlık ismi %95 ile kişi ismi olmuştur. En başarısızda ortalama %72 ile yer varlık ismidir. Türkçe üzerine aynı veri setini kullanan bir çalışma olmadığından sistemin başarısı farklı çalışmalarla karşılaştırılmamıştır. Çalışmanın ileri aşamaları yine dile bağımlı, fakat dokümanın türünden bağımsız olarak daha fazla varlık ismi tanımaya yönelik devam edecektir.

## 6. SONUÇ

Varlık İsmi Tanıma, bilgi çıkarımı ve bilgiye erişim alanında kendisine yer edinse de, dile bağlı olarak geliştirdiğimiz kurallar ile doğal dil işleme altında da kendisine yer bulmuştur. Varlık İsmi Tanıma üzerine yapılan çalışmalar Türkçede sınırlı sayıda olup, çoğu alana bağlıdır. Ayrıca, tanıma çalışılan varlıkların sayısında sınırlıdır. Bu çalışmada resmi olmayan (informal) Türkçe e-postalar üzerinde bazı kurallar çıkarılarak, Şartlı Rastgele Alanlar kullanılıp kişi, kurum ve yer gibi üç farklı varlık isminin tanınması gerçekleştirilmiştir. En yüksek başarı kurumsal e-postalarda kişi varlık ismi ile alınmıştır.

## 7. KAYNAKÇA

- [1] Rau, L.F., “Extracting Company Names from Text”, In *Proc. Conference on Artificial Intelligence Applications of IEEE*, 1991.
- [2] Nadeau, D., Sekine, S., “A survey of named entity recognition and classification”, *Linguistica Investigationes*, vol. 30, no. 1, pp.3-26, 2007.
- [3] Bikel, D.M., Miller, S., Schwartz, R., Weischedel, R., “A High-Performance Learning Name-finder”, In *Proc. Conference on Applied Natural Language Processing*, 1997.
- [4] Sekine, S., “Description of the Japanese NE System Used For Met-2”, In *Proc.Message Understanding Conference*, 1998.
- [5] Borthwick, A., Sterling, J., Agichtein, E., Grishman, R., “Description of the MENE Named Entity System as used in MUC-7”, In *Proc. Seventh Message Understanding Conference*, 1998.
- [6] Asahara, M., Matsumoto, Y., “Japanese Named Entity Extraction with Redundant Morphological Analysis”, In *Proc. Human Language Technology Conference – North American chapter of the Association for Computational Linguistics*, 2003.
- [7] McCallum, A., Li, W., “Early Results for Named Entity Recognition with Conditional Random Fields, Features Induction and Web-Enhanced Lexicons”, In *Proc. Conference on Computational Natural Language Learning*, 2003.
- [8] Minkov, E., Wang, R., Cohen, W., “Extracting Personal Names from e-mail: Applying Name Entity Recognition to Informal Text”, In *Proc. Human Language Technology and*

*Conference on Empirical Method in Natural Language Processing*, 2005.

[9] Brin, S., “Extracting Patterns and Relations from the World Wide Web”, In *Proc.Conference of Extending Database Technology. Workshop on the Web and Databases*, 1998.

[10] Alfonseca, E., Manandhar, S., “An Unsupervised Method for General Named Entity Recognition and Automated Concept Discovery”, In *Proc. International Conference on General WordNet*, 2002.

[11] Cucerzan, S., Yarowsky, D., “Language independent named entity recognition combining morphological and contextual evidence”, In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, 1999.

[12] Tür, G., Hakkani-Tür, D., ve Oflazer, K., “A statistical information extraction system for Turkish”, *Natural Language Engineering*, 9, 2:181-210, 2003.

[13] Bayraktar, Ö., Taşkaya-Temizel, T., “Person name extraction from Turkish Financial news text using local grammar based approach”, In *Proceedings of the International Symposium on Computer and Information Sciences*, 2008.

[14] Küçük, D., Yazıcı, A., “Named entity recognition experiments on Turkish texts”, In *Proceedings of the International Conference on Flexible Query Answering Systems*, 2009.

[15] Dalkılıç, F.E., Gelişli, S., Diri, B., “Türkçe Kural Tabanlı Varlık İsmi Tanıma”, 18. *Sinyal İşleme ve Uygulama Kurultayı*, Diyarbakır, 2010.

[16] Wallach, H.M., “Conditional Random Fields: An Introduction”, *University of Pennsylvania CIS Technical Report*, 2004.

[17] Chang, Y.S., Sung, Y.H., “Applying Name Entity Recognition to Informal Text”, *Stanford University*, 2006.

[18] Sarawagi, S., Conditional Random Field Project Page, <http://crf.sourceforge.net/>