EXPLOITING INFORMATION EXTRACTION TECHNIQUES FOR AUTOMATIC SEMANTIC ANNOTATION AND RETRIEVAL OF NEWS VIDEOS IN TURKISH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

DİLEK KÜÇÜK

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
COMPUTER ENGINEERING

FEBRUARY 2011

Approval of the thesis:

EXPLOITING INFORMATION EXTRACTION TECHNIQUES FOR AUTOMATIC SEMANTIC ANNOTATION AND RETRIEVAL OF NEWS VIDEOS IN TURKISH

submitted by **DİLEK KÜÇÜK** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Computer Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Adnan Yazıcı _____
Head of Department, **Computer Engineering**

Prof. Dr. Adnan Yazıcı _____
Supervisor, **Computer Engineering Dept., METU**

Examining Committee Members:

Prof. Dr. Fazlı Can
Computer Engineering, Bilkent University _____

Prof. Dr. Adnan Yazıcı
Computer Engineering, METU _____

Assoc. Prof. Dr. Cem Bozşahin
Computer Engineering, METU _____

Asst. Prof. Dr. Tuğba Taşkaya Temizel
Informatics Institute, METU _____

Asst. Prof. Dr. Pınar Şenkul
Computer Engineering, METU _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: DİLEK KÜÇÜK

Signature :

ABSTRACT

EXPLOITING INFORMATION EXTRACTION TECHNIQUES FOR AUTOMATIC SEMANTIC ANNOTATION AND RETRIEVAL OF NEWS VIDEOS IN TURKISH

Küçük, Dilek
Ph.D., Department of Computer Engineering
Supervisor : Prof. Dr. Adnan Yazıcı

February 2011, 110 pages

Information extraction (IE) is known to be an effective technique for automatic semantic indexing of news texts. In this study, we propose a text-based fully automated system for the semantic annotation and retrieval of news videos in Turkish which exploits several IE techniques on the video texts. The IE techniques employed by the system include named entity recognition, automatic hyperlinking, person entity extraction with coreference resolution, and event extraction. The system utilizes the outputs of the components implementing these IE techniques as the semantic annotations for the underlying news video archives. Apart from the IE components, the proposed system comprises a news video database in addition to components for news story segmentation, sliding text recognition, and semantic video retrieval. We also propose a semi-automatic counterpart of system where the only manual intervention takes place during text extraction. Both systems are executed on genuine video data sets consisting of videos broadcasted by Turkish Radio and Television Corporation. The current study is significant as it proposes the first fully automated system

to facilitate semantic annotation and retrieval of news videos in Turkish, yet the proposed system and its semi-automated counterpart are quite generic and hence they could be customized to build similar systems for video archives in other languages as well. Moreover, IE research on Turkish texts is known to be rare and within the course of this study, we have proposed and implemented novel techniques for several IE tasks on Turkish texts. As an application example, we have demonstrated the utilization of the implemented IE components to facilitate multilingual video retrieval.

Keywords: information extraction in Turkish, semantic video annotation, semantic video indexing, video retrieval

ÖZ

TÜRKÇE HABER VİDEOLARININ OTOMATİK ANLAMSAL ETİKETLENMELERİ VE ERİŞİMLERİ İÇİN BİLGİ ÇIKARIM TEKNİKLERİNİN KULLANIMI

Küçük, Dilek
Doktora, Bilgisayar Mühendisliği Bölümü
Tez Yöneticisi : Prof. Dr. Adnan Yazıcı

Şubat 2011, 110 sayfa

Bilgi çıkarımının (BÇ) haber metinlerinin otomatik anlamsal indekslenmesinde etkili bir teknik olduğu bilinmektedir. Bu çalışmada, Türkçe haber videolarının anlamsal etiketlenmeleri ve erişimleri için video metinlerinde çeşitli BÇ tekniklerini kullanan metin tabanlı tam otomatik bir sistem önermekteyiz. Sistem tarafından kullanılan BÇ teknikleri isimli nesne çıkarımı, otomatik üstmetin bağlantısı oluşturma, kişi nesnesi çıkarımı ile eşgönderge çözümleme ve olay çıkarımını kapsamaktadır. Sistem, bu BÇ tekniklerini gerçekleştiren bileşenlerin çıktılarını alttaki video arşivlerinin anlamsal etiketleri olarak kullanır. Önerilen sistem, BÇ bileşenleri dışında, bir haber videosu veritabanına ek olarak haber hikayesi bölütleme, kayan yazı tanıma, ve anlamsal video erişimi bileşenlerini de içermektedir. Biz ayrıca sistemin, elle tek müdahalenin metin çıkarımı sırasında gerçekleştiği yarı otomatik bir eşini de önermekteyiz. Her iki sistem de Türkiye Radyo ve Televizyon Kurumu tarafından yayınlanmış videolardan oluşan gerçek video veri kümeleri üzerinde çalıştırılmıştır. Bu çalışma, Türkçe haber videolarının anlamsal etiketlenmeleri ve erişimlerini kolaylaştıran ilk tam otomatik sis-

temi önermesi açısından önemlidir, bununla birlikte önerilen sistem ve onun yarı otomatik eşi oldukça geneldirler ve bu nedenle diğer dillerdeki video arşivleri için de benzer sistemler oluşturmak için uyarlanabilirler. Dahası, Türkçe metinlerde BÇ araştırmalarının seyrek olduğu bilinmektedir ve bu çalışma kapsamında biz Türkçe metinler üzerinde çeşitli BÇ işleri için yeni teknikler önerdik ve gerçekleştirdik. Bir uygulama örneği olarak, biz gerçekleştirilen BÇ bileşenlerinin çok dilli video erişimini kolaylaştırmada kullanımını gösterdik.

Anahtar Kelimeler: Türkçe'de bilgi çıkarımı, anlamsal video etiketleme, anlamsal video indeksleme, video erişimi

To my family

ACKNOWLEDGMENTS

First of all, I would like to express my sincerest thanks to my supervisor Prof. Dr. Adnan Yazıcı for his guidance, motivation, and continuous support throughout this study. Without his valuable ideas and comments, this study would have never been possible.

I am grateful to Assoc. Prof. Dr. Cem Bozşahin and Asst. Prof. Dr. Tuğba Taşkaya Temizel for their valuable comments during my thesis monitoring committee meetings and also for Dr. Taşkaya Temizel's providing their Anadolu Agency data set. I also like to express my gratitude to Prof. Dr. Fazlı Can and Asst. Prof. Dr. Pınar Şenkul for their review of my thesis and their valuable comments on it.

Our periodical meetings with the members of the Multimedia Database Group at the Department of Computer Engineering have also contributed to my thesis, hence I am grateful to all of the group members, particularly to Asst. Prof. Dr. Murat Koyuncu and Asst. Prof. Dr. Mustafa Sert, for the valuable discussions. I gratefully acknowledge the suggestions and ideas of Dr. Selçuk Köprü especially at the very beginning of this study.

I wish to thank Erinç Dikici and Assoc. Prof. Dr. Murat Saraçlar from Boğaziçi University for kindly providing us their sliding text recognizer and news video data set.

I owe thanks to Prof. Dr. Muammer Ermiş and Prof. Dr. Işık Çadırcı for establishing a peaceful and motivating working environment at the Power Electronics Department at TÜBİTAK Uzay Institute. I also like to thank all of my colleagues and all of my friends for their support throughout my study.

Finally, I am grateful to all of the members of my dearest family for their continuous support and encouragement during this study.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ACE	Automatic Content Extraction
ASR	Automatic Speech Recognition
AV	Audio-Visual
IE	Information Extraction
IR	Information Retrieval
JMF	Java Media Framework
LSA	Latent Semantic Analysis
MT	Machine Translation
MUC	Message Understanding Conference
NE	Named Entity
NER	Named Entity Recognition
NL	Natural Language
NLP	Natural Language Processing
OOV	Out Of Vocabulary
SQL	Structured Query Language
TDT	Topic Detection and Tracking

TF	Term Frequency
TF×IDF	Term Frequency×Inverse Document Frequency
TRT	Turkish Radio and Television Corporation
URL	Uniform Resource Locator
VOCR	Video Optical Character Recognition
WER	Word Error Rate

CHAPTER 1

INTRODUCTION

1.1 Motivation

News video archives keep increasing in size everyday as all other types of multimedia [107]. But this increase comes with its yet to be solved problem: How to annotate these news videos to facilitate later retrieval? The users of these video archives are known to be mostly interested in high level semantic concepts in the videos such as the objects and events that are central to the topics of these videos. Manual semantic annotation is not a feasible option since it is a very time-consuming and labor-intensive process to be employed in practical settings. On the other hand, automatic approaches usually facilitate low level feature extraction instead of extracting high level semantic information and they employ the resulting low level information as annotations. This lack of coincidence between the needs of the users and the automatically extracted annotations is commonly referred to as the *semantic gap* [96] and there are several studies that aim to bridge this gap, utilizing various modalities of the videos.

Audio-visual (AV) components of the videos are extensively utilized to arrive at accurate and robust methods for semantic video annotation and retrieval, as surveyed in [97, 110]. Moreover, several plausible studies addressed the need for a standardized set of semantic concepts to be automatically tagged [84, 98], where these concepts are mostly visually or acoustically detectable. However, it is widely acknowledged that an important proportion of semantic concepts in the videos cannot be detected through the AV analysis of the videos alone and that

text cues are proved to be useful, when employed along with the AV features of the videos, in improving semantic video indexing [23, 28, 35, 110]. Hence, a fruitful approach to the problem of automatic semantic video annotation is the employment of video texts such as closed captions, speech transcriptions, sliding texts, and Webcast texts –semi structured natural language texts provided by some sports news sites highlighting important events in sports videos– as an information source for the applicable video domains including news videos and possibly excluding some surveillance videos as pointed out in [30, 47, 65, 78, 90, 97, 110]. For the applicable domains, video texts can usually be obtained through techniques such as automatic speech recognition (ASR), video optical character recognition (VOCR), or sliding text recognition, if not already available as the associated texts. Information extraction from textual data has been extensively studied especially during the last three decades as surveyed in [25, 42, 52, 105] for languages including English, Chinese, Japanese, Spanish, and German. Hence, existing information extraction techniques can readily be applied to video texts to extract semantic information regarding the corresponding videos.

In this study, we propose a text-based fully automated system for semantic annotation[1] and retrieval of news videos in Turkish. The ultimate system exploits several information extraction techniques on the automatically extracted video texts, including named entity recognition, automatic hyperlinking, person entity extraction with coreference resolution, and semantic event extraction. The outputs of the corresponding components are employed as the semantic annotations for the underlying video archives. The system also encompasses several other components to make it a complete automatic semantic annotation and retrieval system: a news story segmenter, a sliding text recognizer for text extraction, a video retrieval interface, and lastly a news video database. The proposed system is fully automated as all of its components including the text extractor operate in fully automated mode. We also propose a semi-automatic counterpart of this system where the sole manual intervention takes place during text extraction which is best applicable to news video archives from which video texts cannot

[1] Throughout the thesis, the words *annotation* and *indexing* are used interchangeably, hence *semantic annotation* and *video annotation* are used interchangeably with *semantic indexing* and *video indexing*, respectively.

be automatically extracted and to those archives for which associated texts are already available. Both of the proposed systems are generic and hence they can be employed for news videos in any language by equipping them with the required components. We build both systems for news videos in Turkish with novel information extraction components for Turkish in addition to generic components applicable to videos in other languages as well, such as the news story segmenter and the video retrieval interface.

The first one of the information extraction components, the named entity recognizer, is a hybrid system for Turkish which is based on a manually engineered system for news texts. The recognizer also has the ability to enhance its information sources through learning from annotated corpora which makes it an extensible hybrid system. After the execution of the named entity recognizer, a subsequent date normalization procedure is carried out to transform possible deictic date expressions output by the recognizer into normalized date expressions. It is known that automatic text extraction techniques from videos such as ASR or sliding text recognition usually result in noisy output. In order to alleviate the effects of the noise in such video texts, the proposed system has an automatic hyperlinking capability through which Web news describing the same event(s) as those in the actual video texts are crawled and aligned with the corresponding news stories. With this capability, the retrieved hypertext data can also be exploited as a more detailed and usually noise-free information source. Utilizing the extracted named entities of type person, location, and organization along with a rule and a string matching based coreference resolution scheme, the system has the capability to extract full person entities together with their occupations, locations/organizations, and aliases from texts. Event extraction is the last technique employed by the overall system which is based on the automatic detection of specific keywords associated with each event type under consideration. To clarify, using an annotated training corpus of news video transcriptions, the most frequent event types and the most frequent keywords for each event type are determined. The ultimate event extraction component makes use of these keywords to extract semantic events, each with a particular confidence value, in unseen news video texts.

Among the other components of the systems, the news story segmentation component basically detects the boundaries of individual news stories using a silence detection procedure on the audio waveforms of the videos. Thereby, the users of the system can access the actual news story segments in addition to full news videos during later video retrieval. The sliding text recognizer [46], as employed by the fully automated system, detects the texts sliding on a text band on the news video frames. This recognizer is executed on the output of the news story segmenter and thereby the news story segments and their corresponding texts are automatically aligned. Next, the resulting story texts are fed into the information extraction components leading to automatic alignment of the semantic information with the story segments within the fully automated system. In the case of the semi-automatic system, the news story segmentation and information extraction procedures are carried out independently and therefore a separate alignment procedure is employed to associate the segments with the extracted information. Production information regarding the raw video files such as their broadcast dates and durations, news story segment information, automatically extracted named entities both in the actual video texts and in the aligned Web texts, full person entity information and coreference information between these entities, together with the extracted events are all stored in a central news video database for later retrieval. Lastly, the semantic video retrieval interface enables access to the underlying news videos and news story segments through boolean queries in which automatically extracted semantic entities and events are utilized as literals. The interface can also act as a natural language interface, if the user chooses it to do so, and thereby semantic queries can well be specified in natural language through the interface.

Considering the previous related work, the studies presented in [30, 47, 78, 90] all utilize ASR tools to extract the texts from the video archives (while the first study considers football videos, the latter three studies target at generic news videos) but ASR is still a hot research topic even for commonly studied languages such as English (the ASR tool employed in [47] reportedly results in an average word error rate (WER) of 29.2%) and this situation leads to considerable error propagation to the other components of the overall semantic

indexing systems. For some languages, including morphologically rich ones such as Turkish, no practical ASR tool exists at all which hinders the applicability of the proposed approaches for video archives in these languages. Similar to the approach presented in [90], the text-based approach in [109] is proposed for the specialized domain of team sports videos and is applied to basketball Webcast texts obtained from a sports news provider site. On the other hand, the semantic indexing systems presented in this study are proposed for the generic domain of news videos. The text extraction procedure employed in the fully automated system is not limited to ASR, while a convenient ASR tool can be utilized in this system if it is available for the language under consideration. If there are other automatic means of text extraction for the considered news video archives such as sliding text recognition (as used in the automatic semantic indexing and retrieval system for Turkish news videos to be described in details in Chapter 4), they may well be employed during the implementation of the system. Moreover, the semi-automatic system can be implemented for news video archives in languages for which no automatic text extraction tool with an acceptable accuracy is available. As a concluding remark on the previous related work, they are implemented for videos in well studied languages of English, German, Dutch, and Italian while our systems are implemented and evaluated for videos in Turkish on which especially ASR and information extraction studies are rare.

1.2 Contributions of the Thesis

Main contributions of this thesis can be summarized as follows:

- Information extraction research on Turkish texts is rare [22]. Within the course of this study, we propose and implement novel techniques for several information extraction tasks including named entity recognition, automatic hyperlinking, full person entity extraction, and semantic event extraction. The implemented components can readily be used on Turkish news texts and can be automatically extended to support other domains (e.g., financial documents and historical texts) as well.

- The information extraction components are successfully exploited for automatic semantic annotation and retrieval of news videos in Turkish. The overall system is also equipped with the other necessary components to make it a practical fully automated system. Some of these components are built from scratch such as the news story segmenter while others are customized and integrated into the proposed system such as the sliding text recognizer. To the best of our knowledge, this is the first fully automated system proposal to facilitate semantic annotation and retrieval of news videos in Turkish.

- As automatically extracted video texts –such as automatic speech transcripts or sliding texts– are often noisy to some degree, the video texts automatically obtained by the sliding text recognizer of the proposed system are noisy as well. Within the course of this study, we evaluate the information extraction components of the system on the noisy and the clean versions of the video texts obtained from a genuine news video corpus where these evaluations constitute the first reports of the performance evaluations of the respective information extraction components for Turkish on such data.

- We also propose a semi-automatic counterpart of the annotation and retrieval system in case the sliding texts are not available for the news videos under consideration. The only manual intervention in this case takes part during speech transcription. However, if the associated texts of the videos are already available, this semi-automatic system can also be executed in fully automated mode as will be clarified in the relevant sections of the study.

- In order to evaluate the performance of the components of the proposed systems, we have compiled and annotated textual and video corpora of considerable size. We have also developed annotation and evaluation tools to be utilized during the evaluation of the components. The lack of publicly available annotated corpora is an important problem which hinders information extraction research on Turkish texts. The annotated corpora created as well as the related tools developed during this study will be

made available to researchers so that sound comparisons of various proposals can be performed.

- By replacing the language-specific components of the proposed systems with components for other languages, these systems can be exploited for other languages as well. For instance, if we want to use the proposed systems for a language, X, silence detection based news story segmenter does not need to be replaced but the named entity recognition component should be replaced with a convenient recognizer for X.

- As an application example, we have also shown the utilization of the semantic annotation and retrieval system for multilingual video retrieval with the incorporation of convenient ASR and machine translation systems.

1.3 Organization of the Thesis

The rest of the thesis is organized as follows: In Chapter 2, relevant literature on information extraction and its employment for semantic video annotation is reviewed. Chapter 3 is devoted to detailed descriptions of the information extraction techniques exploited within the course of this study including named entity recognition, hyperlinking news videos with the related Web news, person entity extraction with coreference resolution, and semantic event extraction. The text-based fully automated semantic annotation and retrieval system proposed for news videos in Turkish is described in Chapter 4 along with an overview of its semi-automatic counterpart. In Chapter 5, evaluation results of various components of the proposed systems and discussions of these results are presented. An application involving the utilization of the information extraction components for multilingual semantic video retrieval is presented in Chapter 6. Finally, Chapter 7 concludes the study and addresses plausible future research directions based on the current study.

CHAPTER 2

RELATED WORK

In this chapter, we first review the relevant literature on textual information extraction and next, on information extraction from Turkish texts. In the last section, we provide an overview of the studies on the employment of information extraction for automatic semantic video annotation.

2.1 Information Extraction

Information extraction (IE) is usually defined as the task of determining important information pieces such as entities, relations, and events in unstructured data sources like free natural language texts and possibly storing the extracted information in predefined fixed formats, very much like database records, thereby making the extracted information more easily processable by the related tasks [39, 41, 42, 52, 61, 81]. Another definition of IE is given in [51] as the creation of stereotypical summaries of the underlying texts. IE differs from information retrieval (IR) in that IR aims at retrieving a set of relevant documents from a large collection of documents, guided by user queries [60, 81, 91, 106]. Nevertheless, since the users still need to examine each of the retrieved documents to determine whether they are relevant or not, it is argued that IE techniques can help improve the retrieval results since it extracts semantic information as opposed to the term based indices maintained and utilized by the IR systems [81]. We leave the discussion on the relationship between IE and IR here by referring interested readers to related studies [60, 81] and continue with our review of the IE literature.

There have been various programs to promote IE research such as Message Understanding Conference (MUC) series [11] and Automatic Content Extraction (ACE) [2]. MUC series were conducted as competitions to promote IE research from 1987 to 1998. The series made considerable contribution to the topic by providing definitions, annotation guidelines, and formal evaluation platforms for various IE tasks in addition to revealing convenient as well as inconvenient approaches to these tasks. Similarly, ACE program has been conducted since 1999 and targets at the development of automatic content extraction technology for natural language texts. Other more specialized programs on IE research [14] include Multilingual Entity Task (MET) conference for Chinese and Japanese [13], Information Retrieval and Extraction Exercise (IREX) for Japanese [94], and Evaluation Contest for Named Entity Recognizers (HAREM) for Portuguese [5]. These programs are significant as they have given rise to related IE research on English as well as on other languages.

There are basically five different IE tasks defined in the MUC series. These tasks are overviewed below following the descriptions in related publications [41, 53]:

1. Named Entity Recognition (NER): It is the task of extracting the names of people, locations, and organizations along with some temporal (time and date) and numeric expressions (money and percent).

2. Coreference Resolution: The task of finding the identity relations between entities. If two entities in a text corefer, it means that they refer to the same real world entity.

3. Template Element Construction: This task involves the extraction of descriptive information for the already extracted named entities and hence filling the templates for these entities where the templates have the necessary slots to be filled.

4. Template Relation Construction: This is the task of identifying relations between the extracted entities.

5. Scenario Template Production: The task of finding the required information regarding specific event scenarios (represented as templates) in texts.

The complexity of the IE task increases as we go over the above list from top to bottom. NER is known to be a solved problem especially for English with state of the art performance above 90%. The best performing NER system in MUC-7 reportedly achieves an F-measure of 93.39% where the performance of human annotators are 97.60% and 96.95% [76]. There also exist studies that propose named entity ontologies [93] so that more detailed types for the extracted named entities can be provided instead of the basic types. The performance rates of the approaches for the remaining four tasks proposed by the participants of the MUC series are comparatively lower. Especially, scenario template production (namely, event extraction) is known to be a difficult task due to the proliferation of the ways in which a particular event can be uttered in natural language [52].

An important point made within the course of the MUC series regarding the IE tasks is that using deep language processing tools like parsers usually yields low performance compared to the employment of shallow processing tools [52]. This result is mainly due to the fact that the output of deep processing tools are usually ambiguous and they require considerable processing time to be employed in practical settings. Similarly, in some studies like [51], it is claimed that successful IE can be achieved without utilizing natural language processing (NLP) tools. Hence, IE is usually classified as a task which does not require full natural language understanding as the types of information that need to be extracted are previously well-defined [49, 60]. In line with this argument, there are several studies that take a pattern matching approach and propose IE systems relying on predetermined or automatically learned IE patterns for structured texts such as Web pages as well as free texts [82]. Proposed pattern based systems include AutoSlog [88], LIEP [56], PALKA [63], RAPIER [36], and SRV [50], among others.

Approaches to address the IE tasks broadly fall into one of the two categories: manually engineered systems and learning systems [25]. There are also studies that try to combine the advantages of these two approaches as well.

As described in [25], manually engineered systems are based on resources such as lexicons, dictionaries, and rule bases, created by the experts of the domain

under consideration. The main drawbacks of this approach are the time spent during resource construction and the need for revisions of the resources when the approach is to be used in other domains, the so called portability problem. In fact, it is this latter drawback that mainly leads researchers to employ learning approaches for IE tasks so that adaptive IE systems can be achieved [105].

Learning systems utilize annotated corpora to automatically create models or learn rules for the domain of interest. Several statistical and machine learning methods have been employed so far to arrive at learning IE systems as surveyed in [105]. These methods include Markov models, maximum entropy models, dynamic Bayesian networks, conditional random fields, hyperplane separators, relational learning (such as inductive logic programming), decision trees, and unsupervised approaches [105]. A well-known unsupervised (or, weakly supervised) learning approach is called bootstrapping method which begins with a seed set of positive examples as its sole resource and iteratively expands the resource with new rules or entries discovered from its own outputs, as employed in studies such as [111]. The main drawback of the learning based IE approaches in general is the cost of compiling annotated corpora for the particular IE task and domain under consideration, which is a labor-intensive and time-consuming process. Although there is considerable amount of annotated text in well studied languages like English for tasks such as NER, unfortunately less studied languages such as Turkish suffer from the unavailability of annotated corpora for the IE tasks which mainly hinders the development of learning IE systems for these languages. Nevertheless, for tasks other than NER, even for languages such as English, there exist comparison and evaluation presentation problems for the various learning approaches employed, as pointed out in [74]. In other words, standard evaluation methodologies to fairly compare the proposed learning approaches for IE tasks are still not well established.

Another direction of research which is closely related to IE research is the emergence of general language engineering frameworks/development environments which aim to facilitate the development of text processing applications. These systems usually supply ready-to-use resources (gazetteers, dictionaries, and grammars, among others) and tools (annotators, evaluation tools, etc.) so

that applications to use the resources/tools are freed from the burden of developing and compiling them from scratch. Well-known development environments of this kind include GATE (General Architecture for Text Engineering) [33, 43], SProUT (Shallow Processing with Unification and Typed feature structures) [48], knowledge and information management platform called KIM for IE and IR applications [87], and TEXTRACT system of IBM which is later migrated to another platform called Unstructured Information Management Architecture (UIMA)[85]. Another related tool is the Voice Transcription Manager (VTM) of IBM again, which facilitates the transcription of voice and the display of the output in several applications by incorporating convenient ASR tools [6].

2.2 Information Extraction in Turkish

IE research on Turkish texts is known to be rare compared to related work for European languages such as English, Spanish, French in addition to Chinese and Japanese. To the best of our knowledge, the first study on the topic is reported in [40] where a language-independent named entity recognizer is proposed and evaluated on Turkish texts along with other texts in Romanian, English, Greek, and Hindi. A study on Turkish noun compounding is presented in [32] where a finite state machine is proposed to extract compounds and the machine can readily be utilized to extract some complex organization names which are in compound form. A statistical IE system for Turkish is presented in [104] which performs various tasks including NER, sentence segmentation, and topic segmentation. It is pointed out in [104] that statistical methods are not directly applicable to Turkish since the highly productive morphology of the language causes data sparseness problems. A person name extractor for financial news texts is proposed in [31] based on the determination of local patterns for extracting person names. The employed local grammar based approach [31] has been proved to be useful for other languages including English [100]. A person mention extractor for Turkish news texts using a set of lexical resources is presented in [67] together with a string matching based coreference resolver to prevent superfluous extraction of the same underlying entities. A rule based NER system

is described in [69, 70] for Turkish news texts, based on a set of lexical resources and sets of rules. The latter NER system is turned into a hybrid system in [71] by equipping it with a learning component so that it extends its resources through learning from annotated corpora if available. The last four studies are carried out within the scope of this thesis. In [22], the authors present an IE architecture for processing business documents in Turkish which is best applicable to restricted domains such as financial documents. Finally, in [38], the authors present their new event detection and topic tracking experiments on Turkish texts. Mainly, a topic detection and tracking (TDT) document collection, called BilCol-2005, is described and various parameters are experimented for the aforementioned two TDT tasks. The cosine and cover coefficients similarity measures are utilized for both tasks along with three stemming options of no stemming, fixed prefix stemming, and lemmatizer-based stemming. Also experimented in [38] is the combination of these similarity measures which yields better results than the individual employment of the measures.

As for information retrieval (IR) research in Turkish, a plausible survey of text-based IR approaches is provided in [37]. Also described in the paper [37] are the results of several IR experiments to investigate the effects of various parameters such as stemming options, collection size, query lengths, matching functions, and using stopword lists on Turkish news texts.

2.3 Information Extraction for Semantic Video Annotation

Video texts are known to constitute an important information source for semi-automatic or automatic video annotation [65, 97, 110]. Mainly following the classification of video texts provided in [110], these texts usually belong to one of the following categories:

1. *Associated Texts (Closed Captions)*: For some video types such as broadcast news videos, it may be possible that the complete speech transcription texts are already available for the corresponding videos. If detailed enough, associated texts are preferable to other types of video texts since most of

the time they are perfect, i.e., no noise exists in the texts. However, if these texts are not time aligned with the videos, then a separate alignment procedure is required possibly utilizing techniques such as text-to-speech alignment.

2. *Automatic Speech Transcripts*: In most of the video types including broadcast news and possibly excluding some surveillance videos, people convey the topic through speaking and hence for such videos speech transcriptions can be utilized for semantic annotation. If the speech transcriptions are not already available, they could be obtained utilizing ASR systems. Successful ASR systems are hard to build and an important problem regarding the ASR outputs is that they are noisy, hence when utilizing ASR output, this point should be taken into account. Moreover, for some languages including Turkish, ASR is still an important research area and it is not currently possible to access and integrate a practical ASR system as a component to larger video annotation systems.

3. *Overlay and Scene Texts*: On images and video frames, texts of nonuniform size and font which are mechanically superimposed are usually termed as overlay texts (also as *graphic texts* or *text overlays*) while texts already existing on the real-world objects or scenes in the images/frames are usually called *scene texts* [113]. They are often detected through VOCR methods. These texts may be useful in some video domains such as football or basketball videos to display important information such as the current score and time, or the statistics regarding the goals or points. However, for other domains such as broadcast news videos, their contribution to the semantic interpretation of the videos may be limited if not misleading.

4. *Sliding Texts*: Especially for some broadcast news videos, it may be possible to have exact speech transcriptions regarding the news stories as uniform sliding texts along a text band on the video frames. Sliding text recognition systems which are based on visual processing of the video frames may achieve higher success rates compared to the ASR systems, as these texts usually have uniform size and font, particularly if the videos have satisfactory resolution. However, similar to the ASR systems, these

recognizers cannot produce perfect output, i.e., the recognized text is usually noisy. Still, an important proportion of the errors made by the sliding text recognizers leads to out-of-vocabulary (OOV) words which can be automatically corrected using tools such as spelling checkers as opposed to the errors in the outputs of the ASR systems.

5. *Webcast (Webcasting) Texts*: These are semi-structured natural language texts mainly describing the flow of events in sports matches provided by some sports sites. Webcast texts are time-aligned and they are easier to process than free natural language texts due to the relative simplicity of the language used in these texts.

6. *Production Texts*: As described in [110], these texts include production metadata such as the titles and broadcast dates of the videos under consideration along with other published information.

IE from video texts to automatically annotate the corresponding video archives is a relatively new topic with some recent studies. Among these, in [44, 90] an architecture is proposed for multimedia indexing through the employment of multilingual IE to annotate semantic events in football videos in English, German, and Dutch. The proposed architecture utilizes ASR tools for these three languages to obtain the speech transcriptions of the videos. The IE components of the architecture include tools for unicode tokenizing, gazetteer lookup, semantic tagging, part-of-speech tagging, parsing, knowledge coding and processing, and coreference resolution. In [30], a system that extracts semantic metadata from TV and radio broadcast news, implemented for Italian over the data from the TV channel RAI, is presented. This system also employs an ASR tool to obtain the video texts, categorizes the individual news items, and acquires semantic metadata for the news items employing IE techniques, such as NER, both on the video texts and on the aligned Web texts [30]. A system, called Rich News, which aims to perform automatic semantic annotation of radio and television news is presented in [47]. The overall system has modules for the following sequential tasks: ASR, topic segmentation, key-phrase extraction, searching the Web for related documents using the key-phrases, manual

annotation to correct segmentation and annotation results, creating the story index documents, semantic annotation of the retrieved Web resources, and finally searching the broadcast news [47]. In [114], a semantic event detection approach for broadcast basketball videos is presented. The approach is based on the analysis of the Webcast texts and the alignment of the texts with the corresponding videos. The Webcast texts for sports videos are analyzed to cluster and detect semantic events using latent semantic analysis (LSA)[1] [45, 72] and it is concluded in the paper that incorporation of the video texts into the event detection process significantly enhances the sports video event detection [114]. The same approach is utilized for event extraction from broadcast sports videos in [109], this time by using probabilistic LSA (pLSA) [55] for clustering Webcast texts. Finally, an approach for the aggregation and retrieval of cross-modal (coming from multiple information sources) multimedia (consisting of multiple types of content) information is presented in [78]. The information sources utilized are online newspaper articles and TV newscasts. The core technologies employed in the implemented prototype system include newscasts detection and segmentation, RSS stream processing, RSS items and news stories aggregation, and derived queries generation [78].

Related work on Turkish texts include previously mentioned [67] where coreferential chains are exploited to prevent superfluous extraction of person entities from political news video texts and [68] where named entities automatically extracted from manually transcribed news video texts are utilized as semantic annotations for the videos. These two studies are carried out within the scope of this thesis. Apart from these, in [26], the challenges of Turkish broadcast news transcription and spoken term detection are addressed and in [103], a rule based event extraction approach for Turkish football Webcast texts is presented.

[1] LSA is a technique that analyzes the associations between a set of documents (or passages) and the terms that they contain [9]. It has many application areas including IR and document classification. Interested readers are referred to [45, 72] for in-depth descriptions of LSA.

CHAPTER 3

INFORMATION EXTRACTION TECHNIQUES

In this section, we describe our proposals for four IE tasks: named entity recognition, automatic hyperlinking of news videos with related Web news, person entity extraction with coreference resolution, and finally event extraction. Novel tools are designed and implemented based on the proposed approaches and these tools are exploited in the fully automated and semi-automatic semantic annotation and retrieval systems for news videos in Turkish to be described in Chapter 4. The evaluation results of these tools are presented in Chapter 5.

3.1 Named Entity Recognition

Named entity recognition (NER) is an important IE task. As previously defined in Section 2.1, it is the task of extracting person, location, and organization names in addition to date, time, money, and percent expressions from texts [83]. The task is also commonly referred to as *named entity recognition and classification (NERC)* [83].

Within the course of this thesis, we first build a rule based NER system for the news domain which heavily relies on a set of lexical resources and pattern bases. Although news texts show considerable diversity and hence most of the resources of the rule based NER system are considerably general-purpose, the system suffers from portability problems, i.e., its performance is considerably hurt when it is evaluated on other text genres such as financial or historical texts. Moreover, the system lacks the ability to keep its already existing resources up-to-date for

the news domain. To address these problems and make the system an extensible one, we improve the system so that it learns from annotated corpora when available and extends its resources accordingly. We implement this ultimate hybrid NER system for Turkish so that it can be employed to extract named entities from Turkish news video texts to be utilized as semantic annotations for the corresponding videos. In order to handle deictic date expressions extracted, a subsequent date normalization procedure is carried out.

In the following subsections, we first describe the core rule based NER system for Turkish, then its ultimate hybrid version is presented, and lastly, the date normalization procedure employed is briefly described.

3.1.1 The Rule Based Named Entity Recognizer

It is a common practice for rule based named entity recognizers to make extensive use of person and organization name lists, gazetteers, along with rule bases or grammars to extract the entities from the input texts. Some of these resources are usually generic such as gazetteers or but most of the time they are proprietary to the domain of interest which enables them to achieve high success rates in the particular domain. In line with this observation, the rule based named entity recognizer for Turkish utilizes a set of lexical resources and pattern bases as two information sources for the extraction of person/location/organization names as well as time/date/money/percent expressions in Turkish news texts.

Turkish is an agglutinative language and hence inflections are very common in Turkish texts. Since named entities may well be inflected, a morphological analyzer for only noun inflections is implemented and utilized by the recognizer to validate the candidate entities.

The input text of the named entity recognizer may be all in upper case or all in lower case, if not, case information may be inherently missing in the input text as is the case in texts obtained from the Web or the outputs of ASR tools. Therefore, we implement the utilization of capitalization information as an additional feature which may be turned on or off. If it is turned on, then

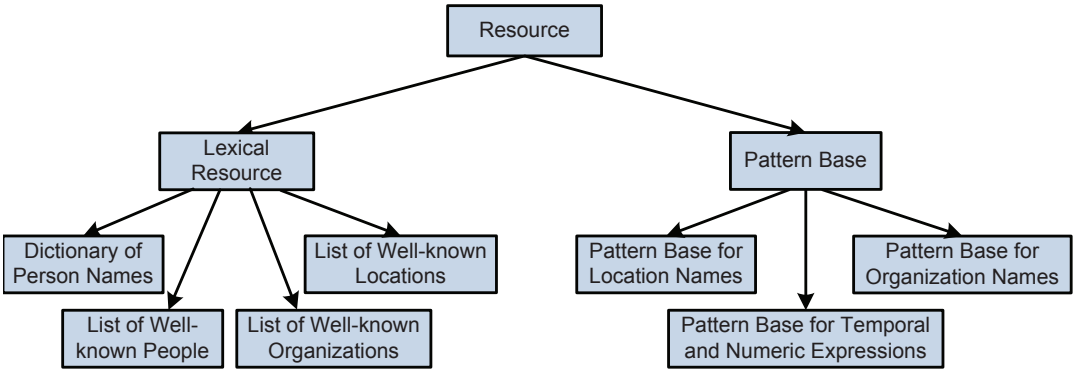


Figure 3.1: The Taxonomy of the Resources Employed by the Rule Based Named Entity Recognizer.

those entity candidates of type person, location and organization name which do not have the initial characters of all of their tokens capitalized are eliminated.

The lexical resources employed by the recognizer comprise a list of person names and lists of well-known people, locations, and organizations. As for the set of pattern bases employed, they contain rules for the extraction of location/ organization names, and temporal/numeric expressions. The total number of entries in the lexical resources is about 12,800 and the total number of patterns in the corresponding pattern bases is 260. We arrive at these resources after examining several sample news articles and try to make their coverage as high as possible. The taxonomy of the resources are given in Figure 3.1. A summary of these resources are provided below where the first four of them correspond to the lexical resources and the remaining three resources are the pattern bases.

1. *List of Person Names*: We employ a set of about 8,300 person names in Turkish so that consecutive tokens each of which is included in the set can be extracted as person names, such as *Abdullah Cevdet*. When employing this resource and the resources to be described thereafter, the morphological analyzer is utilized to validate the information to be extracted. To clarify, when a token in the input text matches one of the entries in the list, the suffixes attached to the entry to form this token are checked by the morphological analyzer. If the token is a valid noun inflection, then this token is validated, otherwise the token is not considered for extraction.

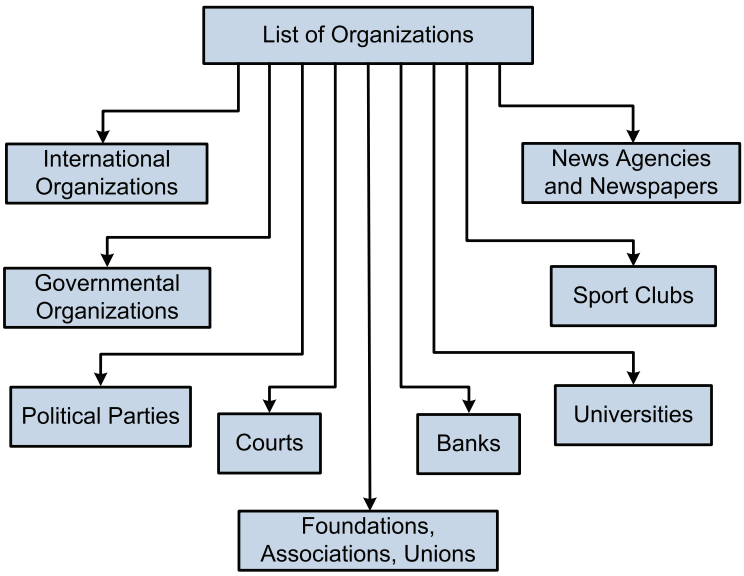


Figure 3.2: The Types of Organizations Included in the *Well-known Organizations List* of the Named Entity Recognizer.

2. *List of Well-known People*: The names of well-known past and present politicians and leaders of some international organizations are compiled and utilized including items like *Recep Tayyip Erdoğan* (current prime minister of Turkey) and *Kofi Annan* (past president of the United Nations).

3. *List of Well-known Organizations*: In this resource, the names of important organizations such as those of political parties, governmental/ international organizations, universities, and banks are included. *Avrupa Birliği* ('*European Union*') and *Adalet Bakanlığı* ('*The Ministry of Justice*') are two example organizations from this resource. The classification of the organization name types included in this resource is provided schematically in Figure 3.2.

4. *List of Well-known Locations*: This resource encompasses some well-known locations in the world and in Turkey. The classification of the locations covered by this resource is demonstrated schematically in Figure 3.3. The types in Figure 3.3 constitute the most common location name types that we come across during the resource compilation for the recognizer.

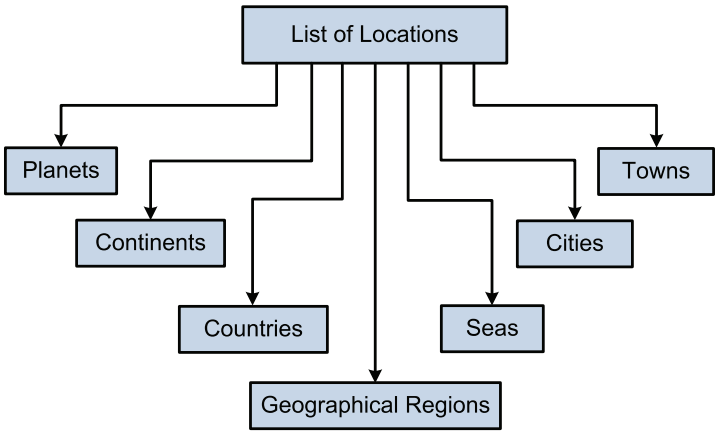


Figure 3.3: The Types of Locations Included in the *Well-known Locations List* of the Named Entity Recognizer.

5. *Pattern Base for Organization Names*: This pattern base includes several patterns for the extraction of organization names. A sample pattern (or, rule) included in this resource is *X Üniversitesi* ('*X University*') where X is the immediately preceding named entity extracted using the lexical resources only if such a named entity exists, and the preceding single token otherwise. An organization name that can be successfully extracted with this pattern is *Süleyman Demirel Üniversitesi* where *Süleyman Demirel* is actually a person name which can be extracted using the list of person names employed by the recognizer, but with the succeeding token *Üniversitesi*, the whole phrase matches the aforementioned pattern and discarding this person name, the recognizer extracts the whole phrase as an organization name. Common organization names to be covered by the included patterns are mostly those organization types for which enumerating all entries is not feasible. Hence, the included patterns aim to extract the names of companies, institutions, faculties, schools, hospitals, prisons, labor/trade unions, sports clubs, police offices, headquarters, and courts. Other sample patterns are given in (1) below:

   (1) *X Grubu/A.Ş./Partisi/Hastanesi/...*
   '*X Group/Inc./Party/Hospital/...*'

6. *Pattern Base for Location Names*: This resource includes several patterns

for the extraction of location names such as the pattern exemplified for the organization name extraction. Similar to the previous resource, the location name patterns mainly aim to cover those not-easily-enumerable/ changing names such as street, road, bridge, river, mountain, station, hotel, tower, palace, museum, campus, airport, shopping/cultural center, and cemetery names. A sample pattern for location name extraction is *X Bulvarı* ('*X Boulevard*'), where X is determined as in the case of the organization name extraction utilizing the corresponding patterns. A case which can be handled appropriately by this pattern is *Adnan Menderes Bulvarı* ('*Adnan Menderes Boulivard*') where *Adnan Menderes* is a person name identifier which can be easily identified by the proper person name dictionary of the lexical resources. Yet, with the following token, *Bulvarı*, the complete identifier satisfies the pattern and hence it is correctly classified as a location name by the named entity recognizer. Similar location patterns employed are provided in (2) below:

(2) *X Sokak/Yolu/Kulesi/Stadyumu/...*
'*X Street/Road/Tower/Stadium/...*'

7. *Pattern Base for Temporal and Numeric Expressions*: This resource includes several rules for the extraction of time, date, money, and percent expressions. A sample and common pattern for time extraction is *X:Y* where *X* and *Y* are numeric expressions and *X* denoting the hour part takes values between 0 (or equivalently, 00) and 23 while *Y* denoting the minute part takes on values between 0 (or equivalently, 00) and 59. It should be noted that deictic date expressions like *bugün* ('*today*'), *dün* ('*yesterday*'), and *yarın* ('*tomorrow*') are also within the scope of the recognizer. Below provided are some date extraction patterns utilized where *X* can be a four digit year name as well as the name of a month.

(3) *X başı/ortası/sonu...*
'*X start/middle/end...*'
'*The start/middle/end... of X*'

In order to increase the precision of person name extraction, which is initially performed through bare list lookups, we add several rules to the above described resources to handle some common surname endings in Turkish and hence correctly classify those satisfying surnames as parts of person names. Some of these common surname ending words and affixes include *oğlu*, *-gil*, and *soy*. This improvement leads to complete and accurate extraction of name-surname pairs such as *Murat Başesgioğlu* and *Sabiha Şensoy* as person names by the recognizer. We also employ a similar strategy to improve the extraction of organization names and hence utilize common endings such as *bank* for the extraction of bank names such as *Akbank* and *spor* for that of the names of sport clubs such as *Gaziantepspor* in addition to above described pattern base for organization name extraction.

The final forms of the resources and pattern bases utilized by the rule based named entity recognizer are made available at `http://www.ceng.metu.edu.tr/~e120329/TurkNERRes.zip` under the Lesser General Public License for Language Resources (LGPL-LR), `http://www-igm.univ-mlv.fr/~unitex/lgpllr.html`.

The details of the rule based NER algorithm are provided in Algorithm 1. This algorithm can be adapted to other morphologically rich languages like Turkish by supplying it with language-specific lexical resources and pattern bases along with a convenient morphological analyzer for the language under consideration. In order to determine the computational complexity of the overall algorithm, first we assume that the sizes of individual elements in the lexical resources and the pattern bases are $O(1)$ in terms of the number of tokens. We also assume that the complexity of morphological analysis of a single token is $O(1)$. Hence, if the size of the input text is $O(n)$ in terms of the number of tokens and if the sizes of the lexical resources and the pattern bases are $O(l)$ and $O(p)$ in terms of the number of elements, respectively, then the computational complexity of the proposed algorithm is $O(n.l) + O(n) + O(n.p) = O(n.(l + p))$, where $O(n.l)$ is the complexity of lines 1–8, $O(n)$ is that of line 9, and lastly, $O(n.p)$ denotes the complexity of lines 10–17.

Algorithm 1 NAMED ENTITY RECOGNITION

Require: Input text *input*, lexical resources and patterns bases.

Ensure: The version of *input* in which named entities are annotated.

1: **for all** element *elt* in the lexical resources **do**
2: **for all** phrase *phr* in *input* matching *elt* **do**
3: *phr* is morphologically analyzed.
4: **if** *phr* is only inflected with noun suffixes (or not inflected at all) **and** *phr* is not already annotated **then**
5: annotate *phr* in *input* with the corresponding named entity, excluding the suffixes at the end, if any.
6: **end if**
7: **end for**
8: **end for**
9: Scan through the resulting text and merge consecutively annotated entities (where the entities apart from the last one should not be inflected) of the same type into single entities.
10: **for all** element *elt* in the pattern bases **do**
11: **for all** phrase *phr* in *input* matching *elt* **do**
12: *phr* is morphologically analyzed.
13: **if** *phr* is only inflected with noun suffixes (or not inflected at all) **and** *phr* is not already annotated **then**
14: annotate *phr* in *input* with the corresponding named entity, excluding the suffixes at the end, if any.
15: **end if**
16: **end for**
17: **end for**

The recognizer utilizes the Standard Generalized Markup Language (SGML) named entity tags (`ENAMEX`, `NUMEX`, and `TIMEX`), proposed and employed in the MUC series [53], to annotate the named entities. Below we provide a sample input (a news text snippet from METU Turkish corpus [92]) and the corresponding output of the recognizer, where the English translation of the input reads as follows: *The committee of Turkish Industrialists' and Businessmen's Association*

(TÜSİAD), who has started a Europe tour in order for Turkey to receive a date for membership negotiations at the Copenhagen Summit to be held in December, has received support at their first stop Athens.

Recognizer input:
Aralıkta yapılacak Kopenhag Zirvesi'nde üyelik müzakereleri için Türkiye'ye tarih verilmesi amacıyla Avrupa turuna çıkan Türkiye Sanayicileri ve İşadamları Derneği (TÜSİAD) heyeti, ilk durağı Atina'da destek buldu.

Recognizer output:
<TIMEX TYPE="DATE">Aralık</TIMEX>ta yapılacak <ENAMEX TYPE="LOCATION">Kopenhag</ENAMEX> Zirvesi'nde üyelik müzakereleri için <ENAMEX TYPE="LOCATION">Türkiye</ENAMEX>'ye tarih verilmesi amacıyla <ENAMEX TYPE="LOCATION">Avrupa</ENAMEX> turuna çıkan <ENAMEX TYPE="ORGANIZATION">Türkiye Sanayicileri ve İşadamları Derneği</ENAMEX> (<ENAMEX TYPE="ORGANIZATION">TÜSİAD</ENAMEX>) heyeti, ilk durağı <ENAMEX TYPE="LOCATION">Atina</ENAMEX>'da destek buldu.

In order to make the recognizer extensible to new text genres, annotated data can be utilized when available so that its information sources can be automatically extended to support these genres. Moreover, the performance of the recognizer on its target genre can also be improved by extending its information sources with new high-confidence entries. In the following section, the extended hybrid version of the rule based recognizer to address these issues is presented.

3.1.2 The Hybrid Named Entity Recognizer

It is widely known that named entities in different genres of text show considerable diversity. For instance, political news texts usually replete with the names of countries, political parties, governmental institutions, and politicians. On the other hand, frequent named entities in financial texts are usually company

names, the names of the heads of these companies as well as that of the governmental institutions. Historical texts mostly include the names of empires, past governments in addition to the names of the royal family members. As a last example, child stories may contain uncommon names for locations and people which are rarely observed in texts about current events such as daily news texts. Usually, locations names seem to demonstrate the least diversity in texts of different genres. Furthermore, a named entity of a certain type in a specific text genre may be a named entity of another type in another text genre. For instance, *Pınar* is a common female first name in generic Turkish texts, however, in financial texts it mostly corresponds to the name of a food company (an organization). Similarly, *Selçuk* is a common male first name in Turkish texts, however, in a historical text, it usually denotes a dynasty (an organization), '*the Seljuks*', ruled in the Middle East from the 11th to the 13th centuries.

Above exemplified diversity of the named entities in texts of different genres results in the observation that the rule based named entity recognizers manually engineered for specific genres usually require manual revisions –which are usually time-consuming and labor-intensive– to adapt the recognizers to other genres of text. But as outlined in [25], rule based systems usually achieve higher success rates for the specific text domain that they are engineered for. On the other hand, learning systems are preferable in cases where a considerable amount of training data is available since they do not require human intervention and hence they are easily extensible to other application domains. With an intention to combine the advantages of the two approaches and make the existing rule based named entity recognizer for Turkish news texts described in the previous section to support other text genres as well, we extend it to learn from annotated corpora when available.

The ultimate hybrid named entity recognizer has the ability to enrich its lexical resources with those that it learns from annotated texts through rote learning. Rote learning is one of the learning approaches experimented to improve IE in [51] where it turns out to yield high precision rates. A rote learner, as described in [51], is a dictionary learner which estimates the probability that a matched fragment is a field instance by calculating the number of times the fragment

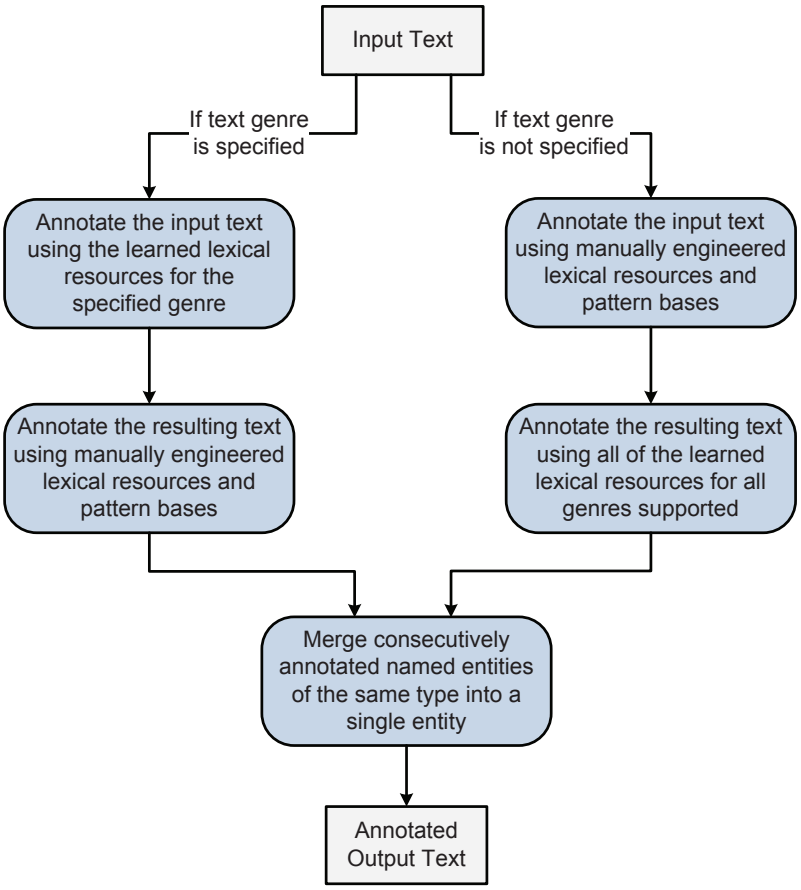


Figure 3.4: The Flow of Execution of the Hybrid Named Entity Recognizer.

appears as a field instance (p) and the total number of occurrences of the fragment (t) where the value p/t is the confidence of the fragment in a prediction. Our rote learning component, as adapted from [51], scans and extracts named entities from annotated corpora and calculates p and t for each extracted entity where, in our case, p denotes the number of occurrences which are annotated as a named entity of the given type and t denotes the total number of occurrences of the entity in the text. Hence, for the annotated training text under consideration, p/t for each extracted named entity can be used as a confidence value for that entity.

We turn the initial recognizer into a hybrid recognizer by equipping it with this learner and with the ability to enrich its information sources with additional lexical resources encompassing those entities which have a confidence value above 0.5, as acquired by the newly added rote learner component.

The flow of execution of the resulting hybrid recognizer is presented schematically in Figure 3.4 and summarized below:

1. If the text genre of the input is specified by the user, then the recognizer first uses the learned lexical resources for that genre to extract the named entities. It annotates the input text with the matching entities and then it annotates the resulting text using the initial lexical resources and pattern bases, i.e., those manually engineered information sources belonging to the rule based recognizer, to carry out a second annotation. Hence, the learned resources for the specified genre are given higher priority.

2. If the input text genre is not specified, the recognizer first executes exactly the same as the rule based recognizer on the text. Then, it employs all learned resources to annotate the resulting text. Hence, in this case, the learned resources are given lower priority. During the annotation process, the morphological analyzer is utilized to validate the candidate entities.

3. As the last step of execution, each set of consecutively annotated entities of the same type (with the constraint that the initial and the intermediate tokens are not inflected) are merged into a single entity. This final step is required since tokens of a single named entity may be annotated individually during the employment of the learned resources and the employment of the manually engineered resources as they are sequentially but independently applied to the input text. To illustrate, assume the input genre is not specified and the second branch of execution is followed which has led to the annotation of only the first name of the person entity during the employment of manually engineered resources followed by the annotation of the surname of the same person entity using the learned resources. The final merging stage prevents the erroneous extraction of this single entity as two partial entities by merging them into a single person entity since they are of the same type and are consecutively annotated in the text.

The computational complexity of the hybrid named entity recognizer (not considering the training procedure) is $O(k_1.n+k_2.n+n)$ where $O(k_1.n)$ corresponds to the complexity of applying the manually engineered information sources to

the input text, $O(k_1)$ being the size of these information sources in terms of the number of entries included and $O(n)$ being the size of the input text in terms of the number of tokens; $O(k_2.n)$ corresponds to the complexity of applying the learned resources to the resulting input text, $O(k_2)$ being the size of these learned resources; and lastly $O(n)$ corresponds to the complexity of the final merging procedure. Therefore, the overall complexity of the hybrid named entity recognizer is $O((k_1+k_2+1).n)$, hence $O((k_1+k_2).n)$. The complexity of the training procedure of the recognizer is $O(m)$, m being the size of the annotated training corpus, as this corpus is scanned to extract and group the annotated named entities and calculate the corresponding confidence values.

The hybrid recognizer is envisioned to support the text genres of financial news texts, historical texts, and child stories in addition to the initial genre of generic news texts. These are the genres for which we could compile and annotate text corpora since we know of no publicly available annotated corpora for named entity recognition on Turkish which is one of the main hindrances against related research since it makes sound comparisons of related experiments impossible.

We also implement a graphical user interface for the proposed hybrid recognizer in Java as shown in Figure 3.5. Through the interface, the users may specify the input text genre among the supported genres of *Child Story*, *Financial News Article*, *Historical Text*, and *News Article*. Also adjustable through the interface are the named entity types to be considered during recognition and whether the capitalization feature will be utilized or not.

3.1.3 Date Normalization

Regarding the date expressions extracted by our named entity recognizers, since deictic ones cannot be considered as semantic information as they are, we implement a date normalization procedure which is applicable to those texts for which the publication date (to be used as a reference date) is known. To illustrate, for news stories, we know the actual broadcast dates of the stories which can be utilized as the reference dates. Based on the reference dates of the prospective inputs, our date normalization procedure executes as follows:

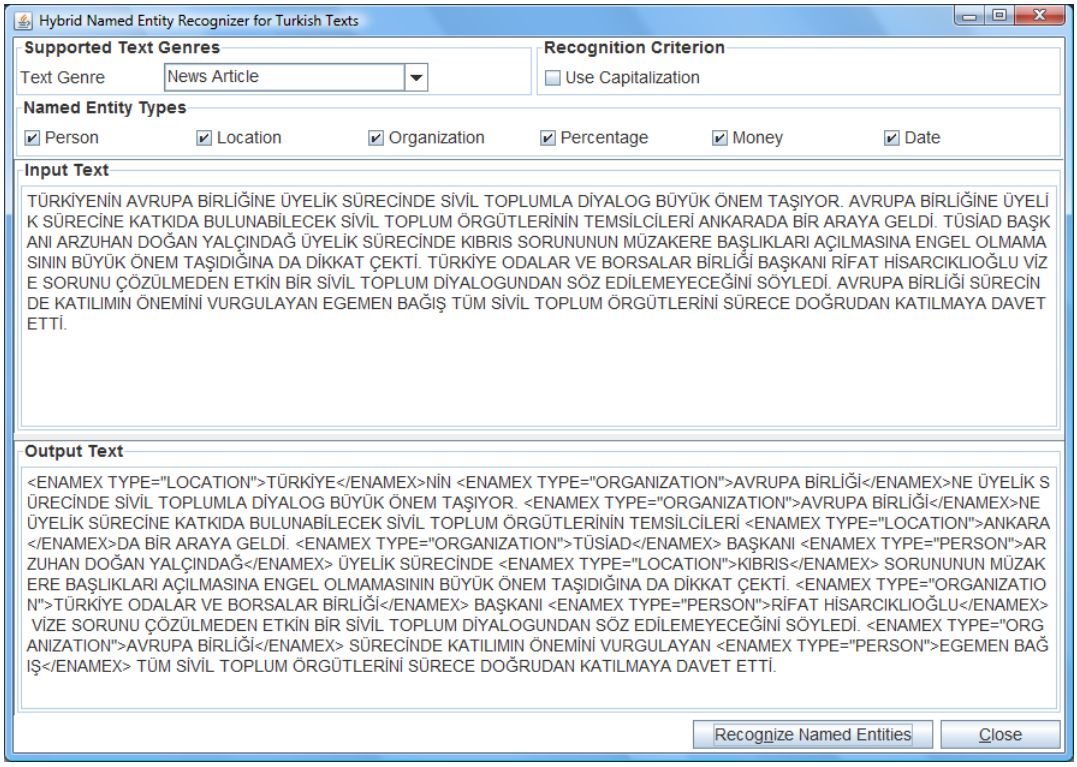


Figure 3.5: A Snapshot of the Hybrid Named Entity Recognizer Interface.

1. Input text is explicitly associated with its reference date, although the reference date may not actually be mentioned in the text.

2. The deictic date expressions are replaced with the corresponding actual date expressions which are calculated with respect to the reference date. For instance, if the extracted date from the text is *bugün* ('*today*'), then the text is not associated with any new date expression, as it is already associated with the reference date but if the extracted date is *yarın* ('*tomorrow*'), then the text under consideration is associated with the date corresponding to the following day of the reference date.

3.2 Hyperlinking News Videos with Related Web News

In several studies on multimedia indexing and retrieval [30, 47, 78], it is pointed out that aligned Web texts can be a useful information source to compensate for the noise in the existing information sources like ASR outputs and to enrich

the already existing semantic information since news articles are usually more detailed than the corresponding broadcasted news. Following this idea, we use Web news articles from a leading Turkish newspaper, *Milliyet* [12], to improve the automatic annotations for the Turkish news videos which are extracted from the corresponding video texts. The process of automatic hyperlinking of news videos with the related Web news articles is also referred to as *Web alignment* and is similar to the *story link detection* task which is one of the main TDT tasks [24].

First, we implement a search crawler to run on the Web site of *Milliyet* for a specified time window. The search criteria include the named entities automatically extracted by the named entity recognizer and associated with their news story segments. The time window corresponds to a three-day interval centered at the actual broadcast date of the news video that the segment under consideration belongs to. The flow of execution of the implemented crawling mechanism for each news story segment is summarized below:

1. The entries in the associated named entity list are searched for in the news articles published in the Web site of *Milliyet* on the broadcast date of the news video under consideration, following the hyperlinks in a breadth-first manner, on the previous and following days, in this order. The Web news articles published by *Milliyet* are available from its homepage, the uniform resource locators (URLs) of which are formatted based on the publication date, as provided below:

   `http://www.milliyet.com.tr/YYYY/MM/DD`

   In the above pattern, YYYY denotes the year, MM denotes the month, and DD denotes the day of the publication date, as exemplified below:

   `http://www.milliyet.com.tr/2007/04/02`

   During crawling, the starting URL is determined utilizing this pattern. Hence, if the date of consideration is April 2nd, 2007, then the above URL is used as the starting URL.

2. For each page (p) in which at least one of the named entities (NEs) mentioned in the corresponding video segment (s) is encountered, a confidence value is calculated as follows:

$$confidence_{s,p} = \frac{number\ of\ NEs\ of\ s\ encountered\ in\ p}{total\ number\ of\ NEs\ in\ s} \tag{3.1}$$

3. All pages with less than 0.5 confidence are eliminated.

4. Among the remaining pages, those ones with confidence less than the maximum confidence are eliminated.

5. Among those pages having different URLs but which happen to have exactly the same content, all but one are eliminated.

6. The remaining pages (which have the same confidence value) are sorted on their detection times and only the top three pages are retained. Since crawling is performed in the broadcast date, previous day and following day order, the pages published on the actual broadcast date are given precedence over the others. If there are already less than three pages, no elimination is carried out.

One drawback of the above procedure is that the crawling process does not make use of any indices or any of the existing libraries, crawling all the new links in the encountered pages, which amounts to a considerable time spent during crawling. This is the case even though we have limited ourselves to crawling a maximum of 1,000 distinct pages, i.e., the crawling for each news video segment is ceased once 1,000 news pages are analyzed.

With this overall automatic hyperlinking procedure, the following features could be integrated into the ultimate news video annotation and retrieval system:

1. The semantic information extracted from the aligned Web articles can be used as semantic annotations as well. Put another way, both the information extracted from actual video texts and those extracted from the Web texts could be used during prospective video retrieval.

2. Aligned Web articles associated with the news videos or video segments can be accessed during news video retrieval in addition to the other required features such as playing the videos or video segments.

3.3 Person Entity Extraction

The NER procedure results in the determination of individual entities of type person, location, and organization, among others. But this information does not cover possible associations between these entities, such as the *managerOf* relation between a person and an organization entity. Moreover, different surface forms of the same underlying entities result in superfluous extraction of these entities as distinct ones without explicit associations between them. In order to alleviate these problems, we implement a slightly modified version of the person mention extraction approach together with the coreference resolution scheme that we previously proposed [67] and utilize their outputs in the ultimate semantic news video annotation and retrieval systems.

To address the first problem, we use the following rule given as a regular expression, similar to the rule given in [67], to extract person entities.

$$((ORG \cup LOC \cup \epsilon)(OCC \cup \epsilon))^{\star} PER$$

The following abbreviations are utilized in the rule: `ORG`, `LOC`, and `PER` for named entities of type organization, location, and person, respectively, and `OCC` for occupation names in Turkish. In order to detect occupation names, we have compiled a set of 136 entries of common occupations encountered in news texts. It should be noted that before searching for entities matching the rule, stopwords in the input text are eliminated. We have extended the stopword list provided in [37] and used this final list provided in Appendix A during all required stopword list elimination processes hereafter.

A text snippet satisfying the above rule is provided below together with the corresponding person template (`Person_1`) filled with the extracted information[1]:

[1] In the text snippet given, POSS stands for the possessive case marker.

```
Maliye  bakanı        Kemal Unakıtan
Finance minister-POSS Kemal Unakıtan
'The minister of Finance Kemal Unakıtan'

Person_1:
Organization: Maliye
Occupation:   bakan
Name:         Kemal Unakıtan
```

The above rule is simpler than the one provided in [67] as the approach in [67] lacks a full-fledged named entity recognizer which leads it to mimic one for political person name extraction, with the more complex rule given below:

$$((C_\epsilon)(GC_\epsilon)(T_\epsilon)(GC_\epsilon)(W_\epsilon)(GC_\epsilon)(P_\epsilon)(PC_\epsilon))^\star((P(PC_\epsilon))^+ \cup N^+ \cup (P(PC_\epsilon))^+N^+)$$

where GC_ϵ is the regular expression $(GC \cup \epsilon)$ and GC is the genitive case marker in Turkish while PC_ϵ is the expression $(PC \cup \epsilon)$ where PC denotes possessive case marker in Turkish. Similarly, the expressions C_ϵ, T_ϵ, W_ϵ, and P_ϵ denote $(C \cup \epsilon)$, $(T \cup \epsilon)$, $(W \cup \epsilon)$, and $(P \cup \epsilon)$, respectively. C is for possible country/continent, T is for city/town, W is for well-known institution, P is for political status, and lastly N is for proper person names in Turkish. The Kleene star $(\star)$ is used to handle the cases where a politician could have multiple status for different countries, cities, or institutions. The subexpression, $((P(PC_\epsilon))^+ \cup N^+ \cup (P(PC_\epsilon))^+N^+)$, covers those politicians who are stated with their political status only, names only, or together with their political status and names in the text [67].

Coreference is a phenomenon in natural language texts where two entities refer to the same real world entity [79, 80]. The determination of those coreferring entities is usually called coreference resolution. A related linguistic phenomenon is anaphora where an entity refers back to another entity in text but they do not necessarily refer to the same real world entity. The referring entity is often called an anaphor while the entity referred to is often called the antecedent, and hence finding the antecedent of an anaphor is called anaphora resolution. Pronouns are known to constitute one of the most common types of anaphors in natural language texts. Interested readers are referred to [79] for details on

the anaphor types and approaches to anaphora resolution in English. There are quite few studies conducted on anaphora in Turkish texts. Among these, in [66] a knowledge-poor pronoun resolution system is presented which relies on a set of constraints and preferences to pinpoint the antecedents of considered pronouns. A syntax-based approach for pronoun resolution in Turkish is described in [101] and a comparison of this approach and the previous one is provided in [102]. Several machine learning approaches are also employed for pronoun resolution in Turkish and these approaches are compared in [62]. Apart from these studies, a string matching based coreference resolution scheme is described in [67] along with the above summarized rule based person mention extraction approach.

For the purposes of the current study, in order to avoid superfluous extraction of person entities as distinct ones, we employ the coreference resolution scheme given in [67].

The approach basically executes as follows: among the list of person entities extracted, beginning with the second entity (based on the position information), in order to form the coreference links, the name of each entity is compared to those of the previously extracted entities in turn to check whether their tokens intersect or not. The intersection is based on the nominal case of all tokens in the name of each entity compared. If the nominal forms of at least one of the tokens in the names of the entities compared match exactly, then they are said to be intersected. The comparison procedure to determine intersection ends when such a match is found which is the most recent match and a coreference link is formed between the entity under consideration and the intersecting entity. Since coreference is a transitive relation, that is, if an entity A corefers with B and B corefers with another entity C, then A corefers with C [54], this procedure eventually results in the identification of all of the desired coreference chains in the text under consideration [67].

Following the above example, the news story text may well include other references to the same person, such as the one given below, and after the coreference resolution procedure, the (`Person_1`) entity is enriched with an alias corresponding this information.

```
bakan    Unakıtan
minister Unakıtan
'The minister Unakıtan'
```

The final form of the entity (`Person_1`) is also shown below where only those slots filled with the corresponding values are shown.

```
Person_1:
Organization: Maliye
Occupation:   bakan
Name:         Kemal Unakıtan
Alias:
      Name:       Unakıtan
      Occupation: bakan
```

3.4 Event Extraction

Automatic semantic event extraction from videos is an important milestone to facilitate semantic video retrieval. Yet, it still remains as a demanding problem as the detection of semantic events in raw video files is a difficult task. A survey of event extraction approaches through the utilization of the audio-visual features is provided in [73]. In [29] and [112], ontology based approaches to semantic event extraction are proposed utilizing objects and relations extracted through the visual processing of the videos.

Text-based event extraction is studied as *scenario template production* in the MUC series [11] and as *event detection and recognition* in the ACE program [2]. Considering the studies exploiting video texts for semantic event extraction from the corresponding videos, we come across two significant approaches presented in [90] and [114], both of which target at team sports videos. In [90], speech transcriptions of football videos are used along with a football event ontology to detect the events occurred in the corresponding videos while Webcast texts of team sports videos are processed in [114] to determine representative keywords

for each significant event type in the video domain, which are then used to extract events from unseen videos.

In TDT research initiative, an event is defined as "something that happens at some specific time and place along with all necessary preconditions and unavoidable consequences" [24]. In our event extraction scheme, we follow this definition, yet, we only target at the extraction of the event types without considering the event attributes such as the entities involved and the time/place of the events.

Our event extractor takes a keyword based approach as it first determines frequent keyword lists for each considered event and next exploits these keywords to decide whether a given input text conveys any of these events. As the training data set from which common event keywords are to be detected, a video collection comprising 35 videos broadcasted by Turkish Radio and Television Corporation (TRT) [19] with a total duration of about four hours[2] is utilized in which the number of distinct news stories is 340. These news stories are annotated with the semantic events amounting to 463 events where news stories conveying general information or health advice are not tagged with any events. Hence, there may be zero to many events annotated in each story and the number of distinct events annotated is 69. The breakdown of these events with their corresponding frequencies are provided in Table B.1 in Appendix B. After the annotation procedure, the story texts that are annotated with the same event are grouped leading to non-disjoint event text groups (since a single story may convey more than one event and hence be included in several groups).

The most frequent events in the training data set turn out to be *Statement*, *Death*, *Trial/Investigation*, *Crash*, *Weather*, *Meeting*, *Attack*, *Injury*, *Election*, and *Operation* which constitute the event types considered by our event extractor. The total number of occurrences of these 10 events (over a total of 69 events) is 303 and hence, these 10 events, which correspond to about 15% (10/69) of all distinct events, represent 65.4% (303/463) of all the annotated events in the training data set.

[2] This data set is revisited as *Video Data Set-1* in Chapter 5 when describing the data sets utilized during the evaluation of the components of the proposed systems.

Below, we outline the coverage of each of these events:

- *Statement*: Events where some people (mostly political figures or authorities) make statements, explain or convey their ideas possibly in press conferences or mass meetings.
- *Death*: Events where individuals are murdered or die by other means such as traffic accidents. This type of event usually coexists with the other event types of *Crash* and *Attack*.
- *Trial/Investigation*: Events about legal trials and related investigations.
- *Crash*: Accidents mostly due to the means for transportation, covering plane crashes, traffic accidents, etc.
- *Weather*: Reports regarding the weather which may well be weather forecasts.
- *Meeting*: Meetings of two or more authorities (mostly politicians) including conferences.
- *Attack*: Events regarding offensive and destructive activities/attacks which may result in death/injury of individuals and economical loss.
- *Injury*: Events in which individuals are injured, which usually coexist with the other event types of *Crash* and *Attack*.
- *Election*: Events regarding all phases of elections including the preparations for the elections and their results.
- *Operation*: Military or legal operations which are carried out by organizations like the police departments or the armed forces.

The frequent keyword detection procedure is carried out as follows for each text group corresponding to the above event types:

1. The text is fed into the named entity recognizer and the extracted named entities are eliminated. Stopwords, occupation, and nationality names among the remaining tokens are also eliminated.

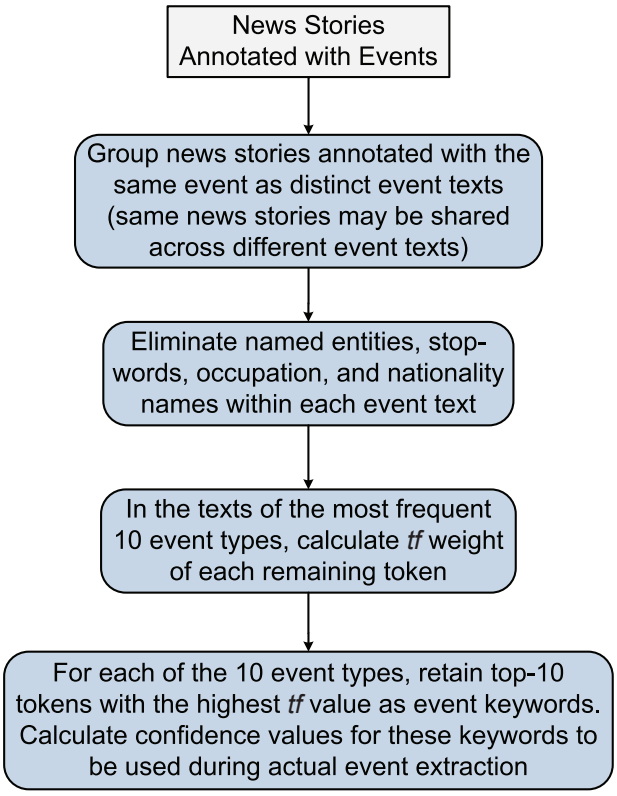


Figure 3.6: The Event Keyword Detection Procedure.

2. For each token in the text, the number of times the token appears (term frequency, tf) is calculated.

3. The most frequent 10 tokens (those with the highest 10 tf values) are retained as the keywords corresponding to the considered event and for each of these keywords (k) of each event type (e), a confidence value (c) is calculated as follows, where $tf_{i,e}$ is the tf of the i^{th} keyword in the text of event e:

$$c_{k,e} = \frac{tf_{k,e}}{\sum_{i=0}^{10} tf_{i,e}} \tag{3.2}$$

With this procedure, which is also given schematically in Figure 3.6, keywords with the associated confidence values are determined for the aforementioned 10 event types. The number of frequent keywords considered for each event type is 10 where this value is determined by examining proposed keywords for sample events to make the coverage of the ultimate event extractor as high as possible while not sacrificing precision too much. Yet, as will be presented in Section

5.2.3.3, the event extractor is also evaluated considering top-5 keywords along with the evaluations considering the top-10 keywords. The most frequent 10 keywords for each of these event types are given in Table 3.1 where the types of the keywords (*V.* for verbs, *N.* for nouns, *ADJ.* for adjectives, and lastly *PREP.* for prepositions), whether they are inflected or not (*inf.* denotes inflection), and approximate English translations are provided in parentheses.

It should be noted that the whole process is automatic (i.e., no manual intervention takes place) during the determination of these keywords, hence some of them are actually erroneous such as the last keyword, *büyük* ('*big*'), for the *Statement* event type where this keyword bears no specific information regarding this event type. Similarly, since some events usually coexist in the news stories, they share quite many keywords as is the case for the events of *Death*, *Injury*, and *Attack*. Another problem is the fact that our training data set is a limited one (extracted from a video data set with a total duration of 4 hours) spanning a continuous time period, leading to the extraction of very specific and hence not generalizable keywords. To illustrate, for the *Trial/Investigation* event type, the name of a trial (*Ergenekon*) is extracted as a keyword due to the fact that it is the topic of a considerable number of news stories broadcasted during the time period of the training data set.

The ultimate event extraction procedure utilizes these keywords to determine whether an event is described in an unseen news story text or not. For each event keyword, if it is encountered in a news story text, then the confidence value for that keyword is added to the particular event confidence score of that story. In other words, the following formula is utilized to determine the confidence score (cs) for each event (e) in each news story (s) in the test data set, where k_i denotes the i^{th} keyword of the event e:

$$cs_{e,s} = \sum_{i=0}^{10} c_{k_i,e}\sigma(k_i, s) \qquad (3.3)$$
$$where\ \sigma(k_i, s)\ is\ 1\ if\ k_i\ appears\ in\ s\ and\ 0\ otherwise.$$

Therefore, the event extraction procedure eventually outputs confidence scores for the considered events in each input news story and events with nonzero

Table 3.1: The Top-10 Representative Keywords for the Considered 10 Event Types Sorted in Descending Order of Their Confidence Values.

STATEMENT	**DEATH**
DEDİ (Vinf. *TOLD*)	KİŞİ (N. *INDIVIDUAL*)
SÖYLEDİ (Vinf. *SAID*)	HAYATINI (Ninf. *(HIS/HER) LIFE*)
BELİRTTİ (Vinf. *CLARIFIED*)	ÖLDÜ (Vinf. *DIED*)
İLİŞKİN (PREP. *REGARDING*)	KAYBETTİ (Vinf. *LOST*)
BİLDİRDİ (Vinf. *INFORMED*)	DÜZENLENEN (ADJ. *ORGANIZED*)
EKONOMİK (ADJ. *ECONOMICAL*)	YARALANDI (Vinf. *INJURED*)
ETTİ (Vinf. *MADE*)	KİŞİNİN (Ninf. *INDIVIDUAL'S*)
AÇIKLADI (Vinf. *EXPLAINED*)	POLİS (N. *POLICE*)
KRİZİN (Ninf. *CRISIS'*)	GAZINDAN (Ninf. *FROM THE GAS*)
BÜYÜK (ADJ. *BIG*)	KARBONMONOKSİT (N. *CARBONMONOXIDE*)
TRIAL/INVESTIGATION	**CRASH**
ERGENEKON (N. *ERGENEKON*)*	UÇAĞIN (Ninf. *PLANE'S*)
TUTUKLU (N./ADJ. *ARRESTED*)	DÜŞEN (ADJ. *CRASHED*)
SORUŞTURMA (N. *INVESTIGATION*)	PİLOT (N. *PILOT*)
KAPSAMINDA (Ninf. *WITHIN THE SCOPE*)	KAZADA (Ninf. *AT THE ACCIDENT*)
KAZI (N. *EXCAVATION*)	KİŞİ (N. *INDIVIDUAL*)
KARAR (N. *DECISION*)	SÖYLEDİ (Vinf. *SAID*)
İDDİA (N./V. *CLAIM*)	UÇAK (N. *PLANE*)
EDİLEN (ADJ. *MADE*)	HAYATINI (Ninf. *HIS/HER LIFE*)
ETTİ (Vinf. *MADE*)	KAZA (N. *ACCIDENT*)
DEVAM (N. *CONTINUATION*)	UÇAĞININ (Ninf. *(ITS) PLANE'S*)
WEATHER	**MEETING**
KAR (N. *SNOW*)	EKONOMİK (ADJ. *ECONOMICAL*)
HAVA (N. *WEATHER*)	ARAYA (Ninf. *INTERVAL*)
YAĞIŞI (Ninf. *RAIN*)	KONFERANSTA (Ninf. *AT THE CONFERENCE*)
ETKİLİ (ADJ. *EFFECTIVE*)	SÖYLEDİ (Vinf. *SAID*)
NEDENİYLE (PREP. *DUE TO*)	ELE (Ninf. *TO HAND*)
YAĞIŞLI (ADJ. *RAINY*)	TOPLANTIDA (Ninf. *AT THE MEETING*)
İÇ (N. *INSIDE*)	ÖNEMLİ (ADJ. *IMPORTANT*)
LODOS (N. *SOUTHWESTER*)	ETTİ (Vinf. *MADE*)
OLUMSUZ (ADJ. *NEGATIVE*)	GELECEK (ADJ. *NEXT*)
ETKİSİNİ (Ninf. *ITS EFFECT*)	GÖRÜŞECEK (Vinf. *WILL MEET*)
ATTACK	**INJURY**
DÜZENLENEN (ADJ. *ORGANIZED*)	KİŞİ (N. *INDIVIDUAL*)
SALDIRGANIN (Ninf. *ATTACKER'S*)	YARALANDI (Vinf. *INJURED*)
SALDIRIDA (Ninf. *AT THE ATTACK*)	DÜZENLENEN (ADJ. *ORGANIZED*)
POLİS (N. *POLICE*)	ÖLDÜ (Vinf. *DIED*)
SALDIRI (N. *ATTACK*)	HAYATINI (Ninf. *(HIS/HER) LIFE*)
BİLDİRDİ (Vinf. *INFORMED*)	KAYBETTİ (Vinf. *LOST*)
SALDIRISINDA (Ninf. *AT THE ATTACK (OF)*)	MEYDANA (Ninf. *TO PLACE*)
FÜZE (N. *MISSILE*)	SALDIRI (N. *ATTACK*)
HAVA (N. *WEATHER*)	AĞIR (ADJ. *SEVERE*)
ROKET (N. *ROCKET*)	PATLAMANIN (Ninf. *EXPLOSION'S*)
ELECTION	**OPERATION**
SEÇİM (N. *ELECTION*)	KİŞİ (N. *INDIVIDUAL*)
SEÇMEN (N. *VOTER*)	GÖZALTINA (Ninf. *TO CUSTODY*)
SEÇİMDE (Ninf. *AT THE ELECTION*)	MEHMETÇİK (N. *MEHMETÇİK*)**
HÜKÜMETİ (Ninf. *GOVERNMENT*)	ALINDI (Vinf. *TAKEN INTO*)
OY (N. *VOTE*)	ARALARINDA (PREP. *AMONG*)
BAŞLAYACAK (Vinf. *WILL START*)	ARAÇ (N. *VEHICLE*)
GİDİYOR (Vinf. *IS GOING*)	BULUNDUĞU (Vinf. *(THAT IT) APPEARS*)
KESİN (ADJ. *DEFINITE*)	BİRLİKLERDEN (Ninf. *FROM THE TROOPS*)
OLDU (Vinf. *HAPPENED*)	DÜZENLENEN (ADJ. *ORGANIZED*)
SAYISI (Ninf. *NUMBER*)	JANDARMA (N. *GENDARMERIE*)

* *ERGENEKON* is the name given to a trial and its frequent occurrence in the training video data set leads to its showing up as the top keyword for the *Trial/Investigation* event.

** *MEHMETÇİK* is a name commonly used to refer to the Turkish Armed Forces.

confidence scores are considered as the semantic events for each news story. We believe that the existence of confidence values for each extracted event is a desirable feature, as during later retrieval, it enables users to specify the least confidence levels for the event types that they utilize in their queries.

The event extraction procedure developed for Turkish news video texts is similar to the one presented in [114] for team sports videos. In the study [114], basketball Webcast texts are first clustered using LSA and keywords in each cluster (corresponding to distinct event types in basketball) are sorted on their $tf \times idf$ weights[3]. According to this ordering, the top-4 keywords in each cluster are used as the representative keywords for the corresponding event type for basketball.

We cannot use the approach in [114] for event detection from generic news video texts as it is, due to the differences between the characteristics of the target text genres. To clarify, although each individual time-aligned Webcast text (usually comprising one and rarely two sentences) usually denotes a single event, a news story (often longer than two sentences) usually conveys more than a single event. Hence, in [114], Webcast texts are automatically clustered using LSA and the resulting text clusters, which are given semantic event labels, are disjoint. On the other hand, we manually annotate the news stories from a training news video data set with the conveyed semantic events leading to non-disjoint text clusters corresponding to events, as news stories annotated with more than a single event are shared across the text clusters of these events. Moreover, the fact that the resulting text clusters are non-disjoint also prevents us from employing $tf \times idf$ for keyword ordering because most of the relevant keywords for the considered events achieve low idf values as these keywords also show up in other event text clusters, leading to low overall $tf \times idf$ values. Hence instead of $tf \times idf$ we use tf which leads to a more plausible ordering of the prospective event keywords.

[3] The $tf \times idf$ weight is commonly employed in IR to rank candidate documents given a user query (comprising the considered terms) and it is usually calculated as follows: $(tf \times idf)_{i,j} = tf_{i,j} \times idf_i$ where $tf_{i,j}$ is the number of occurrences of the term t_i in document d_j and $idf_i = \log(N/n_i)$ where N is the total number of considered documents and n_i is the number of documents in which term t_i appears [60].

CHAPTER 4

TEXT-BASED AUTOMATIC SEMANTIC ANNOTATION AND RETRIEVAL SYSTEM FOR NEWS VIDEOS IN TURKISH

This section is devoted to the description of the text-based semantic annotation and retrieval system for news videos in Turkish. The proposed system executes in fully automated mode and heavily relies on the IE techniques described in Chapter 3 on the video texts to automatically obtain the semantic annotations for the news videos. Apart from the IE components, the system comprises several other components including a news story segmenter, a text extractor (a sliding text recognizer), and a semantic retrieval interface to make it a complete fully-automated system in addition to a news video database to store the extracted information. After the offline annotation of the news videos, the system enables retrieval of the stored videos through the semantic retrieval interface of the overall system where the associated annotations are used to determine the relevant videos.

In the first section below, an overview of the proposed system is provided. The following section presents the details of the main components of the system. The last subsection is devoted to the description of the semi-automatic version of the proposed system which is best applicable to those videos for which video texts cannot be automatically extracted, as the sole manual intervention in this version takes place during speech transcription.

4.1 System Overview

The fully automated semantic annotation and retrieval system is provided schematically in Figure 4.1. The system is generic and can be customized to execute on news video archives in languages other than Turkish as well by replacing the language-specific components such as the ones for IE with convenient components for the language under consideration. The system is fully automated since no manual intervention takes place during the required tasks including the text extraction. In the rest of this section, we will describe the system and the details of its components as they are built for news videos in Turkish.

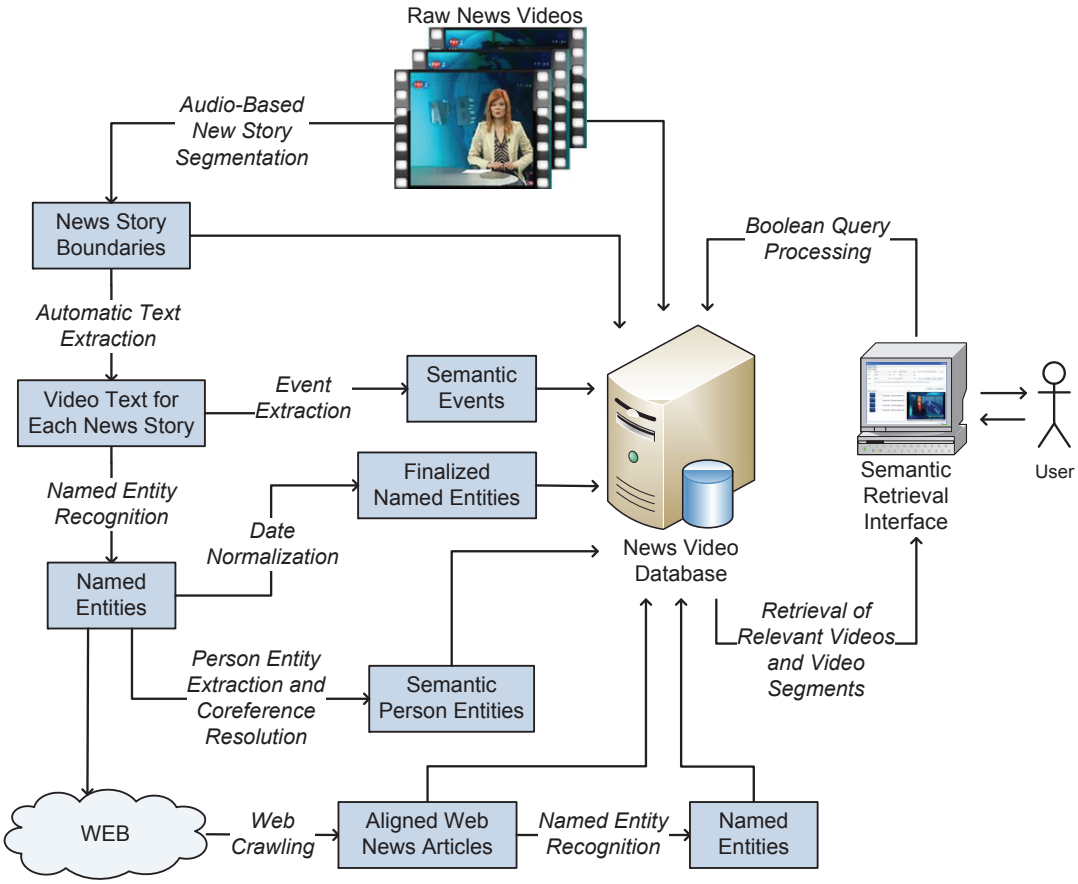


Figure 4.1: The Schematic Representation of the Semantic Annotation and Retrieval System for News Videos in Turkish.

The execution of the overall system is summarized below:

1. Raw news videos are given as input to the news story segmenter of the system. The segmenter detects the boundaries of individual news stories using the audio waveforms of the videos.

2. The video segments corresponding to distinct news stories are individually fed into the text extractor to obtain the news story texts. As the text extractor component, a sliding text recognizer for Turkish is employed since, to the best of our knowledge, there exists no publicly available ASR tool for Turkish, hence the system is fully automated only for those videos in Turkish which have their exact speech transcriptions as sliding texts over the frames.

3. The named entity recognizer is run on the news story texts to obtain the named entity sets of the individual news stories. The temporal information extracted by the named entity recognizer includes deictic expressions. Moreover, there may not be explicit time/date expressions in the news story texts but the story may be about an event/events occurred on the broadcast date of the corresponding news video. Therefore, the date normalization procedure presented in Section 3.1.3 is also carried out to deal with these two cases after the NER process using the broadcast dates of the news stories as the reference dates. After this post-processing procedure, discarding the actual deictic date expressions, the input text is associated with the corresponding normalized date expressions.

4. The extracted named entities are utilized by the Web alignment module to search the Web news articles that describe the same events as the news stories. These Web news are also fed into the named entity recognizer so that named entities mentioned in the Web news are also obtained.

5. News story texts are also given as input to the event extractor component to determine the events that are central to the topics conveyed in the corresponding news stories.

6. Information regarding the raw news videos in addition to the automatically

extracted semantic information are all stored in the news video database of the system.

7. Through the semantic retrieval interface of the system, the users can retrieve and play the videos through boolean query formulations or queries in natural language.

All functional components of the system are implemented in Java except for the sliding text recognizer which was implemented in MATLAB [10].

4.2 System Components

4.2.1 News Video Database

The news video database basically stores two types of information: the first one is the information regarding raw news video files such as their broadcast dates and total durations which may be considered as production metadata, and the second type corresponds to the information automatically extracted by the functional components of the system such as the named entities, events, and the URLs of the aligned Web news, in addition to others. The database is implemented using the open-source and object-relational PostgreSQL database management system [15]. The schema of the database is provided in Figure 4.2 as a Unified Modeling Language (UML) [34] class diagram. The classes in the schema are overviewed below:

- The *NewsVideo* class represents information regarding the raw video files and hence has the necessary attributes to store the paths of the news video files, their broadcast dates, durations, and the numbers of words in the corresponding video texts. It should be noted that the values for the last attribute (*text_word_count*) of the videos are only available after the sliding text recognition process.

- The *NewsVideoSegment* class is used to represent information regarding the news story segments which are obtained after the news story segmen-

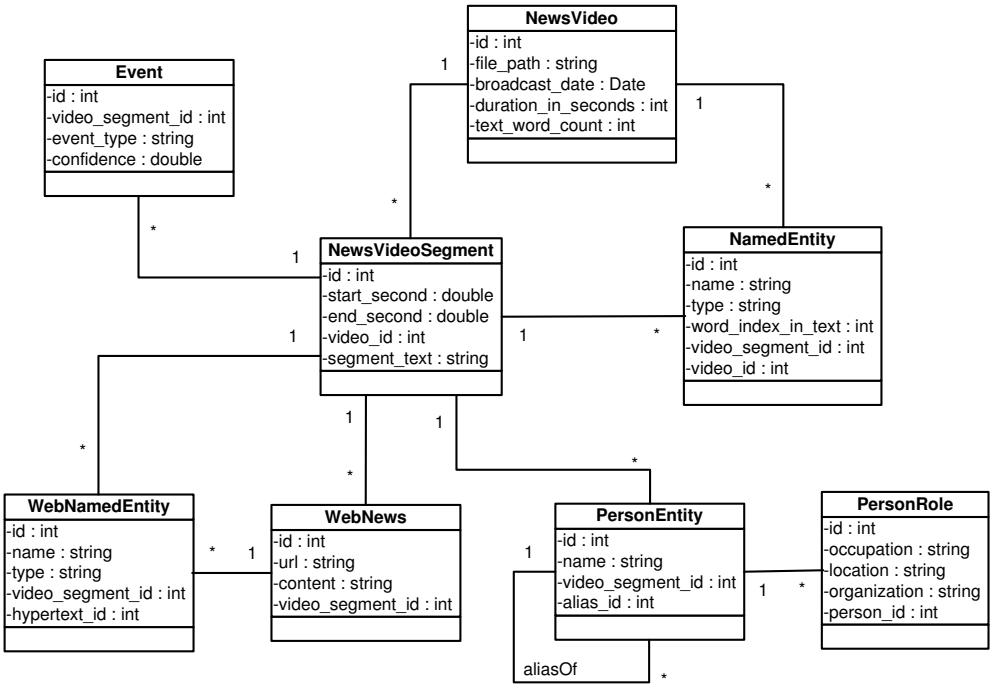


Figure 4.2: The News Video Database Schema as a Class Diagram.

tation process. The class has the attributes to hold the start and end seconds of each story segment.

- The *NamedEntity* class models each of the named entities extracted from the news story texts. It has the necessary attributes to hold the surface form of a named entity (the *name* attribute), its type which can be one of `PERSON`, `LOCATION`, `ORGANIZATION`, `DATE`, `TIME`, `MONEY` or `PERCENT` and the index of the first token of the named entity (the *word_index_in_text* attribute) in the corresponding texts.

- The *WebNews* class is added to the schema to model the aligned Web news articles associated with the news stories. It includes the necessary attributes to hold the URL and content of the Web news.

- The *WebNamedEntity* class represents the named entities extracted from the Web news articles and similar to the *NamedEntity* class, it has the attributes to hold the actual named entity and its type.

- The *PersonEntity* class is for representing each full person entity extracted from the texts. Since each person entity has one or more roles possibly

comprising an occupation, a location, and an organization, these roles are represented with the *PersonRole* class. It should be noted that, coreferential links between person entities are modeled through the *aliasOf* relationship between instances of the *PersonEntity* class.

- Finally, the *Event* class models the events extracted from the news story texts where their types and confidence scores are represented with the corresponding attributes of *event_type* and *confidence*, respectively.

After the news video database is populated with the relevant information including the production metadata and the extracted semantic annotations, it is ready to serve prospective users through the semantic retrieval interface of the overall system which will be described in Section 4.2.5.

4.2.2 News Story Segmenter

News story segmentation is included as a task in the 2003 and 2004 TRECVID conferences [78]. In these tasks, evaluation data is obtained from the data set collected for the TDT research initiative [108] and a news story is defined as a segment with a coherent focus with at least two declarative and independent clauses [27]. Three different resources are utilized for news story segmentation by the participating systems in the TRECVID evaluations: audio, video, and ASR output [27].

Associating the extracted semantic information with the whole video files makes semantic retrieval of these videos through this information possible, yet especially for long videos, it may be tedious for users to find the exact news story they are interested in. Therefore, the semantic information should better be associated with video segments corresponding to individual news stories in which the semantic information is conveyed.

Due to the fact that there is usually a silence period between the audio waveforms of consecutive news stories in a typical broadcast news video, we implement a silence detection algorithm which outputs silence periods to be used as the separators for news stories. We first extract the audio waveforms of each video

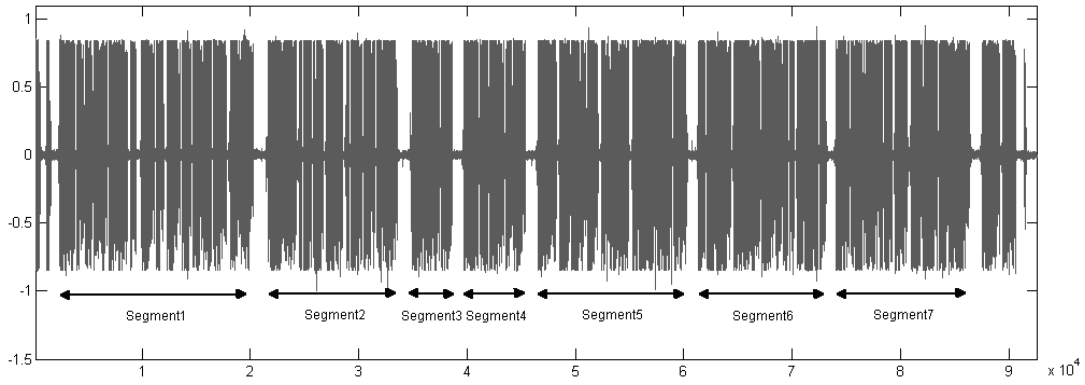


Figure 4.3: The Audio Waveform of a Sample News Video File.

file as audio files in WAVE format sampled at 48 kHz. The undersampled data of one of the channels of the stereo audio file corresponding to a sample news video file is plotted in Figure 4.3, where the segments corresponding the individual news stories are indicated with arrows. Those portions at the beginning and at the end are simply the opening and ending words of the speaker which are not actual news stories and hence they are not considered as distinct segments. Next, we implement a common procedure for silence detection: if there is a sequence of samples for a period longer than 1.5 seconds where each sample is less than one fifth of the maximum sample in the complete waveform, then this sequence of samples is identified as a silence interval together with the beginning and ending seconds. After the detection of the silence intervals, they are used as the boundaries of actual news stories. The first and the last segments and the segments lasting less than five seconds are automatically discarded to avoid the extraction of the beginning and ending portions of the videos which are not considered actual video segments. Those utterances corresponding to the beginning and ending segments are also not considered during text extraction.

We should note that the parameters employed in this segmentation procedure, such as the period limit of 1.5 seconds, are particularly convenient for our news video data sets, broadcasted by TRT, to be described in Chapter 5 and are determined by examining the waveforms of some sample videos. In order to use this segmentation scheme on other news video data sets, the aforementioned parameters should be reviewed and adjusted accordingly. Another plausible approach

to news story segmentation is to carry out of discourse/topic segmentation on the corresponding video texts, as employed in studies such as [47], which is left as a direction of future research.

4.2.3 Sliding Text Recognizer

In order to obtain the speech transcriptions of the videos which are also given as sliding texts over the frames of the videos of our target TRT broadcasted data set, we employ the sliding text recognizer presented in [46] and integrate it to our automatic semantic annotation and retrieval system. This recognizer first converts the video frames into binary images and utilizes the horizontal and vertical histograms of these images to determine the exact text band. Next, connected component analysis is used to segment the text image into individual characters after some noise removal operations. The recognizer then employs template matching for actual character recognition and finally, it carries out a correction procedure based on rules learned through transformation-based learning in order to improve character recognition accuracy [46].

In our semantic annotation and retrieval system, the sliding text recognizer executes on each of the previously determined video segments. Provided below is a sample snippet from the output of the recognizer on a sample news story segment where words with errors are shown in boldface.

TRT VE ÖZEL TELEVİZYONLARDAN DEV İŞBİRLİĞİ. **TÜRHİYE'NİM ONDE** GELEN **7ELEVİZYON KOROLOŞLARI** ORTAK BİR ANTEN **SİŞTEMİ KÜRMAH** İÇİN BİRARAYA GELDİ. **ŞEHİRLERDEHİ GORÜNTÜ** KİRLİLİĞİNE SON VERECEK **ÜVGOLAMAYLA** VATANDAŞ DAHA KALİTELİ YAYIN İZLEME **OLAMAĞINA HAVOŞACAH**, İMZA TÖRENİNDE **KONÜŞAN** DEVLET BAKANI BEŞİR **ATALAV**, İKİ **AV** İÇİNDE 13 BÜYÜK **ŞEMİRDE** ORTAK ANTENLERİN FAALİYETE GEÇECEĞİNİ SÖYLEDİ. ATALAY BU 13 ŞEHİRDE YAYIN YAPAN TELEVİZYONLARIN **LİSANSLANOIRMA** ÇALIŞMALARININ DA KISA SÜREDE TAMAMLANACAĞINI **NAYOET7I.**

As the above output snippet exemplifies, the output of the sliding text recognizer is not perfect and most of the errors are due to OOV words. After

observing such outputs, we employ a two-phase automatic correction procedure on the sliding text recognizer output so that the word error rate (WER) in the output is decreased (as will be observed in Section 5.2.2 where the performance evaluation of the sliding text recognizer is presented):

- In the first phase, some lexical rules are applied to the output text, similar to the ones proposed in [46]. The employed rules basically include the following:
  - If the surrounding characters are not digits, replace 7 with 'T'.
  - If 'O' follows a digit, then replace it with '0'.
  - If '8' is immediately followed by a letter, then replace '8' with 'B'.
  - Replace 'OO', 'OÖ', 'ÖÖ', and 'ÖO' with '00'.
  - If 'O' is immediately followed by a vowel, then replace 'O' with 'D'.
  - If 'Ğ' is the first character of a token, then replace it with 'G'.
  - Replace all tokens of the form 'BO' with 'BU'.
- In the second phase, the spelling checker of the Zemberek [21] open-source NLP library for Turkish is employed to further decrease the WER. Each token is replaced with the first candidate provided by the spelling checker which has the same character count as the token to be corrected.

4.2.4 Information Extraction Components

After the news story texts are obtained, the IE components of the system are executed on these texts. That is, the hybrid named entity recognizer, the Web alignment module, the full person entity extraction and coreference resolution modules, and the event extractor as described in Chapter 3 are all executed on the story texts and the extracted information is stored in the news video database. Aligned Web news are also given as input to the named entity recognizer and the resulting named entities are stored in the database along with the entities obtained directly from the actual news story texts.

Table 4.1: An Overview of the Employed Approaches for the IE Tasks.

IE Task	*Employed Approach*
Named Entity Recognition	First, a rule-based named entity recognizer is proposed for generic news texts. It is then turned into a hybrid system by equipping it with a rote learning component. A subsequent date normalization procedure is also proposed to handle deictic date expressions extracted by the recognizers.
Person Entity Extraction	A rule-based person entity extraction scheme is employed. In order to prevent superfluous extraction of the same person entities as distinct ones, a string matching based coreference resolution scheme is implemented.
Automatic Hyperlinking	Utilizing the named entities extracted from news story texts as search criteria, related Web news articles are crawled and after some elimination operations, high confidence articles are retained and associated with the corresponding news stories.
Event Extraction	A bag-of-words approach is followed. Frequent event types are determined using a training set of video texts. Next, frequent keywords observed in the news stories of the training set conveying the considered events are utilized during actual event extraction.

A summary of the employed approaches targeting at the considered IE tasks is provided in Table 4.1. Evaluation results of the corresponding IE components will be described in detail in Chapter 5 together with some improvement attempts after observing the initial results.

4.2.5 Semantic Video Retrieval Interface

As the final component of the proposed annotation and retrieval system, a semantic video retrieval interface is implemented which enables access to the news video database through boolean queries or queries in natural language over the extracted information. A snapshot of the semantic retrieval interface is given in Figure 4.4. Through this interface, boolean queries can be formulated using the named entities and events as literals and combining them using the boolean

operators of `AND`, `OR`, and `NOT` with parentheses when necessary. A well-formed boolean query (BQ) expression can be generated with the grammar provided below where *ne* denotes a named entity as a surface form/type pair while *event* denotes an event as event_type/$EVENT$[*confidence_value*] where *confidence_value* is used to specify that news videos/video segments associated with events having confidence scores of at least this value should be retrieved. The named entities extracted from the Web news could also be utilized during the formulation and execution of the boolean query expression.

$$BQ \rightarrow (BQ) \mid BQ \textit{ and } BQ \mid BQ \textit{ or } BQ \mid \textit{not } BQ \mid ne \mid event$$

The formulated queries are shown in a text box (named *Query*) on the interface and if they are not well-formed (i.e., they cannot be generated with the above grammar), then the user is informed accordingly. A well-formed query example through the interface is demonstrated in Figure 4.4 where those news stories in which an *Election* event is conveyed and *Sarkozy* (current president of France) is mentioned are queried, with the below query expression. This query also has the constraint regarding the *Election* event that only those stories in which the event is extracted with at least the confidence value of 0.1 should be retrieved, as specified in the square brackets at the end of the $EVENT$ keyword.

```
SARKOZY/PERSON AND ELECTION/EVENT[0.1]
```

The boolean query expressions are processed to transform them into appropriate structured query language (SQL) expressions to be executed on the news video database. During these transformations, an `AND` keyword results in a subexpression in which the operands are combined with the `INTERSECT` keyword in SQL. Similarly, `OR` and `NOT` keywords result in subexpressions which include the `UNION` and `EXCEPT` keywords, respectively.

When handling the criteria of type `PERSON`, information regarding the full person entities extracted (i.e., coreferring person entities represented as aliases) are also utilized in addition to simply using the matching named entities. Doing so, the system successfully handles cases where a person entity is queried by retrieving all news story segments in which the queried person is named differently. Hence,

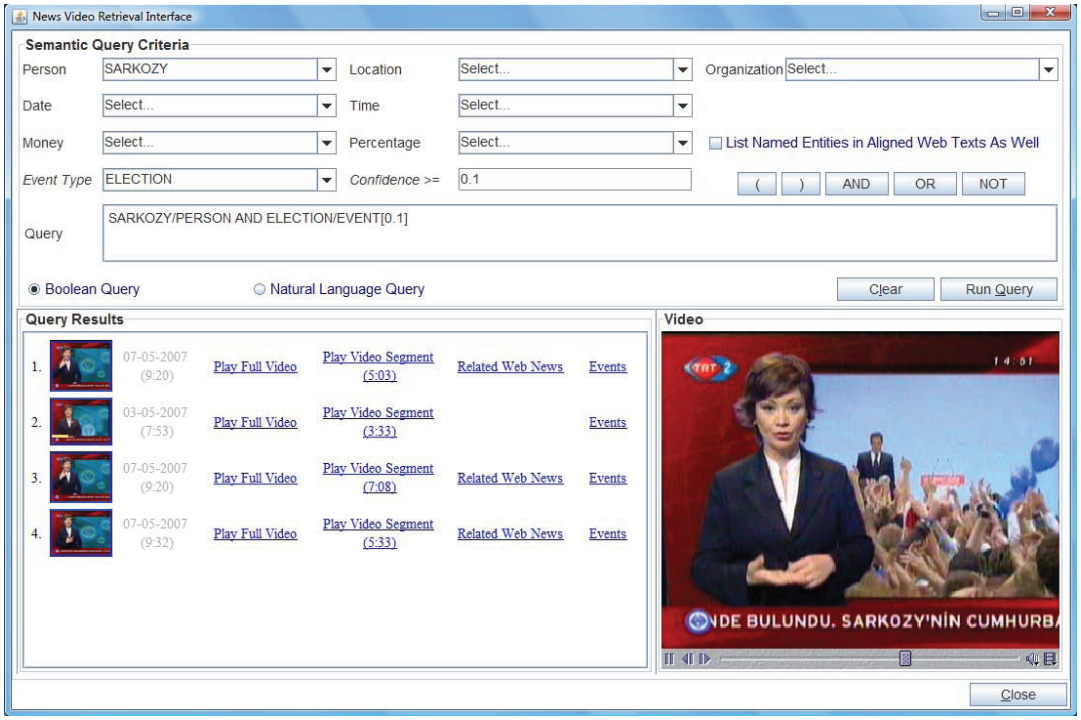


Figure 4.4: A Snapshot of the Semantic Video Retrieval Interface (Boolean Query Example).

during the transformation of the query expression into the corresponding SQL expression, if `X` is a criterion of type `PERSON` encountered in the query expression (as `X/PERSON`), then the following news stories are considered:

- Those news stories which include `X` as a named entity of type `PERSON` (and those named entities whose associated Web articles include `X` as a named entity of type `PERSON`, if specified so).
- Those news stories which does not include `X`, but instead include at least one of `X`'s aliases (entities coreferring with `X`).
- Those news stories which include those entities for which `X` is an alias.
- Those news stories which include those entities which are aliases of an entity for which `X` is also an alias.

The SQL expression for the specified query expression is provided in Appendix C. The tuples in the news video database satisfying the SQL query are retrieved through the interface. The query results are listed on the result panel (titled

Figure 4.5: A Snapshot of the Related Web News Dialog of the Semantic Video Retrieval Interface.

Query Results) with a thumbnail, the broadcast date together with the total duration of the corresponding videos in parentheses, and four hyperlinks. The sorting order of the query results is determined as follows: If the query expression does not contain any positive event criterion, then the results are simply sorted on their broadcast dates, otherwise, the results are sorted in descending order of the sum of the confidence score(s) of the event(s) that are specified in the query expression.

The first one of the hyperlinks for each of the query results, (*Play Full Video*), is for playing the complete video file in the video panel on the right and the second one (*Play Video Segment (XX:XX)*) is for playing the file beginning from the start time of the particular satisfying video segment where the exact start time of the segment is shown in parentheses at the end of the hyperlink. Java Media Framework API (JMF) [8] is utilized to play the videos/video segments through the interface. In order for the system to support several video formats including Audio Video Interleave (AVI), FOBS4JMF JMF plugin [4] is utilized. The third hyperlink (*Related Web News*) is available for those results for which aligned Web news have been obtained and it enables the users to examine these automatically determined Web news through a dialog as shown in Figure 4.5. Lastly, through the *Events* hyperlink for the applicable query results, all of the events extracted from the news story under consideration can be examined as shown in Figure 4.6 with their corresponding confidence values. Also observable through this dialog is related semantic information including extracted person entities, locations/organizations not associated with the person entities (together with their roles and aliases), and date/time expressions. This additional information is important as it can serve to fill in the actor(s), place, and time slots of the extracted events and hence make extracted events more complete.

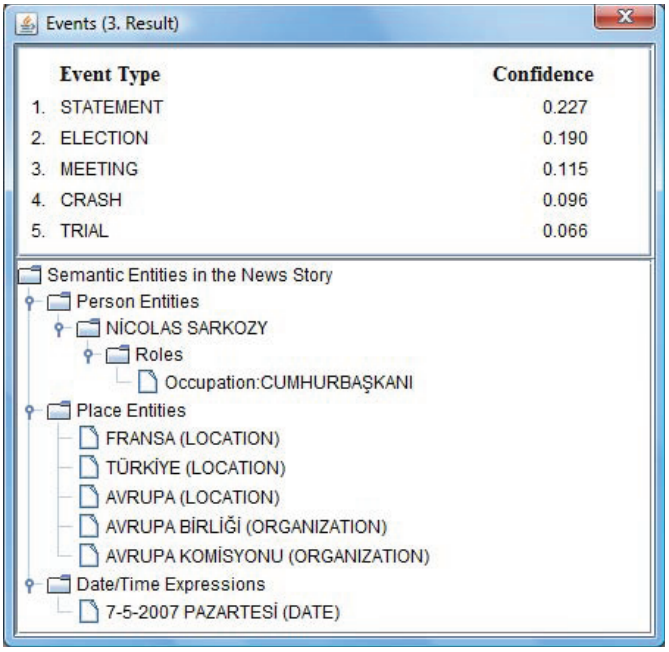


Figure 4.6: A Snapshot of the Events Dialog of the Video Retrieval Interface.

In order to facilitate the retrieval of news videos, we extend the interface to process queries in natural language (NL) as well, hence the ultimate interface also acts as a natural language interface.

A snapshot of the semantic retrieval interface with a sample NL query is demonstrated in Figure 4.7 and the query, which targets at similar videos as does the boolean query in Figure 4.4, is provided below[1].

```
Sarkozy ve  seçimle      ilgili  haberler
Sarkozy and election-INS related news
'News related to Sarkozy and election'
```

It should be noted that we do not utilize deep natural language understanding tools during the processing of the NL queries, but instead, we basically make use of our IE components to transform the query into a boolean query as shown in Figure 4.4 so that thereafter it can be processed in the same way as a boolean query.

[1] In the query expression, *INS* stands for the instrumental/comitative case marker in Turkish.

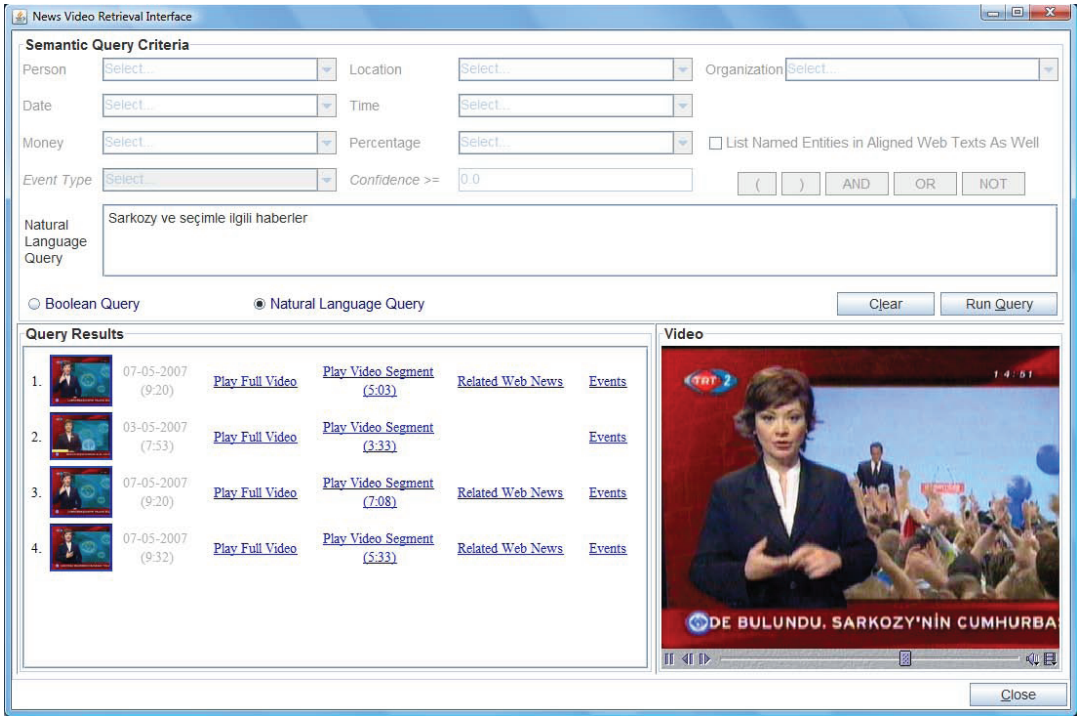


Figure 4.7: A Snapshot of the Video Retrieval Interface (Natural Language Query Example).

The transformation of an NL query into a boolean query is carried out as follows:

1. The NL query is given as input to our named entity recognizer and thereby named entities in the query are determined. In the example NL query in Figure 4.7, *Sarkozy* is extracted as a named entity of type `PERSON`.

2. The top-10 keywords determined for the considered 10 events and the synonyms for these events (if not included in the keyword lists) are used to determine the existence of these events in the NL query. To clarify, for each event type, the top-10 keywords and event synonyms are searched for in the query and if any of them is encountered in the query, then that event type is extracted. To illustrate, in the NL query example above, since *seçim* ('*election*') is one of the top-10 keywords for the `ELECTION` event, this event is extracted from the query. The least confidence levels for the events given in the NL queries are assumed to be 0.

3. In order to determine the connectives between pieces of semantic information (named entities and events) in the NL query, among the tokens

excluding those included in the extracted entities and events, the Turkish phrases of *veya*, *ya da*, *yahut*, *veyahut* meaning '*or*' are searched for and if at least one of them is encountered, then an `OR` operator is used to combine the semantic information extracted before and after the encountered phrase. If no such phrase is encountered, then the extracted entities and events are assumed to be connected with an `AND` operator[2]. After this step, the example NL query is transformed into `SARKOZY/LOCATION AND ELECTION/EVENT[0.0]` which is processed as in the case of an ordinary boolean query and the results are shown in Figure 4.7 exactly the same as in the case of a boolean query, as exemplified in Figure 4.4.

4.3 Semi-Automatic Version of the System

Within the course of this thesis, we have also implemented a semi-automatic version of the fully automated system where the sole manual intervention takes place during the speech transcription to obtain the video texts. The schematic representation of the semi-automatic system is presented in Figure 4.8. This version is proposed since the fully automated version is not applicable to cases where automatic extraction of the video texts is not possible. It should be noted that this semi-automatic version also executes in fully automated mode if the video texts are already available as the associated texts.

The main differences between this version and the fully automated system, the architecture of which is given in Figure 4.1, are that this version lacks the sliding text recognition module and video texts for the whole video is available as opposed to that of the individual news story texts. As described in the previous section, in the fully automated version, video segments corresponding to individual news stories are fed into the sliding text recognizer, hence the semantic information extracted from the news story texts is automatically aligned

[2] The interface currently cannot handle cases requiring the `NOT` operator as this necessitates deeper elaboration of the NL query, which is left as a plausible future research direction. Because, Turkish is a morphologically rich and free word order language in which negativity is conveyed through verb inflection. Hence, in order to accurately determine the negativity and semantic information spanned by it, natural language processing tools such as morphological analyzers and parsers should be employed. Still, using the aforementioned tools does not guarantee the success of the interface as these tools usually produce ambiguous output.

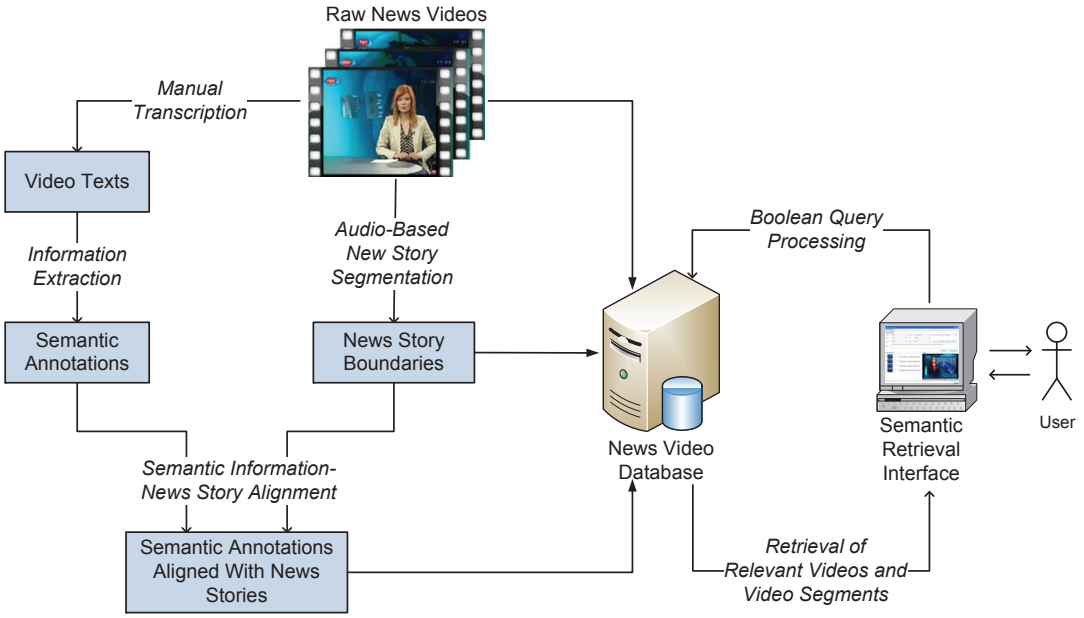


Figure 4.8: The Schematic Representation of the Semi-Automatic System for Semantic Annotation and Retrieval of News Videos in Turkish.

with the news stories. But in this semi-automatic version, whole video texts are fed into the IE components. Therefore, a separate alignment procedure is required to align the extracted information with the news stories, which will definitely not lead to perfect alignment as some of the information will erroneously be associated with the neighboring news stories. The task of associating the texts with their corresponding speech is known as *text-to-speech alignment* in the speech processing community but a tool to perform this task for Turkish is currently unavailable to us and therefore we implement a rough approximation approach to address the alignment of news stories with the named entities. Similar procedures can be employed to align the other extracted information types as well.

The proposed alignment procedure is based on a heuristic that for a named entity included in a particular news story segment, the ratio of word index of the named entity to overall word count of the corresponding video text is most probably included in the ratio interval which begins with the ratio of the start second of the corresponding news segment to the overall video length and ends with the ratio of the end second of that segment to the overall video length, therefore they can be mapped. The alignment procedure is presented in Algorithm 2 where

Algorithm 2 NamedEntity-Segment Alignment

Require: The extracted named entity list *neList*.

Ensure: The value of *video_segment_id* for each *ne* in *neList*.

1: **for all** named entity *ne* in *neList* **do**
2: **let** v be the video file associated with *ne*
3: $video\_start\_sec \Leftarrow min(sg.start\_second)$
4: $video\_end\_sec \Leftarrow max(sg.end\_second)$
5: among all segments sg of v
6: **for all** segment s in the video file v **do**
7: **if** $ne.word\_index\_in\_text/v.text\_word\_count$
8: **between**
9: $(s.start\_second\text{-}video\_start\_sec)/video\_end\_sec$
10: **and** $s.end\_second/video\_end\_sec$ **then**
11: $ne.video\_segment\_id \Leftarrow s.id$
12: **end if**
13: **end for**
14: **end for**

`video_start_sec` and `video_end_sec` are used to rule out the effects of those seconds belonging to the beginning and ending sections of the news videos since their segments are previously discarded and their transcriptions are not included in the final video transcriptions. It should be noted that this procedure is error-prone particularly at the news segment boundaries. At the boundaries, in order to favor the preceding segments to the following ones, the procedure does not subtract the `video_start_sec` from `s.end_second` at line 10, although it is subtracted from `s.start_second` at line 9. Preceding segments are favored due to the fact that during semantic retrieval, the interface enables its users to play videos beginning from the particular news segment returned, and by favoring the preceding segment, we try to make sure to a certain degree that segments are not missed, since even if the actual segment is the following segment, the user has the ability to see that segment anyway with the drawback that she/he is superfluously exposed to the previous segment beforehand but nevertheless, this situation is preferable to not seeing the actual video segment at all. The

computational complexity of the algorithm is $O(n.s)$ where n is the size of the list of named entities extracted and s is the size of the list of news story segments.

To illustrate the idea behind the alignment algorithm, consider a news video with a total length of 473 seconds where the start and end second pairs for a total of 8 story segments of this video are: `(10,69), (72,135), (138,210), (213,269), (273,324), (326,348), (350,384)`, and `(387,467)`. Hence, the start second of the first segment is `10` and the end second of the last segment is `467` which are assigned to *video_start_sec* and *video_end_sec*, respectively as given in lines 3–4 of Algorithm 2. On the other hand, consider a named entity *FRANSA* ('*FRANCE*') of type location name mentioned in this video where the token index of the named entity is `291` among a total of `658` tokens in the whole video. As the ratio of this token index over the total token count ($\frac{291}{658} = 0.44$) is between the ratio of the start second of the fourth segment minus *video_start_sec* over *video_end_sec* ($\frac{213-10}{467} = 0.43$) and the ratio of the end second of the same segment over *video_end_sec* ($\frac{269}{467} = 0.58$), the alignment algorithm (correctly) associates this named entity with this fourth segment of the video. The performance evaluation of the alignment algorithm is provided in Section 5.2.3.1 along with the evaluation results of the named entity recognizers on the texts of the video data sets.

As the automatic hyperlinking module described in Section 3.2 makes use of the list of named entities to search for the related Web news, for the case of the semi-automatic system, the hyperlinking procedure will also be affected from the misalignments performed by this algorithm.

CHAPTER 5

EVALUATION AND DISCUSSION

5.1 Evaluation Data Sets

It is commonly acknowledged that sufficient data sets are indispensable for the implementation of successful knowledge-based systems. For the purposes of the current study, we also need considerable amount of textual and video data in Turkish to build an effective system that utilizes IE techniques for semantic annotation and retrieval of news videos in Turkish. But, to the best of our knowledge, there exists no publicly available annotated text corpus for IE research in Turkish, which mainly hinders related research. Similarly, we know of no available video corpus in Turkish to be utilized in studies regarding video indexing and retrieval.

One of the commonly utilized textual corpora in Turkish is METU Turkish corpus [92] with about 2 million words, but it is not an annotated resource, hence an additional annotation procedure should be employed to make use of this resource during the development of learning systems or during the testing of practical systems for IE. Other resources include Bilkent's TDT collection (BilCol-2005) [38] which comprises Web news articles in Turkish from several news provider sites and again Bilkent's IR test collection [37] consisting of news articles published by the Turkish newspaper, *Milliyet*. These resources have considerably larger sizes and, as indicated, they are mostly convenient for research on TDT and IR, respectively. Therefore, we compile and –when applicable– annotate our own textual and video data sets to serve our needs in various stages of the study.

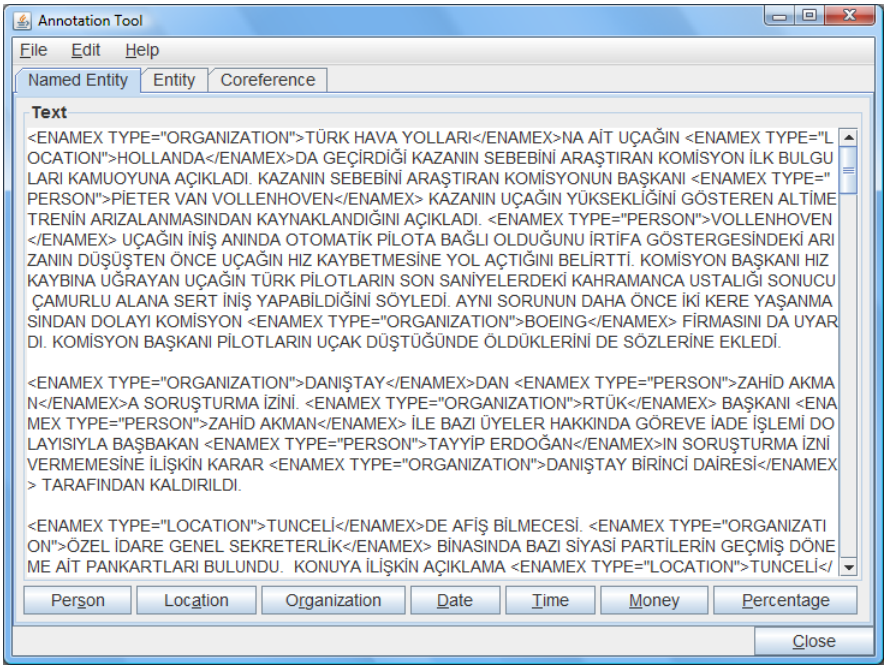


Figure 5.1: A Snapshot of the Named Entity Annotation Tool.

Below described are our textual data sets utilized for training/testing our hybrid named entity recognizer and for testing its rule based predecessor:

1. *News Text Data Set*: This data set contains 50 news articles from the METU Turkish corpus [92]. The total number of words in this set is about 101,700 while the number of distinct words (types) is 29,400. We manually annotate this data set, using an annotation tool that we implement for this purpose, with the aforementioned MUC style named entity tags (`ENAMEX`, `TIMEX` and `NUMEX`) leading to a total of 11,206 named entities. A snapshot of the implemented annotation tool is provided in Figure 5.1. The annotated entities encompass 3,280 person, 2,470 location, 3,124 organization names along with 1,413 date/time and 919 money/percent expressions.

2. *Financial News Text Data Set*: This set comprises 350 financial news articles created by a Turkish news provider, Anadolu Agency [1], which was already annotated with the person and organization names only. We postprocess this data set of about 84,300 words (with 23,850 types) to

make some corrections on the existing annotations and the resulting set contains a total of 5,635 named entities where 1,114 of them are person names and 4,521 of them are organization names.

3. *Child Stories Data Set*: The set of child stories corresponds to two stories by the same author [58, 59] which was already previously utilized in [69, 70]. The set contains about 19,000 words and 7,240 types. The manual annotation process results in the annotation of 1,084 named entities with 836 person, 157 location, 6 organization names in addition to 65 date/time and 20 money/percent expressions[1].

4. *Historical Text Data Set*: Similar to the previous data set, the historical text set was used in [69, 70] and it is composed of the first three chapters of a book on Turkish history [99]. This text contains about 20,100 words (9,150 types) with a total of 1,173 named entities after annotation, with 387 person, 585 location, 122 organization names, and 79 date/time expressions.

Statistical information on the textual data sets is also provided in Table 5.1 as the number of words and the number of named entities that they contain.

Table 5.1: Statistical Information on the Text Data Sets.

Data Set	*Word Count*	*Type Count*	*NE Count*
News Text Data Set [92]	101,700	29,400	11,206
Financial News Text Data Set [1]	84,300	23,850	5,635
Child Stories Data Set [58, 59]	19,000	7,240	1,084
Historical Text Data Set [99]	20,100	9,150	1,173

Having provided summary information about our textual data sets, below we present the details of our two distinct video data sets which are utilized by the proposed fully automated and semi-automatic systems. Both of the data sets comprise news videos broadcasted by TRT and the videos have exact speech transcriptions as sliding texts as well. Information regarding these video data sets is also summarized in Table 5.2.

[1] The annotation processes on child stories and historical text data sets result in slight differences from those provided in [69, 70] mostly due to some minor corrections in the existing annotated versions in addition to some slight scope extensions.

1. *Video Data Set-1*: We have compiled a video corpus of 35 news videos broadcasted by TRT during the February 6, 2009 - March 15, 2009 period with a total duration of about four hours. The number of distinct news stories in the data set is 340. This video data set is a low resolution one and hence could not be automatically transcribed by the sliding text recognizer described in Section 4.2.3. In order to utilize this video data set in the training/testing stages of the components of the proposed automatic and semi-automatic systems, we manually transcribe it and the resulting video text comprising 20,940 words is also annotated with named entities leading to a total of 2,534 named entities (595 person, 1,062 location, and 489 organization names along with 32 time, 241 date, 80 money, and 35 percent expressions). The total number of semantic events annotated in this set is 463 where there are 69 distinct events among them (as previously given in Section 3.4). This low resolution data set is mainly used as the training data set for the applicable components of the fully automated indexing and retrieval system.

2. *Video Data Set-2*: Our second data set comprises 19 videos broadcasted in the fourth, fifth, and sixth months of 2007 again by TRT with a total duration of 1.5 hours. The number of news stories included in the data set is 182. The resolution of this data set is sufficient to be given as input to the sliding text recognizer employed. But in order to assess the word error rate of the sliding text recognition procedure, we have also transcribed this data set and hence we have both the noisy (sliding text recognizer output) and the corresponding perfect transcriptions of the videos. After the annotation of the text of this set to create the answer key for named entity recognition, it turns out that the video text includes a total of 1,408 named entities with 279 person, 534 location, and 373 organization names in addition to 164 date, 21 time, 17 money, and 20 percent expressions. This data set is annotated with 362 events where the number of distinct event types is 55. This data set is mainly exploited as the evaluation data set of the fully automated video indexing and retrieval system.

Table 5.2: Statistical Information on the Video Data Sets.

Data Set	*Duration*	*News Story Count*	*Word Count in Text*	*NE Count*	*Event Count*
Video Data Set-1	4.0 hours	340	20,940	2,534	463
Video Data Set-2	1.5 hours	182	12,043	1,408	362

5.2 Evaluation of the Individual Components of the Fully Automated System and Its Semi-Automatic Counterpart

5.2.1 Evaluation of the News Story Segmenter

The first functional module of the proposed systems, the audio based news story segmenter successfully detects 339 boundaries out of the 340 actual boundaries in *Video Data Set-1* and only misses the remaining single segment by detecting two distinct segments in it. *Video Data Set-2* is also fed into the news story segmenter and similar to the previous evaluation, the segmenter correctly detects the boundaries of 181 out of 182 actual stories and it again detects two distinct stories in the remaining single story. These results indicate that the employed news story segmentation scheme is highly accurate and hence convenient for the proposed systems.

5.2.2 Evaluation of the Sliding Text Recognizer

The news story segments of *Video Data Set-2* are then fed into the sliding text recognizer (Section 4.2.3) to obtain the complete transcriptions and the resulting text output of the recognizer comprising 12,043 words has a WER of 22.25%. The fact that this WER will possibly lead to low success rates during IE and therefore jeopardize the overall purpose of the proposed fully automated system, we execute the two-phase correction approach described in Section 4.2.3 to decrease WER. After the execution of this automatic correction procedure, the error rate drops down to 8.93%. This error rate is highly favorable since to the best of our knowledge, state of the art ASR systems, even for well studied languages such as English, cannot achieve comparable results [16].

Henceforth, discarding the raw output of the sliding text recognizer, this automatically corrected, but still noisy, output is considered as the video texts of *Video Data Set-2*. But its corresponding manually transcribed perfect counterpart is also retained and given as input to various IE components of the system for comparison purposes.

5.2.3 Evaluation of the Information Extraction Components

5.2.3.1 Evaluation of the Named Entity Recognizer

In this section we present the evaluation results of the rule based named entity recognizer and its hybrid successor both on our textual data sets and on the video texts described in Section 5.1.

NER Evaluation on the Textual Data Sets:
The evaluation results of the initial rule based named entity recognizer, proposed for the news domain, on the textual data sets are provided in Table 5.3 in terms of the metrics of precision, recall, and F-measure as utilized in [77]. The metrics are expressed as percentages in all of the tables corresponding to the evaluation results. These metrics, also giving half credit to partial extractions where the type of the extracted named entity is correct but its span is incorrect, are calculated as follows [77]:

$$Precision = \frac{Correct + 0.5 * Partial}{Correct + Spurious + 0.5 * Partial} \tag{5.1}$$

$$Recall = \frac{Correct + 0.5 * Partial}{Correct + Missing + 0.5 * Partial} \tag{5.2}$$

$$F{-}Measure = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5.3}$$

In the above formulae, *Correct* corresponds to the number of named entities extracted by the recognizer which are exactly the same as their counterparts in

the answer key, in terms of their location in text, their named entity types and the tokens they comprise. *Spurious* represents the number of entities spuriously (erroneously) extracted by the recognizer, i.e., they do not have corresponding annotations in the answer key and *Missing* is the number of named entities which are not annotated, hence missed, by the recognizer although they are annotated in the answer key. Lastly, *Partial* denotes the number of named entities extracted by the recognizer which have corresponding entities annotated in the answer key with the same type, hence their types are correct but the tokens they contain are not exactly the same since either some tokens are erroneously missed or included by the recognizer. An example partial extraction occurs when only the name of a person entity is annotated by the recognizer although the consecutive word is the surname of the entity and the answer key correctly contains the name-surname pair collectively annotated as a person entity.

Table 5.3: Evaluation Results of the Rule Based Named Entity Recognizer on the Textual Data Sets (Capitalization Feature is Turned Off).

Data Set	*Precision*	*Recall*	*F-Measure*
News Text Data Set	85.15%	83.23%	84.18%
Financial News Text Data Set	71.70%	50.36%	59.17%
Child Stories Data Set	72.67%	76.81%	74.68%
Historical Text Data Set	53.57%	70.70%	60.96%

Though the data sets are not comparable in size or in terms of the number of named entities that they contain, the recognizer achieves the best results on the news text data set as expected, since news texts constitute its target genre. On news texts, main sources of performance drop include the following:

1. Some common names which happen to be homonymous to some entries in the lexical resources are erroneously extracted as named entities. For instance, common names such as *savaş* (meaning '*war*' as a common name), *barış* (meaning '*peace*'), and *özen* (meaning '*care*') are among the names in our dictionary of person names and erroneously extracted as person names. This situation is one of the main causes of the drop in precision.

2. Although we include some well-known foreign politicians in our lexical resources, other lesser-known foreign names cannot be recognized by our

rule based recognizer and hence this situation leads to a decrease in recall. A named entity recognizer particularly for English can be incorporated to alleviate the effects of this problem, which is currently left as future work.

3. Similar to the first item above, during location and organization name extraction, the recognizer performs erroneous extractions of the phrases such as *anlatmanın yolu* ('*the way to tell*') as location names since they match the location patterns, and similarly phrases like *ilk üniversitesi* ('*first university*') as organization names for they match the organization patterns employed by the recognizer.

4. Organization name recognition also suffers from the erroneous extractions in case of compound organization names. For instance, for the organization name *İstanbul Üniversitesi Siyasal Bilgiler Fakültesi* ('*İstanbul University Political Science Faculty*'), the system extracts two distinct location names of *İstanbul Üniversitesi* and *Bilgiler Fakültesi*.

5. Lastly, an anticipated cause of the performance drop is that some named entities are not covered by the lexical resources and the pattern bases, which results in a certain decrease in the recall of the recognizer.

The financial text data set is close to the news text set in size and evaluation results on financial texts are considerably lower. The main source of the low performance is the frequent occurrence of company names in the text, most of which are not included in the lexical resources and also cannot be extracted using the pattern bases of the recognizer. This situation is revealed as a low recall rate of 50.31% as compared to the considerably better precision rate of 71.47% for the case of the financial news texts.

The performance of the recognizer on child stories is better than its performance on financial texts and historical texts. The main source of performance degradation is the existence of some foreign names annotated in the child stories which cannot be extracted by the recognizer. Actually, the existence of foreign names is a common cause of performance drop for all four data sets, as pointed out previously.

Lastly, similar to the case of financial news texts, the system achieves a very low precision rate when executed on historical texts. The main cause of this result is the homonymy of the names of some empires (organization names) prevalent in the historical text data set with some person or location names. A frequent example case, also addressed in Section 3.1.2, is *Selçuk* which is a common male first name but it is also the name of a historical dynasty ('*the Seljuks*'). Each occurrence of this name is erroneously extracted as a person name by the recognizer instead of an organization name. The comparatively high recall rate on the historical text genre can be attributed to the fact that a considerable proportion of the named entities in the historical text data set are location names and as previously pointed out in Section 3.1.2, location names seem to demonstrate the least diversity across different text genres. Hence, the rule based recognizer achieves a satisfactory performance during the extraction of these location names from the historical text data set, which is revealed as a considerably high overall recall rate for this genre. Above discussion confirms that most of the time the rule based recognizer is prone to low performance on text genres different from generic news texts.

It should be noted that the evaluation results in Table 5.3 are obtained when the recognizer is run without utilizing the capitalization information. When the capitalization feature is turned on, the recognizer extracts those entities included in the lexical resources or those entities satisfying the pattern bases with the capitalization constraint that initial character of each token in the extracted entity should be capitalized. If the initial character of each token in a candidate entity is not capitalized, then this entity is discarded by the recognizer when the capitalization feature is turned on.

The evaluation results on the textual data sets when the capitalization feature is turned on are presented in Figure 5.4. These results reflect that the employment of the capitalization information leads to considerable improvement in precision in all cases as compared to the results presented in Table 5.3 together with a very slight decrease in recall. The increase in precision is an expected result since the homonymity of some common names with some person names in our person name list is a significant cause of the decrease in precision when capitalization is

not utilized. Since proper person names are capitalized but common names are not (apart from those common names which are sentence initial and hence their initial characters are capitalized), employment of the capitalization feature by the recognizer prevents erroneous extractions of these common names as proper person names. Considering the very slight decrease in recall, it is due to the fact that there exist very limited number of named entities in our data sets the initial letters of which are not capitalized although they should be and that these named entities are detectable by the recognizer when the capitalization feature is turned off. Therefore, when the capitalization feature is turned on, these few named entities are eliminated and hence missed by the recognizer leading to the slight decrease in recall.

Table 5.4: Evaluation Results of the Rule Based Named Entity Recognizer on the Textual Data Sets (Capitalization Feature is Turned On).

Data Set	*Precision*	*Recall*	*F-Measure*
News Text Data Set	93.41%	83.12%	87.96%
Financial News Text Data Set	79.02%	50.14%	61.35%
Child Stories Data Set	87.50%	76.47%	81.61%
Historical Text Data Set	79.38%	70.34%	74.59%

Considering the hybrid named entity recognizer, Table 5.5 demonstrates 10–fold cross validation results of the hybrid recognizer on the same data sets. The data sets are randomly divided into 10 equal partitions and in 10 turns, one distinct partition is selected as the test data set and the remaining 9 partitions are used as the training data set for the rote learner component of the hybrid recognizer. The evaluations in each turn are performed after the core rule based recognizer is turned into a hybrid recognizer with the additional lexical resources encompassing entities annotated in the training data set as described in Section 3.1.2. The recall, precision, and F-measure rates provided in Table 5.5 are the averages of the results obtained in these 10 turns.

It should be noted that the training process is separately performed for each data set during the evaluation of the corresponding text genre. To clarify, for each text genre, hence for each text data set, the rote learning component extracts high confidence entities in the training part of the data set and these extracted entities constitute the learned lexical resources of the hybrid recognizer. For instance,

Table 5.5: 10–Fold Cross Validation Results of the Hybrid Named Entity Recognizer on the Textual Data Sets (Capitalization Feature is Turned Off).

Data Set	*Precision*	*Recall*	*F-Measure*
News Text Data Set	85.33%	87.22%	86.25%
Financial News Text Data Set	77.90%	72.06%	74.36%
Child Stories Data Set	77.05%	95.82%	85.34%
Historical Text Data Set	58.16%	79.82%	66.94%

during the evaluation of the historical text data set, the learned resources of the hybrid recognizer only include those entities learned from the training part of the historical data set.

The evaluation results in Table 5.5 demonstrate that the hybrid approach leads to better results as compared to the results of its rule based predecessor in Table 5.3. As the sizes of the data sets are different from each other, we cannot derive a sound conclusion on which data set the recognizer achieves the best improvement. Yet, for all of the text genres, even for the news text data set which belongs to the initial target genre of the recognizer with a modest improvement of 2.07% in F-measure, it can be concluded that the hybrid approach results in considerable improvements in the success rates.

It should be noted that for the news text data set, the precision is almost unchanged when the hybrid recognizer's results in Table 5.5 and that of the rule based recognizer's in Table 5.3 are compared. This finding is due to the fact that the learned resources cause new correct extractions as well as new spurious (erroneous) extractions on the news texts but these extractions do not yield a significant change in the precision although the correct ones result in a significant increase in recall, as recall is not affected by erroneous extractions. Hence, we observe an overall 2.07% improvement in F-measure. Nevertheless, the hybrid recognizer is mainly proposed to alleviate the effects of the porting problem of the rule based recognizer in other text genres distinct from news texts, and the comparison of the results for three other genres in Table 5.5 and Table 5.3 reveals that both the precision and the recall of the hybrid recognizer are considerably better than that of the rule based recognizer for these genres.

Table 5.6: 10–Fold Cross Validation Results of the Hybrid Named Entity Recognizer on the Textual Data Sets (Capitalization Feature is Turned On).

Data Set	*Precision*	*Recall*	*F-Measure*
News Text Data Set	93.38%	87.14%	90.13%
Financial News Text Data Set	83.56%	71.94%	76.80%
Child Stories Data Set	89.72%	95.45%	92.47%
Historical Text Data Set	82.15%	79.53%	80.66%

The evaluation results of the hybrid recognizer when the capitalization information is utilized are provided in Table 5.6. Similar to the point made during the evaluation of the rule based recognizer in the previous subsection, as we compare the results in Table 5.5 and Table 5.6, we can conclude that the employment of the capitalization information results in considerable improvements on the precision and hence on the F-measure of the hybrid recognizer, since recall is almost not affected. Again similar to the points made during the comparison of the cases when the capitalization feature is turned off and on in the previous subsection, slight decreases in recall are observed in Table 5.6 when they are compared to the results in Table 5.5 again due to very few real named entities in the text data sets the initial letters of which are not capitalized although they should be.

Currently, the hybrid named entity recognizer is trained with all the available annotated data and can be executed on the four different text genres of news texts, financial news texts, child stories, and historical texts. However, the current annotated data sets –especially those for child stories and historical texts– are sparse and in order to increase the coverage of the recognizer for these genres, more annotated training data should be fed into the hybrid recognizer.

As emphasized previously, IE research on Turkish texts is rare compared to European languages such as English, German, Spanish as well as Japanese and Chinese. There are even less number of studies on NER in Turkish. While reviewing the relevant literature in order to compare the performance of our rule based recognizer and its hybrid successor, we encounter only two practical studies on NER: the person name extractor for financial news texts in Turkish [31] and the statistical name tagger for Turkish [104]. The former study is par-

ticularly proposed for financial news texts and it extracts person names only, hence it is not a complete named entity recognizer. Moreover, the evaluation of this study [31] requires the analysis of a significant amount of unannotated training data (orders of million tokens) followed by a manual examination phase which is far too time-consuming to be performed in practical settings. These reasons prevent us from making a fair comparison of this person name extractor and our named entity recognizers. To the best of our knowledge, the statistical name tagger proposed in [104] is the sole study that is similar in scope to our named entity recognizer proposals. A main distinction is that the statistical system considers only person, location, and organization names but our recognizers also consider time, date, money, and percent expressions in addition to the former three types. It is reported in [104] that their system is trained on annotated newspaper articles with 492,821 words and evaluated on a test set of 28,000 words, but, since neither the training and testing data sets nor the statistical system itself is available to us to make a comparison of this system with our proposals, we could not evaluate our proposals over this system on the same data set.

Table 5.7: A Qualitative Comparison of the NER Approaches for Turkish Texts.

NER System	*Employed Approach*		*Considered Named Entity Types*					*Target Domain*
	Rule based	*Learning*	*Per.*	*Loc.*	*Org.*	*Num.*	*Time*	
Person Name Extractor [31]	✓	×	✓	×	×	×	×	Financial news texts
Statistical Name Tagger [104]	×	✓	✓	✓	✓	×	×	Domain-independent
Rule Based Named Entity Recognizer	✓	×	✓	✓	✓	✓	✓	News texts
Hybrid Named Entity Recognizer	✓	✓	✓	✓	✓	✓	✓	Domain-independent

A summary of the above discussion regarding the comparison of the proposed approaches is also provided in Table 5.7. In this table, the last two rows correspond to the features of the rule based and hybrid named entity recognizers proposed within the scope of this thesis and the abbreviations of *Per.*, *Loc.*, *Org.*, *Num.*, *Time* are utilized in the table to denote the named entity types of person names, location names, organization names, numeric expressions (money and percent) and temporal expressions (date and time), respectively.

NER Evaluation on the Video Texts:

The transcription texts of the video data sets (*Video Data Set-1* and *Video Data Set-2*) are all in upper case, hence the capitalization feature cannot be utilized during NER evaluation on the video texts. *Video Data Set-1* is mainly utilized as the test data set by the semi-automatic system and when applicable as a training data set for the fully automated system. The evaluation results of the rule based named entity recognizer on *Video Data Set-1* are given in Table 5.8. The results in Table 5.8 are satisfactory considering the fact that capitalization is not utilized.

Table 5.8: Evaluation Results of the Rule Based Named Entity Recognizer on the Text of *Video Data Set-1*.

Precision	*Recall*	*F-Measure*
86.24%	86.77%	86.51%

As we have previously pointed out, named entities and news story segments are independently extracted in the semi-automated system through different modalities hence there is no inherent association between these two information sets although both of them are already associated with the raw video files. In order to address this problem and thereby make the news segment retrieval possible, we have proposed an alignment algorithm in Section 4.3. The named entity-segment alignment procedure presented in Algorithm 2 results in the correct alignment of 86.97% of all the named entities extracted to their news story segments of *Video Data Set-1*.

Table 5.9: Evaluation Results of the Rule Based Named Entity Recognizer on the Text of *Video Data Set-2*.

Data Set	*Precision*	*Recall*	*F-Measure*
Sliding Text	83.92%	80.20%	82.02%
Perfect Transcript	85.42%	87.65%	86.52%

The performance of the rule based named entity recognizer on the automatically corrected video text of *Video Data Set-2*, with the WER of 8.93%, is presented in the first row of Table 5.9. The performance of the recognizer on this video text is satisfactory and as expected considering the WER of the text, that is, the input text is not completely corrected, it is just corrected automatically as described

in the previous section. The evaluation results of the named entity recognizer on perfect transcription of the same data set are given in the second row of Table 5.9. The results in Table 5.9 constitute the first reports of the performance of a named entity recognizer for Turkish on noisy and the corresponding perfect data. A similar performance evaluation is provided in [89] presenting an overview of the first IE evaluation from broadcast news which had been performed as part of a related workshop. The main IE task considered is a multilingual entity extraction task for foreign languages including Spanish, Chinese, and Japanese. It is pointed out that on a perfect transcript, the performance of the best system is 91% in F-measure and when WER is 15%, the performance of the best system is 82% [89].

Video Data Set-1 is utilized as training data to enhance the hybrid named entity recognizer by learning from genuine news video text before the hybrid recognizer is evaluated on *Video Data Set-2*. The performance evaluation of the resulting hybrid recognizer on the text of *Video Data Set-2* is provided in Table 5.10.

When we compare the results in this table and those in Table 5.9, we see that there is an increase of 0.76% in F-measure on the noisy transcript (sliding text) and an increase of 1.41% on the perfect transcript. These increases in performance are, though modest, significant since the whole process is automatic and they also serve to confirm that rote learning could be used to automatically enhance the capabilities of the rule based named entity recognizer when sufficient annotated corpora are available.

Table 5.10: Evaluation Results of the Hybrid Named Entity Recognizer on the Text of *Video Data Set-2*.

Data Set	*Precision*	*Recall*	*F-Measure*
Sliding Text	83.50%	82.07%	82.78%
Perfect Transcript	84.98%	91.10%	87.93%

As the performance of the person entity extraction and coreference resolution tasks described in Section 3.3 depends on the performance of the NER task, we do not carry out any separate evaluation procedures for these tasks.

5.2.3.2 Evaluation of the Automatic Hyperlinker

The Web alignment procedure of the semantic video annotation and retrieval system, described in Section 3.2, is evaluated on the *Video Data Set-2* utilizing the extracted named entities. At the end of Web crawling, a total of 239 Web news are determined for the 149 of 183 segments of the video data set output by the news story segmenter module. Similar to the evaluation of the Web alignment task in [30], we assess the alignment of the news stories and the Web articles using three classes: *Good*, *Fair*, and *Bad*. *Good* means both the Web news article and the associated news story are on the same event but the Web article may include more details. *Fair* corresponds to the case where the main topics in both sources are similar but the actual events are not exactly the same and *Bad* means that the Web article and news story are on different topics. The classification process has taken about 1.5 hours for the 239 URLs determined. We provide the precision and coverage graphs of the Web alignment procedure in Figure 5.2 against the confidence levels. Here, each confidence level denotes all Web articles with at least the specified confidence value. For instance, in case of the confidence level of 0.7, all Web articles having 0.7 or higher confidence are considered. The precision for the *Fair* class is the ratio between the number of Web news classified as *Fair* or *Good* among all news stories for each confidence level. The precision for the *Good* class is the ratio between the number of Web news classified as *Good* among all news stories for each confidence level. Coverage is calculated as the number Web news articles detected for each confidence level over all of the news story segments. As also pointed out in [30], we cannot employ recall measure in this evaluation as it necessitates manual classification of all Web news articles published by *Milliyet* within the broadcast date interval of *Video Data Set-2*.

The precision of the Web alignment procedure is satisfactory, especially for the *Fair* case, for a first attempt to employ this procedure for Turkish news videos. For both cases, there is slight variation in precision especially after reaching its peak value around the confidence levels of 0.7 and 0.8. As we increase our confidence above these levels, the total number of Web articles with these high

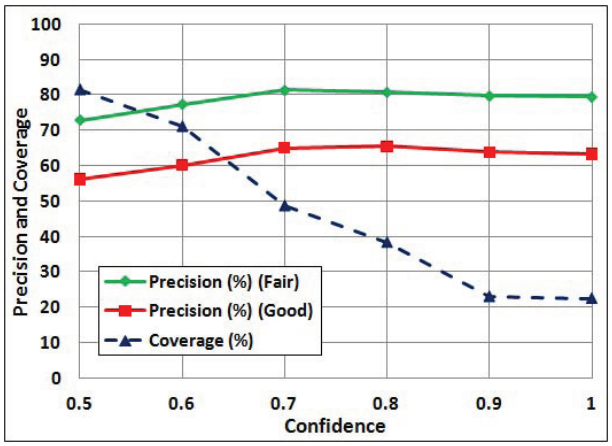


Figure 5.2: The Evaluation Results of the Web Alignment Procedure.

confidences decrease as well as the number of *Fair* or *Good* articles among them. This observation may be attributed to the fact that the ratio of the Web articles classified as *Fair* or *Good* over all Web articles is almost equal for all the considered confidence levels, especially above the levels of 0.7 (for *Good* case) and 0.8 (for *Fair* case) where peak precision values are attained. Put another way, increasing the confidence levels above those corresponding to peak precision values results in the loss of those *Fair* or *Good* articles which have lower confidence and hence precision rates remain almost unchanged.

The results in Figure 5.2 also denote significant error rates which actually lead to noise in our information sources, hence we keep only those URLs classified as *Good*, similar to the manual alignment correction procedure employed in [47]. As mentioned previously, the classification process took about 1.5 hours in our case, which is an acceptable time period for this procedure. But it should be noted that this manual correction procedure is only a decision to be taken or not, and hence if it is not taken, we still keep the automatic nature of the system with some noise introduced into the automatically extracted semantic information.

The resulting Web news articles are associated with the corresponding news story segments. The contents of the aligned Web articles are also fed into the named entity recognizer of the system and the extracted entities are associated with the related news story segments.

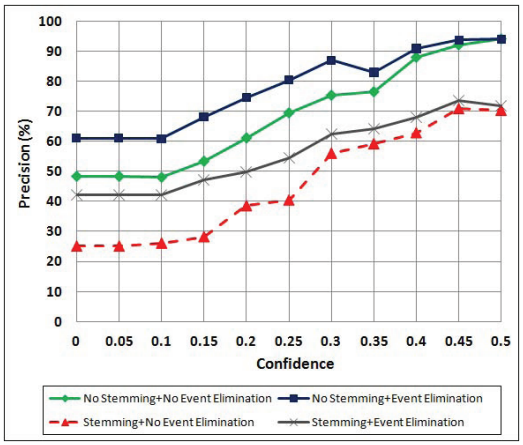
5.2.3.3 Evaluation of the Event Extractor

During the evaluation of keyword based event extraction on *Video Data Set-2*, 16 distinct evaluations are performed utilizing combinations of the values for the following parameters:

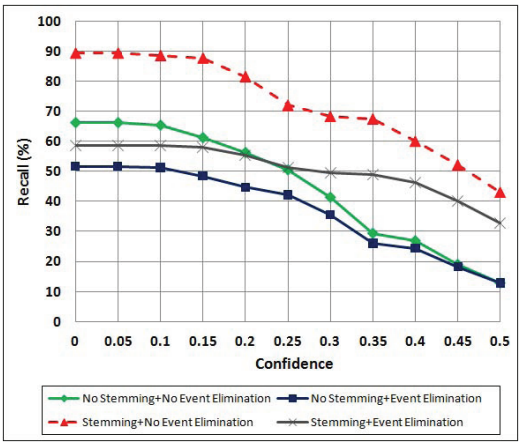
- *The Test Data Set*: Noisy text (sliding text recognizer output after the two-phase automatic correction procedure) or clean text (perfect manual transcription).
- *The Size of the Keywords Utilized*: The most frequent 5 (top-5) or the most frequent 10 (top-10) keywords.
- *Stemming*: Utilizing the keywords as they are or utilizing their stems.
- *The Utilization of Co-occurrence Patterns of Events in the Training Data Set or Not*: This parameter is based on the idea that among the events extracted from a news story, low confidence ones which do not appear together with the higher confidence events in any news story in the training data set may be eliminated. Therefore, the utilization of the co-occurrence patterns of events entails the elimination of such low confidence events extracted from the test data set, as a post-processing step.

The most frequent 10 event types automatically determined from the training data set (*Video Data Set-1*) as described in Section 3.4 include *Statement*, *Death*, *Trial/Investigation*, *Crash*, *Weather*, *Meeting*, *Attack*, *Injury*, *Election*, and *Operation*. The most frequent keywords for these event types are previously given in Table 3.1 in Section 3.4. The text of *Video Data Set-2* is annotated with the events conveyed in the news stories of the videos in this set. For the 182 news stories of this data set, a total of 362 events are annotated, hence similar to the annotation of the training data, each news story may be annotated with more than one event. The breakdown of these 362 events among the 55 distinct event types is provided in Table B.2 in Appendix B.

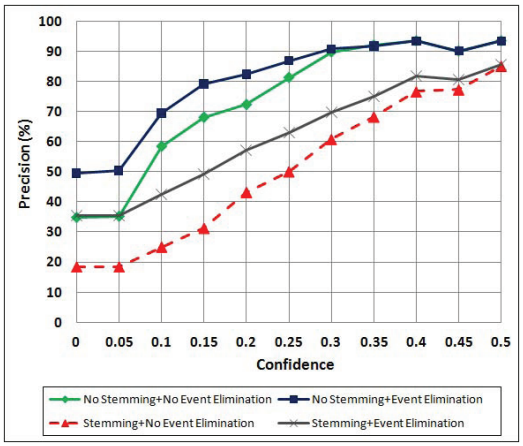
The evaluation results for the distinct parameter settings are presented as distinct precision and recall graphs in Figure 5.3(a) to 5.4(b) for each confidence



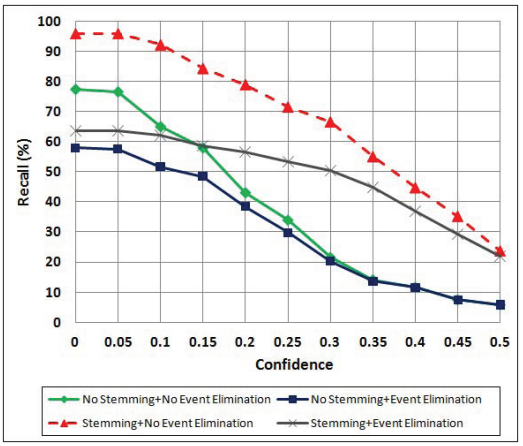
(a) Precision Rates Using Top-5 Keywords on the Noisy News Video Texts.



(b) Recall Rates Using Top-5 Keywords on the Noisy News Video Texts.



(c) Precision Rates Using Top-10 Keywords on the Noisy News Video Texts.

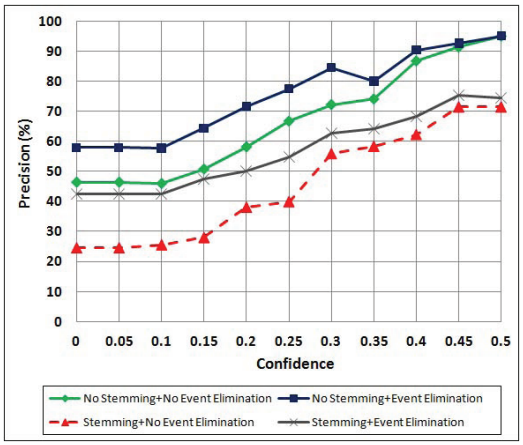


(d) Recall Rates Using Top-10 Keywords on the Noisy News Video Texts.

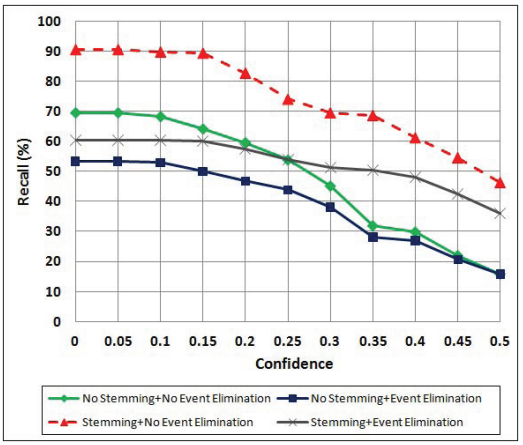
Figure 5.3: The Evaluation Results of Event Extraction from the Noisy News Video Texts.

level from 0 to 0.5. The evaluation results are promising for a first attempt to address automatic semantic event extraction from news videos in Turkish.

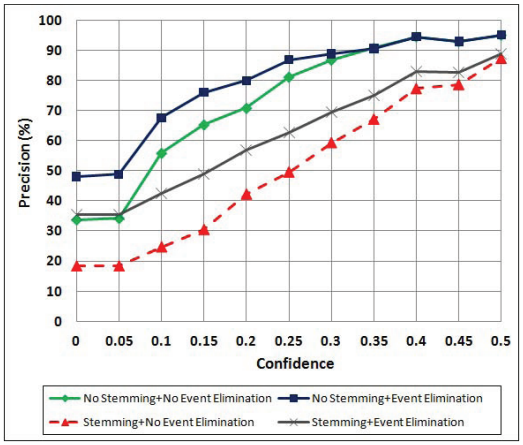
When the graphs in Figure 5.3(a) to 5.4(b) are examined, it is observed that recall values on clean texts are slightly higher than those on noisy texts and precision values are slightly lower for most confidence levels. As we expected, utilizing top-5 keywords against top-10 keywords, using tokens as keywords against using the stems, and the utilization of co-occurrence patterns of events in the training data set, indicated as *Event Elimination* in the figures, against non-utilization of the patterns increase the precision of event extraction, however their employment leads to decrease in the recall of the procedure. It should be noted that the



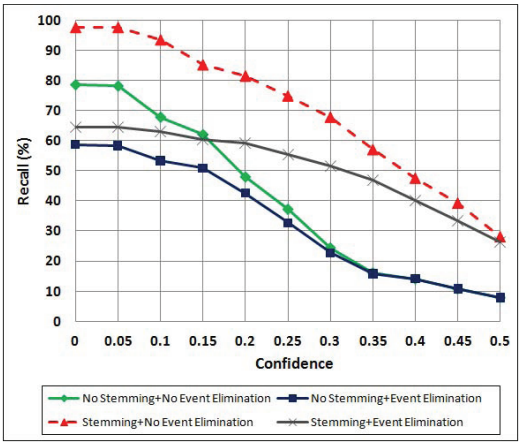
(a) Precision Rates Using Top-5 Keywords on the Clean News Video Texts.



(b) Recall Rates Using Top-5 Keywords on the Clean News Video Texts.



(c) Precision Rates Using Top-10 Keywords on the Clean News Video Texts.



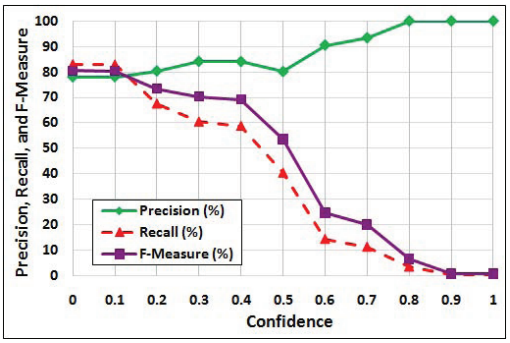
(d) Recall Rates Using Top-10 Keywords on the Clean News Video Texts.

Figure 5.4: The Evaluation Results of Event Extraction from the Clean News Video Texts.

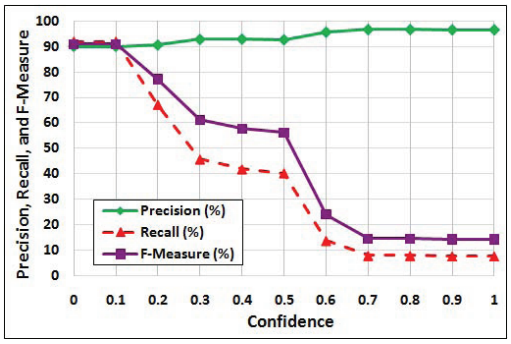
relevant keywords for the 10 considered events are determined by using *Video Data Set-1*, which has a total duration of 4 hours, as the training data set. This is actually a limited set for the task and if a training data set of larger size can be accessed, we expect the results in Figure 5.3(a) to 5.4(b) to get improved.

There are not many studies on semantic event extraction from video texts in Turkish. The sole study that we come across on the topic is presented in [103] where the authors present a rule based event extractor for football Webcast texts. They consider 90 texts corresponding to 90 football matches from Sporx [17] (a sports news provider site) and use 70 of them to engineer the required rules for the 17 event types and the remaining 20 of them are used during the testing of the

rule based event extractor [103]. The considered event types are *Corner*, *Corner-Kick*, *Injury*, *Foul*, *Free-Kick*, *Goal*, *Offside*, *Out (Goal-Kick)*, *Out (Throw-In)*, *Penalty*, *Penalty-Kick*, *Red Card*, *Save*, *Shot*, *Substitution*, *Throw-In*, and *Yellow Card*. It is reported in [103] that the rule based event extractor achieves an overall precision of 97%, a recall of 85%, and hence an F-measure rate of 90.4% on the Webcast texts of the 20 football matches. To compare our keyword based approach to event extraction with this study, we have followed the same strategy and annotated the Webcast texts of 70 football matches from Sporx (the training set) to arrive at the frequent keywords for the 14 events that we consider since we merge the pairs of events (*Corner*, *Corner-Kick*), (*Penalty*, *Penalty-Kick*), and (*Out (Goal-Kick)*, *Out (Throw-In)*) into the single events of *Corner*, *Penalty*, and *Out*, respectively, as the events in these pairs are semantically very close to each other compared to the remaining events in the whole list of 17 types. This time, as the event text groups in the training set are almost disjoint, i.e. each line of the texts usually denotes a single football event, we use the exact $tf \times idf$ weights to order the keywords in each event group as it is done in [114]. We use the top-4 keywords (or top-3 keywords since among the shared keywords for different event types we eliminate the ones with lower confidence) for each event type as the frequent keywords and calculate their normalized confidence values. Next, we test the ultimate event extractor on the text of our test data set corresponding to another 20 football matches and the precision and recall graphs are given in Figure 5.5(a). As can be observed from the figure, the highest rates are achieved when all extracted events (0-confidence level in Figure 5.5(a)) or those events with at least 0.1 confidence are (0.1-confidence level in Figure 5.5(a)) are considered where precision is 77.7%, recall is 83%, and F-measure rate is 80.3% for these levels. Above these confidence levels, precision rates start to increase and the recall rates start to decrease, as expected. Also expected is the fact that these results are lower than those evaluation results reported in [103]. This result can be attributed to the fact that our approach is a fully automated keyword based approach while the approach in [103] relies on manually engineered rules for each event type and manually built systems for specific domains usually achieve better performance compared to fully automated systems. But our approach is easily adaptable to other



(a) The Evaluation Results of Automatic Event Extraction.



(b) The Evaluation Results of Semi-Automatic Event Extraction.

Figure 5.5: The Evaluation Results of Event Extraction from Football Webcasting Texts.

domains provided that annotated data is available while the manual approach is too labor-intensive and time-consuming to be adapted to other domains such as generic news texts.

When the automatically extracted event keywords are manually organized again by a native domain expert and hence make the presented event extraction scheme a semi-automatic one, the results get considerably improved as shown in Figure 5.5(b). This manual intervention includes: (i) rearranging the order and confidence values of the automatically determined keywords, (ii) extending the top-4 keywords by adding new keywords from the ordered set of keywords, (iii) adding compound keywords by merging several keywords in the ordered set of keywords, and (iv) adding new keywords with negative confidence values for some events to make use of those keywords which do not coexist with the considered event (for instance, adding the keyword *kırmızı* ('*red*') with a negative confidence to the *YellowCard* event as this keyword is a sign of a *RedCard* event). Using this ultimate semi-automatic approach, for the confidence levels up to 0.1, the precision and recall rates are 90.1% and 91.9%, respectively, hence an F-measure of 91% is achieved. This is a comparatively less time-consuming and less labor-intensive approach than the manual rule based approach [103] and its F-measure rate is slightly better than that of the manual approach.

CHAPTER 6

EMPLOYMENT OF THE INFORMATION EXTRACTION COMPONENTS FOR MULTILINGUAL VIDEO RETRIEVAL: AN APPLICATION

So far, we have presented our IE components for Turkish and demonstrated their employment within the context of automatic semantic indexing and retrieval of news videos in Turkish. The proposed IE components may also be utilized for semantic indexing and retrieval of video archives in other languages as well provided that we have access to (i) a convenient ASR system for the language of the considered video archive, and (ii) a machine translation (MT) system with an acceptable accuracy from the video language to Turkish. The developed IE components for Turkish then can readily be employed on the output of the MT system and the videos can be annotated with the extracted information. Thereby, multilingual video archives can be semantically indexed and accessed through the extracted semantic information in Turkish.

In this chapter, we will describe an application of the above idea on videos in English by employing an ASR tool for English together with an English-to-Turkish MT system to extract Turkish transcriptions of the original videos. The application involves our named entity recognizer as the IE component to be executed on these transcriptions.

The system corresponding to the implementation of this approach is shown schematically in Figure 6.1.

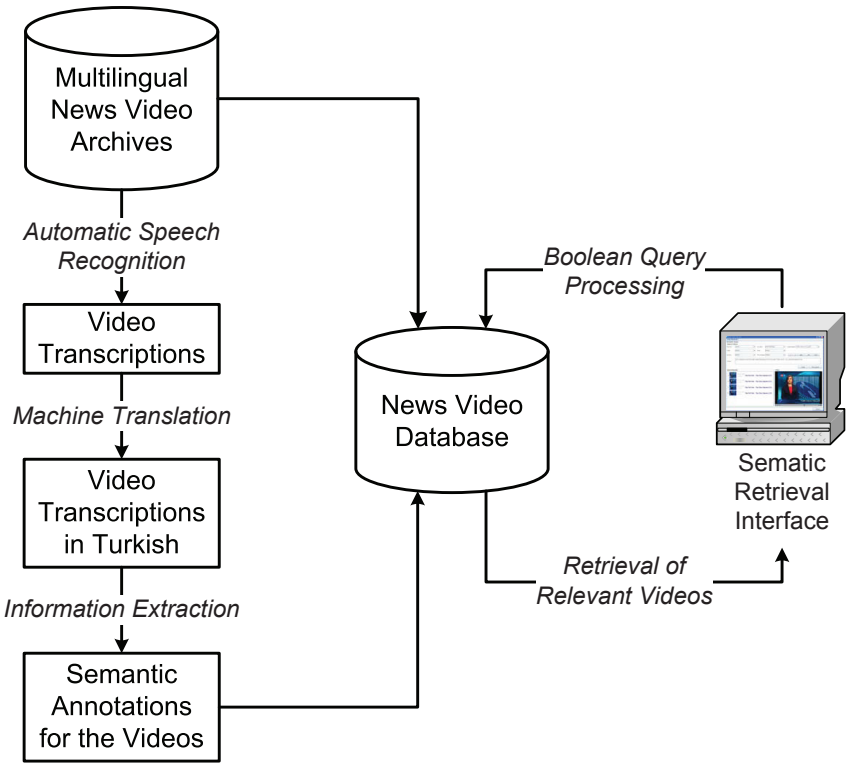


Figure 6.1: The System for Multilingual Video Indexing and Retrieval Employing the Proposed IE Components.

In the following subsections, we first describe the components of the proposed system and next we present information on a sample video data set in English together with the evaluation results of our hybrid named entity recognizer on the Turkish transcriptions of the videos in the set. Finally, we illustrate the retrieval of the videos in English with a query example through the interface of the system in the last subsection.

6.1 Main Components of the System

Apart from the named entity recognizer as the IE component along with the news video database and the semantic retrieval interface which have been described in detail in Chapter 3 and Chapter 4, the multilingual video retrieval system in Figure 6.1 basically encompasses two components: an ASR system to transcribe the videos in English and an MT system to translate the transcriptions from English to Turkish. These two components are briefly described below:

- *The ASR System*: The ASR system employed is the Sphinx system which carries out continuous speech recognition in a speaker independent manner [75]. The Sphinx system has prebuilt language models for several languages including English, Chinese, and French [3]. Hence, it is readily utilized to transcribe videos our video data set in English to be described in the following subsection. The particular version of the ASR system employed is PocketSphinx [57] which reportedly achieves WER values ranging between 9.73–13.95% on a 994-word task where better rates are obtained when speed is sacrificed and lower rates are obtained when the system is improved to execute faster.

- *The MT System*: An English to Turkish MT system [64] is employed to translate the transcriptions to Turkish. The current BLEU score [86] of this MT system is 20.

6.2 Evaluation and Discussion

The video data set on which the presented application is executed comprises 23 videos broadcasted by Youtube[20] which belong to the category of *News & Politics*. The total duration of these videos is about 1 hour. The videos are first given as input to the ASR system described in the previous section and their transcriptions in English are obtained. Next, these transcriptions are translated to Turkish by the English-to-Turkish MT system and hence we arrive at the speech transcriptions of the original videos in Turkish. The total number of tokens in the resulting text is 6,549 and the number of named entities in the text is 270 (42 person, 123 location, and 44 organization names, 52 date and 9 money expressions, with no instances of time or percent expressions). Overall evaluation results of our hybrid NER system on this text are given in Table 6.1 and the evaluation results for each named entity type are provided in Table 6.2.

Table 6.1: Evaluation Results of the Hybrid NER System on the Turkish Transcriptions of the Video Data Set in English.

Precision	*Recall*	*F-Measure*
58.20%	72.34%	64.50%

Table 6.2: Evaluation Results of the Hybrid NER System on the Turkish Transcriptions of the Video Data Set in English for Each Named Entity Type.

Named Entity Type	*Precision*	*Recall*	*F-Measure*
Person	14.14%	32.53%	19.70%
Location	79.92%	84.49%	82.14%
Organization	41.94%	44.83%	43.33%
Date	92.38%	94.17%	93.27%
Money	100%	100%	100%

The overall evaluation results on the translations of the transcriptions of the videos in English are comparatively lower than those results on genuine news video texts in Turkish. This is an expected result basically due to the proliferation of the foreign names uttered in the videos which are not covered by our named entity recognizer for Turkish. A deeper examination of the results in Table 6.2 reveals that the performance is especially hurt when person and organization names are considered. The results are particularly low for person name extraction since foreign person names are not covered by our recognizer (except the names of some well-known political figures) and also there are quite many common names in the translations which are erroneously extracted as person names which is already previously pointed out in Section 5.2.3.1 as a source of the decrease in precision. The performance of the recognizer during the extraction of location names as well as money and date expressions is superior and comparable to the results on genuine Turkish texts presented in Section 5.2.3.1. Though the performance of the NER system on the translations is comparatively lower than its performance on original transcriptions of videos in Turkish, the former results can be improved by extending the lexical resources and the pattern bases utilized by the NER system to cover common foreign named entities.

6.3 A Multilingual Query Example

We store production metadata and named entity information regarding the video data set overviewed in the previous section to the news video database. A query example over this video data set through the semantic retrieval interface of the system is provided in Figure 6.2.

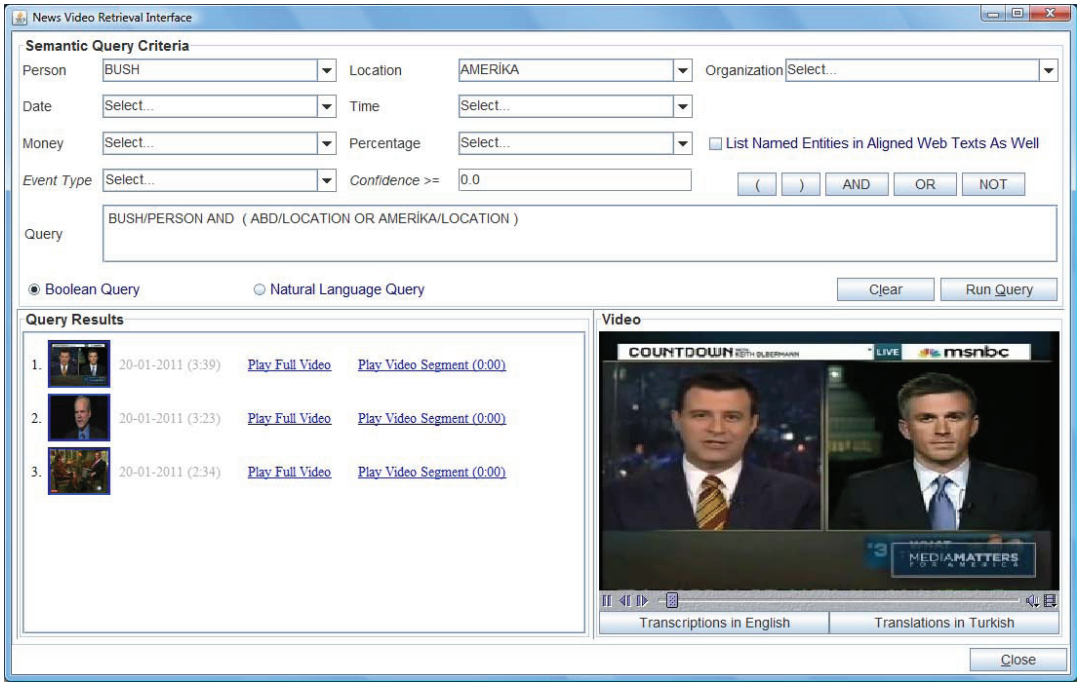


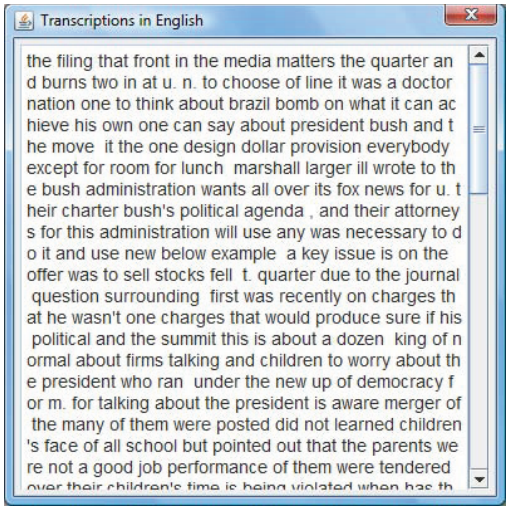
Figure 6.2: A Boolean Query Example Over the Video Data Set in English Through the Semantic Retrieval Interface.

The boolean query illustrated in Figure 6.2 includes the named entities of *Bush* (former president of the USA), *ABD* ('*the USA*'), and *Amerika* ('*America*') which are combined with boolean operators. This query is also provided below and the satisfying videos are listed on the corresponding result panel.

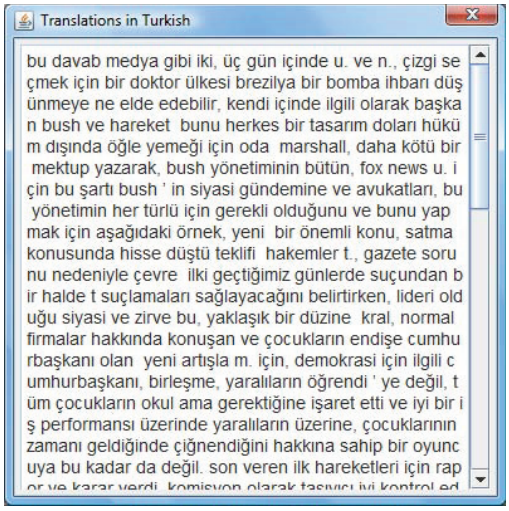
```
BUSH/PERSON AND (ABD/LOCATION OR AMERİKA/LOCATION)
```

We have extended the existing interface to enable the users to examine the English transcriptions of the videos and their translations in Turkish through the corresponding buttons at the bottom of the video panel, as illustrated in Figure 6.3(a) and Figure 6.3(b). With this extension, the users who do not know the language of the original videos (English, in this case) will also be benefiting from these videos through the interface.

To sum up, the named entities extracted from the Turkish translations of the transcriptions of the videos in English provide a plausible means to access these videos over this semantic information. If ASR systems for some other languages and MT systems from these languages to Turkish can be supplied, then this



(a) English Transcriptions.



(b) Translations in Turkish.

Figure 6.3: English Transcriptions and Their Translations in Turkish Corresponding to the Selected Video in Figure 6.2.

system can be implemented for video archives in these languages as well, hence the overall system facilitates multilingual semantic video retrieval.

CHAPTER 7

CONCLUSION AND FUTURE WORK

Automatic semantic annotation of large news video archives is an important research problem in multimedia processing. As audio-visual features automatically extracted from the videos are usually far from being sufficient for semantic video indexing, video texts –if available– stay as a viable candidate information source from which semantic information regarding the videos under consideration can readily be extracted.

In this thesis, we propose a text-based fully automated system for semantic annotation and retrieval of the news videos in Turkish. The system makes use of several IE techniques including NER with date normalization, automatic hyperlinking, person entity extraction with coreference resolution, and event extraction. The outputs of the components implemented for these techniques are utilized as the semantic annotations of the underlying video archives and users can access the videos over these annotations. Apart from these IE components, the proposed system comprises a news video database, a news story segmenter, a sliding text recognizer, and a semantic retrieval interface. Also proposed within the course of this study is a semi-automatic version of the fully automated system for cases where the video texts cannot be obtained automatically. Novel components to carry out the aforementioned IE tasks are implemented and seamlessly integrated into the ultimate systems. For the evaluation of the components of the proposed systems, annotation and evaluation tools are implemented and genuine video corpora broadcasted by Turkish Radio and Television Corporation are used. Moreover, considerable amount of textual corpora as well as the video

texts of the aforementioned video corpora are annotated and used to train and evaluate the components of the systems. The performance evaluations of the components of the systems both on noisy and clean video texts are presented.

The current study is significant since, to the best of our knowledge, it proposes the first fully automated system for semantic annotation and retrieval of news videos in Turkish. The proposed fully automated and semi-automatic systems are also quite generic and by replacing the necessary components of these systems with the necessary counterparts for other languages, the systems can well be employed for videos in these languages. Moreover, novel approaches are proposed and implemented for various IE tasks on Turkish texts which provide a baseline upon which other approaches to the considered IE tasks can be built and compared with the proposed approaches. Finally, we also utilize our hybrid named entity recognizer, news video database, and semantic retrieval interface along with an ASR system for English and an MT system from English to Turkish to facilitate multilingual semantic video retrieval.

Future work based on this thesis include the following:

- The hybrid named entity recognizer for Turkish can be extended to extract other named entity types distinct from the basic types of person, location, organization names and time, date, money, and percent expressions. The extended types may include nationality names, URLs, and e-mail addresses among others. Named entity ontologies such as the one proposed in [93] can be used during the implementation of this extension. Another direction of research is the integration of similar recognizers for languages such as English to our recognizer, to increase its coverage.

- If considerable amount of annotated data can be made available, pure statistical/supervised methods can be employed for NER and the results can be compared with that of the hybrid named entity recognizer and if the performance of any of the newly implemented approaches turns of to be better than our hybrid recognizer, we can replace the hybrid recognizer in the proposed systems with the new best performing NER implementation.

- Regarding event extraction from news video texts, if the size of the training data set can be increased, a wider range of event types could be included in the scope of the event extraction component which currently considers 10 most frequent events in the training data set. Furthermore, related controlled vocabularies such as the *NewsCodes* of the International Press Telecommunications Council (IPTC) can be used to determine the ultimate semantic event types [7].

- The natural language interface of the implemented system for Turkish news videos can be improved to handle a wider range of queries by integrating deeper language processing tools into the interface.

- We have not carried out any improvements regarding the time and space complexity of the implemented approaches as we currently expect our systems to extract semantic annotations offline. But to scale especially the fully automated system to handle news video archives of considerable size, some implementation improvements should be made to decrease the time and space requirements in order for offline information extraction to be performed in reasonable time/space settings.

- Based on the extracted semantic information from news videos, automatic news categorization and summarization facilities can be implemented and integrated into the proposed semantic indexing and retrieval systems.

- Based on the generic system, similar systems for other languages can be built and their performance can be evaluated on convenient news video corpora so that the applicability of the proposed system can further be tested. Doing so, the system can be evaluated on commonly employed multilingual data sets of the multimedia processing community, such as TRECVID news video data sets [95]. To the best of our knowledge, the existing data sets of TRECVID are in English, Dutch, Arabic, and Chinese [18]. Hence, by building semantic indexing and retrieval systems based on the generic system for the aforementioned languages, experimentation on TRECVID data sets and comparisons with other approaches evaluated on these data sets could be possible.

APPENDIX A

STOPWORD LIST UTILIZED DURING FULL PERSON ENTITY AND EVENT EXTRACTION

Table A.1: Stopword List (Slightly Extended Version of the List Provided in [37]).

adeta	bunlar	etmesi	karşın	olmadı	şey
ama	bunları	etti	kendi	olmadığı	şöyle
ancak	bunların	ettiği	kendilerine	olmak	şu
arada	bunu	ettiğini	kendini	olması	şunları
ayrıca	bunun	gibi	kendisi	olmayan	tarafından
az	burada	göre	kendisine	olmaz	üzere
bana	çok	hala	kendisini	olsa	var
bazı	çünkü	halen	kere	olsun	vardı
belki	da	hangi	kez	olup	ve
ben	daha	hatta	ki	olur	veya
beni	de	hem	kim	olursa	ya
benim	defa	henüz	kimse	oluyor	yani
beri	değil	her	mı	ona	yapacak
bile	diğer	herhangi	mi	onlar	yapılan
bir	diye	herkesin	mu	onları	yapılması
birçok	dolayı	hiç	mü	onların	yapıyor
biri	dolayısıyla	hiçbir	nasıl	onu	yapmak
birkaç	edecek	için	ne	onun	yaptı
biz	eden	ila	neden	oysa	yaptığı
bize	ederek	ile	nedenle	önce	yaptığını
bizi	edilecek	ilgili	o	öyle	yaptıkları
bizim	ediliyor	ilk	olan	pek	yeni
böyle	edilmesi	ise	olarak	rağmen	yerine
böylece	ediyor	işte	oldu	sadece	yine
bu	eğer	itibaren	olduğu	siz	yoksa
buna	en	itibariyle	olduğunu	son	zaten
bundan	eski	kadar	olduklarını	sonra	

APPENDIX B

SEMANTIC EVENTS ANNOTATED IN THE TRAINING AND TEST VIDEO TEXTS

Table B.1: Semantic Events in the Training Video Text Corresponding to 340 News Stories (with Frequencies in Parentheses).

Arrangement (6)	Explosion (2)	Research (1)
Arrest (8)	Festival (3)	Response (1)
Attack (18)	Fight (1)	Restoration (1)
Bid (1)	Fire (4)	Retreat (3)
Ceremony (4)	Flood (2)	Reveal (1)
Claim (1)	Funeral (5)	Sentence (3)
Collaboration (1)	Injury (14)	Set Free (3)
Commemoration (1)	Inquiry (2)	Slaughter (1)
Conflict (4)	Intoxication (6)	Startup (2)
Contest (1)	Invasion (2)	Statement (104)
Corruption (2)	Law Acceptance (1)	Stock-market Decline (6)
Crash (20)	Lottery (4)	Sue (4)
Crisis (9)	Meeting (19)	Suicide Attack (2)
Custody (6)	Murder (3)	Takeover (1)
Deal (2)	Objection (1)	Talk (1)
Death (45)	Official Complaint (1)	Testimony (2)
Disaster (6)	Operation (10)	Traffic Jam (1)
Discovery (2)	Poll (1)	Treatment (4)
Discussion (1)	Price Change (6)	Trial/Investigation (43)
Earthquake (2)	Protest (2)	Unemployment (2)
Election (11)	Punishment (3)	Visit (6)
Epidemic (1)	Put Off (1)	War (1)
Examination (5)	Refusal of Law Change (1)	Weather (19)

Table B.2: Semantic Events in the Test Video Text Corresponding to 182 News Stories (with Frequencies in Parentheses).

Arrangement (1)	Examination (1)	Resignation (4)
Arrest (4)	Explosion (2)	Restoration (1)
Attack (1)	Festival (2)	Sabotage (2)
Attack (14)	Flood (2)	Sentence (1)
Ceremony (6)	Funeral (3)	Set Free (2)
Claim (3)	Immigration (1)	Sham Battle (1)
Collaboration (10)	Injury (16)	Sports (3)
Commemoration (1)	Intoxication (1)	Statement (77)
Conflict (5)	Invasion (1)	Sue (1)
Contest (2)	Kidnap (1)	Suicide Attack (2)
Control (1)	Law Acceptance (6)	Talk (2)
Crash (6)	Meeting (22)	Tourism (1)
Custody (7)	Operation (15)	Treatment (4)
Death (26)	Poll (2)	Trial/Investigation (15)
Delegation (1)	Price Change (5)	Unemployment (1)
Disaster (1)	Protest (6)	Visit (7)
Discovery (1)	Punishment (1)	Weather (5)
Election (46)	Refusal of Law Change (3)	
Establishment (4)	Research (3)	

APPENDIX C

SQL EXPRESSION CORRESPONDING TO THE EXAMPLE QUERY POSED THROUGH THE SEMANTIC RETRIEVAL INTERFACE

```
----Query: SARKOZY/PERSON AND STATEMENT/EVENT[0.1]
--To cover the videos including the specified name as
--a person entity.
(select file_path, broadcast_date, duration_in_seconds,
 video_segment_id, start_second
from t_news_video, t_news_video_segment, t_named_entity
where t_news_video.id = t_named_entity.video_id and
 t_news_video_segment.id = t_named_entity.video_segment_id and
 named_entity_type = 'PERSON' and named_entity = 'SARKOZY'
union
--To cover the videos including entities which are aliases
--of the specified name.
select file_path, broadcast_date, duration_in_seconds,
 video_segment_id, start_second
from t_news_video, t_news_video_segment, t_named_entity
where t_news_video.id = t_named_entity.video_id and
 t_news_video_segment.id = t_named_entity.video_segment_id and
 named_entity_type = 'PERSON' and named_entity in
  (select name from t_person_alias
   where person_id in (select id from t_person_composite
                       where name = 'SARKOZY'))
```

```
union
--To cover the videos including entities for which the
--specified name is an alias.
select file_path, broadcast_date, duration_in_seconds,
 video_segment_id, start_second
from t_news_video, t_news_video_segment, t_named_entity
where t_news_video.id = t_named_entity.video_id and
 t_news_video_segment.id = t_named_entity.video_segment_id and
 named_entity_type = 'PERSON' and named_entity in
  (select name from t_person_composite
   where id in (select person_id from t_person_alias
                where name = 'SARKOZY'))
union
--To cover the videos including entities which are aliases
--of entities having the specified name as an alias.
select file_path, broadcast_date, duration_in_seconds,
 video_segment_id, start_second
from t_news_video, t_news_video_segment, t_named_entity
where t_news_video.id = t_named_entity.video_id and
 t_news_video_segment.id = t_named_entity.video_segment_id and
 named_entity_type = 'PERSON' and named_entity in
  (select name from t_person_alias
   where person_id in (select person_id from t_person_alias
                       where name = 'SARKOZY')))
intersect
--To cover the videos including the specified event with
--at least the specified confidence.
select file_path, broadcast_date, duration_in_seconds,
 video_segment_id, start_second
from t_news_video, t_news_video_segment, t_segment_event_noisy
where t_news_video.id = t_news_video_segment.video_id and
 t_news_video_segment.id = t_segment_event_noisy.segment_id
 and event = 'STATEMENT' and confidence >= 0.1;
```

REFERENCES

[1] Anadolu Agency. http://www.aa.com.tr, accessed 21 February 2011.

[2] Automatic Content Extraction (ACE) Program. http://www.itl.nist.gov/iad/mig/tests/ace/, accessed 21 February 2011.

[3] CMU Sphinx Home Page. http://cmusphinx.sourceforge.net/wiki/, accessed 21 February 2011.

[4] Fobs4JMF. http://fobs.sourceforge.net/, accessed 21 February 2011.

[5] HAREM: NER for Portuguese. http://www.linguateca.pt/HAREM/, accessed 21 February 2011.

[6] IBM Voice Transcription Manager. https://vtm.researchlabs.ibm.com/, accessed 21 February 2011.

[7] IPTC NewsCodes. http://www.iptc.org/site/NewsCodes/, accessed 21 February 2011.

[8] Java SE Desktop Technologies - Java Media Framework API (JMF). http://www.oracle.com/technetwork/java/javase/tech/index-jsp-140239.html, accessed 21 February 2011.

[9] Latent Semantic Analysis - Wikipedia. http://en.wikipedia.org/wiki/Latent_semantic_analysis, accessed 21 February 2011.

[10] MATLAB. http://www.mathworks.com/products/matlab/, accessed 21 February 2011.

[11] Message Understanding Conference (MUC). http://www-nlpir.nist.gov/related_projects/muc/, accessed 21 February 2011.

[12] Milliyet Internet. http://www.milliyet.com.tr, accessed 21 February 2011.

[13] Multilingual Entity Task Conference (MET). http://www-nlpir.nist.gov/related_projects/tipster/met.htm, accessed 21 February 2011.

[14] Named Entity Recognition - Wikipedia. http://en.wikipedia.org/wiki/Named_entity_recognition, accessed 21 February 2011.

[15] PostgreSQL DBMS. http://www.postgresql.org, accessed 21 February 2011.

[16] Speech Recognition - Wikipedia. http://en.wikipedia.org/wiki/Speech_recognition#Performance, accessed 21 February 2011.

[17] Sporx. http://www.sporx.com, accessed 21 February 2011.

[18] TREC Video Retrieval Evaluation Home Page. http://trecvid.nist.gov/, accessed 21 February 2011.

[19] Turkish Radio and Television Corporation. http://www.trt.com.tr, accessed 21 February 2011.

[20] Youtube. http://www.youtube.com/, accessed 21 February 2011.

[21] Zemberek Natural Language Processing Library for Turkic Languages. http://code.google.com/p/zemberek/, accessed 21 February 2011.

[22] Serif Adalı, A. Coşkun Sönmez, and Mehmet Göktürk. An integrated architecture for processing business documents in Turkish. In *Proceedings of the Conference on Text Processing and Computational Linguistics (CICLing)*, pages 394–405, 2009.

[23] W. H. Adams, Giridharan Iyengar, Ching-Yung Lin, Milind Ramesh Naphade, Chalapathy Neti, Harriet J. Nock, and John R. Smith. Semantic indexing of multimedia content using visual, audio and text cues. *EURASIP Journal on Applied Signal Processing*, 2003(2):170–185, 2003.

[24] James Allan, editor. *Topic detection and tracking: Event-based information organization*. Kluwer Academic Publishers, Norwell, MA, USA, 2002.

[25] Douglas E. Appelt and David J. Israel. Introduction to information extraction technology. A tutorial prepared for IJCAI-99, Stockholm, Sweden, 1999.

[26] Ebru Arısoy, Dogan Can, Sıddıka Parlak, Haşim Sak, and Murat Saraçlar. Turkish broadcast news transcription and retrieval. *IEEE Transactions on Audio, Speech, and Language Processing*, 17(5):874–883, 2009.

[27] Joaquim Arlandis, Paul Over, and Wesswl Kraaij. Boundary error analysis and categorization in the trecvid news story segmentation task. In *Proceedings of the 4th International Conference on Image and Video Retrieval (CIVR)*, pages 103–112, 2005.

[28] Pradeep Atrey, M. Anvar Hossain, Abdulmotaleb El Saddik, and Mohan Kankanhalli. Multimodal fusion for multimedia analysis: A survey. *Multimedia Systems*, 16(6):345–379, 2010.

[29] Andrew D. Bagdanov, Marco Bertini, Alberto Del Bimbo, Giuseppe Serra, and Carlo Torniai. Semantic annotation and retrieval of video events using multimedia ontologies. In *Proceedings of the International Conference on Semantic Computing (ICSC)*, pages 713–720, 2007.

[30] Roberto Basili, Marco Cammisa, and Emanuale Donati. RitroveRAI: A Web application for semantic indexing and hyperlinking of multimedia news. In *Proceedings of International Semantic Web Conference (ISWC)*, pages 97–111, 2005.

[31] Özkan Bayraktar and Tuğba Taşkaya Temizel. Person name extraction from Turkish financial news text using local grammar based approach. In *Proceedings of the International Symposium on Computer and Information Sciences (ISCIS)*, pages 1–4, 2008.

[32] Ayşenur Akyüz Birtürk and Sandiway Fong. A modular approach to Turkish noun compounding: The integration of a finite-state model. In *Proceedings of the Natural Language Processing Pacific Rim Symposium (NLPRS)*, pages 525–531, 2001.

[33] Kalina Bontcheva, Valentin Tablan, Diana Maynard, and Hamish Cunningham. Evolving GATE to meet new challenges in language engineering. *Natural Language Engineering*, 10(3–4):349–373, 2004.

[34] Grady Booch, James Rumbaugh, and Ivar Jacobson. *The Unified Modeling Language User Guide*. Addison-Wesley, 2nd edition, 2005.

[35] Darin Brezeale and Diane J. Cook. Automatic video classification: A survey of the literature. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 38(3):416–430, 2010.

[36] Mary Elaine Califf and Raymond J. Mooney. Relational learning of pattern-match rules for information extraction. In *Working Papers of the ACL-97 Workshop in Natural Language Learning*, 1997.

[37] Fazlı Can, Seyit Koçberber, Erman Balçık, Cihan Kaynak, H. Çağdaş Öcalan, and Onur M. Vursavaş. Information retrieval on Turkish texts. *Journal of American Society for Information Science and Technology*, 59(3):407–421, 2008.

[38] Fazlı Can, Seyit Koçberber, Özgür Bağlıoğlu, Süleyman Kardaş, H. Çağdaş Öcalan, and Erkan Uyar. New event detection and topic tracking in Turkish. *Journal of American Society for Information Science and Technology*, 61(4):802–819, 2010.

[39] Jim Cowie and Wendy Lehnert. Information extraction. *Communications of the ACM*, 39(1):80–91, 1996.

[40] Silviu Cucerzan and David Yarowsky. 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*, pages 90–99, 1999.

[41] Hamish Cunningham. Information extraction – a user guide. Technical Report CS-97-07, University of Sheffield, UK, April 1999.

[42] Hamish Cunningham. Information Extraction, Automatic. *Encyclopedia of Language and Linguistics, 2nd Edition*, 2005.

[43] Hamish Cunningham, Diana Maynard, Kalina Bontcheva, and Valentin Tablan. GATE: A framework and graphical development environment for robust NLP tools and applications. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 168–175, 2002.

[44] Thierry Declerck, Jan Kuper, Horacio Saggion, Anna Samiotou, Peter Wittenburg, and Jesus Contreras. Contribution of NLP to the content indexing of multimedia documents. In *Proceedings of the International Conference on Image and Video Retrieval (CIVR)*, pages 610–618, 2004.

[45] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the Americal Society for Information Science*, 41(6):391–407, 1990.

[46] Erinç Dikici and Murat Saraçlar. Sliding text recognition in broadcast news. In *Proceedings of IEEE 16th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4, 2008.

[47] Mike Dowman, Valentin Tablan, Hamish Cunningham, and Borislav Popov. Web-assisted annotation, semantic indexing and search of television and radio news. In *Proceedings of the International Conference on World Wide Web (WWW)*, pages 225–234, 2005.

[48] Witold Drozdzynski, Hans-Ulrich Krieger, Jakub Piskorski, Ulrich Schäfer, and Feiyu Xu. Shallow processing with unification and typed feature structures – foundations and applications. *Künstliche Intelligenz*, 1:17–23, 2004.

[49] Ronen Feldman and James Sanger. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, 2007.

[50] Dayne Freitag. Information extraction from HTML: Application of a general learning approach. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI)*, pages 517–523, 1998.

[51] Dayne Freitag. Machine learning for information extraction in informal domains. *Machine Learning*, 39(2-3):169–202, 2000.

[52] Ralph Grishman. Information extraction. In Ruslan Mitkov, editor, *The Oxford Handbook of Computational Linguistics*, chapter 30. Oxford University Press, 2003.

[53] Ralph Grishman and Beth Sundheim. Message understanding conference-6: A brief history. In *Proceedings 16th International Conference on Computational Linguistics (COLING)*, pages 466–471, 1996.

[54] Lynette Hirschman. MUC-7 coreference task definition. Version 3.0. `http://acl.ldc.upenn.edu/muc7/co_task.html`, 1997, accessed 21 February 2011.

[55] Thomas Hofmann. Probabilistic latent semantic indexing. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 50–57, 1999.

[56] Scott B. Huffman. Learning information extraction patterns from examples. In *Proceedings of the IJCAI Workshop on New Approaches to Learning for Natural Language Processing*, pages 127–142, 1995.

[57] David Huggins-Daines, Mohit Kumar, Arthur Chan, Alan W. Black, Mosur Ravishankar, and Alex I. Rudnicky. Pocketsphinx: A free, real-time continuous speech recognition system for hand-held devices. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2006.

[58] Rıfat Ilgaz. *Bacaksız Kamyon Sürücüsü*. Çınar Publications, 2003.

[59] Rıfat Ilgaz. *Bacaksız Tatil Köyünde*. Çınar Publications, 2003.

[60] Peter Jackson and Isabelle Moulinier. *Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization*. John Benjamins Publishing Company, Amsterdam, The Netherlands, 2nd edition, 2007.

[61] Heng Ji. Information extraction. In M. Tamer Özsu, editor, *Encyclopedia of Database Systems*. Springer Science+Business Media, 2008.

[62] Yılmaz Kılıçaslan, Edip Serdar Güner, and Savaş Yıldırım. Learning-based pronoun resolution for Turkish with a comparative evaluation. *Computer Speech and Language*, 23(3):311–331, 2009.

[63] Jun-Tae Kim and Dan I. Moldovan. Acquisition of linguistic patterns for knowledge-based information extraction. *IEEE Transactions on Knowledge and Data Engineering*, 7(5):713–724, 1995.

[64] Selçuk Köprü and Adnan Yazıcı. Lattice parsing to integrate speech recognition and rule-based machine translation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pages 469–477, 2009.

[65] Dilek Küçük, N. Burcu Özgür, Adnan Yazıcı, and Murat Koyuncu. A fuzzy conceptual model for multimedia data with a text-based automatic annotation scheme. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 17(Supp01):135–152, 2009.

[66] Dilek Küçük and Meltem Turhan Yöndem. A knowledge-poor pronoun resolution system for Turkish. In *Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium (DAARC)*, pages 59–64, 2007.

[67] Dilek Küçük and Adnan Yazıcı. Identification of coreferential chains in video texts for semantic annotation of news videos. In *Proceedings of the International Symposium on Computer and Information Sciences (ISCIS)*, pages 1–6, 2008.

[68] Dilek Küçük and Adnan Yazıcı. Employing named entities for semantic retrieval of news videos in Turkish. In *Proceedings of the International Symposium on Computer and Information Sciences (ISCIS)*, pages 153–158, 2009.

[69] Dilek Küçük and Adnan Yazıcı. Named entity recognition experiments on Turkish texts. In *Proceedings of the International Conference on Flexible Query Answering Systems (FQAS)*, pages 524–535, 2009.

[70] Dilek Küçük and Adnan Yazıcı. Rule-based named entity recognition from Turkish texts. In *Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*, pages 456–460, 2009.

[71] Dilek Küçük and Adnan Yazıcı. A hybrid named entity recognizer for Turkish with applications to different text genres. In *Proceedings of the International Symposium on Computer and Information Sciences (ISCIS)*, pages 113–116, 2010.

[72] Thomas L. Landauer, Peter W. Foltz, and Darrel Laham. An introduction to latent semantic analysis. *Discourse Processes*, 25:259–284, 1998.

[73] Gal Lavee, Ehud Rivlin, and Michael Rudzsky. Understanding video events: A survey of methods for automatic interpretation of semantic occurrences in video. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 39(5):489–504, 2009.

[74] Alberto Lavelli, Mary Califf, Fabio Ciravegna, Dayne Freitag, Claudio Giuliano, Nicholas Kushmerick, Lorenza Romano, and Neil Ireson. Evaluation of machine learning-based information extraction algorithms: Criticisms and recommendations. *Language Resources and Evaluation*, 42(4):361–393, 2008.

[75] Kai-Fu Lee and Raj Reddy. *Automatic Speech Recognition: The Development of the Sphinx Recognition System*. Kluwer Academic Publishers, Norwell, MA, USA, 1988.

[76] Elaine Marsh and Dennis Perzanowski. MUC-7 Evaluation of IE technology: Overview of results. In *Proceedings of the 7th Message Understanding Conference (MUC-7)*, 1998.

[77] Diana Maynard, Valentin Tablan, Cristan Ursu, Hamish Cunningham, and Yorick Wilks. Named entity recognition from diverse text types. In *Proceedings of the Conference on Recent Advances in Natural Language Processing (RANLP)*, 2001.

[78] Alberto Messina and Maurizio Montagnuolo. A generalised cross-modal clustering method applied to multimedia news semantic indexing and retrieval. In *Proceedings of the 18th International Conference on World Wide Web (WWW)*, pages 321–330, 2009.

[79] Ruslan Mitkov. *Anaphora Resolution*. Longman (Pearson Education), Edinburgh, UK, 1st edition, 2002.

[80] Ruslan Mitkov, Richard Evans, Constantin Orasan, Catalina Barbu, Lisa Jones, and Violeta Sotirova. Coreference and anaphora: Developing annotating tools, annotated resources and annotation strategies. In *Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium (DAARC)*, pages 49–58, 2000.

[81] Marie-Francine Moens. *Information Extraction: Algorithms and Prospects in a Retrieval Context*. Springer, Dordrecht, The Netherlands, 1st edition, 2006.

[82] Ion Muslea. Extraction patterns for information extraction tasks: A survey. In *Proceedings of the AAAI Workshop on Machine Learning for Information Extraction*, pages 1–6, 1999.

[83] David Nadeau and Satoshi Sekine. A survey of named entity recognition and classification. *Lingvistica Investigationes*, 30(1):3–26, 2007.

[84] Milind Naphade, John R. Smith, Jelena Tesic, Shih-Fu Chang, Winston Hsu, Lyndon Kennedy, Alexander Hauptman, and Jon Curtis. Large-scale concept ontology for multimedia. *IEEE Multimedia*, 13(3):86–91, 2006.

[85] Mary S. Neff, Roy J. Byrd, and Branimir K. Boguraev. The Talent system: TEXTRACT architecture and data model. *Natural Language Engineering*, 10(3–4):307–326, 2004.

[86] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL)*, pages 311–318, 2002.

[87] Borislav Popov, Atanas Kiryakov, Damyan Ognyanoff, Dimitar Manov, and Angel Krilov. KIM - a semantic platform for information extraction and retrieval. *Natural Language Engineering*, 10(3–4):375–392, 2004.

[88] Ellen Riloff. Automatically constructing a dictionary for information extraction tasks. In *Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI)*, pages 811–816, 1993.

[89] Patricia Robinson, Erica Brown, John Burger, Nancy Chinchor, Aaron Douthat, Lisa Ferro, and Lynette Hirschman. Overview: Information extraction from broadcast news. In *Proceedings of DARPA Broadcast News Workshop*, pages 27–30, 1998.

[90] Horacio Saggion, Hamish Cunningham, Kalina Bontcheva, Diana Maynard, Oana Hamza, and Yorick Wilks. Multimedia indexing through multi-source and multi-language information extraction: The MUMIS project. *Data and Knowledge Engineering*, 48(2):247–264, 2004.

[91] Gerard Salton. *Automatic text processing: The transformation, analysis, and retrieval of information by computer*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989.

[92] Bilge Say, Deniz Zeyrek, Kemal Oflazer, and Umut Özge. Development of a corpus and a treebank for present-day written Turkish. In *Proceedings of the 11th International Conference of Turkish Linguistics (ICTL)*, 2002.

[93] Satoshi Sekine. Extended named entity ontology with attribute information. In *Proceedings of the Language Resources and Evaluation Conference (LREC)*, 2008.

[94] Satoshi Sekine and Yoshio Eriguchi. Japanese named entity extraction evaluation: Analysis of results. In *Proceedings of the 18th conference on Computational Linguistics (COLING)*, pages 1106–1110, 2000.

[95] Alan F. Smeaton, Paul Over, and Wessel Kraaij. Evaluation campaigns and TRECVid. In *Proceedings of the ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, 2006.

[96] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta, and Ramesh Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, 2000.

[97] Cees G. M. Snoek and Marcel Worring. Multimodal video indexing: A review of the state-of-the-art. *Multimedia Tools and Applications*, 25(1):5–35, 2005.

[98] Cees G. M. Snoek, Marcel Worring, Jan C. van Gemert, Jan-Mark Geusebroek, and Arnold W. M. Smeulders. The challenge problem for automated detection of 101 semantic concepts in multimedia. In *Proceedings of the 14th ACM International Conference on Multimedia (MULTIMEDIA)*, pages 421–430, 2006.

[99] Ahmet Hamdi Tanpınar. *Beş Şehir*. Dergah Publications, 2007.

[100] Hayssam N. Traboulsi. *Named Entity Recognition: A Local Grammar-based Approach*. PhD thesis, Department of Computing, University of Surrey, 2006.

[101] Pınar Tüfekçi and Yılmaz Kılıçaslan. A syntax-based pronoun resolution system for Turkish. In *Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium (DAARC)*, pages 139–144, 2007.

[102] Pınar Tüfekçi, Dilek Küçük, Meltem Turhan Yöndem, and Yılmaz Kılıçaslan. Comparison of a syntax-based and a knowledge-poor pronoun resolution system for Turkish. In *Proceedings of the International Symposium on Computer and Information Sciences (ISCIS)*, 2007.

[103] Doruk Tunaoğlu, Özgür Alan, Orkunt Sabuncu, Samet Akpınar, Nihan K. Çiçekli, and Ferda N. Alpaslan. Event extraction from Turkish football Web-casting texts using hand-crafted templates. In *Proceedings of the International Conference on Semantic Computing (ICSC)*, pages 466–472, 2009.

[104] Gökhan Tür, Dilek Hakkani-Tür, and Kemal Oflazer. A statistical information extraction system for Turkish. *Natural Language Engineering*, 9(2):181–210, 2003.

[105] Jordi Turmo, Alicia Ageno, and Neus Catala. Adaptive information extraction. *ACM Computing Surveys*, 38(2):1–47, 2006.

[106] Cornelis Joost van Rijsbergen. *Information Retrieval*. Butterworth, London, 1979.

[107] Hal R. Varian. Universal access to information. *Communications of the ACM*, 48(10):65–66, 2005.

[108] Charles L. Wayne. Multilingual topic detection and tracking: Successful research enabled by corpora and evaluation. In *Proceedings of the Language Resources and Evaluation Conference (LREC)*, 2000.

[109] Changsheng Xu, Yifan Zhang, Guangyu Zhu, Yong Rui, Hanqing Lu, and Qingming Huang. Using Webcast text for semantic event detection in broadcast sports video. *IEEE Transactions on Multimedia*, 10(7):1342–1355, 2008.

[110] Rong Yan and Alexander Hauptman. A review of text and image retrieval approaches for broadcast news video. *Information Retrieval*, 10(4–5):445–484, 2007.

[111] Roman Yangarber, Ralph Grishman, Pasi Tapanainen, and Silja Huttunen. Automatic acquisition of domain knowledge for information extraction. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, pages 940–946, 2000.

[112] Yakup Yıldırım, Turgay Yılmaz, and Adnan Yazıcı. Ontology-supported object and event extraction with a genetic algorithms approach for object classification. In *Proceedings of the ACM International Conference on Image and Video Retrieval (CIVR)*, pages 202–209, 2007.

[113] Dongqing Zhang, Belle L. Tseng, Ching-Yung Lin, and Shih-Fu Chang. Accurate overlay text extraction for digital video analysis. In *Proceedings of IEEE International Conference on Information Technology: Research and Education (ITRE)*, pages 233–237, 2003.

[114] Yifan Zhang, Changsheng Xu, Yong Rui, Jinqiao Wang, and Hanqing Lu. Semantic event extraction from basketball games using multimodal analysis. In *Proceedings of the IEEE Conference on Multimedia and Expo (ICME)*, pages 2190–2193, 2007.

VITA

Dilek Küçük was born in Giresun, Turkey in 1982. She received her B.Sc. and M.Sc. degrees in Computer Engineering from Middle East Technical University in 2003 and 2005, respectively. She worked as a computer engineer at Havelsan A.Ş. from August 2003 to January 2007. Since February 2007, she has been working as a senior researcher at TÜBİTAK Uzay Institute. Her research interests include information extraction, shallow natural language processing, knowledge-based systems, multimedia databases, and database applications in engineering domains.

Publications

International Journal Publications

1. D. Küçük and A. Yazıcı. "Exploiting Information Extraction Techniques for Automatic Semantic Video Indexing with an Application to Turkish News Videos". *Knowledge-Based Systems* (Revised).

2. D. Küçük, Ö. Salor, M. Güder, T. Demirci, T. İnan, Y. Akkaya, I. Çadırcı and M. Ermiş. "Assessment of Extensive Countrywide Electrical Power Quality Measurements Through a Database Architecture". *Electrical Engineering* (Revised).

3. T. Demirci, A. Kalaycıoğlu, D. Küçük, Ö. Salor, M. Güder, S. Pakhuylu, T. Atalık, T. İnan, I. Çadırcı, Y. Akkaya, S. Bilgen, and M. Ermiş, "Nationwide Real-Time Monitoring System for Electrical Quantities and Power Quality of the Electricity Transmission System". *IET Generation, Transmission & Distribution* (Accepted for Publication).

4. D. Küçük, T. İnan, Ö. Salor, T. Demirci, Y. Akkaya, S. Buhan, B. Boyraz-

oğlu, Ö. Ünsar, E. Altıntaş, B. Haliloğlu, I. Çadırcı and M. Ermiş. "An Extensible Database Architecture for Nationwide Power Quality Monitoring". *International Journal of Electrical Power and Energy Systems*, Volume 32, Issue 6, pp. 559–570, July 2010.

5. D. Küçük, Ö. Salor, T. İnan, I. Çadırcı and M. Ermiş. "PQONT: A Domain Ontology for Electrical Power Quality". *Advanced Engineering Informatics*, Volume 24, Issue 1, pp. 84–95, January 2010.

6. D. Küçük, N. B. Özgür, A. Yazıcı and M. Koyuncu. "A Fuzzy Conceptual Model for Multimedia Data with a Text-based Automatic Annotation Scheme". *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Volume 17, Suppl. Issue 1, pp. 135–152, August 2009.

International Conference Publications

1. D. Küçük and A. Yazıcı. "A Hybrid Named Entity Recognizer for Turkish with Applications to Different Text Genres". In *Proceedings of the International Symposium on Computer and Information Sciences*. London, UK. E. Gelenbe et al. (Eds.): Computer and Information Sciences, LNEE 62, pp. 113-116, 2010.

2. D. Küçük and A. Yazıcı. "A Text-Based Fully Automated Architecture for the Semantic Annotation and Retrieval of Turkish News Videos". In *Proceedings of IEEE International Conference on Fuzzy Systems*, pp. 1–8. Barcelona, Spain, 2010.

3. D. Küçük and A. Yazıcı. "Named Entity Recognition Experiments on Turkish Texts". In *Proceedings of the International Conference on Flexible Query Answering Systems*. Roskilde, Denmark. T. Andreasen et al. (Eds.): FQAS 2009, LNAI 5822, pp. 524–535, 2009.

4. D. Küçük and A. Yazıcı. "Employing Named Entities for Semantic Retrieval of News Videos in Turkish". In *Proceedings of the International Symposium on Computer and Information Sciences*, pp. 153–158. Güzelyurt, Northern Cyprus, 2009.

5. D. Küçük and A. Yazıcı. "Rule-based Named Entity Recognition from Turkish Texts". In *Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications*, pp. 456–460. Trabzon, Turkey, 2009.

6. D. Küçük and A. Yazıcı. "Identification of Coreferential Chains in Video Texts for Semantic Annotation of News Videos". In *Proceedings of the International Symposium on Computer and Information Sciences*, pp. 1–6. İstanbul, Turkey, 2008.

7. D. Küçük, Ö. Salor, T. İnan and I. Çadırcı. "Building an Ontology for Flexible Power Quality Querying". In *Proceedings of the International Symposium on Computer and Information Sciences*, pp. 1–6. İstanbul, Turkey, 2008.

8. D. Küçük, N. B. Özgür, A. Yazıcı and M. Koyuncu. "A Fuzzy Conceptual Model for Multimedia Data with Application to News Video Domain". In *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pp. 1741–1748. Málaga, Spain, 2008.

9. D. Küçük and M. Turhan-Yöndem. "Automatic Identification of Pronominal Anaphora in Turkish Texts". In *Proceedings of the International Symposium on Computer and Information Sciences*, pp. 1–6. Ankara, Turkey, 2007.

10. P. Tüfekçi, D. Küçük, M. Turhan-Yöndem and Y. Kılıçaslan. "Comparison of a Syntax-Based and a Knowledge-poor Pronoun Resolution System for Turkish" (poster paper). In *Proceedings of the International Symposium on Computer and Information Sciences*. Ankara, Turkey, 2007.

11. D. Küçük, T. İnan, B. Boyrazoğlu, S. Buhan, Ö. Salor, I. Çadırcı and M. Ermiş. "PQStream: A Data Stream Architecture for Electrical Power Quality". In *Proceedings of the International Workshop on Knowledge Discovery from Ubiquitous Data Streams of ECML/PKDD*. Warsaw, Poland, 2007.

12. D. Küçük and M. Turhan-Yöndem. "A Knowledge-poor Pronoun Resolution System for Turkish". In *Proceedings of the Discourse Anaphora and Anaphora Resolution Colloquium*, pp. 59–64. Lagos, Portugal, 2007.

National Conference Publications (in Turkish)

1. Ö. Salor, D. Küçük, M. Güder, T. Demirci, Y. Akkaya, I. Çadırcı, M. Ermiş. "Türkiye Elektrik İletim Sisteminde Harmonik Bozulma ve Kırpışma Parametrelerinin Oluşturulan Güç Kalitesi Veritabanı Yapısıyla Değerlendirilmesi (Assessment of Harmonic Distortions and Flicker in the Turkish Electricity Transmission System Based on the Developed Database Architecture)". *3. EMO Enerji Verimliliği ve Kalitesi Sempozyumu (EVK)*. Kocaeli, Türkiye, 2009.

2. T. Demirci, A. Kalaycıoğlu, Ö. Salor, S. Pakhuylu, T. İnan, D. Küçük, M. Güder, T. Can, Y. Akkaya, S. Bilgen, I. Çadırcı, M. Ermiş. "Türkiye Elektrik İletim Sistemi için Yurt Çapında Güç Kalitesi İzleme Ağı ve Veri Değerlendirme Merkezi: Güncel Gelişmeler (National Power Quality Monitoring Network and Assessment Center for Turkish Electricity Transmission System: Recent Developments)". *IEEE 16. Sinyal İşleme, İletişim ve Uygulamaları Kurultayı (SIU)*. Didim, Türkiye, 2008.

Thesis

1. D. Küçük. "A Knowledge-poor Pronoun Resolution System for Turkish". Master's Thesis. Department of Computer Engineering, Middle East Technical University, 2005.