Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 112 (2017) 195–204

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

The role of temporal inferences in understanding Arabic text

Hajer Omri [a*], Zeineb Neji [a], Marieme Ellouze [a], Lamia Hadrich Belguith[a]

Faculty of Economics and management, Tunisia, Sfax3042

Abstract

Inference approaches in Arabic question answering systems are in their first steps if we compare them with other languages. Evidently, any user is interested in obtaining a specific and precise answer to a specific question. Therefore, the challenge of developing a system capable of obtaining a relevant and concise answer is obviously of great benefit. This paper deals with answering questions about temporal information involving several forms of inference.

Keywords: temporal inference, Arabic Language, several forms of inference, Natural language processing.

Peer-review under responsibility of KES International

1. Introduction

Over the years, the problems of temporal analysis in Natural Language Processing (NLP) have been addressed with various approaches, ranging from heavily inference-oriented to mostly NLP motivated ones [1]. This analysis varies from one study to another. Although all of these approaches have their own strengths in providing the infrastructure required for representing and reasoning about time in NLP, few of them can at the same time deal with sophisticated inferences.

Advances in Natural Language Processing (NLP), Information Retrieval techniques (IR) and Information Extraction (IE), have given Question/answering systems (QA) a strong boost.QA have started incorporating NLP techniques to parse natural language documents, extract entities, resolve anaphora, and another language ambiguities [2]. In order to develop question answering capabilities, we believe that large corpuses of questions and answers that are based on temporal information should be discovered. In this paper, we focus on the task of question answering in Arabic by thinking of an approach which can improve the performance of traditional Arabic question answering

* Corresponding author..
E-mail address: hajer.omri2010@gmail.com

1877-0509
Peer-review under responsibility of KES International
10.1016/j.procs.2017.08.228

systems for handling temporal inference. Obviously, any user is interested in obtaining a specific and precise answer to a specific question [3]. Therefore, the challenge of developing a system capable of obtaining a relevant and concise answer is obviously of great benefit.

The challenge becomes huge when we try to automatically process a complex natural language, such as Arabic. This complexity is mainly due to the inflectional nature of Arabic. The situation gets worse. When we talk about the considerable lack of resources in general and the Arabic NLP (ANLP) community which still suffers from the lack of free available annotated corpora. In this paper, we propose an approach for the resolution of temporal information in particular, we focus on putting forward a practical way for handling, temporal inference. The proposed method in this paper is based on the Arabic Wikipedia.

2. Particularity of time constraints in the Arabic language

The possibility of automatically tracking temporal information, capturing changes relating to events from text documents, is a new challenge in the field of information retrieval (IR), specifically temporal information retrieval (TIR) and natural language processing (NLP). There are many research studies on the extraction of temporal information in other languages that use Latin alphabets, such as English, French, or Spanish, however the Arabic language is still not well supported in TIR and therefore needs to be inquired.

In Arabic text documents, extraction of temporal entities is a hard task due to the morphology of the Arabic language and the way that the temporal entities are expressed.

According to Ernest Renan, the Arabs boast of having 80 words for honey, 200 for the snake, 500 for the lion, 1000 for the camel and the sword, and up to 4400 to make the idea of misfortune. Overall vocabulary includes more than 60,000 words.

The notion of time itself encounters a variety of representation. Hence, the need for a system of equivalence between the representations set which designates the same temporal information to resolve any ambiguity.

As an example of an ambiguity, one can cite a simple and meaningful example, which is the designation of the months of the year when one can note a significant variety of this word.

Firstly, in Arabic language, there are two types of months one is said to be lunar, which is specific for Muslims and the other is called solar. In fact, each of the Arab countries has its own term designation.

Table 1. The list of names for a month in the Arabic language

English	Arabic
January	جانفي/ jAnfy , يناير / ynAyr محرم/ mHrm , كانون الثاني/ kAnwn AlvAny
February	فيفري/ fyfry , فبراير / fbrAyr صفر/ Sfr
March	مارس/ mArs ,ربيع الاول/ rbyE AlAwl
April	أفريل / >fryl ,أبريل / >bryl ربيع الثاني/ rbyE AlvAny
May	ماي/ mAy , مايو / mAyw جمادى الاول/ jmAdY AlAwl
June	جوان/ jwAn , يونيو / ywnyw جمادى الثاني/ jmAdY AlvAny