

ParsTime: Rule-Based Extraction and Normalization of Persian Temporal Expressions

European Conference on Information Retrieval (ECIR 2018)

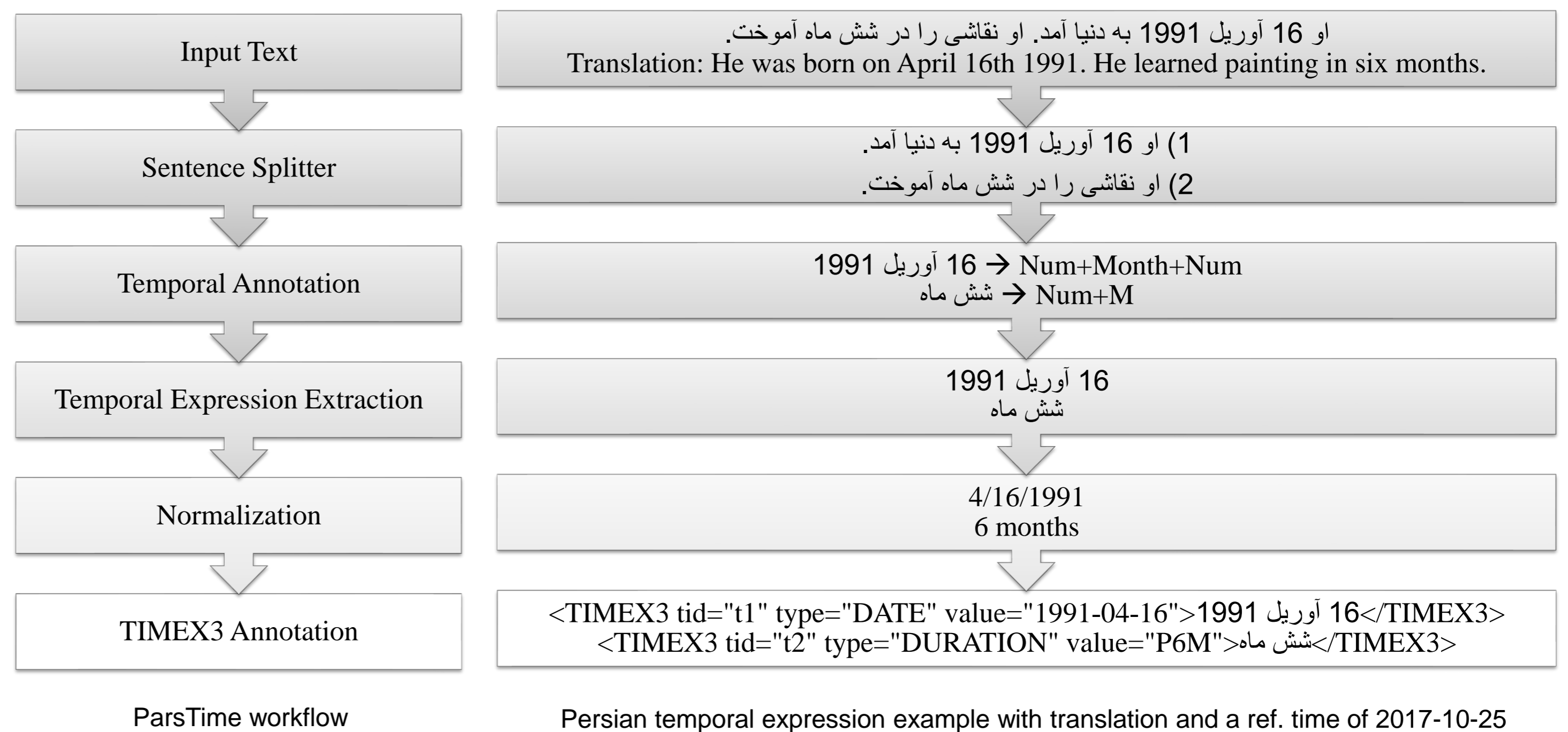
March 26-29 2018 | Grenoble, France

	Behrooz Mansouri	M.Sadegh Zahedi	Mojgan Farhodi		Ricardo Campos		Maseud Rahgozar
	WebAzma lab, Information Technology Faculty, Iran Telecommunication Research Center, Tehran, Iran				Polytechnic Institute of Tomar, INESC TEC – LIAAD, Portugal		University of Tehran, Iran
{b.mansouri, s.zahedi, farhodi}@itrc.ac.ir			ricardo.campos@ipt.pt		rahgozar@ut.ac.ir		

Introduction

Extraction and normalization of temporal expressions are essential for many NLP tasks. While a considerable effort has been put on this task over the last few years, most of the research has been conducted on the English domain, and only a few works have been developed on other languages. In this paper, we present ParsTime, a tagger for temporal expressions in Persian (Farsi) documents. ParsTime is a rule-based system that extracts and normalizes Persian temporal expressions according to the TIMEX3 annotation standard. Our experimental results show that ParsTime can identify temporal expressions in Persian texts with an F1-score 0.89. As an additional contribution we make available our code to the research community.

Approach



Types of Temporal Expressions

Date	Example: هفده سپتامبر 2017
	Translation: September seventeenth 2017
	Normalized value: 2017-09-17
Time	Example: یک ربع به هفت
	Translation: a quarter to seven
	Normalized value: 2017-10-25T06:45
Interval	Example: چهار روز آینده
	Translation: the next four days
	Normalized value: P4D
Set	Example: عصر هر سهشنبه
	Translation: every Tuesday afternoon
	Normalized value: XXXX-WXX-2TAF

Results

Dataset	Extraction			Attribute	
	Precision	Recall	F1-score	Value	Type
Hamshahri	0.92	0.86	0.89	0.84	0.93
Parsijoo	0.91	0.85	0.88	0.82	0.92

ParsTime performance on Hamshahri corpus and Parsijoo query log records

Temporal Tagger	Extraction			Attribute	
	Precision	Recall	F1-score	Value	Type
ParsTime	0.92	0.86	0.89	0.84	0.93
SuTime (English)	0.88	0.96	0.92	0.82	0.92
HeidelTime (English)	0.90	0.82	0.86	0.85	0.96
HeidelTime (Spanish)	0.96	0.84	0.90	0.85	0.87

Effectiveness of ParsTime and other temporal taggers

Datasets

- 2000 random news articles from Hamshahri news dataset
- 1000 web queries containing temporal expressions from Parsijoo query log



<http://dbrg.ut.ac.ir/ParsTime/>



ParsTime Source Code

<https://github.com/BehroozMansouri/ParsTime>

Conclusions

In this paper, we presented ParsTime, the first temporal tagger for extracting and normalizing Persian temporal expression from texts. ParsTime is a rule-based system that can extract different types of temporal expressions including date, time, duration and set. Our experimental results, over two newly created TIMEX3 annotated datasets, show that ParsTime achieved high F1-score. As an additional contribution to the research community we also make available a Java version of our method. This will enable researchers to use our system despite guaranteeing the reproducibility of our research. In future work, we plan to provide resources for detecting implicit temporal expressions, such as "Rio Olympics", which implicitly refer to 5-21 August 2016.

Literature cited

- Campos, R., Dias, G., Jorge, A., and Jatowt, A. (2014). Survey of Temporal Information Retrieval and Related Applications. In CSUR, 47(2). Article No.: 15.
- Mansouri, B., Zahedi, M., Rahgozar, M., Oroumchian, F., Campos, R. (2017). Learning Temporal Ambiguity in Web Search Queries. In ACM Conference (CIKM), pp-6-10.
- Mansouri, B., Zahedi, M., Rahgozar, M., & Campos, R. (2017). Detecting Seasonal Queries Using Time Series and Content Features. In Proceedings of the 2017 ACM on International Conference on the Theory of Information Retrieval ACM. pp. 279-300.
- Strötgen, J., Armiti, A., Van Canh, T., Zell, J., & Gertz, M. (2014). Time for more lan-guages: Temporal tagging of Arabic, Italian, Spanish, and Vietnamese. ACM Transactions on Asian Language Information Processing (TALIP), 13(1), 1.
- Chang, A. X., & Manning, C. D. (2012). Sutime: A library for recognizing and normalizing time expressions. In LREC .Vol. 2012, pp. 3735-3740.
- Li, H., Strötgen, J., Zell, J., & Gertz, M. (2014). Chinese Temporal Tagging with Heidel-Time. In EACL Vol. 2014, pp. 133-7.

Further information

- www.ccc.ipt.pt/~ricardo/
- <http://dbrg.ut.ac.ir/>

Download this Poster

