TISCO: Temporal Scoping of Facts

Anisa Rula*
University of Milano-Bicocca
Milan, Italy
anisa.rula@unimib.it

Matteo Palmonari
University of Milano-Bicocca
Milan, Italy
matteo.palmonari@unimib.it

Simone Rubinacci
University of Oxford
Oxford, United Kingdom
s.rubinacci@stats.ox.ac.uk

Axel-Cyrille Ngonga Ngomo
University of Paderborn
Paderborn, Germany
axel.ngonga@upb.de

Jens Lehmann
University of Bonn
Bonn, Germany
lehmann@cs.uni-bonn.de

Diego Esteves
University of Bonn
Bonn, Germany
esteves@cs.uni-bonn.de

Andrea Maurino
University of Milano-Bicocca
Milan, Italy
andrea.maurino@unimib.it

ABSTRACT

Some facts in the Web of Data are only valid within a certain time interval. However, most of the knowledge bases available on the Web of Data do not provide temporal information explicitly. Hence, the relationship between facts and time intervals is often lost. A few solutions are proposed in this field. Most of them are concentrated more in extracting facts with time intervals rather than trying to map facts with time intervals. This paper studies the problem of *determining the temporal scopes of facts*, that is, deciding the time intervals in which the fact is valid. We propose a generic approach which addresses this problem by curating temporal information of facts in the knowledge bases. Our proposed framework, Temporal Information Scoping (TISCO) exploits evidence collected from the Web of Data and the Web. The evidence is combined within a three-step approach which comprises matching, selection and merging. This is the first work employing matching methods that consider both a single fact or a group of facts at a time. We evaluate our approach against a corpus of facts as input and different parameter settings for the underlying algorithms. Our results suggest that we can detect temporal information for facts from DBpedia with an f-measure of up to 80%.

CCS CONCEPTS

• **Information systems** → **Resource Description Framework (RDF)**; *Incomplete data*; • **Computing methodologies** → *Information extraction.*

KEYWORDS

Temporal Information Extraction; Temporal Semantic Web; Temporal Scoping; Fact Checking

*Principal corresponding author.

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.
WWW '19 Companion, May 13–17, 2019, San Francisco, CA, USA
© 2019 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.
ACM ISBN 978-1-4503-6675-5/19/05.
https://doi.org/10.1145/3308560.3316524

ACM Reference Format:
Anisa Rula, Matteo Palmonari, Simone Rubinacci, Axel-Cyrille Ngonga Ngomo, Jens Lehmann, Diego Esteves, and Andrea Maurino. 2019. TISCO: Temporal Scoping of Facts. In *Companion Proceedings of the 2019 World Wide Web Conference (WWW '19 Companion), May 13–17, 2019, San Francisco, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3308560.3316524

1 INTRODUCTION

The Web of Data can be regarded as a dynamic environment where information can change rapidly and cannot be assumed to be static [6]. Changes in the Web of Data sources should reflect changes in the real world, otherwise data can soon become outdated. Some facts are not time variant and thus do not change over time e.g. `<CristianoRonaldo, bornIn, Portugal>` while others need to be checked for their *veracity* [3]. In this work, we focus on a particular case of veracity of facts that is the *validity time* which refers to the time interval in which the fact is true, e.g. `<CristianoRonaldo, playFor, ManchesterUnited>` refers to a fact valid from 2003 to 2009.

Most of knowledge bases store facts under a historical perspective. Facts in this knowledge bases have been true sometime until the current time. We refer to the representation of these facts in the knowledge bases as to *historification of dynamic facts*. Historification of dynamic facts, which is frequently found in many and prominent datasets in Linked Data (LD), fails to provide details about the time interval when the facts have been true. These knowledge bases adopt a *temporal flattening approach* of representing dynamic facts. The incompleteness and the inaccuracy of temporal information in LD [9] is often due to the information extraction process (that can be error prone) or to the representation model (that requires very sophisticated meta-modeling strategies to represent *versioning metadata* in RDF). For instance, in DBpedia it is not possible to know the time interval of the fact `<CristianoRonaldo, playFor, ManchesterUnited>` since all time points are associated directly with the entity rather than the fact and the semantics of the predicate is the same for the starting and the ending time points (e.g. `<CristianoRonaldo, year, 2009>`).

Despite the importance of the relationship between facts and time intervals, very few solutions are proposed. Most of them are concentrated more in extracting facts with time intervals from text [5, 8, 12] rather than trying to map facts with time intervals. More recent works [2, 7] provide an enrichment of time intervals in the knowledge bases. The system CoTS provided in [11] is similar to our system since it also detects validity time of facts. In contrast to our approach, CoTS relies on document metadata such as its creation date, to assign time intervals to facts. To the best of our knowledge, this is the first work employing both local and global matching approach, and a system for mapping facts to time intervals.

To map those facts to the correct time intervals, we have to address two main challenges. First, the set of time intervals is created as a combination of all time points extracted from the knowledge base for each entity where the starting time point is smaller than the ending time point. This set needs to be reduced since it contains also noisy intervals such as all intervals starting with the birth year. Second, to find the correct intervals, we need to extract supporting evidence from external sources that indicates how often a fact occurs with each time point, and subsequently predict the possible time interval of each fact based on the acquired evidence.

In this work, we focus on curating time intervals associated with facts. We introduce an approach for detecting the temporal scope of facts referred to by triples (short: the temporal scope of the triples). Given a fact (i.e., an RDF triple), our approach aims to detect the time points at which the temporal scope of the triple begins and ends. Two sources can be envisaged for gathering such information: the document Web and LD. Our approach is able to take advantage of both: the Web is made use of by extending upon a fact validation approach [4], which allows detecting Web documents which corroborate a triple. In contrast to typical search engines, the system does not just search for textual occurrences of parts of the statement, but tries to find webpages which contain the actual statement phrased in natural language. The second source of information for time scopes is the Web of Data itself. Here, we use DBpedia, for possible time scopes and devise an algorithm for combining the results extracted from Web documents with those fetched from RDF sources. The algorithm consists of three main steps: First, the evidence extracted from Web documents is *matched* against a set of relevant time intervals to obtain a significance score for each interval. Second, a small set of more significant intervals is *selected*. Finally, the selected intervals are *merged*, when possible, by considering their mutual temporal relations. The set of disconnected intervals [1] returned by the algorithm defines the temporal scope of the fact. We also propose two *normalization* strategies that can be applied to the data extracted from Web documents before running the algorithm, to account for the significance of dates appearing in the documents corroborating the input fact.

This article makes the following contributions: i) We present an approach for modelling a space of relevant time intervals for a fact starting from dates extracted from RDF triples. ii) We devise a three-phase algorithm for temporal scoping, i.e. for mapping facts to sets of time intervals, which integrates the previous steps via matching, selection and merging. iii) We describe two matching methods that consider facts in isolation or cluster them according to the main entity. iv) We provide TISCO, a running prototype; the first system able to provide temporally annotated facts which are modelled according to a relationship-centric perspective [9].

We provide an extension of [10] as: i) We describe in detail more alternative solutions, including an additional function in the matching phase of our approach and a normalization function of occurrences of dates; ii) We developed a prototype for annotating facts with temporal information and show how all the different matching functions and their combinations can be integrated in one framework; iii) We extend the experimental results showing the effectiveness of our results.

2 CONCLUSIONS

This paper studies the problem of determining and mapping time intervals to dynamic facts. We proposed a framework comprising several functions and configuration parameters that can efficiently provide the matching and the selection of the set of time intervals that maximize the effectiveness of our approach. In addition, we proposed a running prototype, TISCO that will support the users in exploring facts with temporal scopes and simplify the testing of new algorithms for matching and selection functions. We evaluated our approach on facts extracted from DBpedia by using cleaned-up temporal scopes extracted from Yago2. We provide the benchmarking dataset to support the users in testing their approaches since it is difficult to perform the task at hand. Our approach is also tested on google books dataset and it outperforms one of the state-of-the-art approaches applied on the same dataset.

REFERENCES

[1] James F Allen. 1983. Maintaining knowledge about temporal intervals. *Commun. ACM* 26, 11 (1983), 832–843.

[2] Melisachew Wudage Chekol, Giuseppe PirrÃš, Joerg Schoenfisch, and Heiner Stuckenschmidt. 2017. Marrying Uncertainty and Time in Knowledge Graphs.. In *AAAI*, Satinder P. Singh and Shaul Markovitch (Eds.). AAAI Press, 88–94.

[3] Diego Esteves, Anisa Rula, Aniketh Janardhan Reddy, and Jens Lehmann. 2018. Toward Veracity Assessment in RDF Knowledge Bases: An Exploratory Analysis. *J. Data and Information Quality* 9, 3 (2018), 16:1–16:26.

[4] Daniel Gerber, Diego Esteves, Jens Lehmann, Lorenz Bühmann, Ricardo Usbeck, Axel-Cyrille Ngonga Ngomo, and René Speck. 2015. DefactoâĂŤtemporal and multilingual deep fact validation. *Web Semantics: Science, Services and Agents on the World Wide Web* 35 (2015), 85–101.

[5] Dirk Hovy, James Fan, Alfio Gliozzo, Siddharth Patwardhan, and Chris Welty. 2012. When did that happen?: linking events and relations to timestamps. In *13th EACL*. 9.

[6] Tobias Käfer, Ahmed Abdelrahman, Jürgen Umbrich, Patrick O'Byrne, and Aidan Hogan. 2013. Observing Linked Data Dynamics. In *ESWC*. 213–227.

[7] Erdal Kuzey, Jannik Strötgen, Vinay Setty, and Gerhard Weikum. 2016. Temponym Tagging: Temporal Scopes for Textual Phrases. In *Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion)*. International World Wide Web Conferences Steering Committee, 841–842.

[8] Xiao Ling and Daniel S Weld. 2010. Temporal information extraction. In *25th AAAI*.

[9] Anisa Rula, Matteo Palmonari, Andreas Harth, Steffen Stadtmüller, and Andrea Maurino. 2012. On the Diversity and Availability of Temporal Information in Linked Open Data. In *11th ISWC*.

[10] Anisa Rula, Matteo Palmonari, Axel-Cyrille Ngonga Ngomo, Daniel Gerber, Jens Lehmann, and Lorenz Bühmann. 2014. Hybrid Acquisition of Temporal Scopes for RDF Data. In *11th ESWC*. 488–503. https://doi.org/10.1007/978-3-319-07443-6_33

[11] Partha Pratim Talukdar, Derry Tanti Wijaya, and Tom Mitchell. 2012. Coupled temporal scoping of relational facts. In *5th WSDM*. 73–82.

[12] Naushad UzZaman and James F Allen. 2010. TRIPS and TRIOS system for TempEval-2: Extracting temporal information from text. In *SemEval*. ACL, 276–283.