Temporal Knowledge Representation and Exploitation for the Henri Poincaré Correspondence Corpus

Born in Nancy, France, Jules Henri Poincaré (1854-1912) is considered one of the major scientists of his time. He remains better known for his contribution in mathematics (automorphic forms, topology) and physics (3-Body problem resolution) but he also played a significant role in the development of philosophy as shown by his book *La Science et l'Hypothèse* which had a major international impact. Henri Poincaré held many roles within French administrations and learned societies such as the French Academy of Sciences or the *Bureau des longitudes*. He was also involved in international societies — mainly as a corresponding member — such as the Dutch Society of Sciences in Haarlem or the American Philosophical Society.

The Henri Poincaré correspondence corpus is composed of around 2200 letters which constitute scientific, administrative and private exchanges. It is an important source of information for historians both for societal and scientific aspects. Semantic Web technologies have been chosen to represent and to exploit corpus data: the RDF model, the RDFS knowledge representation language and the SPARQL query language [1]. Corpus data is structured using an ontology whose aim is to represent knowledge related to historical corpora, by modeling documents (articles, books, reports, letters, etc.), persons and institutions, places, etc.

The first corpus representation did not consider temporal knowledge related to the statements described in the RDF graph. However, this temporal knowledge is essential for the study of historical corpora. As an example, consider the fact that Henri Poincaré was a member of the French Academy of Sciences learned society. The related statement should be considered valid from his election date, on January 31, 1887. Integrating temporal knowledge into RDF graphs is not an easy task and this field of research has aroused significant interest from the Semantic Web community [2]. Indeed, RDF relies on the definition of triples which correspond to binary relations between elements. Adding a temporal element (an instant or a temporal interval) correspond to the creation of a ternary relation. Beyond the mere integration of temporal elements, another objective is to be able to set up reasoning in order to extract knowledge from corpus data.

This communication aims at presenting the temporal knowledge representation issues encountered in the exploitation of the Henri Poincaré correspondence corpus. Through examples, it details the implementation of the approach chosen for this corpus which is based on the definition of n-ary relations [3]. This approach seems to be one of the most intuitive and has been recommended by the W3C. No extension of the semantics of RDF is necessary and the number of additional triples introduced by the representation remains low compared to other models. A form-based interface has been developed to assist users in corpus data querying. This interface is used to generate SPARQL query and to retrieve facts with or without the associated temporal elements. This system also relies on the definition and the use of custom inference rule. As an example, a rule can be defined to restraint the birth date of a person if the life dates of one its parent is known. Another rule can represent the fact that, if a letter B has been identified as a response to letter A, then it is admitted that its writing date is earlier than that of letter A. Alongside the use of Allen's interval algebra included in the OWL-Time ontology [4], this system enables to define various rules to exploit corpus temporal knowledge.

References:

[1] Bruneau O., Lasolle N., Lieber J., Nauer E., Pavlova S. & Rollet L. (2021). Applying and Developing Semantic Web Technologies for Exploiting a Corpus in History of Science: the Case Study of the Henri Poincaré Correspondence, *Semantic Web – Interoperability, Usability, Applicability* 12(2) (2021), 359–378.

[2] Analyti A., & Pachoulakis, I. (2012). A survey on models and query languages for temporally annotated RDF. *International Journal of Advanced Computer Science and Applications* 3(9), 28-35.

[3] Noy, N., Rector, O., Hayes, P. & Welty C. (2006). Defining n-ary Relations on the Semantic Web, *W3C working group note*, *12*(4). https://www.w3.org/TR/swbp-n-aryRelations/.

[4] Cox S., Little C., Hobbs, J. R. & Pan, F. (2020). Time Ontology in OWL, https://www.w3.org/TR/owl-time/.