Dealing with Data Imperfection in OWL 2 - Application to Alzheimer's patients Software

Nassira Achich[1,2]

[1] CEDRIC Laboratory, Conservatoire National des Arts et Métiers (CNAM), France
nassira.achich.auditeur@lecnam.net
[2] MIRACL Laboratory, University of Sfax, Tunisia
nassiraachich@fsegs.u-sfax.tn

Abstract. Data is mostly tainted with various kinds of imperfections such as imprecision, uncertainty and incompleteness. This explains the emergence of different approaches to deal with data imperfections in semantic web. Given the importance of the temporal dimension in many information sources and considering that temporal data are one of the most affected by several kinds of imperfection, we concentrate, at a first level, on treating temporal data imperfection. In this work, we introduce our typology of temporal data imperfection as well as an approach that handles imprecise dates and time clocks. Furthermore, we highlight our approach to deal with uncertain temporal data. To evaluate our proposed approaches, multiple prototypes are implemented and integrated into an ontology-based memory prosthesis for Alzheimer's patients to handle temporal data imperfection.

Keywords: Data Imperfection · Temporal Data Imperfection · Alzheimer Disease · Ontology.

1 Problem Statement

Captain Memo [20], a memory prosthesis, is proposing to palliate mnesic problems of people with early-stage Alzheimer's disease. It is based on an OWL 2 ontology that enables modeling and reasoning about interpersonal relationships (e.g., mother, neighbor) and people description (e.g., lived events). Captain Memo makes early stage Alzheimer's patients active in entering data (e.g., name, geographical information, temporal data, links,...) to improve their autonomy. However, these data entered by these particular users, living in uncertainty, are mostly imperfect (i.e., imprecise, uncertain,...). Dealing with data imperfection in ontology is a challenging task. This task becomes more complicated when it concerns temporal data that forms the first part of our interest in this thesis.

As a first step, we decided that creating a typology is essential to helping in gathering all imperfections that may affect temporal data. Several typologies of data imperfections have been proposed. However, these typologies cannot be applied to temporal data because of the complexity of this type of data and the specificity that it contains. Besides, to the best of our knowledge, there is no typology of temporal data imperfections. To solve this problem, we propose a typology of temporal data imperfection [3]. It is divided into direct imperfections, indirect imperfections that can be deduced from the direct ones and other factors that may interfere in specifying the imperfection type.

As a second step, we address some imperfections defined in our typology. We specifically deal with imprecise temporal data and uncertain temporal data. Indeed, imprecision and uncertainty are the frequent imperfection types affecting data entered by Alzheimer's patients in the context of the Captain Memo. In the semantic web field, several approaches have been proposed to represent and reason about "perfect" temporal data (i.e., precise/certain temporal data). However, most of them handle only time intervals and associated qualitative relations (i.e., they are not intended to handle time points and qualitative relations between a time interval and a time point or two time points). Therefore, we propose an approach to deal with precise and imprecise dates and time clocks in OWL 2 [2]. We also propose an approach to treat uncertain temporal data imperfection in OWL 2 [1]. This work is accepted (but still not published).

2 Importance

As we mentioned, a memory prosthesis is being proposed, called Captain Memo to help Alzheimer's patients to palliate mnesic problems. It supplies a set of services. Among these services, one is devoted to help users to remember their convivial relatives and surroundings. Data are structured using an OWL 2 ontology, called PersonLink [16]. However, these data inputs by Alzheimer's patients and/or their surroundings, which are mostly imperfect (i.e., imprecise, uncertain, wrong,...) and could be particularly numerous in the context of a memory prosthesis, are not supported by PersonLink. This latter needs to be extended to handle these kinds of imperfections. This work is a part of the VIVA[3] project ("*Vivre à Paris avec Alzheimer en 2030 grâce aux nouvelles technologies*").

3 Related Work

Related work includes two parts. First, we review typologies of data imperfection. Then, we discuss temporal data representation and reasoning in Semantic Web.

3.1 Typologies of Data Imperfection

We distinguish two types of typologies: generic typologies of data imperfections and domain-specific typologies of data imperfections.

We identify four generic typologies. [21] defines four concepts which are uncertainty, imprecision, ambiguity and generality. [7] distinguishes three types of imperfection which are uncertainty, imprecision and incompleteness. [18] propose a typology of data uncertainty divided into fuzziness and ambiguity. [24] classifies the imperfections into imprecision, inconsistency and uncertainty.

[3] http://viva.cnam.fr/

Many domain-specific typologies of data imperfections have been proposed. We cite, due to the page limitation, some of them. [11] propose a typology of uncertainty of geographic data. They classify the data into a well or a badly defined data. The well defined data is subject to uncertainty. In other cases, the imperfection is due to imprecision, ambiguity and/or incompleteness. [9] establishes a typology of data imperfection resulting from the economic activity. It is divided into uncertainty, imprecision and error. [10] rely on the typology of [11] to propose a typology of imperfection adapted to the context of archaeological data. They classify imperfections into uncertainty, imprecision, ambiguity and incompleteness. [25] propose a classification of imperfections to characterize spatial data. This taxonomy distinguishes three types of imperfection: Imprecision, inconsistency (conflict or incoherence in values), and uncertainty. [26] proposes several types of imperfect data during the process of information retrieval and data integration in smart cities, such as uncertainty, imprecision, and vagueness.

Temporal data can have more imperfections compared to the ones proposed in the existing typologies. Indeed, it can be numeric or natural language-based and can be subject to several factors that may interfere in specifying the imperfection type. For instance, in the following input "The first day of the week, we will have a meeting", the temporal data indicates a "circumlocution" which does not exist in any of the existing typologies. To the best of our knowledge, there is no typologies that consider the specificities of temporal data imperfection.

3.2 Representing and Reasoning about Temporal Data in the Semantic Web Field

Representing temporal data in ontology is a pressing need. However, ontology languages such as OWL, provide a minimal support since they are all based on binary relations that simply connect two instances. This explains the emergence of many approaches for representing temporal data in ontology.

Representing Temporal Data in Semantic Web We classify the approaches into two categories: *(i)* approaches that extend OWL or RDF syntax by defining new OWL or RDF operators and semantics to incorporate temporal data, and *(ii)* approaches that are implemented directly using OWL or RDF to represent temporal data without extending their syntax. The first category includes Temporal Description Logics [5], Concrete Domains [19] and Temporal RDF [15]. Temporal Description Logics extends the standard description logics with new temporal semantics such as "until". This approach retains decidability and does not suffer from data redundancy. However, it is considered as an avoidable solution since it requires extending OWL or RDF, which is a tedious task [14]. Concrete Domains requires introducing additional data types and operators to OWL. Temporal RDF uses only RDF triples. It does not have all the expressiveness of OWL. The second category includes Versioning [17], Reification [8], N-ary Relations [22], 4D-Fluents [29] and Named Graphs [28]. Versioning is described as the ability to handle changes in ontology by creating different variants of it.

In this approach, all the versions are independent from each other. This requires exhaustive searches in all of them [30]. Reification is a technique for representing N-ary relations when only binary relations are allowed. Whenever a temporal relation has to be represented, a new object is created. N-ary relations proposes to represent an N-ary relation as two properties each related with a new object. It maintains property semantics. The Named Graphs approach represents each time interval by exactly one named graph, where all triples belonging share the same validity period. The mentioned approaches suffer from data redundancy. The 4D-Fluents approach represents time intervals and their evolution in OWL. It maintains a full OWL expressiveness and reasoning support [23].

All the reviewed approaches handle only perfect temporal data and neglect imperfect ones. They are not intended to handle time points and qualitative temporal relations between a time interval and a time point or two time points. Our approach should rely on existing OWL constructs. Thus, we exclude the Temporal Description Logic, Concrete Domain and Temporal RDF approaches. We exclude the Named Graphs approach as it does not support OWL and it is not a W3C compliant solution. We choose to extend the 4D-fluents approach to represent imprecise quantitative temporal data and associated qualitative temporal relations. Compared to the Reification, N-ary relations and Versioning, the 4D-fluents approach minimizes data redundancy as the changes occur on the temporal parts and keep the static part unchanged. 4D-fluents approach introduces two classes named “TimeSlice” and “TimeInterval”, and four properties named “tsTimeSliceOf”, “tsTimeIntervalOf”, “HasBeginnig”, and “HasEnd”.

Allen’s Interval Algebra: Definition and Extensions 13 qualitative relations between precise time intervals are proposed by Allen. Their definitions are expressed in Table1. A characteristic of Allen’s algebra is that we can deduce new relations through the composition of other ones. For instance, “Before(A,B)” and “Equals(B,C)” gives “Before(A,C)”. Allen’s interval algebra is not dedicated to represent imprecise time intervals. Furthermore, it does not relate neither a time point and a time interval nor two time points. Many approaches have been extended this algebra such as [29], [28], [27], [12] and [6]. However, these extensions are based on theories related to imperfect data and cannot be supported in the context of crisp ontology. Furthermore, most of these extensions do not preserve all the properties of the original Allen’s algebra [22].

4 Research Questions About

The main problem statement of our research is to deal with different kinds of data imperfection. The application of our work is in the context of data input by Alzheimer’s patients in Captain Memo prothesis. Thus, the general research question proposed in my thesis is “How to deal with different kinds of data imperfection in Semantic Web?” Two sub-questions are proposed:

– How to deal with temporal data imperfection: imprecision, uncertainty,... ?

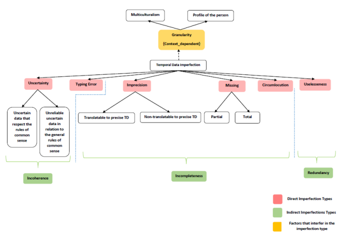

Fig. 1. Our typology of temporal data imperfection.

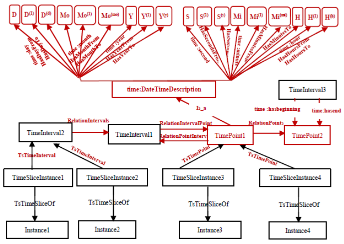

Fig. 2. Our 4D-fluents approach extension.

– How to deal with other data (names, links, events,..) that may be imperfect?
– Can we apply our work to other related fields (e.g., Geographic data)?

5 Preliminary Results

5.1 A Typology of Temporal Data Imperfection

We introduce our typology of temporal data imperfections illustrated by Fig. 1. Our typology is based on the studied mentioned typologies and collected real examples. We divide our typology into direct imperfections, indirect imperfections and factors that may interfere in determining the type of imperfection. The direct ones are those which can be deduced directly from the given data: uncertainty, typing error, imprecision, missing and uselessness. The Indirect imperfections are those that can be deduced from the direct ones: *(i)* The incoherence can be generated from the uncertainty and typing error. *(ii)* The incompleteness can be generated from the imprecision and the missing. *(iii)* The redundancy can be generated from the uselessness.

5.2 Dealing with Imprecise Temporal Data: Dates and Time Clocks

We propose a crisp-based approach, depicted by Fig. 2, to represent and reason about precise and imprecise temporal data (Dates and time Clocks) in ontology.

We extend the 4D-fluents approach to represent: *(i)* precise and imprecise time points and *(ii)* imprecise time intervals. Some of the introduced components are already defined in OWL-Time ontology such as the class named “time:DateTimeDescription” to express the dates and time clocks, and the datatype properties “time:day”, “time:month” and “time:year” to relate, respectively, the “time:DateTimeDescription” class with the values of the day, month and year. Some others that we define, do not exist, such as the class “TimePoint” to represent a precise or an imprecise time point, and the datatype properties: “HasDayFrom” and “HasDayTo”, “HasMonthFrom” and “HasMonthTo”, “HasYearFrom” and “HasYearTo” to represent the disjunctive ascending sets representing an imprecise date. Four temporal relations may exist between time points and time intervals: We assign four crisp object properties: "RelationPoints", "RelationIntervalPoint", "RelationPointInterval" and "RelationIntervals".

For the reasoning, our approach consists of extending the Allen's interval algebra to: *(i)* reason about precise and imprecise quantitative temporal data to infer qualitative temporal relations and *(ii)* to reason about the qualitative temporal relations to infer new ones. We define temporal relations in a crisp way. At the beginning, we propose qualitative temporal relations between imprecise time intervals. Then, we adapt these relations to relate a time interval and a time point or two time points. The proposed temporal relations are based on orderings between the time points contained in the intervals. They may be expressed using time point comparators like the ones proposed by Vilain and Kautz's Algebra. When considering precise time intervals, our approach reduces to Allen's interval algebra. We redefine the Allen's relations to propose temporal relations between imprecise time intervals. We adapt the qualitative temporal relations between time intervals to propose relations between a time interval and a time point. We adapt these relations between time intervals to propose relations between time points. All the redifined relations can be consulted in our publication [2].

Based on our extensions of the 4D-fluents approach and Allen's interval algebra, we implement our OWL 2 temporal ontology (`http://cedric.cnam.fr/isid/ontologies/files/CrispTimeOnto.html`). We instantiate the object properties "RelationIntervals", "RelationIntervalPoint", "RelationPointInterval" and "RelationPoints" based on our Allen's extension. A set of SWRL rules are proposed to infer missing qualitative temporal relations.

5.3 Dealing with Uncertain Temporal Data in OWL 2

We propose an approach to represent and reason about uncertain temporal data in terms of qualitative (for example, "before") and quantitative (intervals and time points) relationships. This approach is based on classical ontology and does not use a probabilistic one. It consists of three parts. *(1)* The first part concerns the representation of certain and uncertain temporal data in OWL2 using the 4D-fluent approach. We extend it with new ontological components to represent: *(1.1)* certain quantitative temporal data (time points) and uncertain (time points and time intervals), and *(1.2)* qualitative temporal relationships between time intervals and time points. *(2)* The second part concerns the reasoning about certain and uncertain temporal data by extending Allen's interval algebra. We propose qualitative temporal relationships between uncertain time intervals. They retain important properties of the original algebra. We adapt the resulting interval relations to propose temporal relations between a time interval and a time point, and two time points. *(3)* The third part consists in proposing an OWL 2 ontology called "UncertTimeOnto" (`http://cedric.cnam.fr/isid/ontologies/files/UncertTimeOnto.html`) which can be integrated into other ontologies to manage certain and uncertain temporal data. It is implemented based on the proposed extensions. The inferences are made using SWRL rules.

6 Evaluation

To validate our approach, we introduce a prototype based on our proposed ontologies. It is implemented based on JAVA. It uses JENA API and SPARQL-DL API for managing and querying crisp ontology. First, the user instantiates our ontology. After each new temporal data input, the "Qualitative Temporal Data Inference" component is automatically executed to infer missing data. It is based on the proposed SWRL rules. Currently, we are working on defining an experimentation protocol for our approaches.

7 Discussion and Future Work

Data is mostly subject to imperfections, especially if these data is inserted by Alzheimer's patients. To this end, our aim is to treat data imperfections in this context. We started by treating temporal data imperfection. We introduced our typology which is classified into direct imperfections, indirect imperfections and factors that may interfere in specifying the imperfection type. Then, we focused on the imprecision and uncertainty. We proposed a crisp-based approach for representing and reasoning about precise and imprecise dates and time clocks in ontology. Then, we proposed an approach to deal with uncertain temporal data. Future work will be devoted to handle temporal data which are "uncertain and imprecise" at the meantime.

Acknowledgements I would like to thank Pr. Elisabeth Métais, Dr. Fayçal Hamdi, Pr. Faiez Gargouri and Dr. Fatma Ghorbel for their valuable supervising.

References

1. Achich, N., Ghorbel, F., Hamdi, F., Metais, E., Gargouri, F. Approach to Reasoning about Uncertain Temporal Data in OWL 2. 24th International Conference on Knowledge Based and Intelligent Information and Engineering Systems (KES), Sep 2020, verona, Italy.
2. Achich, N., Ghorbel, F., Hamdi, F., Metais, E., Gargouri, F. (2019, August). Representing and Reasoning About Precise and Imprecise Time Points and Intervals in Semantic Web: Dealing with Dates and Time Clocks. In: International Conference on Database and Expert Systems Applications (pp. 198-208). Springer, Cham.
3. Achich, N., Ghorbel, F., Hamdi, F., Metais, E., Gargouri, F. (2019). A Typology of Temporal Data Imperfection. https://doi.org/10.5220/0008168403050311
4. Allen, J. F. (1983). Maintaining Knowledge about Temporal Intervals. *Commun*, **26**(11), 832-843.
5. Artale, A., Franconi, E. (2000). A Survey of Temporal Extensions of Description Logics. *Annals of Mathematics and Artificial Intelligence*, **30**(1-4), 171-21.
6. Batsakis, S., Tachmazidis, I., Antoniou, G. (2017). Representing Time and Space for the Semantic web. *International Journal on Artificial Intelligence Tools*, **26**(03).
7. Bouchon-Meunier, B. (1995). La logique floue et ses applications.
8. Buneman, P., Kostylev, E. (2010). Annotation Algebras for RDFS. *In Workshop on the Role of Semantic Web in Provenance Management.*

9. Casta, J. F. (2009). Incertitude et comptabilité.
10. Desjardin, E., Nocent, O., De Runz, C. (2012). Prise en compte de limperfection des connaissances depuis la saisie des donnees jusqu'a la restitution 3d. established by: Mauro Cristofani and Riccardo Francovich, (supplemento **3**):385–396.
11. Fisher, P., Comber, A., RA (Richard) Wadsworth. (2005). Nature de l'incertitude pour les données spatiales.
12. Gammoudi A., Hadjali A., Yaghlane B. B. (2017). Fuzz-TIME: an intelligent system for managing fuzzy temporal information. *Intelligent Computing and Cybernetics*, **10**(2), 200-222.
13. Gavignet, E., Leclercq, E., Cullot, N., and Savonnet, M.(2016). Raisonner en logique modale sur l'incertitude liee aux donnees-application en archeologie. Revue Internationale de Geomatique, 26**(4)**:467–490
14. Ghorbel, F., Métais, E., Hamdi, F. (2018, December). A Crisp-Based Approach for Representing and Reasoning on Imprecise Time Intervals in OWL 2. In International Conference on Intelligent Systems Design and Applications (pp. 640-649).
15. Gutierrez, C., Hurtado, C., Vaisman, A. (2005). Temporal RDF. *In European Semantic Web Conference* (pp. 93-107).
16. Herradi, N., Hamdi, F., Metais, E., Ghorbel, F., Soukane, A. (2015). Personlink: an ontology representing family relationships for the captain memomemory prosthesis. In International Conference on Conceptual Modeling, pages 3–13. Springer
17. Klein, M. C. A., Fensel, D. (2001). Ontology Versioning on the Semantic Web. *In Semantic Web Working Symposium* (pp. 75-91), Stanford University, USA.
18. Klir, G. J. and Yuan, B. (1995). Fuzzy sets and fuzzy logic: theory and applications, volume 574. Prentice Hall PTR New Jersey.
19. Lutz, C. (2003). Description Logics with Concrete Domains - A Survey. *In Advances in Modal Logics* (pp. 265-296).
20. Métais, E., Ghorbel, F., Herradi, N., Hamdi, F., Lammari, N., Nakache, D., ... Soukane, A. (2012). Memory prosthesis. Non-pharmacological Therapies in Dementia, **3**(2), 177.
21. Niskanen, V. A. (1989). Introduction to imprecise reasoning, uncertainty, decision making and knowledge engineering.
22. Noy, N., Rector, A., Hayes, P., Welty, C. (2006). Defining N-Ary Relations on the Semantic-Web. *W3C Working Group Note*, **12**(4).
23. Preventis, A., Petrakis, E. G., Batsakis, S. (2014). Chronos Ed: A Tool for Handling Temporal Ontologies in PROTÉGÉ. *IJAIT*, **23**(04).
24. Smets, P. (1997). Imperfect information: Imprecision and uncertainty, uncertainty management in information systems: from needs to solutions. pages 225–254.
25. M. Snoussi, P.-A. D. (2014). Methodological proposals to handle imperfect spatial and temporal information in the context of natural hazard studies. 495–517. International Journal of Geomatics and Spatial Analysis 3-4, 23.495–517
26. Sta, H. (2016). Quality and the efficiency of data in smartcities. Future Generation Computer Systems.
27. Sadeghi, K. M. M., Goertzel, B. (2014). Uncertain Interval Algebra via Fuzzy/Probabilistic Modeling. *In IEEE International Conference on Fuzzy Systems.*
28. Tappolet, J., Bernstein, A. (2009). Applied Temporal RDF: Efficient Temporal Querying of RDF Data with SPARQL. *In European Semantic Web Conference* .
29. Welty, C. A., Fikes, R. (2006). A Reusable Ontology for Fluents in OWL. *In Formal Ontology in Information Systems* (pp. 226-236).
30. Zekri, A., Brahmia, Z., Grandi, F., Bouaziz, R. (2016). tOWL: A Systematic Approach to Temporal Versioning of Semantic Web Ontologies. *Journal on Data Semantics*, **5**(3), 141-163.