The ISC Int'l Journal of Information Security

August 2019, Volume 11, Number 3 (pp. 59–65)

http://www.isecure-journal.org

Selected Paper at the ICCMIT'19 in Vienna, Austria

Considering Uncertainty in Modeling Historical Knowledge☆

Fairouz Zendaoui *, and Walid Khaled Hidouci

Laboratoire de la Communication dans les Systémes Informatiques, Ecole nationale Supérieure d'Informatique, BP 68M, 16309, Oued-Smar, Alger, Algérie.

ARTICLE INFO.

Keywords:
Belief, Event, History, Information Imperfection, Knowledge Representation.

Abstract

Simplifying and structuring qualitatively complex knowledge, quantifying it in a certain way to make it reusable and easily accessible are all aspects that are not new to historians. Computer science is currently approaching a solution to some of these problems, or at least making it easier to work with historical data. In this paper, we propose a historical knowledge representation model taking into consideration the quality of imperfection of historical data in terms of uncertainty. To do this, our model design is based on a multilayer approach in which we distinguish three informational levels: information, source, and belief whose combination allows modeling and modulating historical knowledge. The basic principle of this model is to allow multiple historical sources to represent several versions of the history of a historical event with associated degrees of belief. In our model, we differentiated three levels of granularity (attribute, object, relation) to express belief and defined 11 degrees of uncertainty in belief. The proposed model can be the object of various exploitations that fall within the historian's decision-making support for the plausibility of the history of historical events.

1 Introduction

Different versions of the same historical event can reach us through various testimonies or more general sources (such as archives, documents, vestiges, etc.). Historians often find themselves confronted with abundant and scarce data sources, and the information they contain tends to be imperfect, uncertain, imprecise, ambiguous, and incomplete. However, to our knowledge, none of the works that have represented historical knowledge has considered the quality of the imperfection of historical information and has been content to represent only one version of the history of an event, the one retained by the historian as a definitive version deduced by supporting a historical methodology.

In this paper, we propose a conceptual model to represent historical knowledge by taking into account the quality of information imperfection in terms of uncertainty. Our model allows multiple sources to represent several versions of the history of a historical event by assigning degrees of belief. It is based on a multilayer approach that distinguishes information, source and belief, whose combination allows modeling and modulating historical knowledge. We represent

☆ The ICCMIT'19 program committee effort is highly acknowledged for reviewing this paper.

* Corresponding author.

Email addresses: f_zendaoui@esi.dz, w_hidouci@esi.dz

ISSN: 2008-2045

uncertainty at three levels of granularity (attribute, object, relation) and define 11 degrees of belief to express uncertainty.
This article is organized as follows. In Section 2, we present some works that represented historical knowledge. Then, we introduce in Section 3 the concept of history with a focus on historical data. We study the quality of information imperfection and introduce the notion of belief. In Section 4, we articulate our contribution where we propose our model of representation of historical knowledge. We conclude our paper in Section 5 and state some perspectives in Section 6.

2 Related Work

Historical research is currently undergoing major changes in its methodology largely due to the advent and availability of high-quality digital data sources. Without a doubt, the most difficult element for historical research is information management. Several researchers have attempted to represent historical knowledge to support the production and sharing of knowledge in history. In the majority of works, interest and efforts have focused on the semantic representation of history, especially with the emergence of the semantic web that shook the paradigm of publishing and sharing research data.
The paper [1] presented a very rich state of the art works dealing with the use of methods and technologies of the semantic web in historical research. Among these works, we mention the project SYGMOGIH a collaborative platform based on a geographic information system for geo-historical resource sharing [2]. Another project called SEGRADA still based on semantic networks represents a historical crossroads exploiting a semantic database allowing the preservation of the information and the semantic links that connect them [3]. CIDOC CRM [4] is a semantic reference model constituting an information ontology relating to cultural heritage defined as an international standard by ISO since 2006. This standard was applied in order to develop event ontology for the First World War [5]. The paper [6] has structured data from Thailand's historical events to improve learning innovatively. The authors, in [7] and [8], have developed an ontology and proposed an intelligent virtual environment dedicated to history and heritage of industrial landscapes.

3 Historical Information Imperfection and Belief

3.1 Concepts and Definitions

The term history comes from the Latin Historia, not only "narrative of historical events" and "object of historical narrative" but also "fabulous story, nonsense", itself borrowed from the Greek Historia "research, investigation, information" and "result of an investigation", hence "narrative" [9].
The event is about what happens at a certain date and place. It is, therefore, spatiotemporal. Moreover, it is what distinguishes itself from every day, from the banal, from the ordinary. As such, it is a fact but differs from other facts of the same nature. Then, an event is what is true for a group, people, a culture, a large group of men. It has a proper meaning that is not only mechanical, like the facts. It is also, what modifies in the long term and irreversibly a certain human reality [10]. An event can be as much political, scientific, natural as historical.
The role of the historian is to explain historical events, to establish their relations, to define their causes and their consequences, in short, to reveal the coherence, the organization of a historical period, its continuity or its discontinuity. It thus determines a trend, a logic in the apparent chaos of events. The historian must, therefore, identify only the historical elements that make sense to answer a previously established problem and thus indicate the meaning of history [9].
Historical information will go through several distinct stages representing the historical research process. The authors of [11] proposed a life cycle of historical information consisted of six phases namely creation, enrichment, editing, retrieval, analysis and presentation. In the middle of the historical information life cycle, three aspects are identified which are not only central to history and computing, but also in the humanities in general: durability, usability, and modeling.

3.2 Data, Information and Knowledge in History

The use of the terms "information", "data", and "knowledge" in our context may be misleading. Several authors have attempted to define these three complementary concepts. According to [12], data is the raw material of information; it becomes information through a process of interpretation that gives signification sense. For the author of [13], knowledge is the result of apprehension, perception of reality. Thus, knowledge is a way of appropriating an object, of transforming the information perceived through its manifestation into something meaningful. For example, "the 15th century" is data. The phrase "This road dates from the 15th century" asserts information, while the phrase "This road is very old" represents knowledge.
In the field of historical information, historical data is a conceptual and semantic representation of spatiotemporal facts from multiple historical sources describing the unfolding of historical events. Histori-

cal information is a simplified model of reality often fraught with errors and imperfections. In order to make a decision, the historian must take cognizance of the contents of the sources (observations), criticize the authenticity of the source, then criticize the text itself, interpret its meaning and place it in the context of its production. After that, he proceeds to abstraction in order to create a cognitive model [14] and to propose his interpretation. This process is a source of divergence between the data produced and the data desired by the user for a given application [15]. Another source of this difference, according to [14] is the errors that can alter the data throughout their production process.

3.3 Information Imperfection

Not considering the imperfection of historical information can lead to misinterpretation [16]. It must, therefore, be integrated throughout the process, from data acquisition to restitution of assumptions. This requires being able to identify, model, and quantify it. In the works [17] and [18], the authors identified the following four major categories of imperfection. Knowing that these categories are by no means exclusive, and the data are simultaneously prone to imperfections of many kinds [19].

- **Uncertainty**: There is a doubt about the knowledge validity. The object is well defined, but its realization is uncertain. This situation is often linked to the random aspect of the measurement of a physical phenomenon.
- **Imprecision**: There is a difficulty in expressing knowledge clearly. This time, the object is not sufficiently defined. There is a lack of precision in the definition of the object. This may be due to the vague aspect of either the semantics of the concept or its limits.
- **Ambiguity**: There is a difficulty in agreeing and a doubt about how to define an object or a phenomenon. More specifically, there is a conflict if at least two contradictory classifications for a single object are possible. When a definition of a relation or an object can lead to several senses, or when the scale of the analysis is likely to lead to multiple interpretations, we speak of non-specificity.
- **Incompleteness**: There is missing or partial knowledge. Incompleteness is the absence of knowledge or lacuna knowledge. Absence is the phenomenon that occurs when in a database, values are missing from the description of certain objects. The lacuna is the fact that one or more objects in the database only partially describe a structure encompassing them.

3.4 The Degrees of Belief

In the broadest sense, a belief is a specific mental state that leads to give its assent to a certain representation or to make a judgment whose objective truth is not guaranteed and which is not accompanied by a subjective feeling of certainty [20]. In this sense, belief is synonymous with opinion, which does not imply the truth of what is believed, and is opposed to knowledge, which implies the truth of what is known. Because the truth of what is believed is only possible, and the adherence of the mind to the content of a belief may be more or less strong, the meaning of belief varies according to the degree of objective guarantee given to the representation and according to the degree of subjective confidence that the subject feels as to the truth of this representation [20].

- When the objective guarantee of opinion is very weak, or null, although the one who asserts it may have a very strong conviction to the contrary, “belief” is synonymous with a false or doubtful opinion, and is expressed as prejudice, illusion, enchantment or superstition.
- When beliefs are likely to be true or to have some objective foundation or are pending verification or justification, there is the talk of suspicions, presumptions, suppositions, previsions, estimates, assumptions or conjectures.
- When one wants to designate beliefs based on a strong subjective feeling but whose objective foundation is not guaranteed, one speaks of convictions, doctrines or dogmas.
- Lastly, we speak of belief in the last sense, to designate an attitude which is not, like opinion, proportionate to the existence of certain data and certain guarantees, but which goes beyond what these data or guarantees make it possible to affirm. It is in this sense that one speaks of the belief in someone or something to designate a form of trust or faith. In this case, the degree of subjective certainty is extreme, although the degree of objective guarantee can be very low.

4 Historical Knowledge Representation Model

In this section, we will first present the conceptual framework of our historical knowledge representation model. Then, we will expose the modeling results and discuss some possible exploitations.

4.1 Principles of The Model

The basic idea of our model is to represent knowledge about historical events by taking into account the imperfection of information. Specifically, we consider the quality of uncertainty of historical information. To

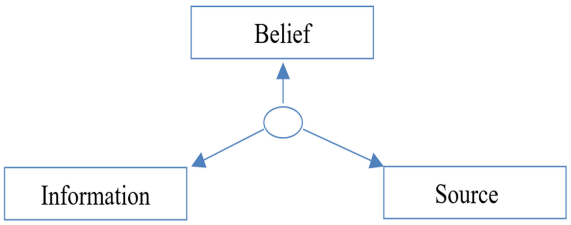

Figure 1. Relations between Layers

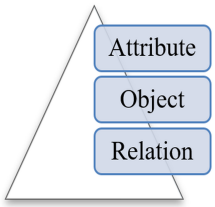

Figure 2. Granularity Levels of Beliefs

each knowledge, the unit is associated with a degree of belief corresponding to the degree of confidence that one grants it. We modulated the degrees of belief according to the level of certainty expressed by the historian. We summarize the main objectives of our model in the following points:

- Conceptual and semantic representation centered on historical events.
- Conservation of multiple versions of the history of an event from a multi-source environment.
- Representation of knowledge with degrees of certainty.

4.2 Multilayer Design of The Historical Knowledge Base

The principle of the multilayer approach that we propose consists of distinguishing three informational levels (layers) in each historical knowledge unit:

- **Information layer**: This is the first classic layer; it concerns the basic information necessary and useful to represent historical events, namely events, actors, objects, places, and dates.
- **Source layer**: It is a layer describing the directly lower layer and which completes the information in terms of references to the historical information. It is used to describe the source and date of reference.
- **Belief layer**: This layer will allow us to connect the information layer and the source layer with taking into account the imperfection of information in terms of uncertainty (see Figure 1). We associate with historical information a value qualifying the degree of belief that is assigned to it.

Assuming the information I equal to "a source S on date D asserts that an event E occurred at place P". Event and place belong to the information layer. The source and the date of reference are relating to the source layer since they are not directly related to the course of the event but rather to its reference. The information I combines two pairs of data; the first one provides information on the event course, a couple involving the event and another entity of the same information layer (event E, place P), the second one describes the first pair by the reference of this information (reference source S and reference date D). To represent the belief granted to such information, we connect these two pairs (event, place) and (reference source and date) by relation and associate with it a degree of belief. This relation reconstructs the knowledge unit involving the four data and introduces the third layer (belief), which will serve to represent the beliefs.
In our model, each knowledge unit describing the history of an event will be represented by a quadruple always involving the event, the source of reference and the date of reference as well as another entity of the information layer like event date, event place, actor or object. This quadruple is associated with a degree of belief preserved in an instance of this relation at the belief layer. The degree of belief as we have represented qualifies knowledge in its integrity, that is, the belief that one grants it is relative to that source, and not in an absolute way. This way of conceiving knowledge will enable us to simultaneously represent several testimonies about the course of each event while qualifying this knowledge via belief values.

4.3 Representation of Belief at Multiple Granularities

To strengthen and enrich our model, we assigned beliefs at three levels of granularity (Figure 2):

- **Attribute imperfection**: Belief masses can define some attributes in the conceptual model. For example, a historian may express his or her doubt about the area of a country.
- **Object imperfection**: An object can belong to a class with a degree of belonging (belief mass). Sometimes, the historian is squarely in doubt about the existence of an object. For example, it challenges the actual existence of a city affirmed in a historical narrative.
- **Relation imperfection**:The relation between classes is subject to a degree of certainty (mass of belief). For example, the historian expresses doubt about the place in which a war began. In this case, the imperfection is relative to the relation between the war as an event and the

Table 1. The 11 Degrees of Certainty Used in The Model

Certainty level	Certainty degree
Certainty	1
Extremely strong presumption	0.9
Strong presumption	0.75
Presumption	0.6
Favorable hesitation	0.55
Doubt	0.5
Unfavorable hesitation	0.45
Low probability	0.4
Very low probability	0.25
Extremely low probability	0.1
Certainty of negation	0

place.

4.4 Definition of Belief Degrees

In our model, we admit that a historian expresses his beliefs with a certain degree of certainty. To represent belief at each of the granularity levels mentioned above, we have established a set of degrees of certainty (see Table 1). These degrees are ranging from the certainty of knowledge that represents the 100% percentage to the certainty of the negation of the knowledge having 0% belief, passing in the middle by doubt representing equality between belief and non-belief with a percentage of 50%. The historian thus has several possibilities and can express his belief as to the truth of his knowledge according to the degree of belief towards which he converges the most. The knowledge unit that has not received a belief assignment is considered as certain for the historian and therefore will benefit from 100% belief.

4.5 Modeling Historical Knowledge

The entities that we deemed necessary to represent for covering the historical event are event, place, date, actor, object, and source. We respected the constraint that an event involves one or more places, is affected by one or more temporal entities, is concerned by one or more actors, and involves one or more objects. Figure 3 represents a simplified and centered event-restricted UML class diagram modeling the course of an event by abstracting other aspects such as relations between events, between places, between actors, and between sources. This diagram illustrates the quadruple relations that each time link event to another entity (of the information layer) by always involving the source layer (source and date of reference). Two attributes are provided for each quadruple relation (belief layer): one to represent a semantic concept that comes to make sense, the other to assign a degree of belief that represents the uncertainty expressed by the historian. This diagram shows the belief at the highest level of granularity, which is relation granularity.

4.6 Possible Exploitations of The Proposed Model

The way we modeled historical knowledge makes our model a terrain of very interesting exploitations and uses, such as:

- **History rewriting**: From the different versions of event history, one can reconstruct a single most plausible version for example by applying mathematical theories of the uncertain for the beliefs fusion and the decision making, such as the theory of belief functions of Dempster Shafer [21].
- **Events causal inference**: Such a knowledge representation model can also be used to deduce causal inferences that establish the causal links between events in works treating coherence, and relevance.
- **Conflict analysis and treatment**: The fact of keeping several simultaneous versions for the same event allows one to reflect the orthographic and semantic conflicts that exist between the sources and to detect the similarities and concordances.
- **Source classification**: The proposed model provides a field of reflection on historical data sources. It makes it possible to classify sources according to different aspects such as influence and coverage. It also allows the identification of competing currents, schools of thought or even intellectual traditions.
- **Historical knowledge edition and sharing**: This model can be implemented by a historical knowledge base shared by the historian community and representing a crossroads of historians where they can write their versions of the history of events and express their mutual beliefs on these and all the shared historical knowledge.

5 Conclusion

In this article, we have set ourselves the objective of representing historical events by keeping multiple versions of its history qualified by degrees of belief corresponding to the certainty degrees expressed by the historian. Our contribution is summed up in a multilayer design approach that distinguishes three informational levels namely information, source, and belief whose combination allows modeling and modulating historical knowledge. Further, we differentiated

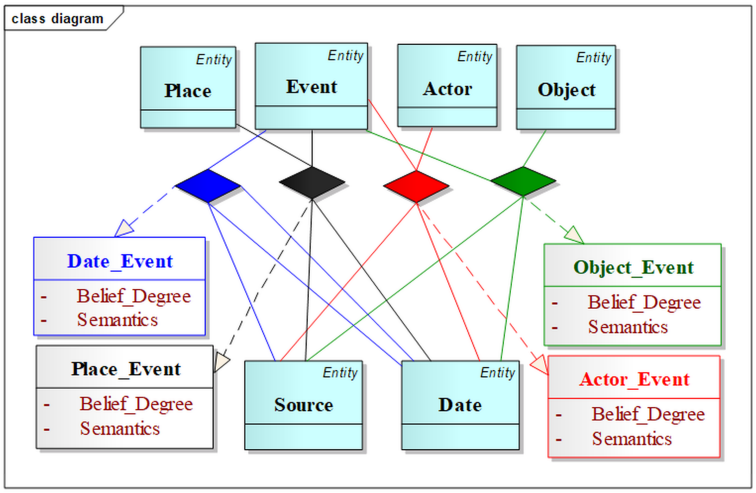

Figure 3. Event-centered Class Diagram

three levels of granularity (attribute, object, relation) to express belief, and we defined 11 degrees of uncertainty in belief. We also highlighted several possible areas of exploitation and very interesting uses of our model.

6 Perspectives

In this article, we have tried to represent historical knowledge by taking into account its imperfection in terms of uncertainty. In future work, we will use the proposed model to implement a collective platform for publishing and sharing historical knowledge dedicated to the historian community. It is a crossroads between historians, where they can introduce their versions of the events' history and can express their beliefs by certainty degrees about any historical knowledge kept in the knowledge base. We will also try to merge the beliefs expressed by historians about historical knowledge to decide as to their plausibility. For this, we will apply the theory of belief functions, also called evidence theory as a mathematical model dedicated to uncertain reasoning.

References

[1] Meroño-Peñuela A., Ashkpour A., Erp M., Mandemakers K., Breure L., Scharnhorst A., Schlobach S., and Harmelen F. Semantic technologies for historical research: A survey. In *Semantic Web*, number 6.6, pages 539–564, Lille, France, 2015.

[2] Beretta Francesco and Butez Claire-Charlotte. The SyMoGIH project and Geo-Larhra: A method and a collaborative platform for a digital historical atlas. In *Digital Humanities 2014*, Digital Humanities 2014. Conference Abstracts, pages 117–119, Lausanne, Switzerland, July 2014. EPFL, Lausanne / UNIL, Lausanne.

[3] Kalus M. Semantic networks and historical knowledge management: Introducing new methods of computer-based research. volume 10, 2007.

[4] Doerr Martin. The cidoc crm - an ontological approach to semantic interoperability of metadata. *Ai Magazine - AIM*, 24, 01 2003.

[5] Hyvönen E., Lindquist T., Törnroos J., and Mäkelä E. History on the semantic web as linked data - an event gazetteer and timeline for world war i. In *Geographical information systems 1*, 2012.

[6] Ekkachai S.J. and Kulthida T. Development of semantic web for thai historical events. In *Proceedings of the Doctoral Consortium at the 18th International Conference on Asia-Pacific Digital Libraries, and Asia-Pacific Forum of Information Schools*, 2016.

[7] Laubé Sylvain, Garlatti Serge, Querrec Ronan, and Rohou Bruno. Any-Artefact-O : an ontology developed for history of industrial cultural landscape. In *2nd Data for History workshop*, Lyon, France, May 2018. Pôle histoire numérique (Digital history department) of the LARHRA laboratory.

[8] Laubé S., Querrec R., Abiven M. M., and Garlatti S. Lab in virtuo : A intelligent virtual environment dedicated to history and heritage of industrial landscapes. In *#dhnord2018: Matérialités de la recherche en sciences humaines et sociales*, Lille, France, 2018.

[9] Ollier-Pochart E. Les écritures de l'histoire dans les romans québéquois de la décennie 1980-1990. 2012.

[10] Granda Shirley. *Critique et analyse des bases de données sur les catastrophes naturelle pour la caractérisation des évènements hydro-météorologique a forts impacts socio-économiques.* PhD thesis, Laboratoire d'Études des Transferts

en Hydrologie et Environnement (LTHE), Bâtiment OSUG-B, Domaine universitaire, BP 53, 38041 Grenoble cedex 09, 2014., 2014.

[11] Boonstra Onno, Breure Leen, and Doorn Peter. Past, present and future of historical information science. *Historical Social Research/Historische Sozialforschung*, pages 4–132, 2004.

[12] Reix R. Systèmes d'information et management des organisations. In *2ème édition. Vuibert*, 1998.

[13] Pornon H. Sig la dimension géographique du système d'information. In *InfoPro, Dunod*, 2011.

[14] Devillers Rodolphe. *Conception d'un système multidimensionnel d'information sur la qualité des données géospatiales*. PhD thesis, Université de Marne la Vallée, 2004.

[15] Bédard Y. Uncertainties in land information systems databases. In *The Proceedings of Eighth International Symposium on Computer-Assisted Cartography (AutoCarto 8)*, pages 175–84, 1987.

[16] Deitrick Stephanie A. Uncertainty visualization and decision making: Does visualizing uncertain information change decisions. In *Proceedings of the XXIII International Cartographic Conference*, pages 4–10, 2007.

[17] Fisher Peter F. Models of uncertainty in spatial data. In *Geographical information systems 1*, pages 191–205, 1999.

[18] De Runz Cyril, Herbin Eric, Frederic Piantoni, and Michel Herbin. Towards handling uncertainty of excavation data into a gis. In *Proceedings of the 36th CAA Conference*, volume 2, page 6, 2008.

[19] Desjardin E., Nocent O., and De Runz C. Prise en compte de l'imperfection des connaissances depuis la saisie des données jusqu'à la restitution 3d. *established by: Mauro Cristofani and Riccardo Francovich*, (supplemento 3):385–396, 2012.

[20] Engel Pascal. Les croyances. In *D. Kambouchner, Les notions de philosophie, Gallimard*, volume II, pages 9–101, 1995.

[21] Wilson N. Systèmes d'information et management des organisations. In *Kohlas J., Moral S. (eds) Handbook of Defeasible Reasoning and Uncertainty Management Systems. Handbook of Defeasible Reasoning and Uncertainty Management Systems*, volume 5. Springer, Dordrecht, 2000.

Fairouz Zendaoui received the diplôme d'ingénieur d'état and the master degree both in computer science from the Ecole nationale Supérieure d'Informatique (ESI) in Algiers, Algeria, in 2010 and 2014, respectively. She joined the Institut National de la Poste et des Technologies de l'Information et de la Communication (INPTIC) in Algiers, as Master Assistant B between 2015 and 2018, then as Assistant Master A from 2018 until today. She is a PhD student affiliated to the Laboratory of Communication in Computer Systems (LCSI) of ESI since 2015 and a member of the Laboratory of Research in ICT (LABORTIC) of INPTIC. Her research interests include knowledge representation, uncertain reasoning, belief functions, artificial intelligence, algorithmic and programming, and information systems.

Walid-Khaled Hidouci is a professor in computer science at Ecole nationale Superieure d'Informatique (ESI) in Algiers. He leads the Advanced Database Systems team in the LCSI research laboratory. His main topics of interests are: database systems, data structures, artificial intelligence, operating systems, and parallel programming.