The DICTOMAGRED case study

1. Introduction

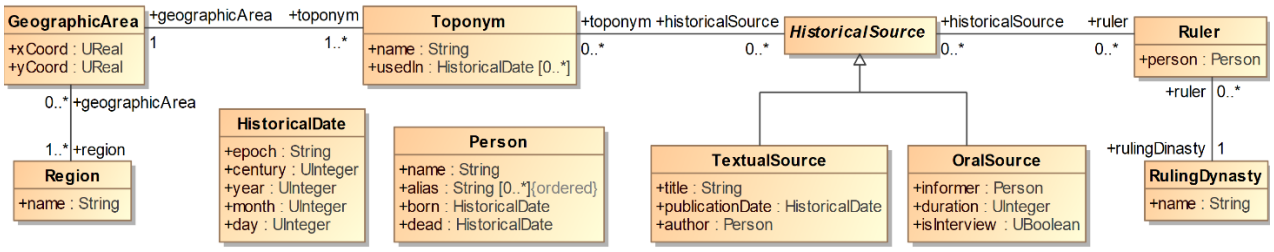

Figure 1 Toponyms – Domain model

This case study is taken from the DICTOMAGRED project [31] in the digital humanities domain, which analyses historical sources (including, e.g., oral testimonies and legal documents), most of them in Arabic. Such sources contain geographical references describing routes through different areas in the Maghreb, their place names (toponyms), and other related historical events. The main goal of the project is "to provide a software tool for humanities specialists to retrieve information about the location of toponyms in North Africa as they appear in historical sources of medieval and modern times". This project has been extensively used in other proposals [32, 33, 34] that employ domain models to represent the inherent vagueness of the type of information managed by the project. For example, some of the sources are not reliable, provide imprecise or incomplete descriptions and details, or may have been altered over the years, with different versions of the same facts (as with oral sources). Our goal is to illustrate here how our proposal enables the specification of the vagueness associated to DICTOMAGRED models and show how we can even be more expressive than these previous proposals by allowing different users to express different opinions and analyzing how these disparate opinions could be reconciled.

Let us begin with a conceptual model that represents the types of entities and relationships that capture the toponyms of the region known today as the Maghreb. Figure 1 shows that model. Toponyms are located in a geographical area that belongs to one or more regions (regions and their borders have frequently changed over the years). The names of the toponyms have been mentioned in different historical sources, dated when certain rulers where governing particular dynasties.

Suppose now that two researchers, Anne (a Geographer) and Brandon (a Librarian) have independently created two different object models about the *Ashir* toponym, after consulting different sources. These models are shown in Figure 2 and Figure 3. Both Anne and Brandon agree on some details, such as the geographic area and the toponym name (i.e., the content of the objects Maghreb, area1 and Ashir), and also about the existence of Hammad as the ruler of the BanuHammad dynasty. However, their models differ in other aspects such as missing sources and rulers.

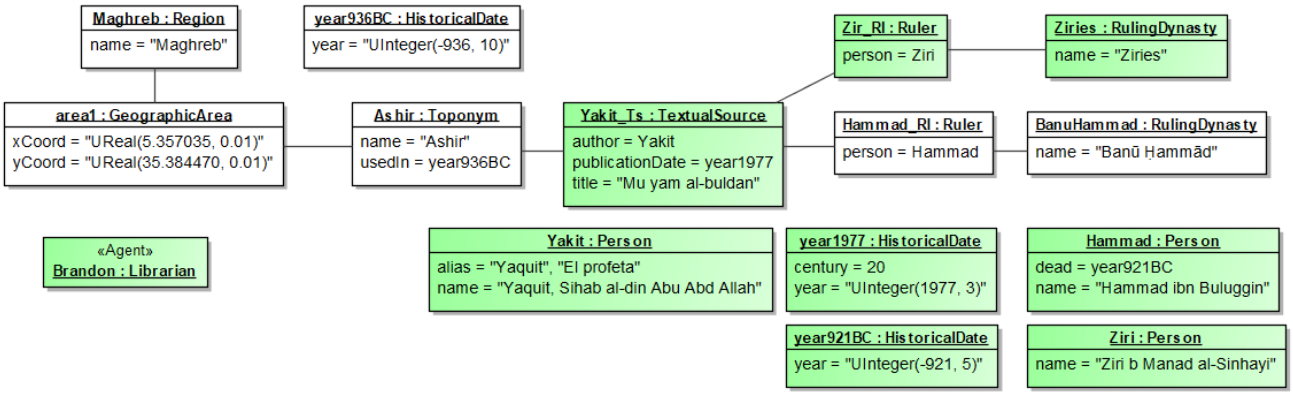

Figure 2 Model developed by Brandon about the Ashir toponym

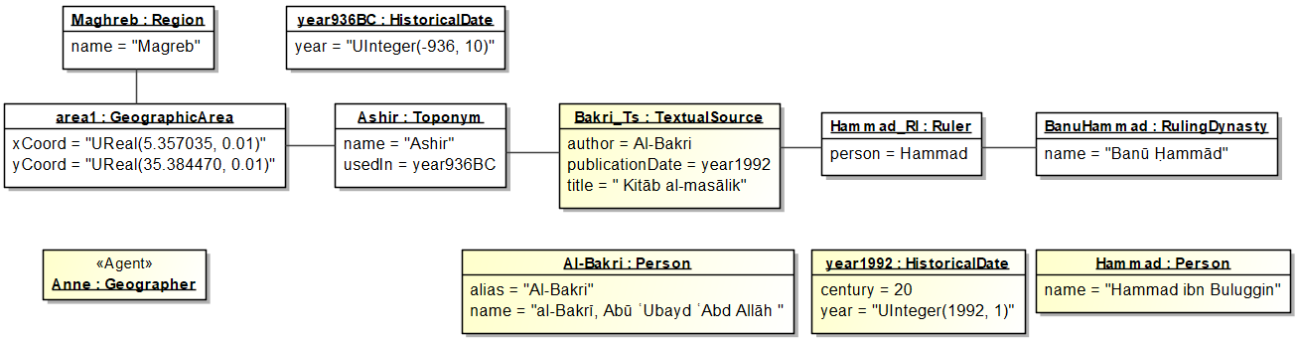

Figure 3 Model developed by Anne about the Ashir toponym

2. Enriching the model with subjective opinions

Once they have developed their models, they could add their *degree of certainty* (or uncertainty) to some of the model elements about which they are not completely sure.

Figure 4 and Figure 5 show the previous models but enriched with their individual opinions about some of the elements of their models. For example, Brandon is not completely sure whether the reference mentioned in the source Yakit_Ts actually refers to the toponym Ashir, and therefore assigns one subjective opinion, namely b_Yakit, to the link between the toponym and the source. Note that neither the toponym nor the source are questioned, and therefore neither of them has an associated opinion (which is equivalent to assigning them a true dogmatic opinion). He also assigns opinions to objects Ziri_RI and Hammad_RI because he is not sure that they were the rulers at the time of the source (although he is certain about the existence of these two persons, and of the two dynasties). Finally, he also expresses some doubts about the year in which Hammad died.

Similarly, Anne's doubts about the elements of the model she has designed are shown in Figure 5. Remember that model elements with no assigned belief are assumed to be dogmatic trues, i.e., SBoolean(1,0,0,1).

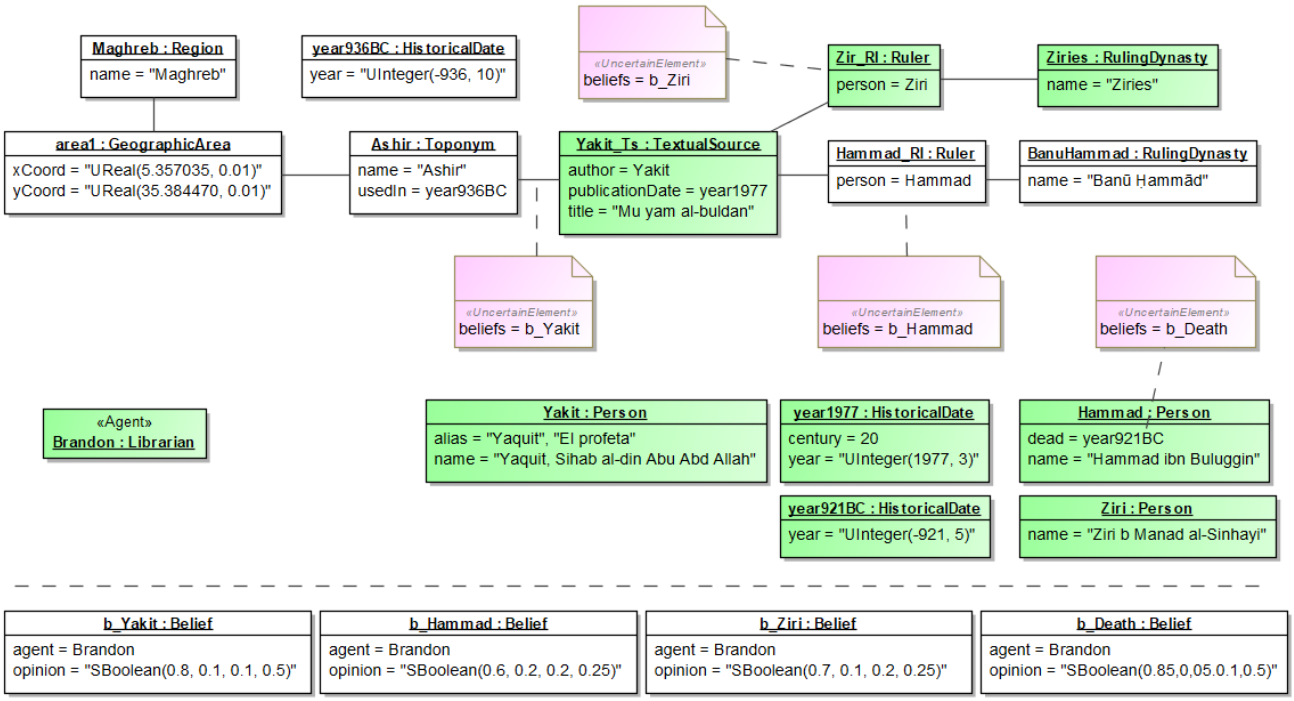

Figure 4 Model developed by Brandon about the Ashir toponym, with his opinions.

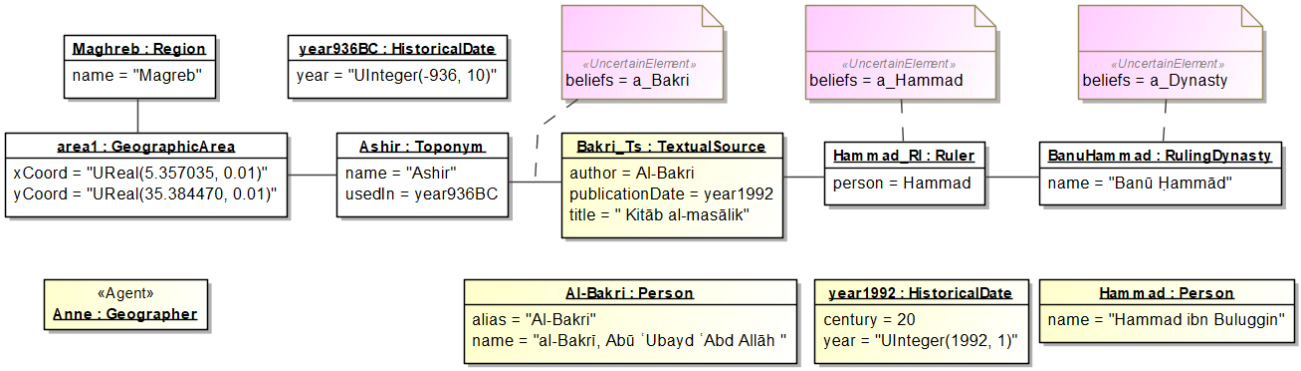

Figure 5 Model developed by Anne about the Ashir toponym, with her opinions.

To solve their disagreements on some of the most controversial topics, they also rely on the opinion of Carol, a well-known expert in the field, who will not create an object model but provide her opinion on the models proposed by Anne and Brandon.

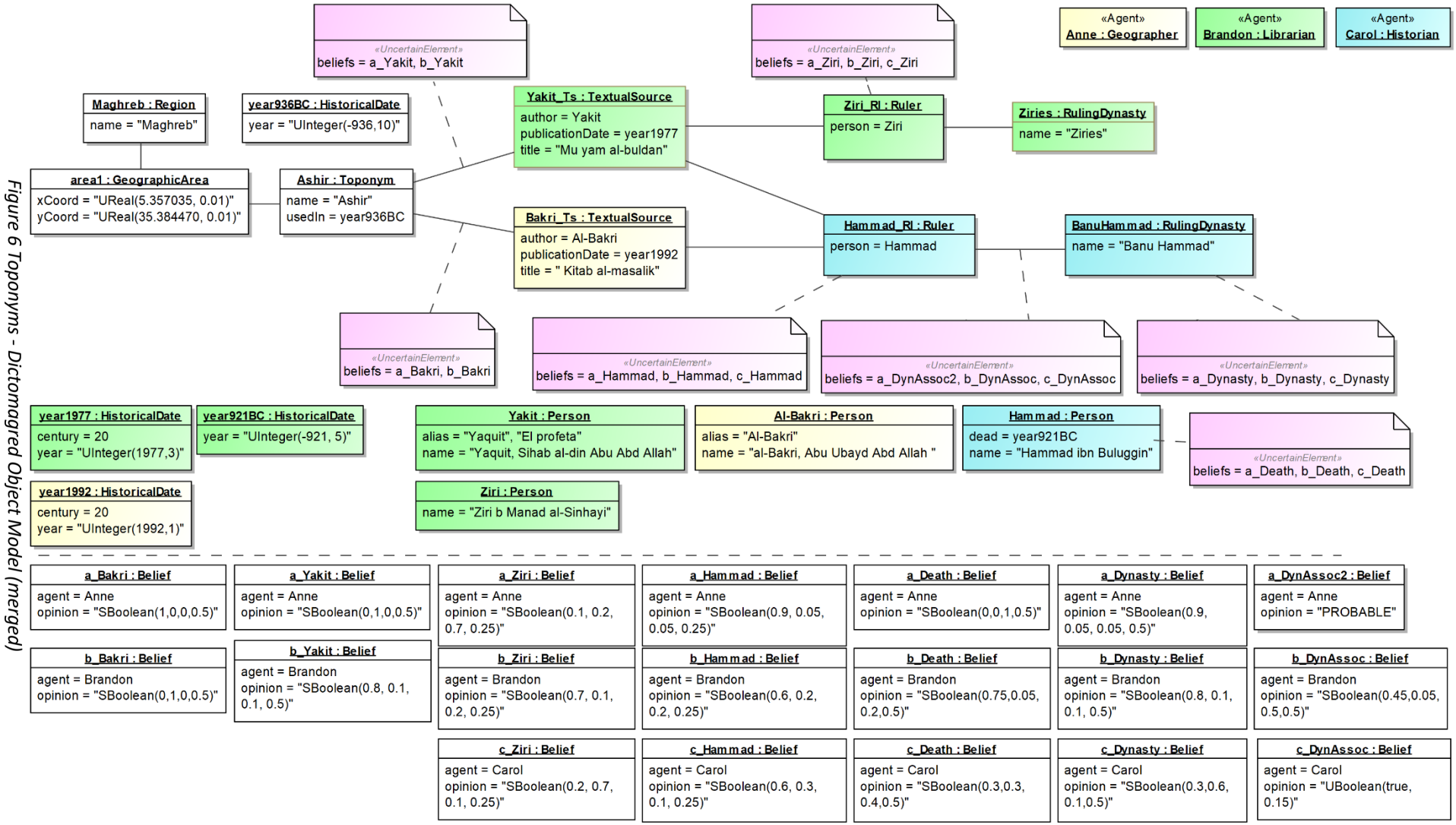

Figure 6 Toponyms - Dictomagred Object Model (merged)

Suppose now that Anne and Brandon decide to collaborate in order to develop a more accurate model of the Ashir toponym and its related sources. Then, they share their models and try to agree on a common one. To do this, they merge their two models and tag their elements with their subjective opinions, as shown in **Error! Reference source not found.**.

Once we have a model enriched with the opinions of individual agents, the question now is how to combine them to produce an opinion that can be assumed to better reflect the truth than each belief in isolation. This is called Belief fusion [25], and in general it is a challenging process because different belief fusion situations can occur in practice and may require different fusion operators depending on the purpose and nature of the fusion process [26, 52].

3. Fusing opinions

3.1. Situation 1

Context:
Anne and Brandon have studied the same toponym, but they have used two different textual sources. While Anne has used *Kitab al-masalik*, Brandon opted for *Mu yam al-buldan*, which has led to contradictory information. They try to agree on what is the best source to keep studying together, but in case they doubt about the usefulness of the source that the other person has chosen, they will keep using the one they originally selected. Note that they do not doubt about the existence of the textual source but about its usefulness to study the toponym, hence they place their subjective opinion on the association between the toponym and the source.

Fusion operator: In this case, the fusion operator that matches the situation the best is Belief Constraint Fusion (BCF): no room for compromise in case of totally conflicting arguments, hence the fusion result is not defined in that case.

Result: After revisiting the class model, Anne is stubbornly sure that the source she is using the only text that contains true information and that Brandon's source is completely mistaken. Therefore, she has assigned SBoolean(0,1,0,0.5) to the existence of the association between Ashir and Bakri_Ts; and SBoolean(1,0,0,0.5) between Ashir and Yakit_Ts to express that she does not believe that Bakri_Ts is a textual source to study the toponym, while Yakit_Ts is.

In contrast, Brandon considers that Anne's source is not correct and that his is better. Therefore, he assigns SBoolean(0.8,0.1,0.1,0.5) and SBoolean(0,1,0,0.5) to the association between Ashir and Yakit_Ts and Bakri_Ts, respectively.

Note that, since there are two textual sources (Yakit_Ts and Bakri_Ts), each textual source is a binary domain (it can be either picked or not) and therefore the prior probability (i.e., the base rate) of both sources is 0.5.

The result of the BCF operator on the opinions on the association between Ashir and Bakri_Ts is undefined (due to the contrary opinions); and the result of the BCF operation on the opinions on the association between Ashir and Bakri_Ts is SBoolean(0, 1, 0, 0.5). Since they cannot agree on studying only one source, they decide that each will keep studying their own and both are kept in the object model.

3.2. Situation 2

Context: Even studying different sources, both Anne and Brandon seem to have discovered that someone called Hammad was a ruler. Nevertheless, Brandon also thinks that there was another ruler during that period of time named Ziri. Anne, Brandon and Carol annotate the model with their subjective opinions about the existence of the two rulers.

Fusion operator: After inviting Carol to give her opinion, the amount of independent evidence increases (one more agent/source of knowledge) and the degree of uncertainty must be reduced (due to her expertise). Therefore, the Aleatory Cumulative Belief Fusion (ACBF) operator can be applied.

Result: The possible options in this case are that the ruler was Ziri, the ruler was Hammad, both Ziri and Hammad were rulers, or none of them was. Then, we are facing a quaternary domain for which for which their opinions' base

rate must be 0.25. After applying the operator on the opinions of both objects, the results are:

```
ACBF(b_Ziri, a_Ziri, c_Ziri)       = SBoolean(0.39, 0.54, 0.07, 0.25)
ACBF(b_Hammad, a_Hammad, c_Hammad) = SBoolean(0.82, 0.15, 0.03, 0.25)
```

The group's belief of the ruler Ziri is not very much supported with a belief below 0.4, a disbelief of almost 0.54 and not much uncertainty; while there seem to be a consensus on the existence of the ruler Hammad, supported by a belief of 0.82. The three agents decide to agree that Hammad was a ruler while Ziri was not.

3.3. Situation 3

Context: Anne was unable to set a death date for the ruler Hammad, but Brandon found some hints on the textual source that he studied that make him think that Hammad may have died around 921BC. While Anne is unsure about this fact (SBoolean(0,0,1,0.5)), Brandon supports it (SBoolean(0.75,0.05,0.2,0.5)), and Carol finds it sensible (SBoolean(0.3,0.3,0.4,0.5)).

Fusion Operator: To try to agree on whether 921BC was the date of Hammad's death or not, the agents may decide to apply the Averaging Belief Fusion (ABF) operator, to give each opinion the same weight; or the Weighted Belief Fusion (WBF) operator, to give greater weight to more confident opinions.

Result: Since the result of both operators support the death date more strongly than not, the agents agree to keep the date of Hammad's death in the final model.

```
ABF(b_Death, a_Death, c_Death) = SBoolean(0.53, 0.12, 0.35, 0.5)
WBF(b_Death, a_Death, c_Death) = SBoolean(0.63, 0.12, 0.25, 0.5)
```

3.4. Situation 4

Context: While the group has agreed to keep studying the ruler called Hammad, they discuss about the dynasty he belongs to. Despite Anne's certainty about the fact that it belongs to the *Banu Hammad* dynasty (as reflected in her opinions), Brandon and Carol are not so sure that this was the case. Brandon beliefs that the dynasty existed but has some concerns about the fact that Hammad was part of it (i.e., his concerns are reflected in the association). In contrast, Carol doubts about both facts: the existence of the object BanuHammad and of the association. Note also that Carol has not specified a Subjective Opinion about the association but her opinion with a probability (i.e., Credence), which will be automatically lifted to the corresponding Subjective logic value SBoolean(0.15,0.85,0,0.15) when needed; for instance, to operate with other SBoolean values.

The group would like to make their findings public, but before going to the press with a statement like “there was a ruler named Hammad who belonged to the Banu Hammad dynasty”, they need to see what their collective belief is about this.

Fusion operator: To compute the collective belief, we need to aggregate each individual belief about the different model elements and then combine the resulting beliefs with an appropriate fusion operator (⊗). The individual opinions are aggregated with using an and operator. Hence, the formula is:

```
⊗ (b_Hammad and b_DynAssoc and b_Dynasty, a_Hammad and
    a_DynAssoc and a_Dynasty, c_Hammad and c_DynAssoc and c_Dynasty)
```

In this situation, the three agents have the same credit and they are observing the same fact, and hence the most suitable fusion operator is the Averaging belief fusion (ABF).

Result: Using the ABF operator, the result of the formula above is SBoolean(0.26,0.37,0.37,0.05), whose projection is 0.28, and its degree of uncertainty is 0.37. This is not sufficiently convincing and thus the three experts decide to continue studying the toponym before making any press release.

Note how these operators can be used in standard OCL expressions, given that they form part of the extended UML and OCL type system that we have developed, and where SBoolean is just another primitive datatype [36].

References

[25] Audun Jøsang. 2016. Subjective Logic - A Formalism for Reasoning Under Uncertainty. Springer. https://doi.org/10.1007/978-3-319-42337-1

[26] Audun Jøsang, Dongxia Wang, and Jie Zhang. 2017. Multi-source fusion in subjective logic. In Proc. of FUSION'17. IEEE, 1–8. https://doi.org/10.23919/ICIF.2017.8009820

[31] Miguel Angel Manzano, Helena de Felipe-Rodríguez, and Laura Gago-Gómez. 2018. DICTOMAGRED: Diccionario de Toponimia Magrebí. https://dictomagred.usal.es/

[32] Patricia Martín-Rodilla and Cesar Gonzalez-Perez. 2018. Representing Imprecise and Uncertain Knowledge in Digital Humanities: A Theoretical Framework and ConML Implementation with a Real Case Study. In Proc. of TEEM'18. ACM, 863–871. https://doi.org/10.1145/3284179.3284318

[33] Patricia Martín-Rodilla and Cesar Gonzalez-Perez. 2019. Conceptualization and Non-Relational Implementation of Ontological and Epistemic Vagueness of Information in Digital Humanities. Informatics 6, 2 (2019), 20. https://doi.org/10.3390/informatics6020020

[34] Patricia Martín-Rodilla, Martin Pereira-Fariña, and Cesar González-Perez. 2019. Qualifying and Quantifying Uncertainty in Digital Humanities: A Fuzzy-Logic Approach. In Proc. of TEEM'19. ACM, 788–794. https://doi.org/10.1145/3362789.3362833

[36] Anne Muñoz, Carol Burgueño, Victor Ortiz, and Brandon Vallecillo. 2020. Extending OCL with Subjective Logic. Journal of Object Technology 19, 3 (2020), 3:1–15. https://doi.org/10.5381/jot.2020.19.3.a1

[52] Rens Wouter van der Heijden, Henning Kopp, and Frank Kargl. 2018. Multi-Source Fusion Operations in Subjective Logic. In Proc. of FUSION'18. IEEE, 1990–1997. https://doi.org/10.23919/ICIF.2018.8455615

Appendix A

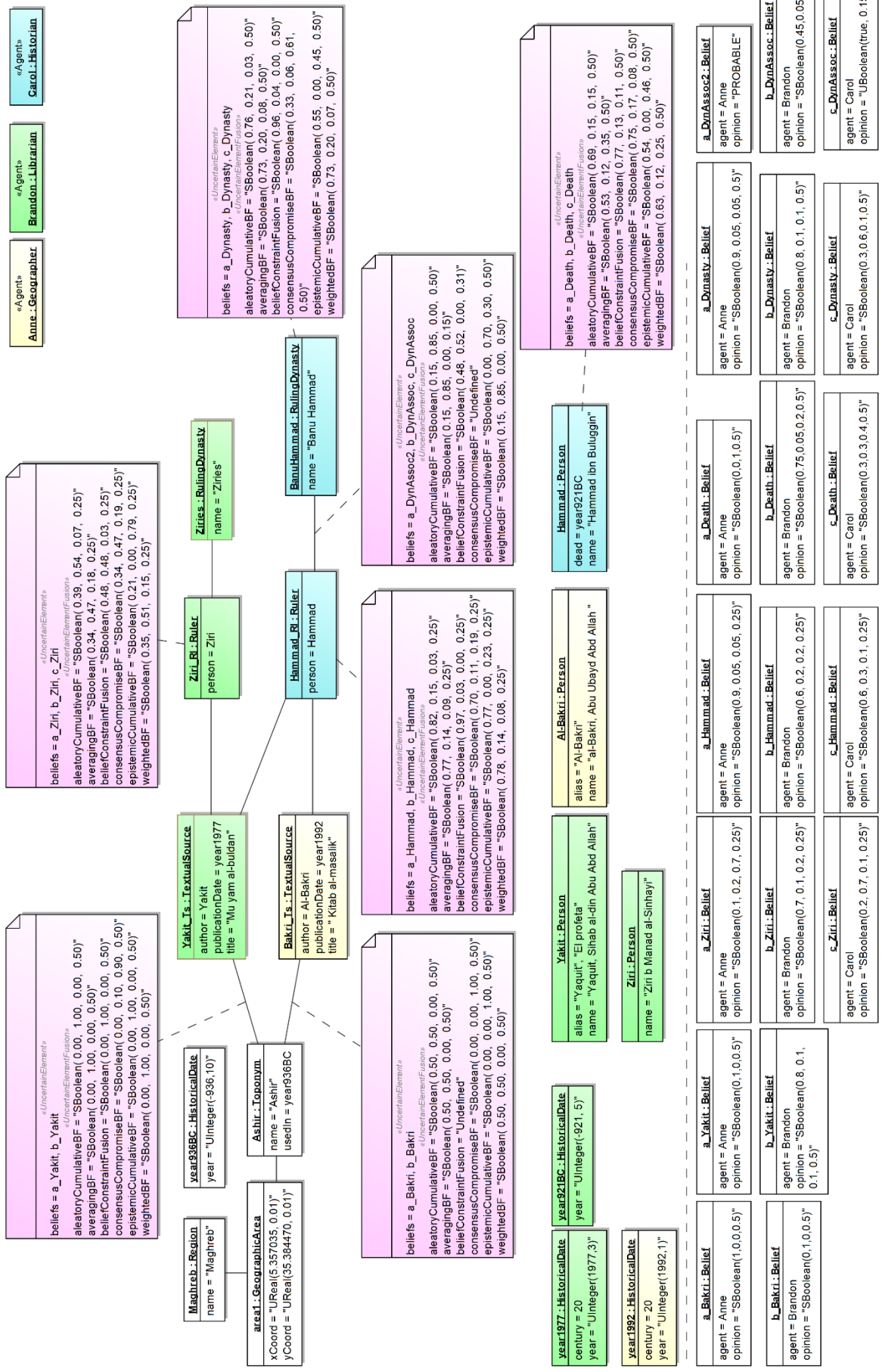

Figure 7. Toponyms - Dictomagred Object Model (merged and with opinions fused)