Geological Information Capture with Sketches and Ontologies

Yuanwei Qu[✉]

SIRIUS Centre, University of Oslo, Oslo, Norway
quy@ifi.uio.no

Abstract. Recent groundbreaking research [3] has shown the viability and usefulness of *qualitative information processing* in the digitisation of geosciences, as well as the possible role of semantic knowledge representation as an underlying framework. While there is significant work on e.g. geological ontologies and the integration of geological data, means of data entry that are suitable for geologists remain a bottleneck. Form and table data entry methods are good matches for an ontology-based data entry system; however, they cannot fulfill geologists' needs: to use sketches to express their idea and knowledge. This PhD project will lead to a novel user interface that will allow geologists to use sketches as an information entry method to input and store qualitative geological information in RDF format, which will enable machine-readable qualitative geological data query and reasoning in the future, and bring geology digitization one step forward. In this research work, assuring that qualitative geological information is captured completely and correctly is one of the main challenges.

Keywords: Semantic web · Ontology · Geology digitization

1 Introduction

Geology is the subject of studying the Earth that feeds humankind, both theoretically and practically. The resources, energy, and environment needed for human society rely on a deep understanding of the Earth. Thus, the development of geology is essential to the whole society.

In recent years, the use of information technology in geology has increased, many tools and systems have been developed to assist geologists in overcoming obstacles [7]. Machine learning has assisted geologists in interpreting and predicting results from quantitatively-oriented geological data [15,19]. But there is also a growing body of work on non-numerical information.In semantic technologies, semantics experts and geoscientists have been working together to address the problem of geological data silos [10,21]. Domain ontologies have been provided to avoid ambiguity and to support collecting heterogeneous data [1,6,22].

Recent work like that on the GeologicalAssistant [3] and SiriusGeoAnnotator [9] has shown that qualitative reasoning on geological information and image

data annotation are also possible, important, and interesting to the industry. Compared with machine learning, the decision support made by qualitative reasoning is an analogue of human inference, which is explainable and reversible. Moreover, it is applicable when extensive numerical data is not available. However, digitizing qualitative geological information is still problematic. In contrast to the maturity of numerical data collection and storage, infrastructure methods for capturing and storing digital qualitative geological information still need investigation. Massive qualitative geological information within sketches, illustrations, and geological photos is still waiting to be digitized.

Multiple easy-to-use front-end applications and user interfaces have been proposed to support RDF data entry and maintenance. It is often pointed out that such ontology-based interfaces for data entry can adapt arbitrary input ontologies, and their user interfaces are novice-friendly [2,5,12]. But apparently, these systems cannot fully meet geology users' needs. Being generic, the interfaces use elements like forms, or reflect the knowledge graph structure, etc. While these interfaces *could* be used for arbitrary domains, they are not necessarily appropriate for the intended users. Geologists prefer to use drawings to express their ideas and knowledge. Geologic sketch tools like [11,13] are designed from on ad hoc needs and aiming for visualisation. The gap between the geological qualitative information entry and geology digitization is still recognizable.

What geologists need is an easy-to-use tool that can allow them to draw sketches as an information entry method and store the information in a format that is machine-readable, suitable for inference and ready to support the rising trend of geology digitization. For this work, we are focusing on building an intelligent system for capturing the geological qualitative information by drawing sketches, and the system will generate knowledge graphs to store captured geological information. An example of this idea has been illustrated in Fig. 1, a schematic geological scenario drawn by a user and a corresponding knowledge graph generated by the system. By providing this system, this work aims to allow geologists to input and update geological information easily, pave the path for digital geoscience query, qualitative reasoning, and support supervised machine learning for both academia and industry.

2   State of the Art

In this section, related work consists of three major parts: 1) how semantics experts tried to offer user interfaces to improve data input efficiency and enhance the user experience for the arbitrary domain. 2) how geology and semantics researchers worked together to harmonize geological data in the context of semantic technologies. 3) how geologists and computer scientists build artifacts to assist researchers in conducting their studies.

2.1   User Interface for RDF Data Entry

UTILIS is a method presented by Hermann et al. [8] that aims to utilize existing objects, their properties, and known information of new objects and assist users
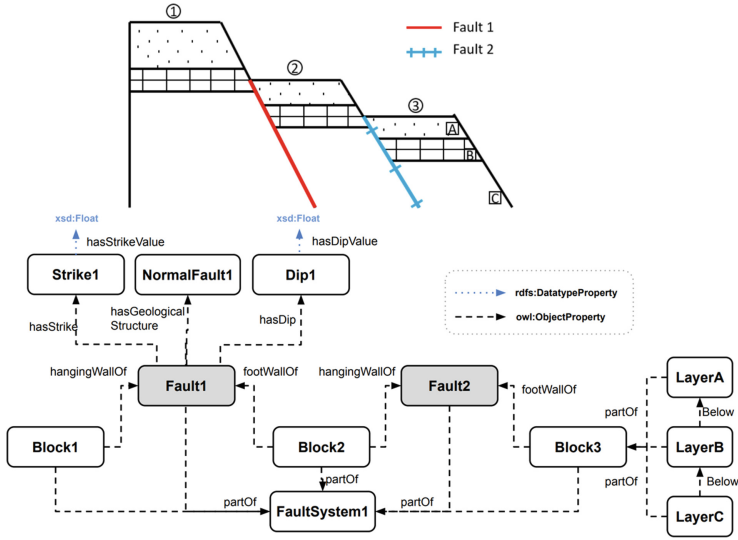
Fig. 1. Above: a sketch containing two geological faults (Fault*n*), three blocks (Block*n*), and each block has layer A, B, and C. Fault 1 is in red and Fault 2 in blue. Below: part of the knowledge graph of the geological scenario above (Color figure online)

when they are adding or updating semantic web data. This system matches similar data objects and uses their properties as suggested descriptions to provide to users when adding new data. Aiming to make casual users create their own RDF data, Butt et al. [2] propose ActiveRaUL, which can automatically generate web form-based user interface from input ontology with no domain-specific limitation. With a similar purpose, Frischmuth et al. [5] present OntoWiki, a user interface for RDF knowledge graphs integrated with data management and visualization method.

Based on the method of UTILIS, FORMULIS was developed by Maillot et al. [12], a form-based user interface for knowledge graph editing. FORMULIS can not only give suggestions while the user is adding or updating RDF data, but also offer users an easy-to-use interface without the need for IT experts have to set it up first, and extended possibility depends on users' needs. With a similar idea of reducing reliance on IT experts in the context of data retrieval, Soylu et al. [17] render an overview and discuss the achievability of ontology-based data access, visual representation and interaction for users in an ontology-based system, and potential user roles within such system. However, all these methods and systems still cannot fully meet geology users' needs.

2.2 Geological Data Integration

Integrating geological data is a well-known challenge in this subject, and diverse approaches have been applied to address this issue. Therein, semantic technolo-

gies have been achieved several successes. By applying a shared conceptualization model to describe geological map objects and their properties in the model to bring semantic unification, Laxton [10] has successfully deployed a system to integrate geological map data across several nations.

Holding the vision of breaking the data silos of geological data, the Deep-Time Digital Earth Program was proposed by the International Union of Geological Science and several associations, surveys, and institutions. Scientists are trying to use such platform to link and integrate data in existing databases, and serve for more future knowledge graphs in geoscientific usage [21]. Aiming to integrate subsurface geological data within a digital modelling flow and let experts in diverse disciplines involved, Verney et al. [14] presented their works using designed ontologies as knowledge representations to characterize and correlate subsurface geological structures, and record parameters of characterized targets in the system. Having the goal of integrating multi-source early geological data, Wang et al. [20] proposed a semi-automatic method based on ontology and natural language processing to reconstruct low-cost vector geological profiles.

2.3   Software Assistance for Geological Research

Nowadays, geoscientists, especially geologists, still prefer to use pens to draw sketches on paper to illustrate their ideas or concepts. Following this preference, both Lidal et al. [11] and Natali et al. [13] have presented their approaches for sketching, drawing, and visualizing geological models in a time-efficient way that allows geologists to interact with their sketch model and to communicate and share their conceptualization model with others.

Images are usually considered as a type of critical data for geoscience, because images can contain a large amount of information that is useful for researchers. Aiming to support geoscientists to annotate geological information within the image and make those content accessible, SiriusGeoAnnotator has been deployed [9]. This artifact offers users an interface to annotate geological information. With the help of embedded ontology, the user's annotations will transfer to RDF format data, which can increase image query efficiency.

Din et al. [3] proposed GeologyAssistant, a logic-based formal system for geological reasoning to assist exploration work in the energy industry and reduce the laborious work of geoscientists. The system is designed to infer and generate multiple sound possibilities for an uncertain subsurface interpretation by taking qualitative geological information as input data and combing the formalized geological knowledge in first-order logic. This work proves that qualitative geological reasoning is also executable and essential. But where to find such formalized geological qualitative information?

3   Problem Statement

As mentioned in the previous section, the gap between the geological qualitative information entry and geology digitization is still recognizable. And this work

aims to propose an intelligent system for geological qualitative information entry and storage in the RDF format.

Our core hypothesis of this work is that geo-user with no semantic knowledge and experience can use sketches and drawing as satisfactory and efficient information entry options to easily and precisely input and update qualitative geological information in the RDF format, without being highly dependent on data experts to set up and maintain data evolution. Other hypotheses: this intelligent system can be modified to adapt knowledge from different geology sub-domains; the stored geological qualitative information can be used in data query, qualitative reasoning, and supervised machine learning.

3.1   Research Questions

The previous section discussed the need for geologists to entry of geological information, and this research's primary objective has been settled. To achieve this goal, the following research questions have been derived:

RQ1. How can we achieve loose coupling between the generic ontology-driven components and the specialised sketching component?

It should be possible to use the system with different geological ontologies, e.g., covering different areas of geology, using different upper ontologies, etc. A menu-based interface can be made to automatically adapt to any given ontology, but a sketching interface is more intimately tied to the intended domain. Representations of entities as lines, areas, colours cannot be read from the ontology, any more than useful modes of interaction with these graphical elements. The challenge is therefore how to bridge the relatively rigid sketch-based part with the generic ontology-driven components.

RQ2. How can the system's information entry method be sufficient and accessible for users to express their knowledge?

Sufficiency is about allowing users to express all information required for the task at hand. The system should permit users to express their domain knowledge without being hindered by a lack of expressiveness. Accessibility is about ensuring that the information entry methods can easily be adapted to domain users and meet their needs.

RQ3. How can this geological information entry method be precise and avoid users' missing input?

Compared to forms or tables, using a drawing tool as a geological information entry method increases makes it more difficult to ensure that the system store the correct information. What the users draw and what the system stores are may not be the same, the system should be able to confirm its stored information with users. For the users, they shouldn't be expected to understand RDF, their work is to describe scenarios. Checking the RDF quality is not users' but semantic experts' work.

4   Research Methodology

This research work is still in its initial stage. As the research progresses, the current methodologies might be changed to fit evolving and emerging questions and challenges.

The initial idea of designing such a system is inspired by the work of the Geological Assistant [3]. This work has shown the viability and usefulness of qualitative information processing in the digitisation of geology, but where is the formalised geological qualitative information? The geology community needs an easy-to-use method to prepare data for qualitative information processing. The SiriusGeoAnnotator [9] took the first step to annotate image information, based on this, we want to make the system one step forward to allow users to sketch, instead of only annotating. CogSketch [4] is a sketch tool with a knowledge base for cognitive science research purposes, which also inspires us on what could a sketch tool with an ontology look like.

The system consists of an ontology-driven part based on similar concepts to previous ontology-based systems and a graphical part to allow drawing. But designing such a system to balance two parts is a question. Besides, validating the system's usefulness is also a challenge. Due to different domains of geology are having diverse needs, the scope of this work also needs to limit to a suitable sub-domain of geoscience to prove the concept of this work.

To develop the system, first step is to determine the essential elements that need to be implemented for an ontology-based geological sketching tool. This can be done by talking with domain experts and doing a literature review. Once the fundamental elements have been settled, the work on the project can continue.

Research question 1 concerns constructing two parts of the system, bridging the gap between the sketch tool and the geology domain ontology. Compared with current ontology-based systems that use tables or forms as data entry methods, using drawing as geological information entry method requires a more complex and sophisticated ontology-based system. To address this question, we are considering Reasonable Ontology Templates (OTTR) [16] as a technical foundation. OTTR templates are a high-level language that focuses on modeling patterns for building and maintaining ontologies. By providing designed ontology templates, OTTR allows system designers to describe the mapping from high level description like ≪there is a fault through this formation≫, into RDF triples and in a maintainable way. And modifying domain ontology and bridging rigid and nonrigid parts of the system will also be possible by applying OTTR.

The causes and relations of RQ2 and RQ3 are illustrated in Fig. 2. In this figure, the user's intention is about what the user *wants* or tries to do; the user's idea of the system is about the user's understanding of what they have expressed or not expressed in the system; the actual stored data stands for the information that is actually stored by the system. The yellow arrow in Fig. 2 represents the scenario when users want to input some information but soon realize that the system lacks the expressiveness and cannot fulfill their needs. (RQ2). In order to answer RQ2 and to avoid poor usability, a competency question-driven domain-
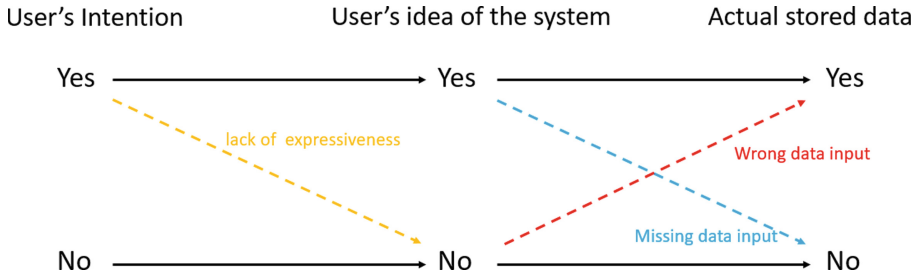
Fig. 2. This figure illustrates the connection between the user's intention, user's idea of what is represented in the system and the actual stored information, and what consequences will wrong connections lead to. (Color figure online)

specific ontology should be provided. Many ontologies relevant to geology have been presented in the literature. Though, most of these ontologies were designed for various purposes and disciplines. Thus, we need specific criteria to evaluate, modify and reuse these available domain ontologies to fit our purpose. Based on this need, a user case survey needs to be designed and conducted to collect the most critical and frequent questions that target domain users ask. After evaluations and modifications that are based on collected competence questions, a question-driven ontology will be presented, which contains sufficient knowledge to answer those questions. Potential users' drawing preferences will also be collected to meet their needs for the graphic drawing part of the system. Thus, the system's expressivity will be made to fit users' intentions and expectations as closely as possible.

As for RQ3, mistakes in the captured information can occur in two ways:

– The user thinks that their sketches should lead to some information being stored, but the actual data does not reflect this information (blue arrow in Fig. 2). For example, in Fig. 1 Fault 1 is on the left of Fault 2. In the missing data input scenario, a user draws this figure to express the fact that geological fault 1 is to the West of fault 2, but the system stores only that there are two faults, not their relative location.
– The user enters a sketch, and the system misinterprets or over-interprets the meaning and stores information that was not intended (red arrow in Fig. 2). For example, in Fig. 1 the user draws the Fault 1 to the left of Fault 2 without intending to express anything about the relative location, but the system stores data to represent that Fault 1 lies to the west of Fault 2.

In order to address this question, the user needs a clear understanding of what the system can do, and they need to be supported by proper information entry methods. In addition to having a tailored domain-specific ontology and providing the instruction book and some demonstrations, the system should also provide users with clear instructions to help users double-check the actual stored information. The work of detailed quality assurance should leave to semantic experts. For the user side, the system can provide a certain degree of query or

reasoner to help users test their stored data. Besides, a detailed evaluation study will be performed to distinguish under what circumstances users' information entry will lead to wrong or missing data input, which is in Sect. 5.

5   Evaluation Plan

To validate the usefulness of the proposed system and make sure the proposed research questions are addressed, use cases examinations and qualitative empirical methods such as design workshops, interviews, and observations will be implemented in the evaluation plan. Since this research work is interdisciplinary, geologists and semantics experts will be involved.

Before inviting users, concrete use cases will be used to examine the expressiveness and correctness of the system and its ontology. These real word geological cases will be selected from industrial and academic structural geology analysis publications and reports. The application domain will first concentrate on carbon capture and storage, petroleum exploration, and production. We will first test the correctness and completeness of our ontology. If the ontology can describe scenarios, then the system will be deployed to draw sketches to describe use cases and check the quality of the stored knowledge graphs.

Qualitative empirical evaluation has two main parts. One part is to design and organize workshops with geologists. Before the workshops, several use cases with geological sketches that contain critical geological information will select as entry material. Geologists will be asked to use this system to input geological information in the sketches with a short system introduction. After the information entry, a prepared qualitative survey will give to users for collecting as much feedback as possible concerning the overall solution, and the satisfaction measurements will based on the system usability scale. Then, a group of semantics experts will assess the correctness of stored information. Any mismatch between the captured information and the users' expectations and intentions will lead to a discussion among semantic and domain experts, and they shall solve the problem together.

Another part will be to invite geologists who have attended the workshops to apply this system in day-to-day work, especially in their fieldwork. Raw geological information entry will bring more challenges to the system, allowing us to assess the usability.

The current work focuses on structural geological faults. It is easy to have envisioned that this system could extend to other structural geological subjects or even relevant geologic domains, and be implemented in energy, mining, or construction industries. More implementation possibilities will lead to more future evaluation plans.

On a broader level, based on Verne's analysis [18] of how digitization and automation influences users' experiences, we will take the following aspects into account during the evaluation:

– are domain competency questions covered and answered by using the system?
– is there any essential aspect is not covered by using this system?

– is there any new task, either positive or negative, that is brought by using this system?
– does any new challenge appear outside this system?

6 Conclusions and Further Work

The current semantics-based table/form data entry user interfaces and traditional geological information entry methods cannot fulfill the needs of geology digitization. This work will result in a system that takes digital geoscience a step forward. It allows geologists to input qualitative information in RDF format in a convenient way. The sketching entry method ensures that geologists follow the conceptualizations in their minds to precisely enter geological information, which is convenient for them and keeps the completeness of their ideas. Thus, qualitative geological information will no longer be limited to sketches and figures, it is captured by the system and stored in RDF format. The stored RDF qualitative information will increase the geological information query efficiency. It can also be used for other purposes, such as qualitative information reasoning and supervised machine learning in various industrial domains.

The availability of geological information in RDF format will enable new digitization in the geology domain and support machine-readable geological information query and reasoning. Besides, this work handles the issue of bridging the gap between the ontology part of the system and the graphic drawing part of the system, which can transfer to other ontology-based information systems as well as other interdisciplinary work between geology and semantic technologies.

References

1. Abel, M., Lorenzatti, A., Fiorini, S.R., Carbonera, J.: Ontological analysis of the lithology data in PPDM well core model. In: 19th International Conference on Petroleum Data Integration and Data Management (2015)
2. Butt, A., Haller, A., Liu, S., Xie, L.: ActiveRaUL: automatically generated web interfaces for creating RDF data. Semant. Web **2013** (2013)
3. Din, C.C., Karlsen, L.H., Pene, I., Stahl, O., Yu, I.C., Østerlie, T.: Geological multi-scenario reasoning. NIK: Norsk Informatikkonferanse (2019)
4. Forbus, K., Usher, J., Lovett, A., Lockwood, K., Wetzel, J.: CogSketch: sketch understanding for cognitive science research and for education. Top. Cogn. Sci. **3**(4), 648–666 (2011)
5. Frischmuth, P., Martin, M., Tramp, S., Riechert, T., Auer, S.: Ontowiki-an authoring, publication and visualization interface for the data web. Semant. Web **6**(3), 215–240 (2015)
6. Garcia, L.F., Abel, M., Perrin, M., dos Santos Alvarenga, R.: The GeoCore ontology: a core ontology for general use in geology. Comput. Geosci. **135**, 104387 (2020)

7. Gil, Y., Pierce, S.A., Babaie, H., Banerjee, A., Borne, K., Bust, G., Cheatham, M., Ebert-Uphoff, I., Gomes, C., Hill, M., et al.: Intelligent systems for geosciences: an essential research agenda. Commun. ACM **62**(1), 76–84 (2018)

8. Hermann, A., Ferré, S., Ducassé, M.: An interactive guidance process supporting consistent updates of RDFS graphs. In: ten Teije, A., et al. (eds.) EKAW 2012. LNCS (LNAI), vol. 7603, pp. 185–199. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33876-2_18

9. Jiménez-Ruiz, E., et al.: SiriusGeoAnnotator: ontology-driven knowledge graph population for geological image annotatio. https://sws.ifi.uio.no/project/sirius-geo-annotator/. Accessed 06 Dec 2021

10. Laxton, J.: Geological map fusion: OneGeology-Europe and INSPIRE. Geolog. Soc. London Special Publ. **408**(1), 147–160 (2017)

11. Lidal, E.M., Hauser, H., Viola, I.: Geological storytelling: graphically exploring and communicating geological sketches. In: Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling, pp. 11–20 (2012)

12. Maillot, P., Ferré, S., Cellier, P., Ducassé, M., Partouche, F.: FORMULIS: dynamic form-based interface for guided knowledge graph authoring. In: Ciancarini, P., et al. (eds.) EKAW 2016. LNCS (LNAI), vol. 10180, pp. 140–144. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58694-6_18

13. Natali, M., Klausen, T.G., Patel, D.: Sketch-based modelling and visualization of geological deposition. Comput. Geosci. **67**, 40–48 (2014)

14. Perrin, M., Thonnat, M., Rainaud, J.F., Verney, P.: A knowledge-based approach of seismic interpretation: horizon and dip-fault detection by means of cognitive vision. In: 2008 SEG Annual Meeting. OnePetro (2008)

15. Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M.: Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol. Rev. **71**, 804–818 (2015)

16. Skjæveland, M.G., Lupp, D.P., Karlsen, L.H., Forssell, H.: Practical ontology pattern instantiation, discovery, and maintenance with reasonable ontology templates. In: Vrandečić, D., et al. (eds.) ISWC 2018. LNCS, vol. 11136, pp. 477–494. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00671-6_28

17. Soylu, A., Giese, M., Jimenez-Ruiz, E., Kharlamov, E., Zheleznyakov, D., Horrocks, I.: Ontology-based end-user visual query formulation: why, what, who, how, and which? Univ. Access Inf. Soc. **16**(2), 435–467 (2016). https://doi.org/10.1007/s10209-016-0465-0

18. Verne, G.B.: Adapting to a robot: Adapting gardening and the garden to fit a robot lawn mower. In: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction, pp. 34–42 (2020)

19. Waldeland, A.U., Jensen, A.C., Gelius, L.J., Solberg, A.H.S.: Convolutional neural networks for automated seismic interpretation. Lead. Edge **37**(7), 529–537 (2018)

20. Wang, B., et al.: A semi-automatic approach for generating geological profiles by integrating multi-source data. Ore Geol. Rev. **134**, 104190 (2021)

21. Wang, C., et al.: The deep-time digital earth program: data-driven discovery in geosciences. Natl. Sci. Rev. **8**(9), nwab027 (2021)

22. Zhong, J., Aydina, A., McGuinness, D.L.: Ontology of fractures. J. Struct. Geol. **31**(3), 251–259 (2009)