frontiers | Frontiers in Earth Science

TYPE Original Research
PUBLISHED 06 May 2025
DOI 10.3389/feart.2025.1530004

OPEN ACCESS

EDITED BY
Fan Xiao,
Sun Yat-sen University, China

REVIEWED BY
Antony Mamuse,
Midlands State University, Zimbabwe
Feng Han,
Guangxi Minzu University, China

*CORRESPONDENCE
Sandra Paula Villacorta Chambi,
villacortasp@gmail.com

RECEIVED 18 November 2024
ACCEPTED 31 March 2025
PUBLISHED 06 May 2025

CITATION
Villacorta Chambi SP, Lindsay M, Klump J, Gessner K, Gray E and McFarlane H (2025) Assessing named entity recognition by using geoscience domain schemas: the case of mineral systems.
Front. Earth Sci. 13:1530004.
doi: 10.3389/feart.2025.1530004

Assessing named entity recognition by using geoscience domain schemas: the case of mineral systems

Sandra Paula Villacorta Chambi[1]*, Mark Lindsay[2,3,4], Jens Klump[2], Klaus Gessner[5], Erin Gray[5] and Helen McFarlane[2]

[1]IAEG Peruvian Group, Lima, Peru, [2]CSIRO, Mineral Resources, Kensington, WA, Australia, [3]The School of Earth Sciences, The University of Western Australia, Crawley, WA, Australia, [4]ARC centre for Data Analytics for Resources and Environments (DARE), Perth, WA, Australia, [5]Department of Energy, Mines, Industry Regulation and Safety, Geological Survey of Western Australia, East Perth, WA, Australia

Named Entity Recognition (NER) is crucial for accurately extracting and classifying specialized domain terms from textual data. This study introduces the Schema for Mineral Systems (SMS), designed through domain characterization, word disambiguation, taxonomy development, and expert input to refine NER approaches in geosciences. SMS, featuring nine geological and five general entity classes, enhances the precision of term identification in mineral system texts. Utilizing domain-specific dictionaries and schema-linked annotations, the schema facilitates the distinct recognition of unique terms, underscored by iterative expert validation to refine NER accuracy. Applied to iron and lithium deposit corpora in Western Australia, SMS highlights the challenges and effectiveness of context-specific schemas in specialized knowledge extraction and accurate entity recognition within complex domains.

KEYWORDS

knowledge management, NLP, NER, geological terminology, ontologies

1 Introduction

Named Entity Recognition (NER) has evolved significantly since Grishman and Sundheim, (1996) foundational work, where NER was defined as "the task of identifying and classifying proper names in text into predefined categories," such as Persons, Organizations, Locations, Dates, and Times. NER's importance has been recognized in various fields, including geosciences, where it helps in the automatic extraction and classification of geological terminology from unstructured scholarly literature and reports. A critical component of developing NER systems is the use of text corpora—large, structured sets of texts used for linguistic analysis and model training. Corpora provide empirical data essential for refining lexical and grammatical theories, as well as supporting the development of Natural Language Processing (NLP) models (Biber et al., 1998).

In the context of geosciences, the adoption of Information Extraction (IE) technologies has played a vital role in automating knowledge discovery and reducing the need for manual intervention. Angeli et al. (2015) emphasize that IE technologies have been instrumental in extracting valuable information from large datasets, which is particularly relevant for managing the vast amount of unstructured data in the geoscience domain. With the advent of advanced Machine Learning (ML) and NLP techniques, especially with models

like BERT (Devlin et al., 2018) and GPT-3 (Brown, 2020), NER tasks have enhanced capabilities in capturing language context and semantics. Building on these technological advancements, it is recognized that specific challenges exist in geological text processing, such as ambiguity and variability in the analysed text (Huber and Klump, 2015; Qiu et al., 2019). Ambiguity can arise from poorly written text or a lack of sufficient context, making it difficult for NER models to capture meaning successfully. Variability relates to subtle but important differences between domains (for example, terms like 'formation' or 'basin'). They recognise the need to develop flexible and scalable models tailored to the unique characteristics of the geoscience domain. To address these challenges, the use of structured frameworks such as ontologies has emerged as a crucial strategy for establishing and defining relationships among concepts within a specific domain. Ontologies like OntoGeonous (Lombardo et al., 2018), GeoCore (Garcia, 2020) and the GeoScience Ontology (GSO) (Brodaric and Richard, 2020) illustrate how these frameworks enable systematic NER and contextual understanding in the field of geosciences. OntoGeonous integrated semantic technologies for geologic mapping across various geological concepts (Mantovani et al., 2020). GeoCore's structured approach enabled categorization and retrieval of geoscientific information, demonstrating how well-defined ontological frameworks can facilitate semantic consistency across diverse datasets. GSO, developed in Canada, exemplifies a structured representation of key geoscience knowledge through a three-layer framework, enabling comprehensive representation and customization for specific requirements. Despite this advancements, there are downsides in fully capturing the dynamic and complex nature geosciences. For example, Babiae et al. (2023) illustrates in the case of mineral systems that even well-designed ontologies are not suitable for direct application in NER tasks without substantial transformation and adaptation. These ontologies also are not readable accessible for the public research. OntoSimilarly, GeoCore and GSO faced challenges in integrating knowledge with data usage and adapting to emerging terminologies across various geological sub-schemas (cf. sub-disciplines and sub-categories used elsewhere). Problems arise when updating ontologies, as geoscience is constantly evolving. Incorporating new subschemas requires expert validation and periodic revisions to include new interdisciplinary terms that often do not fit into existing categories, complicating integration efforts. Additionally, reliance on foundational frameworks creates difficulties in adapting to new terminology and connecting with other domain-specific ontologies, resulting in a labor-intensive process (Garcia, 2020; Brodaric and Richard, 2020). Addressing these challenges, Qiu et al. (2023) introduced a geological domain ontology with over 50,000 terms across twenty-three sub-categories, representing a significant step forward in enhancing NER in geosciences. Their research classifies geological entities into six main types, including geological time and structures, facilitating systematic labeling of academic literature in Chinese language.

Schemas form the core structure of these ontologies enabling the organization of data, the definition of relationships between words, and the recognition of domain-specific terminology. The development of effective schemas for NER in geosciences has been highlighted by Ma (2022) and further explored by Wang et al. (2022) who discussed the use of schemas in structuring geoscientific data, stressing the importance of having clear objectives and focused classification types relevant to the field. Despite these advancements, a significant challenge in achieving optimal NER precision in geosciences, is the need for large annotated corpora verified and validated by experts (Villacorta et al., 2024).

Geoscience-specific controlled vocabularies, hosted by commissions and national or state surveys, exist for many concepts, such as stratigraphic rank (Cox and Richards, 2015) and lithology, further supporting the organization and standardization of geoscientific data. These controlled vocabularies provide an ontological framework emphasizing the importance of clarity and consistency in communication within specialized topics like mineral exploration (Lindsay et al., 2024), further illustrating the need for domain-specific tools and frameworks in NER tasks.

While significant progress has been made in NER for geosciences, comprehensive, ontology-driven approaches remain a critical challenge. Such approaches are essential for enhancing the capabilities of automated annotation systems and for effectively exploring the intricate relationships between geological entities. To contribute to addressing this challenge, building geological knowledge graphs offers a new approach to structuring complex geoscience texts and provides a practical visual analysis of the insights from geoscience papers (Zhou et al., 2021). Addressing these gaps offers an opportunity to broaden the scope of NER research to encompass a wider variety of entities and relationships, thereby increasing its relevance and applicability to geoscience research and exploration.

This study aims to contribute to these ongoing efforts on improving NER by analyzing the use of specialized geological schemas tailored for this field. Building on prior research in corpora creation, this research explores avenues for achieving a semantic understanding of geoscientific language. Specifically, our research focuses on the development and application of three distinct geological schemas: OzRock, GeoIElite_rev, and SMS, applied to corpora concerning iron and lithium mineral deposits in Western Australia. This approach exemplifies ML applications in geological contexts and contributes to the understanding and processing of geoscientific language, targeting a notably underexplored area in geosciences. The primary objectives are to enhance geoscientific data processing and knowledge representation, thereby optimizing the extraction of information from geoscientific texts. Such improvements are crucial for facilitating more efficient knowledge discovery and data management within the field.

In the following sections of this article, we will show that assessing NER in the geoscience domain enables more reliable results consistent with geological reasoning. Accordingly, adhering to the methodologies outlined in this paper provides a practical approach to assess the effectiveness of NER in this intricate field.

2 Methodology

This section outlines the creation of a domain-specific schema, emphasizing its key role in enhancing NER and classification. Developing such a schema is connected to addressing specific research questions for understanding complex domain issues, and schemas are more effective when they are aligned with the domain's processes and when they can capture lexical entities and their semantic relationships. In response to the identified gaps within

existing ontological frameworks as discussed in the introduction, we are including the process of adapting these frameworks into a functional schema tailored specifically for the NER system for this paper. This adaptation involved customization to align vocabularies with the unique lexical and semantic challenges presented by texts on mineral systems, ensuring the final schema could effectively support entity recognition and classification. The exploration of the OntoLex-Lemon model—the primary mechanism for representing lexical data on the Semantic Web, demonstrates how detailed semantic relationships, context-specific usage, and multilingual representation can be effectively captured (McCrae et al., 2017). This model defines lexical concepts with metadata about usage contexts, capturing nuanced differences in word usage across domains and languages, including specialized terms. Through two use cases, representing multilingual dictionaries and the WordNet Collaborative Interlingual Index, McCrae et al. (2017) illustrate how the model addresses complex linguistic structures and domain-specific terminology. Such a semantic approach when designing the ontological framework is reinforced by Bikaun et al. (2024) who developed a schema for the maintenance domain. It focuses on maintenance work, order texts generated by technicians during engineering tasks, primarily describing equipment conditions. Their schema is structured around critical questions, such as 'who is performing what action on what component, and why?' For instance, 'who?' refers to a technician, 'what action?' could be 'replacing,' 'what component?' could be 'a broken alternator bolt,' and 'why?' would be 'due to failure.' By organizing the schema around these questions, Bikaun et al. (2024) emphasize the importance of a question-driven design in improving information extraction and knowledge representation. This approach ensures technical robustness and relevance to research objectives, leading to more precise and meaningful entity recognition outcomes while avoiding the ambiguity that arises from non-specific constraints, such as excessive class options that confuse AI systems when selecting the correct class in a given context.

2.1 Steps to create a schema in a specialized domain

Adhering to the methodology proposed by Lamparter et al. (2004) and incorporating the insights provided by Qiu et al. (2023), the procedure entails the following stages:

- Domain Characterization: This involves identifying and defining the scope and relevant concepts within the specific domain to ensure comprehensive coverage and precision in entity recognition. Collaboration with domain experts to capture the intricacies of the selected field and create annotation guidelines and a domain ontology are considered.
- Word Disambiguation: This step is crucial for distinguishing between multiple meanings of terms, which is common in technical fields like geoscience. It involves deciding the most appropriate meanings and improving clarity in entity recognition to ensure that the schema accurately reflects the intended connotation of terms. This potentially reduces ambiguity and enhances precision in domain-specific entity classification.

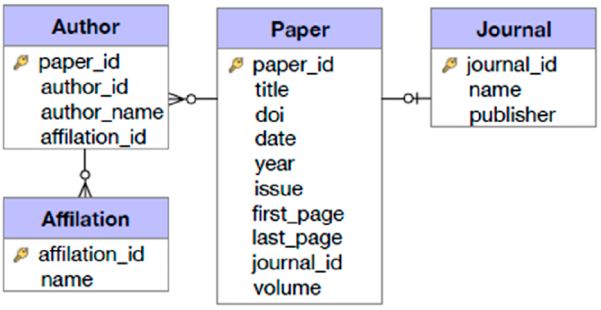

FIGURE 1
Typical academic papers database schema (from Guo et al., 2024).

- Taxonomy Creation: Develop a hierarchical organization by identifying and structuring the domain's classes, entities, and relations. This taxonomy forms the backbone of the schema, facilitating systematic classification and information retrieval (Figure 1). It requires defining parent-child relationships, attribute hierarchies, and cross-references among entities. The taxonomy should be flexible enough to accommodate new findings and scalable to manage large datasets. Tools like ontology editors can assist in visualizing and managing this complex structure.
- Identification of Other Relations: Beyond hierarchical classifications, capturing complex interactions within the analysed data is essential for accurately representing the nuances of geological information. This step involves identifying and defining relational attributes that illustrate how different entities interact or influence one another, including temporal relationships, spatial dependencies, and causal links. These relations enrich the schema, allowing for more dynamic querying and analysis of geoscientific data. The approach of Qiu et al. (2023) in leveraging a knowledge graph to capture these intricate relationships can serve as a model for this process.

2.2 Tools and libraries

The following tools are common in these kinds of applications:

- NLTK (Natural Language Toolkit): This library is used for text tokenization, particularly for breaking down the extracted text from PDF documents into individual sentences (Loper and Bird, 2002). It is widely recognized for its extensive collection of text-processing libraries suitable for tokenization, parsing, and classification tasks.
- Pdfplumber and Pdfminer: These libraries are employed to extract text from PDF files. *pdfplumber* (Singer-Vine, 2020) offers robust capabilities for extracting text, tables, and other data from PDFs, while *pdfminer* (Shinyama and Guglielmetti, 2014) handles exceptions related to PDF parsing errors.
- Flair: This library is popular for its good performance when training and applying the NER model for

specialized domains (Akbik et al., 2019). Flair provides a simple interface for training and applying state-of-the-art sequence taggers, such as NER models.
- Pandas: A data manipulation and analysis library used here to load and handle annotated datasets from CSV files (McKinney, 2010). *Pandas* is essential for managing and preprocessing structured data efficiently.
- Scikit-learn: This library provides ML and statistical modelling tools, including the *confusion_matrix* and *classification_report* functions used to evaluate the NER model's performance (Pedregosa et al., 2011). These functions are fundamental for generating performance metrics that offer insights into the model's accuracy and error rates.
- Matplotlib and Seaborn: These libraries are utilized for data visualization, specifically for plotting the confusion matrix. *Matplotlib* is a versatile plotting library, while *Seaborn* builds on *Matplotlib* by providing an interface for creating informative statistical graphics.

2.3 Validation and performance evaluation

The validation process evaluates the suitability of schemas when combined with the NER model for recognizing and classifying specific domain entities. This aims to assess NER model performance and identify areas for improvement. It also helps to identify common misclassifications of the NER algorithm and understand the underlying reasons for these errors. Key steps of this process are:

- Annotation and Benchmarking: These involve manually annotated benchmark datasets. These datasets contain specific-domain entities that experts annotate (verify) for correct classification using the selected schema. They serve as a reference for evaluating the performance of the NER models.
- Evaluation Using Confusion Matrices: The model's predictions must be compared to the benchmark dataset to visualize the correspondence between the NER model's predictions and the verified dataset categories. The visual representation helps identify areas where the model performed well and struggled when classifying entities. Darker diagonal cells in the matrices indicate correct predictions, while lighter, non-diagonal cells highlighted misclassifications.
- Weighted F1 score: This is an evaluation metric that combines precision and recall assessing the performance of NER systems, especially in scenarios where class distribution may be uneven or complex (Tjong Kim Sang and De Meulder, 2003). Unlike the standard F1 score, the weighted F1 score calculates the F1 score for each class and then takes a weighted average based on the number of instances of each class. This ensures that the score reflects the model's performance across all entity classes, not just the most frequent ones.

2.4 Script pipeline

The application starts by converting PDF documents into corpora and then applies a pre-trained NER model to finalize evaluating its performance using a confusion matrix and classification report. It includes the following parts (Figure 2):

- Pre-processing: The script begins by extracting text from PDF files located in a specified directory on the virtual environment (workspace). The *pdfplumber* library is used to open and read the text from each PDF. The text is then tokenized into sentences using *nltk.sent_tokenize*, which facilitates subsequent processing by the NER model. This step converts the raw textual data into a format the model can process (Villacorta and Lindsay, 2023).
- NER: A pre-trained NER model (*best-model.pt*), previously obtained using an annotated dataset on a domain-specific corpus, is loaded using the Flair library to recognize entities relevant to the geosciences.
- Sentence Extraction: Sentences are extracted from the PDF files stored in the workspace for further processing. Using pandas, an annotated dataset (validated by experts) containing manually labelled sentences is loaded from a CSV file. These annotations serve as a benchmark for evaluating the model's predictions.
- Entity Classification: The NER model is applied to each sentence from the annotated dataset. The model identify terms and assigns entity labels for each sentence.
- Evaluation: The evaluate function compares the previously identified entity labels to the true labels in the annotated dataset. It calculates and prints a classification report, which includes precision, recall, and F1-scores for each entity class. A confusion matrix is also generated to visually represent the model's performance. This step is essential for assessing the model's accuracy.

3 Case study: comparing the efficacy of the NER Flair model using geological schemas and geoscience papers on iron and lithium deposits in Australia

This section outlines the creation of the Schema for Mineral System (SMS), designed from a controlled vocabulary for mineral exploration. This schema incorporates the critical components of mineral systems unifying cross-discipline geoscientific concepts to capture various physical processes and spatial-temporal elements associated with the formation of economically viable mineral resources (Lindsay et al., 2024). By using SMS we illustrate the benefits of domain-specific ontological approaches and ML tools for entity recognition and classification of complex terminology. In this study, we use the SMS to structure corpora derived from academic literature related to iron and lithium deposits and to assess the NER of mineral systems terminology. Following the method explained in Section 2, while any kind of question can guide schema development, they need to specifically focus on the 'what', 'how', 'when', or 'why' of geological processes and phenomena, rather than broad, no-process-oriented questions. For example, general 'who' questions are not relevant here, as our schema is centered on natural processes rather than human actions. Considering that, three specific questions were selected to benchmark our schema design

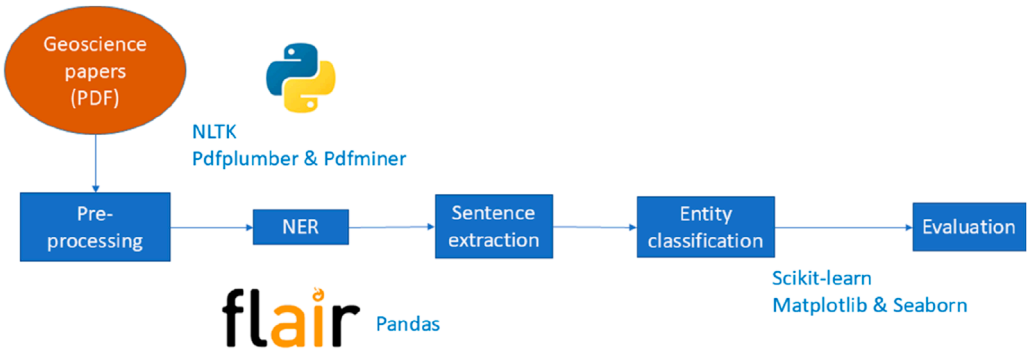

FIGURE 2
Pipeline of the NER process.

and evaluate its effectiveness in capturing relevant terminology (geological entities) in the context of the mineral deposits we have chosen:

- "What are the important tectonic processes for lithium-bearing deposits?";
- "What are the important structures for iron deposits in Western Australia?";
- "What are the important mineralogical associations for iron and lithium deposits in Australia?"

These questions, serving as benchmarks rather than research inquiries, set specific objectives for the schema. By assessing how accurate the schema identifies and classifies terms related to these questions, we can have an approximation on its ability to cover essential terminology (critical geological concepts relevant to mineral systems), as well as the contextually appropriateness of the selected entity classes.

The tools summarized on Section 2.2 were integrated into the EASI Hub high-performance cluster (Woodcock et al., 2018), which has a Tesla V100 GPU, a high-performance processor designed specifically for deep learning and parallel computing tasks. The complete script with usage instructions is included in Supplementary Appendix 1.

3.1 Training of the Flair NER model

In this study, Flair was employed to read geological corpora using three geological schemas: OzRock (Enkhsaikhan, 2021), GeoIElite_rev (Villacorta et al., 2024), and SMS. Flair's effectiveness in domain-specific applications is well-documented across diverse fields such as the biomedical sector for extracting entities like diseases and genes (Patel, 2020), the legal field for identifying statutes and case law (Mathis, 2022), and the business sector for recognizing financial entities like organizations and currencies (Bhattacharya, 2023). The training process involved fine-tuning the Flair NER model on annotated datasets associated with OzRock, GeoIElite_rev and SMS. Each of these schemas was selected for its characteristics wich make them adequate within geoscientific text processing. OzRock offers a comprehensive overview of general geological entity classes and serves as the baseline for understanding common geological terms and categories relevant to mineral exploration (Enkhsaikhan, 2021). GeoIElite_rev was developed to delve deeper into specialized geological entity classes and focuses on processing academic papers concerning iron deposits in Western Australia. It enhances the granularity of geological classifications beyond the foundational OzRock. Complementing these, the SMS schema was developed by the research group involved on writing this paper to address the most complex and nuanced aspects of geoscientific terminology, particularly those associated with mineral systems. The training process involved fine-tuning the Flair NER model on annotated datasets corresponding to these schemas.

3.1.1 Datasets

- Iron and lithium deposits in Western Australia hold considerable economic and environmental importance, influencing global markets, particularly in steel production and battery manufacturing domestically and internationally (Angerer et al., 2015; Perring et al., 2020; Greim et al., 2020). Western Australia's geological setting and tectonic development are prospective for these deposits and thus worthwhile for scientific investigation. This study explores academic literature about these deposits within the framework described here using three geoscience schemas: OzRock, GeoIElite_rev, and SMS.
- *OzRock* (Enkhsaikhan, 2021) was generated from a corpus of hundreds of documents and is focused on mineral exploration. In this dataset, the geological entities are categorised into six types (Table 1). For our research, the OzRock Evaluation set, the annotations based on which comprise 83,838 sentences and 3,238 entities, was utilized as a schema to train Flair NER to produce a customized NER model for the geoscience domain. This model enabled recognition of geological classes from the explored geoscience papers related to iron and lithium deposits. This dataset, publicly available on GitHub (https://github.com/majiga/OzROCK), was already annotated by domain experts, allowing us to utilize it directly for model training without additional annotation. The wide coverage and

TABLE 1 Description of OzRock entity types (Enkhsaikhan (2021).

Label (class)	Description	Example
MINERAL	Mineral	Copper, fire opal, goethite, gold, iceland spar, magnesite, iron, natural salt, silica
ROCK	Lithology	Conglomerate, sandstone, felsic volcanic rock, migmatite, volcaniclastic sedimentary rock
ORE_DEPOSIT	Ore types	Channel iron deposit, iron ore, nickel ore, silver ore
TIMESCALE	Geological time	Archean, Lower Proterozoic, Paleoproterozoic, Triassic, Upper Cretaceous
STRAT	Stratigraphy	Angas Hills Formation, Bingy Bingy Basalt Member, Marra Mamba Iron Formation
LOCATION	Geographical location	Kalgoorlie Terrane, Kimberley Craton, Perth, Pilbara, Pilbara Craton, Western Australia

extensive representation of general geological entities provided a robust base for the model to learn from well-defined classes.

- GeoIElite_rev (Villacorta et al., 2024) was developed to compare other geological classes with those included in OzRock. It was constructed from concepts detailed in 20 PDF papers focused on iron deposits (list of papers in Supplementary Appendix 2). This dataset encompasses eighteen distinct entity classes (Table 2) and required manual annotation to ensure its entity classes were appropriately applied to this project. The annotation count for GeoIElite_rev includes 5,400 sentences and 5,028 entities.
- SMS: The Schema for Mineral System (SMS) was developed to evaluate NER within the domain of mineral systems literature. Following the steps indicated in Section 2, the SMS schema was defined collaboratively with geoscientists, and defines complex mineral systems terminology through literature review. Critical entities and relationships were identified, and a hierarchical taxonomy was created to organize the schema's 14 classes (Table 3), providing flexibility to accommodate updates for large datasets. After several discussions, subject matter experts selected nine from the twenty-four geological classes defined to be part of this schema. These include the specific terminology associated with the selected questions considered as relevant. Additionally and following the methodology outlined by Ding et al. (2021), specific categories such as Country, Province/State, and City were consolidated into a single class, GPE (Geopolitical Entity), to address context-based ambiguities. Additionally, general domain classes, such as Person-Scholar (PS), were included to capture mentions to researchers (for example, geologists, biologists, and palaeontologists) and ensure the semantics are understandable to the machine. Similarly to the case of GeoIElite, this dataset required manual annotation to ensure it was appropriately applied this project. The annotations count is 910 sentences and 832 entities.

3.1.2 Training process

The training process involved fine-tuning the Flair NER model individually for each geological schema resulting in three distinct best-model.pt files, each tailored to its specific entity classes. For each schema, the model was loaded and fine tuned on its respective annotated dataset, integrating each schema's structured vocabulary to specialize in recognizing terms relevant to each schema. For instance, the SMS schema emphasized classes like TECTONIC_SETTING, critical for analyzing mineral systems, while the OzRock schema covered broader geological categories like MINERAL and ROCK. This structured approach enabled the model to differentiate and contextualize geological terms according to each schema's focus, ensuring tailored recognition capabilities across the three different geological datasets.

3.2 Validation and performance evaluation

The validation process for the geological schemas used in this project involved testing the Flair NER model on our annotated datasets. These annotations were performed by subject matter experts consisting of the co-authors of this paper along with additional colleagues from CSIRO Minerals Resources. Their expertise delivered a rich depth of knowledge necessary for accurately tagging geological entities in the corpora.

The methodology for annotation was jointly developed by the authors of this paper. Given the logistical challenges of working in different locations and the limitations of open-access annotation tools, which were not suitable for handling large corpora, we as annotators, opted for a more flexible approach. Annotations were made directly within online Google Sheets documents, which facilitated easy access editing. This approach allowed for real-time collaboration, ensuring all annotators could participate effectively despite geographical disparities.

The expert-validated annotations and the Flair NER model's predicted classifications were continuously compared using the collaborative Google Sheets documents as online platform. This setup ensured a dynamic and responsive validation process, allowing for immediate expert inputs and adjustments.

To visually represent the accuracy of classifications, confusion matrices were generated. These matrices showcased the alignment between the model's predictions and the expert-validated categories, highlighting any discrepancies and common misclassifications. The evaluation also included calculating weighted F1 scores, providing a detailed measure of the model's precision and recall, particularly for handling the diverse and occasionally rare classes within the SMS schema. This metric was crucial for assessing the nuanced performance of the NER system across different geological terminologies.

TABLE 2 Description of GeoIElite_Rev entity types (Villacorta et al., 2024).

Label (class)	Description	Example
FORMATION	Geological formation	Angas Hills Formation, Bingy Bingy Basalt Member, Marra Mamba Iron Formation
AGE	Age of the rocks	4,000 to 2,500 million years ago, 2,500–541 million years ago
TIMESCALE	Geological time	Archean, Lower Proterozoic, Paleoproterozoic, Triassic, Upper Cretaceous
MINERAL	Mineral	Fire opal, goethite, martite, Iceland spar, natural salt
ROCK	Lithology	Conglomerate, sandstone, felsic volcanic rock, migmatite, volcaniclastic sedimentary rock, metamorphic gneisses
PROCESS	Geological process	Deposition, erosion, basin development, mountain building, volcanism, weathering, hydrothermal alteration and mineralization, karst formation
ELEMENT	Metal/elements	Iron, gold, nickel, lithium, bauxite, copper, zinc, lead, cobalt, rare earth elements, tantalum and niobium, vanadium, platinum group elements, uranium, manganese
CHARACTERISTIC	Geological feature	Fractured, metamorphosed, pelitic, altered, folded, weathered, intruded, granitic, foliated, sheared, veined
LOCATION	Geographical location	Countries, cities, states, places like: Kalgoorlie, Kimberley, Pilbara, Pilbara, Western Australia
ORE_DEP_REG	Locations where mineral resources have been discovered or explored	Mines, exploration sites like: Kalgoorlie Terrane, Kimberley Craton, Perth, Pilbara, Pilbara Craton, Western Australia
LANDFORM	Geomorphological forms	Channel, cratons, mountain, basin, hill, ophiolite (represents ancient oceanic crust and upper mantle rocks), karst systems, river, lava flows, lakes, dunes, regolith, pluton
TYPE	Type of ore deposits	Banded iron, nickel sulfide, volcanogenic massive sulfide, copper–gold porphyry
METHOD	Methods of exploration activities	Drilling, sampling, or testing
YEAR	Year of exploration activities	1970, 1980, 1990, 2000, 2003, 2005
COMPANY	Company responsible for exploration/production	BHP Group, Rio Tinto, Fortescue Metals Group, Gold Fields, Western Areas, IGO Limited, Pilbara Minerals, Woodside Energy
INSTITUTION	Government entity involved	Department of Energy, Mines, Industry Regulation and Safety, Geological Survey of Western Australia, Australian Government Department of Industry, Science and Resources, Minerals Research Institute of Western Australia Environmental Protection Authority (EPA) of Western Australia, Western Australian Planning Commission, Aboriginal Lands Trust, Office of the Environmental Protection Authority, Water and Environmental Regulation Department
PERM_LIC	Permissions/licenses for exploration and production	Exploration License, Mining Lease, Prospecting License, Retention License, Miscellaneous License, General Purpose Lease, Program of Work Approval, Environmental Approvals, Native Title Agreements, Water License, Cultural Heritage Clearances
IMPACT	Environmental impact of exploration	Impact on water, air, or land

4 Results

The comparison of F1 scores across the OzRock, GeoIElite_rev, and SMS schemas reveal significant variations in performance (Figures 3–8). Figures 3, 5, 7 present Flair NER confusion matrices and F1 scores for the three geoscience schemas applied to iron deposit literature. These matrices utilize a blue palette to indicate the count of predictions made by the model, with darker

10.3389/feart.2025.1530004

TABLE 3 Description of SMS entity types.

Label (class)	Description	Example
LITHOSPHERIC ARCHITECTURE	The geometric structure of the solid Earth (Earth's crust and lithospheric mantle) as defined by domains of similar chemical composition and the discontinuities that separate them. Lithospheric architecture is the result of geodynamic processes	Chemical compositional change, age, geometry, physical property change, mineralogy, domain
TECTONIC HISTORY	The temporal sequence of events that forms lithospheric architecture, such as magmatism, deformation, metamorphism, subsidence or exhumation. These events can be interpreted in paradigmatic frameworks and attributed to geodynamic processes such as subduction, seafloor spreading, mantle plumes, large igneous provinces, and the resulting geodynamic environments	Magmatism, deformation, metamorphism, subsidence, uplift
TECTONIC SETTING	Lithospheric region deformed by contiguous geodynamic conditions resulting in characteristic geological processes	Deformational regime, thermal regime
TEMPORAL EXTENT	Period during which processes responsible for forming or developing a particular mineral system occur	Time units (chronostratigraphic units): Archean, Lower Proterozoic, Paleoproterozoic, Triassic, Upper Cretaceous
MINERAL	Inorganic elements or compounds (apart from liquid mercury and a few organic minerals) and defined by their chemical composition and crystal structure	Fire opal, goethite, martite, Iceland spar, natural salt, quartz, magnetite, columbite, monazite
ROCK	Solid mass of aggregate of minerals (lithology)	Conglomerate, sandstone, felsic volcanic rock, migmatite, volcaniclastic sedimentary rock, metamorphic gneisses
GEODYNAMIC ENVIRONMENT	Dynamic setting characterized by planetary-scale events and physical processes in the Earth's crustal and mantle envelopes. Differs from tectonic settings in time and space	Boundary condition, subduction, seafloor spreading, mantle plume, large igneous province, volcanism, mountain building, basin formation, hotspot
SOURCE	A volume of rock, fluid or magma that, by its chemical composition, acts as an origin for a particular chemical compound (ion, ligand, crystal or lithic fragment) that is subsequently transported from its primary site to a secondary site (of mineralization)	Sedimentary pile, felsic magmas, mafic magmas, source rock, crystalline basement, metal, fluid, chemical species: iron, gold, nickel, lithium, bauxite, copper, zinc, lead, cobalt, rare earth elements, tantalum and niobium, vanadium, platinum group elements, uranium, manganese
ROCK DEFORMATION	The change of shape or the displacement of a mineral aggregate through crystal-plastic (ductile) processes or by fracturing (brittle) due to mechanical failure	Crystal-plastic deformation, brittle deformation
GPE (GEOGRAPHIC LOCATION)	Geographical location of the place entities as represented by latitude and longitude values	Countries, cities, states, geographical coordinates
DATE	Absolute or relative dates or periods (general domain)	5000 BC, 1750 AD, 20th Century, 18th Century, 1990s, 1500s, 2010, 2011, 2012, 2013, etc.
LOC	Non-GPE locations, mountain ranges, bodies of water	Places like: Kalgoorlie, Kimberley, Pilbara, Pilbara, Western Australia
PS (PERSON-SCHOLAR)	Researchers' names (for example, geologists, biologists, and palaeontologists)	During, Perring, Ramanaidou, Thorne, Angerer, Rodger, etc.
QUANTITY	Measurements, as of weight or distance	km, metric, tons, degrees, mm, litres, percent, etc.

shades of blue representing higher frequencies of classifications within each category. Similarly, Figures 4, 6, 8 use the same visual representation for the analysis of lithium deposit research papers. The following results provide insights into the strengths and limitations of each schema and highlight areas for further enhancement. Previous research (Villacorta et al., 2024) indicated that increasing the number of papers does not improve the F1 score. Hence, we compared schemas based on different class types and

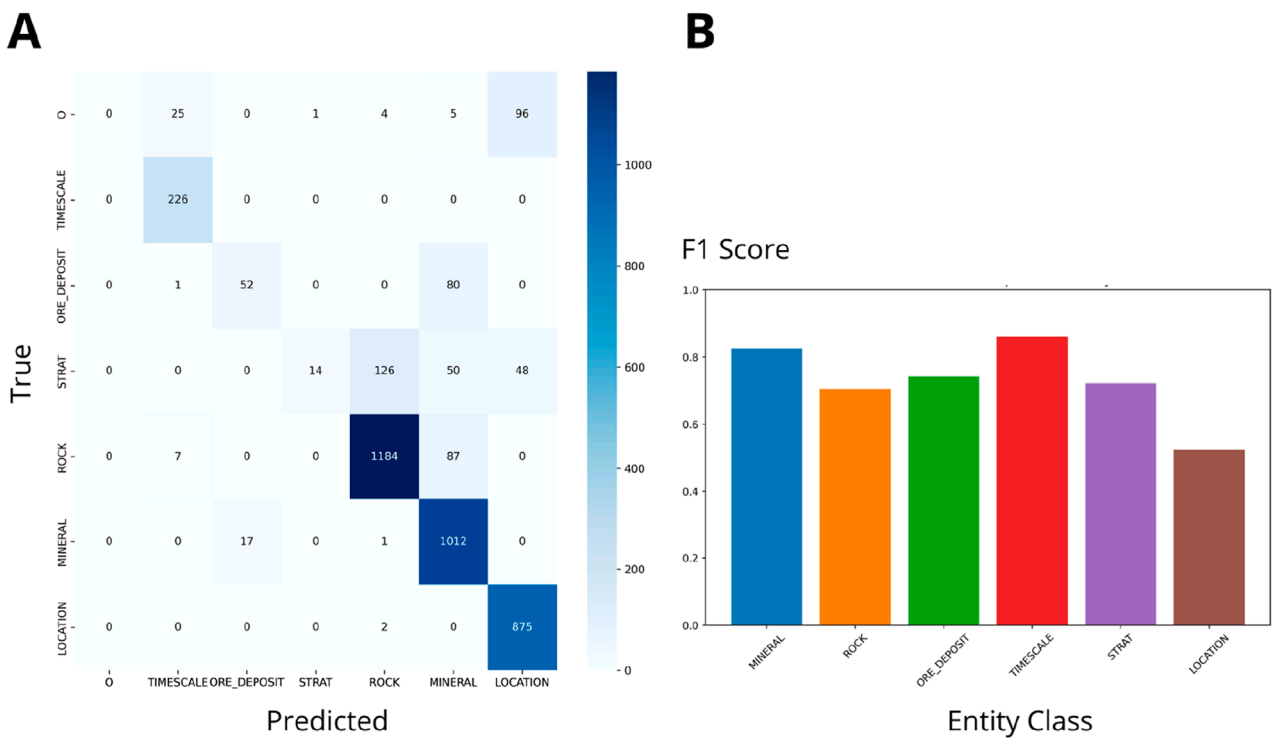

FIGURE 3
(A) Ozrock Confusion matrix (Wgt F1: 0.72). **(B)** F1 scores for NER on Iron Deposits Datasets by class.

numbers. Note in the figures that an 'O' was used to indicate tokens that do not belong to any entity like "by" or "and".

4.1 Dataset comparisons

- OzRock (Figures 3, 4) confusion matrices and F1 score bars show that it performs robustly in identifying entity classes like "MINERAL", "ROCK", and "ORE_DEPOSIT". The high weighted F1 scores (0.72 and 0.71) indicate better precision and recall balance than the other schemas in categorizing geological terms.
- GeoIElite_rev (Figures 5, 6) presents slightly lower weighted F1 scores of 0.69 and 0.70. This result suggests moderate effectiveness, due to the schema's expansive inclusion of diverse entity types, which might introduce complexity in accurately tagging less distinct classes such as "PROCESS" and "METHOD".
- SMS (Figures 7, 8) depicts a considerable drop in weighted F1 scores to 0.27 and 0.35, indicating challenges in entity recognition.

4.2 Entity class performance

- The F1 scores across different entity classes reveal that core geological categories ("MINERAL", "ROCK", "TIMESCALE") consistently achieve higher accuracy.
- Lesser-defined classes, such as "PROCESS", "METHOD" and "IMPACT", demonstrate lower F1 scores.

5 Discussion

5.1 Challenges and limitations

The limitations of NER models in recognizing annotated geological classes and its misclassification patterns are closely tied to the complexity of geoscientific terminology, the challenges in design and annotating geological schemas and the difficulties of ensuring high-quality annotations. These factors impact model performance and highlight the need for continuous refinement of schemas and training datasets.

5.1.1 Schema classes selection

While the detailed processes involved in the development and application of the SMS schema are outlined in the case study section, it is crucial to emphasize the broader implications of our findings here. The use of the SMS schema shows the critical need for ontological resources in geosciences that are not only scientifically rigorous but also adaptable to the evolving landscape of geological research. The encountered challenges highlight the importance of developing frameworks that can be easily updated and refined to accommodate new scientific insights and terminologies.

The selection of geological classes in the SMS schema was guided by their relevance to characterizing iron and lithium deposits, illustrates a targeted approach to ontology design. For

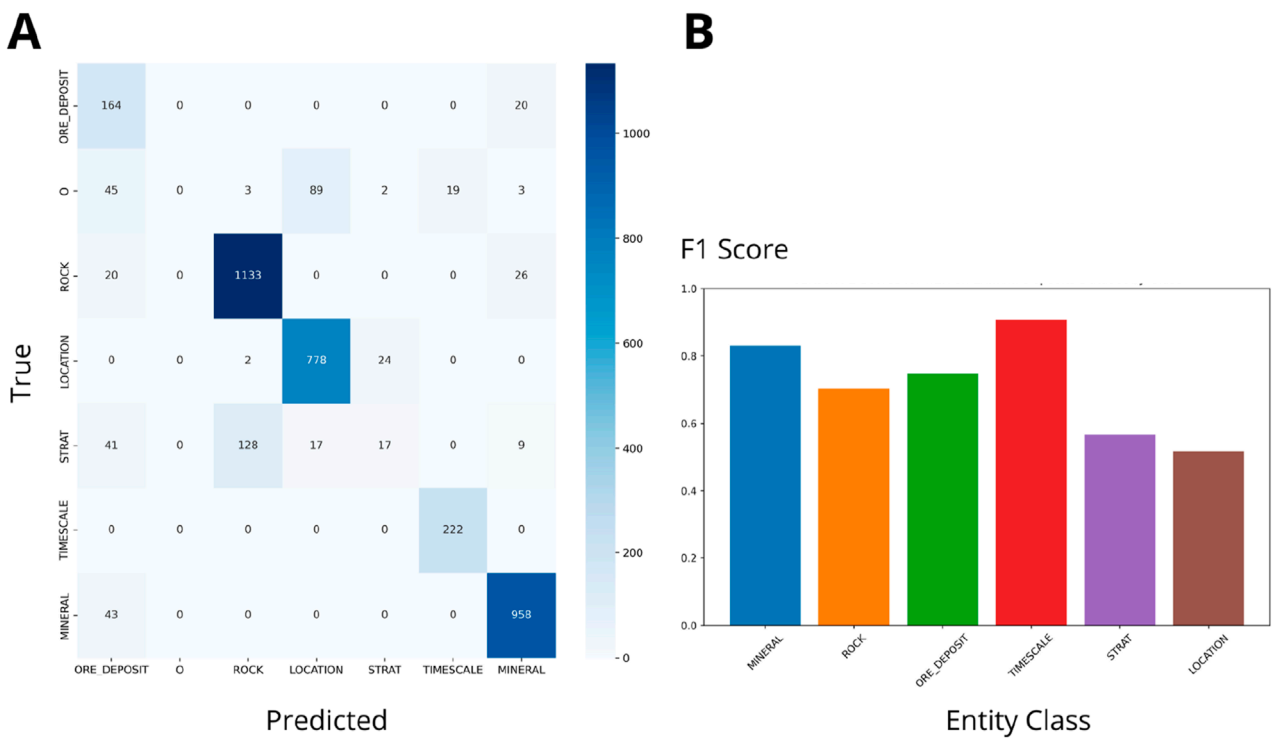

FIGURE 4
(A) Ozrock Confusion matrix (Wgt F1: 0.71). **(B)** F1 scores for NER on Lithium Deposits Datasets by class.

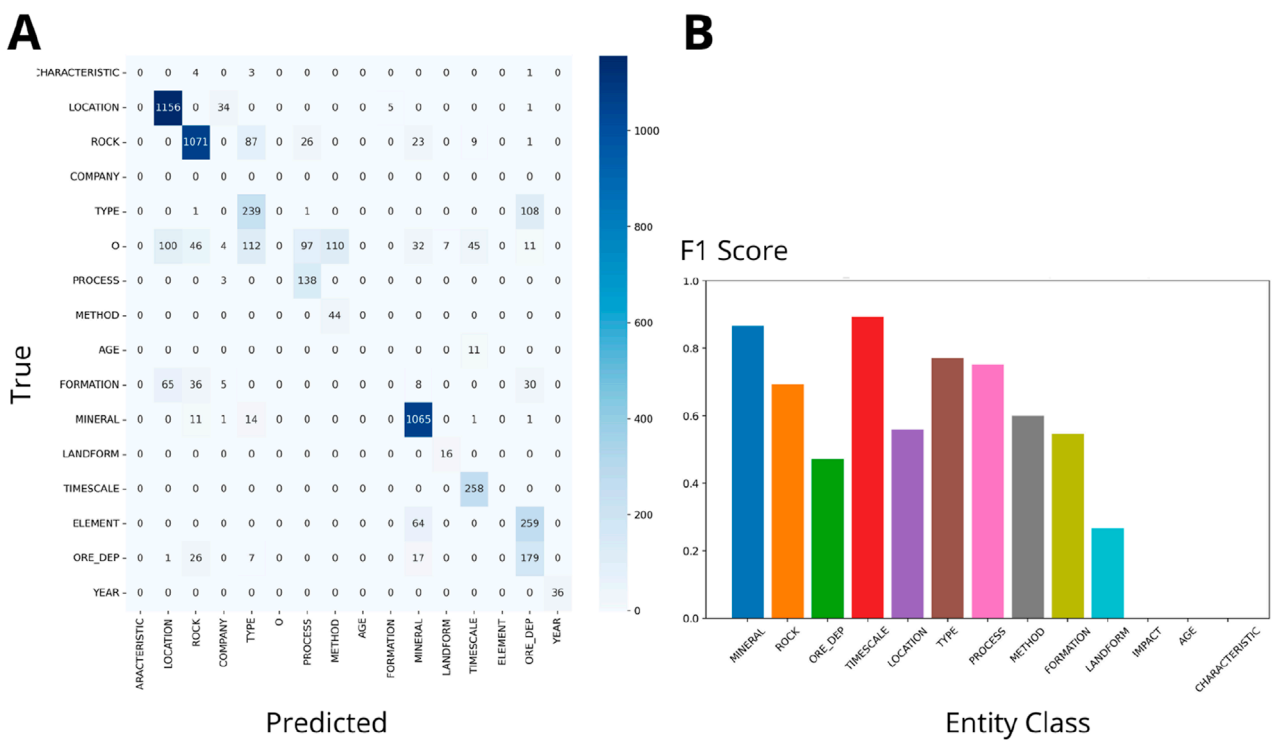

FIGURE 5
(A) GeolElite_rev Confusion matrix (Wgt F1: 0.69). **(B)** F1 scores for NER on Iron Deposits Datasets by class.

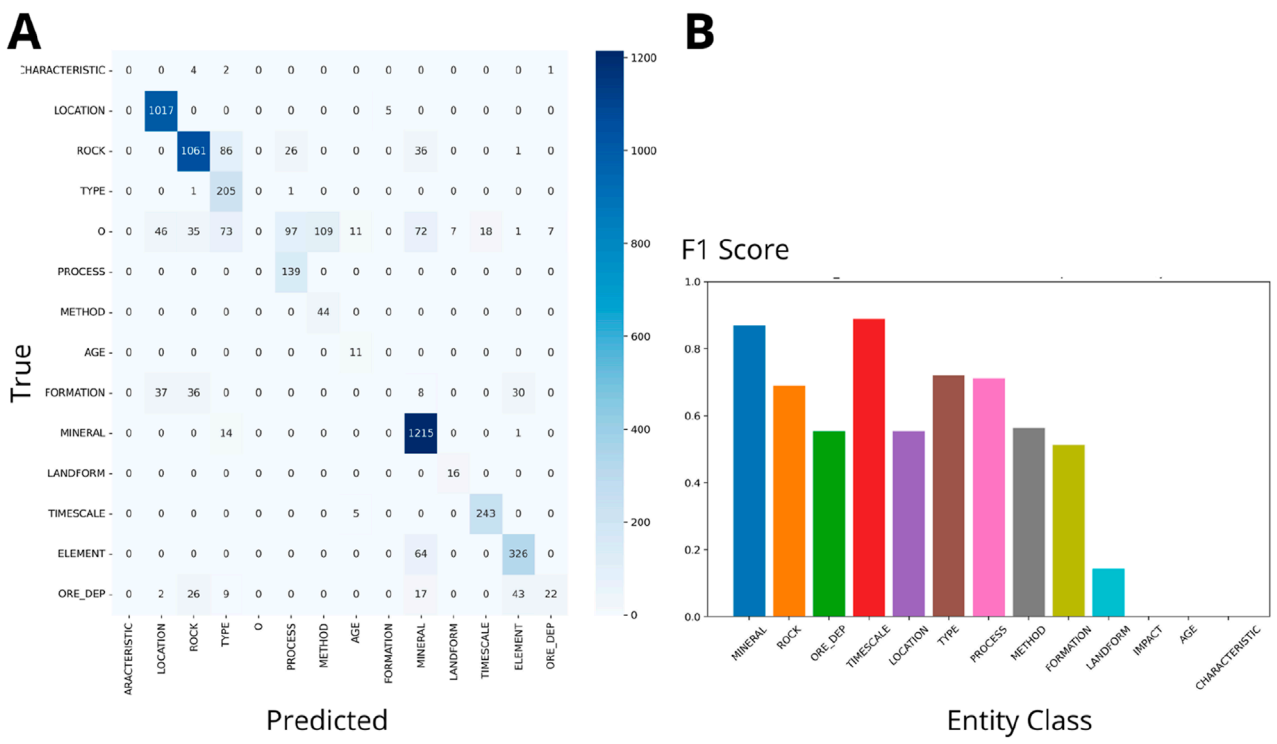

FIGURE 6
(A) GeolElite_rev Confusion matrix (Wgt F1: 0.70). **(B)** F1 scores for NER on Lithium Deposits Datasets by class.

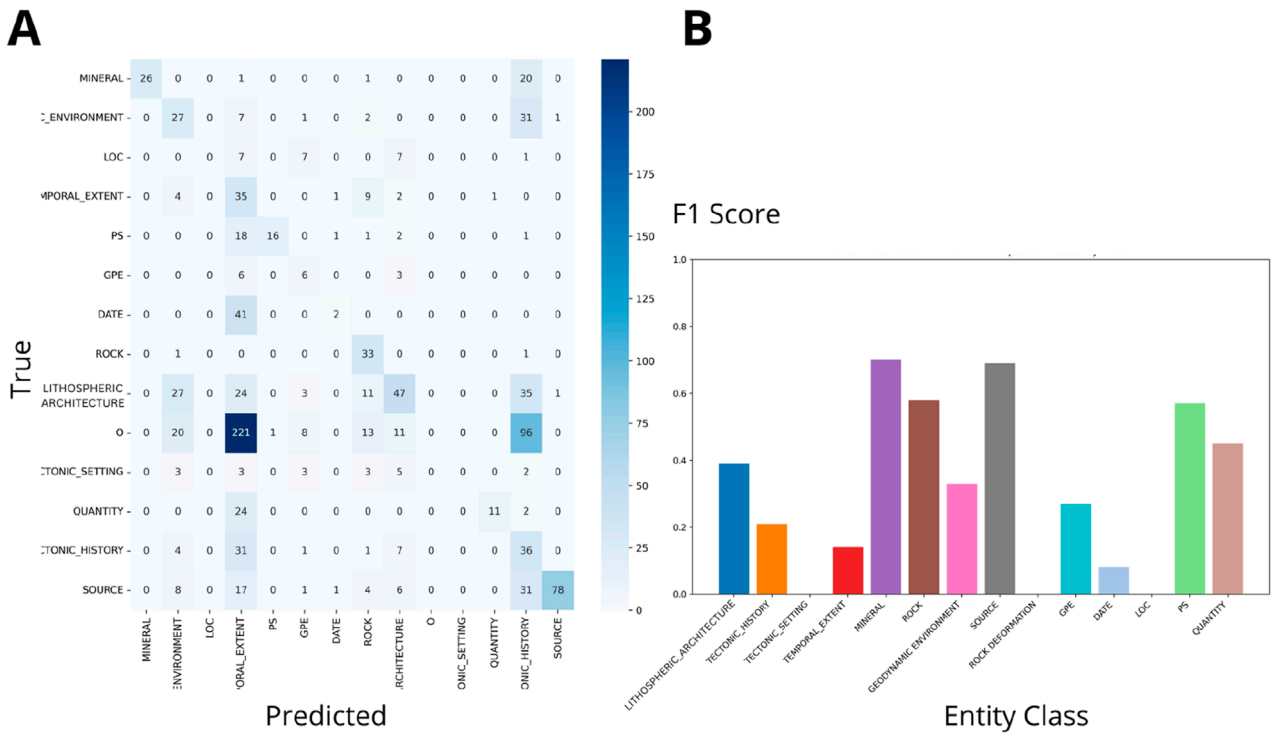

FIGURE 7
SMS **(A)** Confusion matrix (Wgt F1: 0.27). **(B)** F1 scores for NER on Iron Deposits Datasets by class.

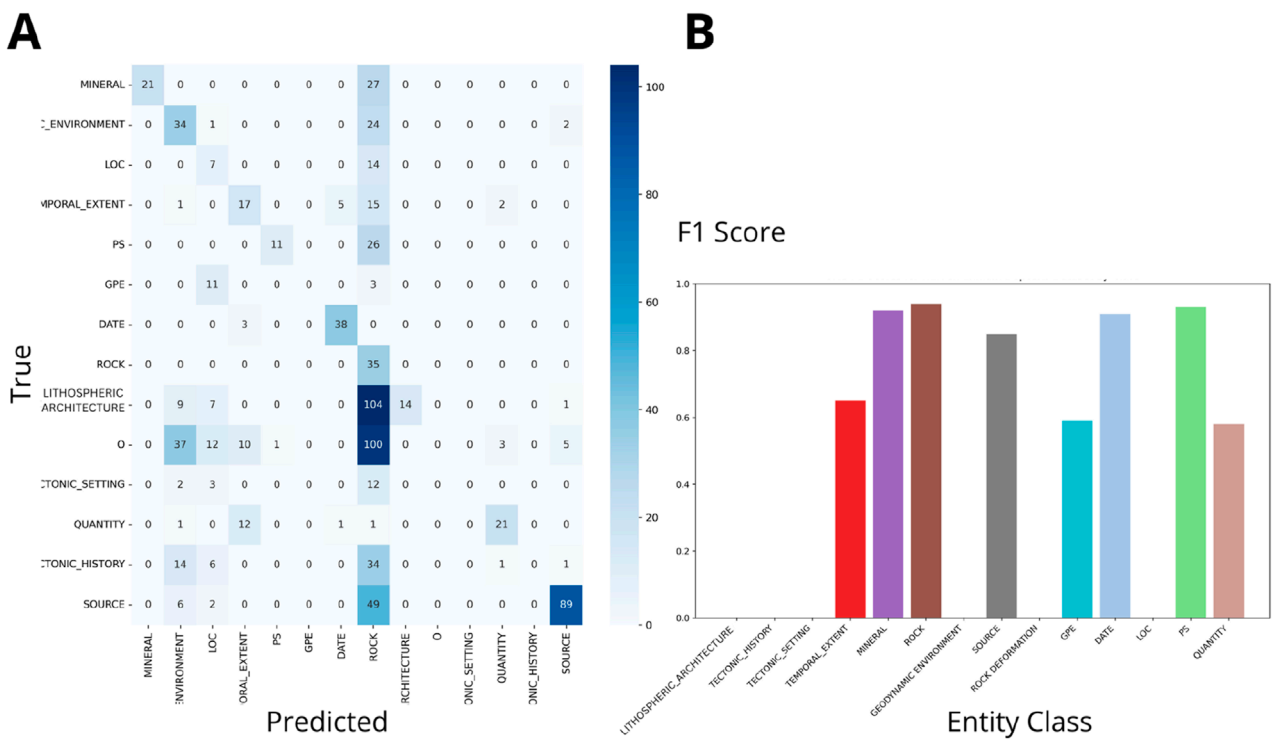

FIGURE 8
SMS **(A)** Confusion matrix (Wgt F1: 0.34). **(B)** F1 scores for NER on Lithium Deposits Datasets by class.

instance, the inclusion of LITHOSPHERIC_ARCHITECTURE and TECTONIC_SETTING helped to understand the formation of iron deposits, such as banded iron formations (BIFs), which are influenced by regional tectonic activity and large-scale lithospheric processes. This specificity in class selection is crucial for enhancing the precision of NER tasks in complex domain like geosciences, where the accuracy of terminology recognition directly impacts the quality of data extracted from scholarly texts.

Similarly, the MINERAL and ROCK classes are key terms for both deposits, as they determine the feasibility of extraction by directly influencing the concentration and accessibility of valuable minerals. For lithium pegmatites, minerals like spodumene and lepidolite are pivotal for extraction viability, while the mineralogy and composition of iron banded formations play a critical role in determining the grade and recoverability of iron. Additionally, TEMPORAL_EXTENT aids in understanding the timeframes of geological processes critical to the formation of these deposits.

The analysis of confusion matrices and F1 scores for the SMS schema (Figures 7, 8) reveals that while classes such as LITHOSPHERIC_ARCHITECTURE and GEODYNAMIC_ENVIRONMENT were recognized with reasonable accuracy, others like TEMPORAL_EXTENT and TECTONIC_SETTING experienced significant misclassification. These findings highlight the challenges in distinguishing closely related or complex classes, particularly in geosciences, where terms often have nuanced and overlapping meanings. Overly detailed categories could possible have overwhelmed the Flair model, underscoring the importance of schema simplicity and specificity in achieving accurate NER performance.

The model demonstrated higher performance for schemas with general categories, as reflected by the higher F1 scores for MINERAL and ROCK in the tree geoschemas. However, the low F1 scores for SMS suggest that a more nuanced definition of certain classes is necessary to capture the complexity of mineral system vocabularies. As noted by Qiu et al. (2019), an effective schema starts with a focused set of terms that are representative of the domain-specific entities. Although the SMS schema was tailored to address geological questions related to lithium pegmatites and iron deposits in Western Australia, the results indicate that further refinement is required.

The findings suggest that the challenge lies not only in the number of classes but also in selecting foundational and contextually relevant ones. Expert input and iterative validation are critical to ensure the schema maintains classification consistency and accurately reflects geoscientific terminology, ultimately improving NER performance for specialized domains.

5.1.2 Annotation

Annotating large corpora for geoscience NER presents considerable challenges due to the need for substantial manual efforts. Automated tools, such as Python-based packages and specialized NER models like Flair, offer potential solutions; however, the complexity of automatic annotation and expert validation remains significant (Villacorta et al., 2024; Bikaun et al., 2022). This study corroborated the major challenge when validating annotations across extensive geological datasets. Despite efforts to comprehensively verify the annotations in a corpus of twenty papers, only one was fully annotated and validated due to time constraints and the lack of specialized annotators. This partial validation served as a foundation

for automating the annotation of the rest of the corpus. The automated process faced limitations; the Flair NER model recognized only a subset of the annotated classes. The weighted F1 scores for the SMS dataset were 0.27 and 0.34 for both datasets (iron and lithium deposits), indicating variability in the model's performance across different entity types. Specifically, the confusion matrices revealed that classes such as 'DATE' and 'PS' achieved high accuracy with minimal misclassifications (accuracy of 0.98 and 0.96, respectively). The geological classes like 'TECTONIC SETTING' and LITHOSPHERIC ARCHITECTURE' are often misclassified (Figures 7, 8).

In addressing these challenges, Qiu et al. (2023) implemented a systematic approach to annotation and validation in the geological domain. Their annotation platform allowed input from domain experts, categorizing entities into six main types. Using a specialized Python-based annotation tool, it was facilitated manual annotation and iterative consistency checks, achieving a high level of annotation consistency and expert involvement enabling the authors to construct large-scale, high-quality corpora in the Chinese language. To mirror Qiu et al.'s efforts, we can adopt a similar approach in the future by developing domain-specific annotation guidelines in collaboration with experts. Additionally we can utilize specialized tools such as INCEpTION, an open-source platform that supports collaborative and interactive annotation in general domains (Klie et al., 2018). Implementing this approach, with automated checks and expert validation, can potentially produce high-quality data for training and validating NER models in geoscience, significantly improving their accuracy and reliability.

A	B
Word	**Tag**
Carina	B-LITHOSPHERIC_ARCHITECTURE
Iron	B-SOURCE
Yilgarn	B-LITHOSPHERIC_ARCHITECTURE
Craton	I-LITHOSPHERIC_ARCHITECTURE
Australia	B-PERSON
B	B-PERSON
.	O
E	B-PERSON
.	O
Nicolson	B-PERSON
Kettlewell	B-PERSON
Carina	B-LITHOSPHERIC_ARCHITECTURE
iron	I-LITHOSPHERIC_ARCHITECTURE
ore	B-SOURCE
105	B-TEMPORAL_EXTENT
Yilgarn	B-LITHOSPHERIC_ARCHITECTURE
Iron	I-LITHOSPHERIC_ARCHITECTURE
Ore	B-SOURCE
40	B-ORDINAL
km	B-QUANTITY
iron	B-SOURCE
Mt	B-PERSON
Walton	B-PERSON
May	B-DATE
2007	B-DATE

FIGURE 9
Automatic annotation using as base the SMS schema, highlighting misclassification of geological entities.

5.2 Misclassification cases

The analysis of the SMS schema's misclassification patterns reveal three primary types of misclassification: overlap due to complex terminology, context dependency, and the underrepresentation of rare classes. Figures 7, 8 provide a detailed visual representation of these misclassification patterns. Several factors contribute to this misclassification, including the complexity and variability of language in the corpus, nuanced distinctions between similar classes, and potential inconsistencies in initial manual annotations. Addressing these challenges requires refining iterative annotation schemas, improving ML algorithms, and potentially expanding the manually validated sample size to improve the model's accuracy and coverage.

5.2.1 Overlap due to complex terminology

Certain classes, such as LITHOSPHERIC_ARCHITECTURE and TECTONIC_SETTING, have nuanced meanings that the model struggled to capture without expert guidance. For example, as seen in Figure 9, terms like "Yilgarn" were misclassified as LITHOSPHERIC_ARCHITECTURE instead of LOC (Location), likely because they refer to geological regions. This demonstrates the model's difficulty in distinguishing between geological structures and geographic regions, where context plays a significant role. Another notable example is TECTONIC_SETTING, which was occasionally misclassified as ROCK (confusion matrices 7A, 7B) due to overlapping terminology with geological formations, as shown in Figure 9. This frequent overlap, exemplified by terms such as 'continental collision,' which may reference both tectonic settings and rock-associated processes, highlights a significant challenge in AI: the generation of spurious concept relationships or hallucinations. Jiang et al. (2024) emphasize that such errors in entity recognition can propagate through subsequent stages of analysis, compounding inaccuracies in data interpretation. To address this issue, further refinement of our schema is essential. By enhancing its capacity to distinguish between closely related terms, and incorporating advanced AI techniques that apply deep contextual analysis, we can improve the accuracy of entity recognition.

5.2.2 Context dependency

The Flair NER model's limited capacity to incorporate contextual cues significantly complicates its handling of context-dependent classes within the SMS dataset. For instance, as depicted in Figure 9, the terms "Iron" and "Ore" frequently receive the label 'SOURCE' instead of the more appropriate 'MINERAL' or 'ROCK', contingent upon the surrounding textual context. This inadequacy in context processing not only underscores the model's struggle with polysemous terms but also impacts its ability to deliver precise geological classifications. The inclusion of an additional class category like 'ELEMENT' has potential to enhance the model's discernment of such nuances. However, the fundamental resolution involves not just technological enhancements but also rigorous expert validation to ensure the accuracy and consistency of labeling, critical in the domain of geosciences where the exactitude of each term holds substantial implications. Further, the misclassification of terms like 'TECTONIC_SETTING', which

might be incorrectly annotated due to overlapping or ambiguous context, can severely distort the geological interpretations essential for addressing specific research inquiries, such as understanding relevant tectonic processes. This limitation is crucial because accurate classification directly influences the integrity and utility of data used in determining geological dynamics, which are foundational to mineral exploration and geological mapping strategies. Orellana et al. (2020) and Hu et al. (2024) emphasize the importance of enhancing NER systems' contextual comprehension to mitigate AI-induced misinformation and improve the reliability of information extraction processes. They advocate for the adoption of advanced NLP strategies to deepen the contextual understanding of NER models, which would enhance their precision and recall. This is crucial as these metrics are essential for validating the effectiveness of entity recognition and classification within complex, domain-specific datasets.

5.2.3 Rare classes

Rare classes in the SMS dataset, such as PROCESS, METHOD, and IMPACT, demonstrated lower F1 scores, highlighting the challenges of limited representation in training data. The SMS dataset, with only 91 sentences and 832 entities, provided insufficient data for these classes, resulting in reduced generalization capability. For comparison, the OzRock dataset, which contained 83,838 sentences and 3,278 entities, offered broader coverage and higher F1 scores for general classes like ROCK and MINERAL (Figures 3, 4). However, even within OzRock, nuanced or less frequent classes were more prone to misclassification. The disparity in performance between these datasets highlights how insufficiently diverse or narrowly scoped training data can lead to suboptimal model performance, particularly for complex or infrequent classes. Integrating broader datasets and continuous expert feedback into the training process can help address these shortcomings by enhancing the diversity and representativeness of the training data, thus reducing the incidence of AI-induced errors and improving the overall reliability of the model.

5.3 Future research

The limitations of automated recognition with the SMS schema highlights the need for diverse, representative training datasets, refined schemas to capture domain-specific nuances, and model adaptations to address the complexities of geological terminology and context, enhancing entity extraction accuracy. Increasing the number of annotated examples for rare classes, expanding the diversity of training data, and enhancing model adaptability to context would improve classification accuracy and the utility of NER models in geoscientific research.

Using advanced techniques such as few-shot learning (Hofer et al., 2018) can improve NER model's ability to recognize less frequent or underrepresented classes. Few-shot learning is a ML technique that allows NER models to generalize with a limited number of labelled examples which is common in specialized domains like geoscience. Liu et al. (2022) have been pioneers exploring few-shot learning in geosciences. They used GeoBERT and Few-shot learning approach for recognizing long geological terms using a minimal amount of annotated datasets. They fine-tuned a pre-trained model using a geological domain thesaurus achieving an F1 score of 0.80.

Additionally, a continuous feedback loop, where domain experts validate and refine the model's outputs, can help improve its accuracy and reliability over time. This aligns with our previous findings (Villacorta et al., 2024), which highlighted the importance of schema training data diversity in enhancing NER model performance in geosciences. Specifically, it was noticed that GeoIElite achieved a modest F1 score compared to OzRock, which performed better due to its broader linguistic diversity and the inclusion of hundreds of documents. The narrower scope and fewer entity classes in GeoIElite contributed to its lower scores. The analysis suggested that limited annotated data scope, as seen with GeoIElite, hinders robustness. Expanding the diversity and context of annotated data can improve contextual recognition. Additionally, Flair's performance tends to decline with an increasing number of entity classes, while F1 scores indicate that corpus size (7 PDFs vs. 20 PDFs) has a limited impact on overall NER accuracy. Misclassifications, particularly between geological entities such as 'ORE_DEPOSIT' and 'MINERAL,' emphasize the need for schema refinement. This can be addressed by ensuring that classes are well-defined and distinct, or in some cases, merging similar classes to reduce ambiguity and improve classification accuracy. Models with a generalized LOCATION class (GeoIElite_rev, OzRock) show different F1 scores, suggesting that class generalization may impact model accuracy. Expert validation which is reflected in the annotated dataset is crucial for creating schemas due to the challenges of integrating automated tools. Future research will focus on expanding and diversifying the annotated datasets to cover additional geological subdomains and terminology.

Large language Models (LLMs) like GPT-4, BERT, and others have shown significant potential in processing and analyzing geoscientific texts (Touvron et al., 2023). These models, trained on vast amounts of diverse data, can capture complex language patterns and contextual nuances, making them well-suited for handling the specialized terminology and varied contexts found in geoscientific literature. LLMs offer opportunities for processing geoscientific texts, some examples are GeoBERT (Liu et al., 2022) which promises improvemed NER tasks in geology; GeoGalactica (Lin et al., 2023), which was fine-tuned using geoscience-specific data to improve knowledge extraction, document classification, and question answering, and the use of LLM in analyzing climate-related questions (Bulian et al., 2023), enhancing understanding of environmental changes. However, their effectiveness vary significantly based on the specificity of the training data and the domain-specific challenges they are tailored to address. To fully harness their potential in geosciences, ongoing efforts are needed to increase the diversity and representativeness of training datasets, refine domain-specific schemas and ontologies, and develop ML techniques to enhance model performance with minimal data.

Moreover, collaboration with domain experts is essential to validate and improve model outputs, ensuring that LLMs can provide accurate, reliable insights in geoscientific research. The challenges and limitations of LLMs are that they may struggle with rare or underrepresented geological terms or concepts if the training data lacks diversity. This limitation can lead to incomplete or inaccurate entity recognition. While LLMs are powerful, they may still face challenges in generalizing across different geological contexts, particularly when

encountering less common terms or unique geological formations or acronyms in geoscience contexts. As observed in previous discussions, this can result in a model that recognizes only a subset of the relevant geological classes as observed in this research. Yet, applying LLMs in geosciences requires significant computational resources, particularly when fine-tuning models on domain-specific data, which can be considered a limiting factor for smaller research teams or projects with constrained budgets.

While the initial phase of the project focused on the steps for implementing schemas, future developments of the SMS plan to integrate relational attributes, such as temporal relationships and spatial dependencies. Capture more complex relationships within geological entities, will enhance contextual recognition in this specialized domain. Also is planned to investigate how the mentioned LLM can be fine-tuned and integrated with our developed schemas to improve entity recognition accuracy and reduce errors related to context misinterpretation and ambiguous terminologies.

6 Conclusion

This research highlights the potential and current limitations of automated annotation tools using open-access NER models, tailored for geoscience literature. The introduction of the Schema for Mineral Systems (SMS) has provided insights into the classification and recognition of geological entities, particularly emphasizing the schema's capability to detail the nuanced aspects of complex mineral systems.

Our findings demonstrate that while schemas such as OzRock and GeoIElite_rev establish essential frameworks for geological entity recognition, they occasionally fall short in capturing the more detailed and subtle geological features that SMS excels in identifying. However, our results also highlight a critical challenge: the detailed and comprehensive nature of SMS, while beneficial, can sometimes introduce complexities that hinder the effectiveness of NER systems. This intricacy necessitates significant fine-tuning and expert validation to achieve reliable performance.

The analysis of confusion matrices and performance evaluations from the datasets reveals a stark contrast in the effectiveness of different schemas. OzRock and GeoIElite_rev showed robust performance in general geological categorization, whereas SMS, despite its detailed approach, showed variability in its effectiveness, particularly struggling with classes that require deep contextual understanding or are less represented in the training data.

From this study, it is evident that achieving optimal NER performance requires a balance between schema detail and simplicity. Future research should thus focus on refining schema definitions to ensure they capture essential geological nuances without overwhelming the NER systems. Incorporating diverse and high-quality training data, along with leveraging advanced machine learning strategies such as few-shot learning and domain-specific language models, will be crucial in enhancing the precision and utility of NER systems for geoscientific applications.

Continued collaboration with domain experts is imperative to ensure the relevance and accuracy of schema classifications. Such partnerships are vital for aligning the schemas with evolving geological concepts and maintaining the high standards necessary for automated knowledge extraction from geoscientific literature.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

SV: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review and editing, Validation. ML: Methodology, Resources, Supervision, Validation, Writing – review and editing, Conceptualization, Project administration. JK: Conceptualization, Supervision, Writing – review and editing. KG: Writing – review and editing, Methodology, Validation. EG: Writing – review and editing, Validation. HM: Writing – review and editing, Validation.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by the CSIRO ResearchPlus Science Leader program.

Acknowledgments

The authors express their gratitude to Mario Iglesias and Marta Sośnicka for their assistance in annotating the SMS dataset. We appreciate Ryan Noble's insightful feedback on entity class classification and Behnam Sadeghi's contributions to the manuscript's approach. Thanks are also to Andy Wilkins and Tadro Abbot for their thorough review within the CSIRO peer review system. We are grateful to the Executive Director of the Geological Survey of Western Australia for granting K Gessner and E Gray the permission to participate in this study, which was pivotal for our research. Marta Sośnicka deserves additional acknowledgment for her comprehensive review and input on the practical applications of our findings in mineral exploration research. A special thanks to the journal reviewers, Dr. Antony Mamuse and Dr. Feng Han, for their constructive feedback and thoughtful suggestions, which greatly contributed to improving the quality of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. To assist in editing the manuscript and enhancing its readability. AI is not credited as an author of the manuscript; it was solely utilized for summarizing text and reducing redundancy. All content edited with the help of Generative AI has been verified for factual accuracy and checked for plagiarism.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/feart.2025.1530004/full#supplementary-material

References

Akbik, T. B., Blythe, D., Rasul, K., Schweter, S., and Vollgraf, R. (2019). "FLAIR: an easy-to-use framework for state-of-the-art NLP," in Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics (demonstrations), 54–59.

Angeli, G., Premkumar, M. J. J., and Manning, C. D. (2015). "Leveraging linguistic structure for open domain information extraction," in Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 344–354.

Angerer, T., Duuring, P., Hagemann, S. G., Thorne, W., and McCuaig, T. C. (2015). A mineral system approach to iron ore in archaean and palaeoproterozoic BIF of Western Australia. *Geological Society, London, Special Publications* 393 1, 81–115.

Babaie, H. A., Davarpanah, A., and Elliott, W. C. (2023). Ontology of the complex rare-earth elements mineral system. *Special Pap. Geol. Soc. Am.* 558, 29–44. doi:10.1130/2022.2558(03

Bhattacharya, A. (2023). *Custom named construct recognition in the business and management literature*. Ottawa, ON, Canada: Carleton University. Doctoral dissertation.

Biber, D., Conrad, S., and Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press. Available online at: https://academic.oup.com/dsh/article-abstract/14/2/305/936240 (Accessed February 2, 2025).

Bikaun, T. K., French, T., Stewart, M., Liu, W., and Hodkiewicz, M. (2024). "MaintIE: a fine-grained annotation schema and benchmark for information extraction from maintenance short texts," in Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 10939–10951.

Bikaun, Stewart, M., and Liu, W. (2022). "Quickgraph: a rapid annotation tool for knowledge graph extraction from technical text," in *Proceedings of the 60th annual meeting of the association for computational linguistics: system demonstrations*, 270–278.

Brodaric, B., and Richard, S. M. (2020). The geoscience ontology. Abstract retrieved from AGU Fall Meeting Abstracts 2020 (IN030 07).

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *arXiv: 14165*. doi:10.48550/arXiv.2005.14165

Bulian, J., Schäfer, M. S., Amini, A., Lam, H., Ciaramita, M., Gaiarin, B., et al. (2023). Assessing large language models on climate information. *arXiv Prepr. arXiv:2310.02932*.

Cox, S. J., and Richard, S. M. (2015). A geologic timescale ontology and service. *Earth Science Informatics* 8. 5–19.

Devlin, M.-W., Chang, Lee, K., and Toutanova, K. (2018). Bert: pre-training of deep bidirectional transformers for language understanding. *arXiv Prepr. arXiv:1810.04805*.

Ding, N., Xu, G., Chen, Y., Wang, X., Han, X., Xie, P., et al. (2021). Few-nerd: a few-shot named entity recognition dataset. *arXiv* [Preprint]. *arXiv:2105.07464*.

Enkhsaikhan, M. (2021). *Geological knowledge graph construction from mineral exploration text*. Doctoral thesis (UWA: University of Western Australia).

Garcia, L. F., Abel, M., Perrin, M., and dos Santos Alvarenga, R. (2020). The GeoCore ontology: a core ontology for general use in Geology. *Comput. and Geosciences* 135, 104387. doi:10.1016/j.cageo.2019.104387

Greim, P., Solomon, A. A., and Breyer, C. (2020). Assessment of lithium criticality in the global energy transition and addressing policy gaps in transportation. *Nat. Commun.* 11 (1), 4570. doi:10.1038/s41467-020-18402-y

Grishman, R., and Sundheim, B. M. (1996). "Message understanding conference-6: A brief history," in *COLING 1996 volume 1: The 16th international conference on computational linguistics*.

Guo, Z., Wang, C., Zhou, J., Zheng, G., Wang, X., and Zhou, C. (2024). GeoKnowledgeFusion: a platform for multimodal data compilation from geoscience literature. *Remote Sens.* 16 (9), 1484. doi:10.3390/rs16091484

Hofer, M., Kormilitzin, A., Goldberg, P., and Nevado-Holgado, A. (2018). Few-shot learning for named entity recognition in medical text. *arXiv Prepr. arXiv:1811.05468*.

Hu, Z., Hou, W., and Liu, X. (2024). Deep learning for named entity recognition: a survey. *Neural Comput. Appl.* 36 (16), 8995–9022. doi:10.1007/s00521-024-09646-6

Huber, R., and Klump, J. (2015). Agenames a stratigraphic information harvester and text parser. *Earth Sci. Inf.* 8, 125–134. doi:10.1007/s12145-014-0171-5

Jiang, G., Luo, Z., Hu, C., Ding, Z., and Yang, D. (2024). Mitigating out-of-entity errors in named entity recognition: a sentence-level strategy. *arXiv Prepr. arXiv:2412.08434*.

Klie, J. C., Bugert, M., Boullosa, B., De Castilho, R. E., and Gurevych, I. (2018). "The inception platform: machine-assisted and knowledge-oriented interactive annotation," in In Proceedings of the 27th international conference on computational linguistics: System demonstrations (Santa Fe, NM: Association for Computational Linguistics), 5–9.

Lamparter, S., Ehrig, M., and Tempich, C. (2004). "Knowledge extraction from classification schemas," in On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2004, Agia Napa, Cyprus, October 25–29, 2004 (Springer Berlin Heidelberg), 618–636.

Lin, Z., Deng, C., Zhou, L., Zhang, T., Xu, Y., Xu, Y., et al. (2023). Geogalactica: a scientific large language model in geoscience. *arXiv Prepr. arXiv:2401.00434*.

Lindsay, M., Villacorta, S. P., McFarlane, H., Gessner, K., and Gray, E. (2024). Geo-semantics and ontologies: an approach to decode gold mineral systems using controlled vocabularies. *arXiv*. doi:10.5281/zenodo.15151900

Liu, H., Qiu, Q., Wu, L., Li, W., Wang, B., and Zhou, Y. (2022). Few-shot learning for name entity recognition in geological text based on GeoBERT. *Earth Sci. Inf.* 15 (2), 979–991. doi:10.1007/s12145-022-00775-x

Lombardo, V., Piana, F., and Mimmo, D. (2018). Semantics–informed geological maps: conceptual modeling and knowledge encoding. *Comput. and Geosciences* 116, 12–22. doi:10.1016/j.cageo.2018.04.001

Loper, E., and Bird, S. (2002). Nltk: the natural language toolkit. *arXiv Prepr. cs/0205028*.

Ma, X. (2022). Knowledge graph construction and application in geosciences: a review. *Comput. and Geosciences* 161, 105082. doi:10.1016/j.cageo.2022.105082

Mantovani, A., Piana, F., and Lombardo, V. (2020). Ontology-driven representation of knowledge for geological maps. *Comput. and Geosciences* 139, 104446. doi:10.1016/j.cageo.2020.104446

Mathis, B. (2022). Extracting proceedings data from court cases with machine learning. *Stats* 5 (4), 1305–1320. doi:10.3390/stats5040079

McCrae, J. P., Bosque-Gil, J., Gracia, J., Buitelaar, P., and Cimiano, P. (2017). "The Ontolex-Lemon model: development and applications," in Proceedings of eLex 2017 conference, 19–21.

McKinney, W. (2010). "Data structures for statistical computing in Python," in *Proceedings of the 9th Python in science conference*. Editors S. van der Walt, and J. Millman (Austin, TX: SciPy), 51–56. doi:10.25080/Majora-92bf1922-00a

Orellana, M., Fárez, C., and Cárdenas, P. (2020). "Evaluating Named Entities Recognition (NER) tools vs algorithms adapted to the extraction of locations," in

2020 International Conference of Digital Transformation and Innovation Technology (Incodtrin) (IEEE), 123–128.

Patel, H. (2020). Bionerflair: biomedical named entity recognition using flair embedding and sequence tagger. *arXiv Prepr. arXiv:2011.01504*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. Available online at: https://www.jmlr.org/papers/v12/pedregosa11a.html.

Perring, C., Crowe, M., and Hronsky, J. (2020). A new fluid-flow model for the genesis of banded iron formation-hosted martite-goethite mineralization, with special reference to the north and south flank deposits of the Hamersley Province, Western Australia. *Econ. Geol.* 115 (3), 627–659. doi:10.5382/econgeo.4734

Qiu, Q., Tian, M., Xie, Z., Tan, Y., Ma, K., Wang, Q., et al. (2023). Extracting named entity using entity labeling in geological text using deep learning approach. *J. Earth Sci.* 34 (5), 1406–1417. doi:10.1007/s12583-022-1789-8

Qiu, Q., Xie, Z., Wu, L., and Tao, L. (2019). GNER: a generative model for geological named entity recognition without labeled data using deep learning. *Earth Space Sci.* 6 (6), 931–946. doi:10.1029/2019ea000610

Shinyama, Y., and Guglielmetti, P. (2014). *pdfminer.six* (Version 20240706). *GitHub*. Available online at: https://pypi.org/project/pdfminer.six/ (Accessed September 4, 2024).

Singer-Vine, J. (2020). Pdfplumber (Version 0.11.0). *GitHub*. Available online at: https://pypi.org/project/pdfplumber/ (Accessed September 4, 2024).

Tjong Kim Sang, E. F., and De Meulder, F. (2003). "Introduction to the CoNLL-2003 shared task: language-independent named entity recognition," in Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, 142–147.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., et al. (2023). Llama: open and efficient foundation language models. *arXiv Prepr. arXiv:2302.13971*.

Villacorta, S. P., Lindsay, M., Klump, J., and Francis, N. (2024). "Assessing named entity recognition efficacy using diverse geoscience datasets," in 2024 International Conference on Machine Intelligence for GeoAnalytics and Remote Sensing (MIGARS) (IEEE), 1–3.

Villacorta, and Lindsay, M. (2023). "Exploring the importance of preprocessing operations in geoscience knowledge graphs through the application of a machine learning approach," in Proceedings of the 26th World Mining Congress, Brisbane, Australia, 177–188.

Wang, C., Chen, J., and Li, Y. (2022). Named entity annotation schema for geological literature mining in the domain of porphyry copper deposits. Abstract retrieved from AGU Fall Meeting Abstracts (IN12C-0276).

Woodcock, R., Paget, M., Wang, P., and Held, A. (2018). "Accelerating industry innovation using the Open Data Cube in Australia," in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, 8636–8638.

Zhou, C., Wang, H., Wang, C., Hou, Z., Zheng, Z., Shen, S., et al. (2021). Geoscience knowledge graph in the big data era. *Sci. China Earth Sci.* 64 (7), 1105–1114. doi:10.1007/s11430-020-9750-4