ARTICLE IN PRESS

Geoscience Frontiers xxx (xxxx) xxx

Contents lists available at ScienceDirect

Geoscience Frontiers

journal homepage: www.elsevier.com/locate/gsf

A unified framework of temporal information expression in geosciences knowledge system

Shu Wang [a], Yunqiang Zhu [a,b,*], Yanmin Qi [a,c], Zhiwei Hou [d], Kai Sun [a], Weirong Li [a,e], Lei Hu [a,e], Jie Yang [a], Hairong Lv [f]

[a] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[b] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
[c] University of Nottingham Ningbo China, Faculty of Science and Engineering, Ningbo 315154, China
[d] Guangzhou Institute of Geography, Guangdong Academy of Sciences, Guangzhou 510070, China
[e] University of Chinese Academy of Sciences, Beijing 100101, China
[f] Department of Automation, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Article history:
Received 14 February 2022
Revised 17 August 2022
Accepted 31 August 2022
Available online xxxx

Keywords:
Unified Time Framework (UTF)
Time ontology
Geosciences knowledge system
Time conversion
Time scale
Big data

ABSTRACT

Time is an essential reference system for recording objects, events, and processes in the field of geosciences. There are currently various time references, such as solar calendar, geological time, and regional calendar, to represent the knowledge in different domains and regions, which subsequently entails a time conversion process required to interpret temporal information under different time references. However, the current time conversion method is limited by the application scope of existing time ontologies (e.g., "Jurassic" is a period in geological ontology, but a point value in calendar ontology) and the reliance on experience in conversion processes. These issues restrict accurate and efficient calculation of temporal information across different time references. To address these issues, this paper proposes a Unified Time Framework (UTF) in the geosciences knowledge system. According to a systematic time element parsing from massive time references, the proposed UTF designs an independent time root node to get rid of irrelevant nodes when accessing different time types and to adapt to the time expression of different geoscience disciplines. Furthermore, this UTF carries out several designs: to ensure the accuracy of time expressions by designing quantitative relationship definitions; to enable time calculations across different time elements by designing unified time nodes and structures, and to link to the required external ontologies by designing adequate interfaces. By comparing the time conversion methods, the experiment proves the UTF greatly supports accurate and efficient calculation of temporal information across different time references in SPARQL queries. Moreover, it shows a higher and more stable performance of temporal information queries than the time conversion method. With the advent of the Big Data era in the geosciences, the UTF can be used more widely to discover new geosciences knowledge across different time references.

1. Introduction

Time, as an essential reference system for human cognition, serves as a fundamental framework for the geosciences knowledge system, comprising different time elements, such as time descriptions, units, coordinate systems, recording uncertainties, and dating methods. (Levine, 2016; Saïs et al., 2020). Different temporal information uses different time reference systems with corresponding time elements. For example, "D.C. 2022" in the year-based solar calendar system and "Early Jurassic" in "Ma"--based geological time system are two totally different time systems. The accuracy of time element representation is a prerequisite for both knowledge inference and calculation, which directly affects the possibility and accuracy of geoscience knowledge discoveries, especially in interdisciplinary studies (Ma et al., 2011; Wang et al., 2018, 2019; Zhang et al., 2018). Currently, significant geoscience knowledge discoveries still rely on massive temporal

* Corresponding author at: State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.

E-mail address: zhuyq@lreis.ac.cn (Y. Zhu).

https://doi.org/10.1016/j.gsf.2022.101465
1674-9871/

Please cite this article as: S. Wang, Y. Zhu, Y. Qi et al., A unified framework of temporal information expression in geosciences knowledge system, Geoscience Frontiers, https://doi.org/10.1016/j.gsf.2022.101465

information calculations with their domain-specific knowledge bases. For example, massive fossil time interval calculation via the Paleozoic marine dataset revealed Paleozoic marine biodiversity and the extinction event (Fan et al., 2020). Complicated temporal information calculations through the use of late Quaternary mammal datasets revealed the relationship between global fire activities and late Quaternary herbivore extinctions (Karp et al., 2021). Nevertheless, these specific datasets have content limitations as they contain only domain knowledge and are not linked to a vast amount of other relevant information. Geosciences, however, is a complex and comprehensive system that requires more comprehensive and multi-domain datasets to reveal and validate more intricate discoveries than any other discovery (Bergen et al., 2019; Normile, 2019; Zhou et al., 2021). Thus, current geoscience knowledge discoveries with a domain-specific knowledge base are insufficient. Moreover, different domain knowledge bases may use different time reference systems, which make the records have different time descriptions, time units, time uncertainties, and other time elements (Lehmann et al., 2015; Fan et al., 2020; Tanon et al., 2020). These time elements are usually ignored because they are consistent within the specific dataset and are omitted without any impact on their calculations. The situation, by contrast, is entirely different in datasets with multiple different time references. These temporal information calculation issues across different time references significantly affect the geoscience knowledge calculations across different datasets and restrict the progress of significant geosciences discoveries in the Big Data era. Consequently, accurate and efficient calculation of temporal information across different time references assumes great significance for future geosciences discoveries.

At present, all the methods of temporal information calculation across different time references use the basic idea of time conversion (Arboit et al., 2021; Fu et al., 2021; Wang et al., 2021b). It indicates temporal information expressions in different datasets with different time references have been converted to the same target. Time standards and ontologies in many domains have been constructed to systematically elucidate time elements and their relationship. For example, ISO19108 published a temporal information expression standard of the solar calendar that defines "TM_Instand" and "TM_Period" as the essential elements to represent geographic temporal information and their relationships (The International Organization for Standardization, 2002). International (Chrono)Stratigraphic Chart (ISC) has been constructed to express geological time systematically by considering the features of geological temporal information, such as time uncertainty, time boundaries, time relationships, and time versions (Alekseev, 2015; Jiang and Liu, 2021). With different versions of ISC being released, the version ontologies have been constructed to align ages in different ISC versions (Ma and Fox, 2013; Ma et al., 2020). Additionally, several relevant time ontologies such as the time-zone ontology (Pan and Hobbs, 2004; Frasincar et al., 2010), the dating ontology (Perrin et al., 2011), and the geological time scale ontology (Cox and Richard, 2015; Cox, 2016) have also been constructed. These standards and ontologies provide a systematic description of the elements and their relationships by defining commonly used time elements from the viewpoint of different domains. Besides, they also contribute to the clarification of time elements in different time reference systems in the course of time transformation.

However, there are still some issues with both the time conversion idea and current efforts in the calculation of massive temporal information across different time references that restrict automatic knowledge calculations on computers. First, different time systems have different ways of expressing time, which form domain standards or ontologies, making it difficult to apply calendar time to geological time and other time systems. For example, "TM_Instand" in ISO19108 is represented as a point value, which is not compatible with the geological time, as each geological time needs an uncertainty value to describe the Point time, such as "BaseLowerTriassic" that stores "251.0" with an uncertainty value of "±0.4". These time elements of standards and ontologies built for domain demands cannot apply to all geoscience knowledge. It is therefore necessary to develop a unified framework for the expression of temporal information for the whole geosciences knowledge system.

Second, temporal information cannot be calculated automatically on a computer because time conversion needs expert experience and human involvement. For example, every temporal information calculation needs a time conversion process, including the critical steps of time parsing, time complement, and time alignment. Time parse is to analyze and understand the time elements; for example, to judge whether a record of "2000" refers to "CE 2000" or "BCE 2000", which needs expert experience to fix. Time complement is to fill up omitted time elements to achieve time calculation, for example, to add the unit "year" or "million years" to the record of "180" by considering the relevant information in a specific dataset, which involves expert experience or extra descriptive information from file headers to fix. Moreover, time alignment is to convert different time references into a single time reference by using the previous time elements; for example, converting "Base Lower Triassic" into "251.0 ± 0.4 Ma" to compare with "260 Ma". This process also requires manual or programming code conversions, which rely on expert experience to complete. To achieve automatic time calculations across different time references on a computer, it is necessary to handle the essential basic steps mentioned above, which currently need to be done manually. Thus, temporal information expression must be designed to fit all time references, calculated directly without complex conversions. In short, current time conversion methods have domain limitations and experience dependence, which restrict accurate and efficient temporal information calculation across different time references on a computer.

To address this issue, this paper designs a Unified Time Framework (UTF) in geosciences knowledge system with systematic time elements representation to support accurate and efficient temporal information calculation across different time references on a computer. The remainder of this paper is organized as follows: Section 2 illustrates the designed UTF in Geosciences Knowledge System and its ideas. Section 3 shows how UTF works compared with current time conversion processes. Section 4 discusses the advantages, influencing factors, and prospects of the UTF. Finally, the conclusions and future works are stated in Section 5.

2. Methodology

2.1. Problems and basic idea

The key to solving the above problems is to remove the domain limitations and experience dependence in the calculation process. In other words, temporal information with different expressions and structures is heterogeneous because they lack a unified system to express various time elements. Current time calculation procedures across different databases refer to steps of time parsing, time complement, and time alignment, as detailed in Fig. 1a. These steps all have domain limitations and experience dependence because it has to consider the target time reference when calculating. These relationships lead to different outcomes from the above three steps in the calculations with different target time references. Thus, the time conversion process must remove the dependency of the target time reference if the current problem is to be solved.

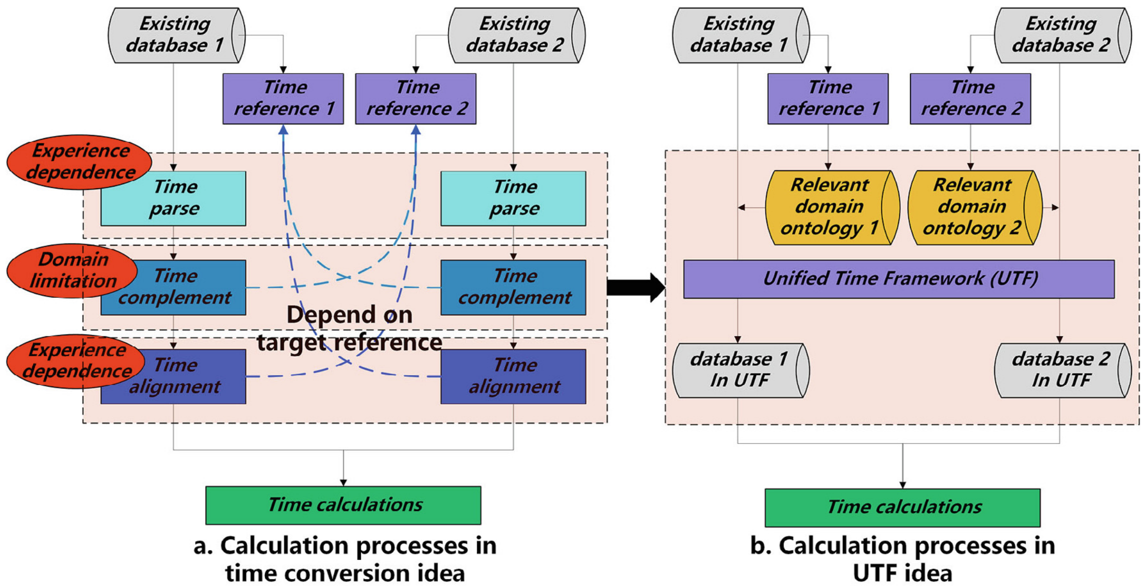

Fig. 1. The processes of temporal information calculation across different time references. (a) In time conversion idea; (b) in UTF idea.

The reason for the dependency on the target time reference during the time calculation process is that different time references have different time elements and structures. If a UTF is available to accommodate different time elements and structures, the time conversion process can be eliminated (Fig. 1b). The calculation structure based on the UTF idea only needs to refer to the relevant domain ontologies. Furthermore, most of them have already been constructed in recent two decades, such as geological time ontologies (Perrin et al., 2011; Ma et al., 2012; Cox and Richard, 2015; Hou et al., 2018; Ma et al., 2020), solar calendar ontologies (Hobbs and Pan, 2006; Zhang et al., 2011; Cox and Little, 2021), lunar calendar ontologies (Wong et al., 2009; Zou and Park, 2011), and over 1600 domain time ontologies in the open-sourced database (DBpedia community, 2021a). The massive existing domain time ontologies pose a formidable challenge in dealing with the different time elements and structures in the UTF.

2.2. Design principles

To address the issues caused by different time elements and structures, the UTF needs to follow two core design principles to suit the existing different time ontologies.

2.2.1. Comprehensiveness in time elements

The UTF needs to accommodate all types of time elements, including time types, time expressions, time references, time units, and other element types from different time ontologies to suit diverse time ontologies. Having comprehensive time elements gives UTF the ability to express the entire existing time ontologies. Examples include the "time reference" element in geological time ontologies, the "time dating" element in lunar calendar ontologies, and the "time type" element in general time ontologies. Moreover, sub-elements of "time reference" also need to be considered comprehensively, including "time reference version", "time reference boundaries", and "time reference dating method". The different levels of time elements need to be organized as well, which implicates the second principle.

2.2.2. Systematicness in time structures

The time structure in the UTF should be organized systematically, taking into account the different semantics, levels, and relationships of the time elements, since these determine the time calculation performance when querying on a computer. A systematic time structure would therefore include three aspects. First, a general "Time" element shall be designed to link its object since the entire time elements need to be organized. By using a general "Time" element, other information will no longer be queried when performing time calculations. Second, the UTF needs to select the minimum number of core time elements to ensure the efficiency of time calculations. The core time elements (including time type, time reference, time expression, time dating, and time relation) guarantee the integrity of the fast semantic calculation. And the minimum number of time elements determines the query efficiency. Third, more detailed time elements need to be designed in a hierarchical structure to store different sub-elements in different element groups, such as "time reference" includes "time reference version", "time reference boundaries", "time reference dating method", etc.

Guided by these two core principles, the details of the UTF can be designed well.

2.3. Unified time Framework (UTF)

According to this design philosophy, the detailed Unified Time Framework (UTF) is designed in Fig. 2. In general, it consists of four parts: a root class to link all kinds of relevant time elements about the object, some general classes, such as time type, time reference, time relation, and et al., several description attributes in each class, for example, "Point_time" class has the attributes of point_id, point_value, point_unit, point_uncertainty, and et al., and some attribute interfaces to link to external ontologies, such as the reference_system attribute can link to existing geological time ontologies.

Point_time, Inverval_time, and Combination_time are known to be the primary time expressions for recording temporal information in the UTF. Point_time, such as BaseJurassic, is a point to store an absolute time flash. Inverval_time, such as Jurassic, is an interval to express time. Combination_time, a mixed set of Point and Interval times, is a form of expressing hybrid complex temporal information. For example, the glacial period consists of several Interval times and is hence suitable for expression in terms of Combination_time. Combination_time is an effective supplement

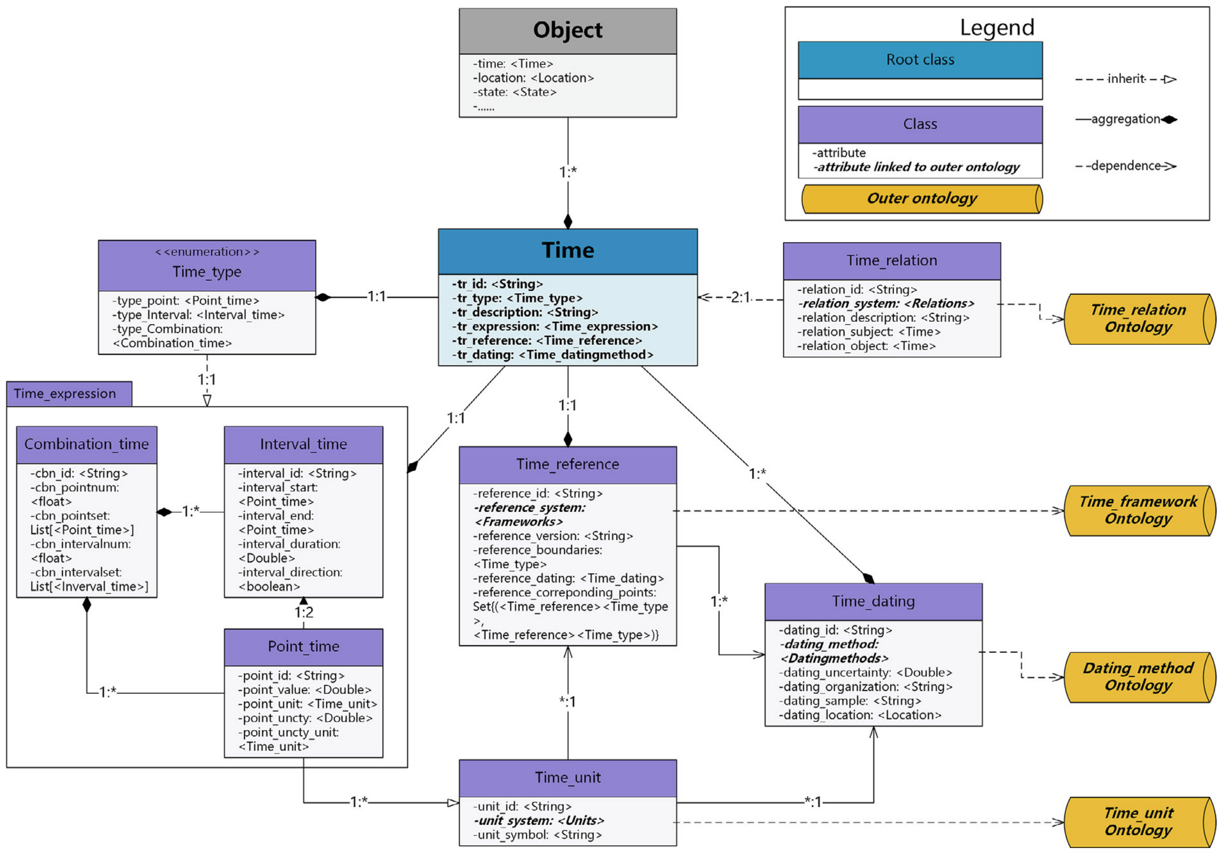

Fig. 2. The logic structure of UTF.

of Point_time and Interval_time in complex time forms. It is noted that Combination_time can cover the expressions of Point_time and Interval_time when the quantity of Interval_time or Point_time is zero, respectively. For instance, the Jurassic can be seen as the Combination_time, which contains the intervals of Early, Middle, and Late Jurassic, without Point_time. The structure of the Combination_time (“cbn_intervalset: List <Interval_time>”) is similar to that of the Interval_time in this case and has no effect on the temporal information calculations. Thus, users can choose the time expression type according to actual requirements. More definitions and examples of time elements in the UTF are listed in the Supplementary data Table S1.

According to the diverse time elements, the UTF structure has five typical features to ensure an accurate and efficient expression of unified temporal information.

2.3.1. An independent time root node

To avoid irrelevant attributes when time calculating, the UTF designs an independent time root node to link to the object. All the relevant time elements organize a time root node, which can be accessed and calculated directly. The time root includes essential attributes, which encompass the id, type, description, expression, reference, and dating method information.

2.3.2. Different time types

To adapt to the time expression of different geoscience disciplines, the UTF designs three types of time expressions: Point, Interval, and Combination. The Point time refers to the time moment, which stores the single time value and unit, along with the uncertainty and uncertainty units of the time value, and is the fundamental statement of all geosciences time expressions. The Interval time includes two Point times: a start and an end instance time. It is important to point out that since Interval time in different directions cannot be calculated, interval directions also need to be recorded to describe the order of the Point times. It is because the Interval times with different directions cannot be calculated. The third type is the Combination time that indicates a set of points and intervals. It consists of id, point_number, point_set, interval_number, and interval_set.

2.3.3. Same time nodes and structures

To achieve time calculation across different time elements, the UTF designs same time nodes and structure to record the time elements. It is because each Time has to record its reference, unit, dating method, and other time elements to express its semantic meaning integrally. This unified structure supports that all the time calculations have the same queries without estimating whether the nodes indicate the same time elements, significantly improving the time calculation performance across different time elements.

2.3.4. Quantitative relationship definitions

To ensure the accuracy of time expressions, the UTF defines the direct quantitative relationships between time elements. The explicit quantitative relationships can promote two key aspects. On one hand, it rules the topological relations among different time elements, thereby clarifying relationships such as sequence and overlap among them. On the other hand, it defines the quantitative relationships between different time elements, thereby improving the standardization ability of time expressions and avoiding errors among numerous time expressions.

2.3.5. Adequate interfaces to external ontologies

To link to the external ontologies, the UTF sets some interface attributes that can cite the required and existing ontologies. For example, the attribute of reference_system can cite different external framework ontologies. Moreover, relation_system, unit_system, and dating_method can also link to the required ontologies. This structure makes the UTF a platform to link to current ontologies.

In terms of physical storage, the UTF uses RDF (Resource Description Framework) triples to store the knowledge. All the semantic information can be recorded in S—P—O (<Subject (S), Predict (P), Object (O)>) triple. For instance, "Time has a tr_reference property." can be represented in (<utf:tr_reference, rdfs:subPropertyOf, utf:Time >) where "utf" and "rdfs" are short for the schema prefix "@prefix utf: <https://www.ddeworld.org/ddeKG/base#>" and "@prefix rdfs: <https://www.w3.org/2000/01/rdf-schema#>", respectively. Thus, the UTF can be stored in standard concept RDF triples (Fig. 3 (a)). The entire RDF files can be accessed from the Github platform (https://github.com/shuwang8951/Unified-Time-Framework-UTF-ontology). Accordingly, the instances in UTF are shown in Fig. 3b. The example describes the time value information of a DBpedia object Gryponyx. It has a specific time root node with associated time root properties including tr_id, tr_description, and tr_expression, and corresponding Point time properties, such as id, value, uncertainty, unit, and uncertainty unit.

3. Experiment and result analysis

To verify the validity and improvement of the UTF in time calculation across different time references, the experiment compares the time query procedures in SPARQL (SPARQL Protocol and RDF Query Language), a typical RDF query language, between the time conversion and the UTF structure method. It is noted that time query is the most frequent process among the time calculation. This section follows experimental data, experimental procedures, and comparative analyses.

3.1. Experimental data

The original experimental dataset is extracted from the entire DBpedia dataset (DBpedia community, 2021b), which is available on the Github platform (https://github.com/shuwang8951/Unified-Time-Framework-UTF-ontology). It stores the relative fossil time information using the predicate of "<…/fossilRange>" and other attributes (see Fig. 4). The dataset is chosen as the fossil description data in DBpedia contain multiple time types across different time references, such as geological time descriptions as "Eocene" and "Early_Jurassic", calendar time descriptions as "1900", "635" and "130_Ma", and combination time descriptions as "Late_Emsian_to_Middle_Devonian", "Early_Cretaceous,_135-130_Ma", and "(Molecular_clock_evidence_for_origin_between_1050_and_800Ma)". These complex time descriptions create an equally significant challenge for both the time conversion structure and the UTF structure for time calculations.

Specifically, the dataset contains 106,358 triples, including 12,342 time description triples. These time description triples have five main types: simple phrase, simple number, interval phrase, phrase and number, and complex phrase, which cover all geological time types. More classified information and examples are listed in Table 1. Examples indicate that the time query process in the experimental dataset requires consideration of multiple time descriptions.

In short, this experimental dataset with multiple time descriptions is challenging to calculate time across different time references, so it can be used as the benchmark to compare the performance of the time conversion and the UTF structure.

3.2. Experimental procedures

To compare the time query procedures based on time conversion and the UTF structure, a specific query task is designed based on the experimental dataset as a way of answering the question "What fossils predate than the Selmasaurus?", which requires considering complex time descriptions across different time

```
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix utf: <https://www.ddeworld.org/ddeKG/base#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

utf:Time
  rdf:type owl:Thing ;
  rdf:type skos:Concept ;
  rdf:type skos:Collection ;
  rdfs:subPropertyOf rdf:Property ;
  rdfs:label "Unified Time Framework in Deep-time Digital
Earth IUGS Big Science Program"@en ;
  skos:prefLabel "Unified Time Framework in Deep-time
Digital Earth IUGS Big Science Program"@en ;
.
utf:tr_id
  rdf:type rdf:Property ;
  rdfs:subPropertyOf utf:Time ;
  rdfs:range xsd:string ;
  skos:prefLabel "Time root id"@en ;
.
utf:tr_type
  rdf:type rdf:Property ;
  rdfs:subPropertyOf utf:Time ;
  rdfs:range utf:Time_type ;
  skos:prefLabel "Time root type"@en ;
.
```

(a)

```
@prefix utf: <https://www.ddeworld.org/ddeKG/base#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/#> .
@prefix geo: <http://www.opengis.net/ont/geosparql#> .
@prefix gts: <http://resource.geosciml.org/ontology/timescale/gts#> .
@prefix isc: <http://resource.geosciml.org/classifier/ics/ischart/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rank: <http://resource.geosciml.org/ontology/timescale/rank/#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix db: <http://dbpedia.org/#> .
@prefix thors: <http://resource.geosciml.org/ontology/timescale/thors#> .
@prefix time: <http://www.w3.org/2006/time#> .
@prefix ts: <http://resource.geosciml.org/vocabulary/timescale/#> .
@prefix vann: <http://purl.org/vocab/vann/#> .
@prefix void: <http://rdfs.org/ns/void#> .
@prefix xkos: <http://rdf-vocabulary.ddialliance.org/xkos#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

<db:Gryponyx> <utf:Time> <utf:Gryponyx_timeroot>.
<utf:Gryponyx_timeroot> <utf:tr_id> <utf:99baa434-63ea-11ec-9b34-020017000b7b>.
<utf:Gryponyx_timeroot> <utf:tr_description> <Lower_Jurassic>.
<utf:Gryponyx_timeroot> <utf:tr_expression> <utf:Point_time>.
<utf:Gryponyx_timeroot> <utf:Point_time> <utf:Gryponyx_Point_time>.
<utf:Gryponyx_Point_time> <utf:point_id> <utf:cae69067-d399-3fef-8999-a32eb1bb9f2f>.
<utf:Gryponyx_Point_time> <utf:point_value> 199.6.
<utf:Gryponyx_Point_time> <utf:point_uncty> 0.6.
<utf:Gryponyx_Point_time> <utf:point_unit> <utf:Ma>.
<utf:Gryponyx_Point_time> <utf:point_uncty_unit> <utf:Ma>.
```

(b)

Fig. 3. The examples of specific physical contents of the UTF RDFs. (a) Concepts relationship RDFs in UTF; (b) instance RDFs in UTF.

<http://dbpedia.org/resource/Selmasaurus> <http://dbpedia.org/property/imageCaption>
<Selmasaurus_johnsoni_mounted_skull_in_the_Rocky_Mountain_Dinosaur_Resource_Center_in_Woodland_Park,_Colorado.>.
<http://dbpedia.org/resource/Selmasaurus> <http://dbpedia.org/property/authority> <Wright_&_Shannon,_1988>.
<http://dbpedia.org/resource/Selmasaurus> <http://dbpedia.org/property/taxon> <Selmasaurus>.
<http://dbpedia.org/resource/Selmasaurus> <http://dbpedia.org/property/fossilRange> <Late_Cretaceous,_Campanian>.
<http://dbpedia.org/resource/Semantoridae> <http://dbpedia.org/property/imageCaption> <Restored_skeleton_of_Puijila_darwini>.
<http://dbpedia.org/resource/Semantoridae> <http://dbpedia.org/property/authority> <Orlov,_1931>.
<http://dbpedia.org/resource/Semantoridae> <http://dbpedia.org/property/taxon> <Semantoridae>.
<http://dbpedia.org/resource/Semantoridae> <http://dbpedia.org/property/fossilRange> <Late_Oligocene_-_Late_Miocene,>.
<http://dbpedia.org/resource/Semantoridae> <http://dbpedia.org/property/subdivision>
<*_Semantor_(Orlov,_1931__type)_*_Necromites_(V._V._Bogachev._1940)_*_Potamotherium_(Geoffroy,_1833)_*_Puijila_(Rybczynski_et_al.,_2009)>.
<http://dbpedia.org/resource/Semantoridae> <http://dbpedia.org/property/subdivisionRanks> <http://dbpedia.org/resource/Genus>.

Time description triples Attribute triples

Fig. 4. The examples of experimental datasets about Selmasaurus and Semantoridae.

Table 1
The classified information about different types of time description triples.

Types	Quantity	Examples of time description part
Simple phrase	9177	<Middle_Cambrian>
Simple number	165	<~310_Ma>
Interval phrase	2493	<Eocene_to_Middle_Miocene>
Phrase and number	459	<Middle_Pleistocene_to_0.3_Ma>
Complex phrase	48	<Late_Cretaceous-Early_Eocene,__Santonian-Ypresian>

references. Two different calculation procedures are occurred by using the Time Conversion (TC) method and the UTF method (Fig. 5).

In TC procedures, geological time descriptions and calendar time descriptions are parsed into the corresponding dataset using the relevant ontologies, such as calendar-based data, Geological Time Scale 2004 (GTS04) based data, and Geological Time Scale 2018 (GTS18) based data. Then, different datasets are converted into the target time reference; for example, GTS04 to GTS18 are converted into calendar references in Fig. 5 using programming codes with expert experiences. And finally, TC-based fossil data can be calculated by the TC SPARQL query. Fig. 6 shows the representative data in the TC procedures including the original experimental dataset, TC-based fossil data, and corresponding SPARQL query. It is noted that different target references have different structures to organize the temporal information. Fig. 6 also uses different colors to represent the structure, relevant data, and corresponding query. While the structure changed, the relevant TC-based fossil data and SPARQL query will also be replaced according to the target reference structure. Thus, the TC procedure will be changed all the time depending on target reference and expert experience to convert.

In UTF procedures, the dataset construction is different from the TC procedures, and all the experimental data are parsed into the

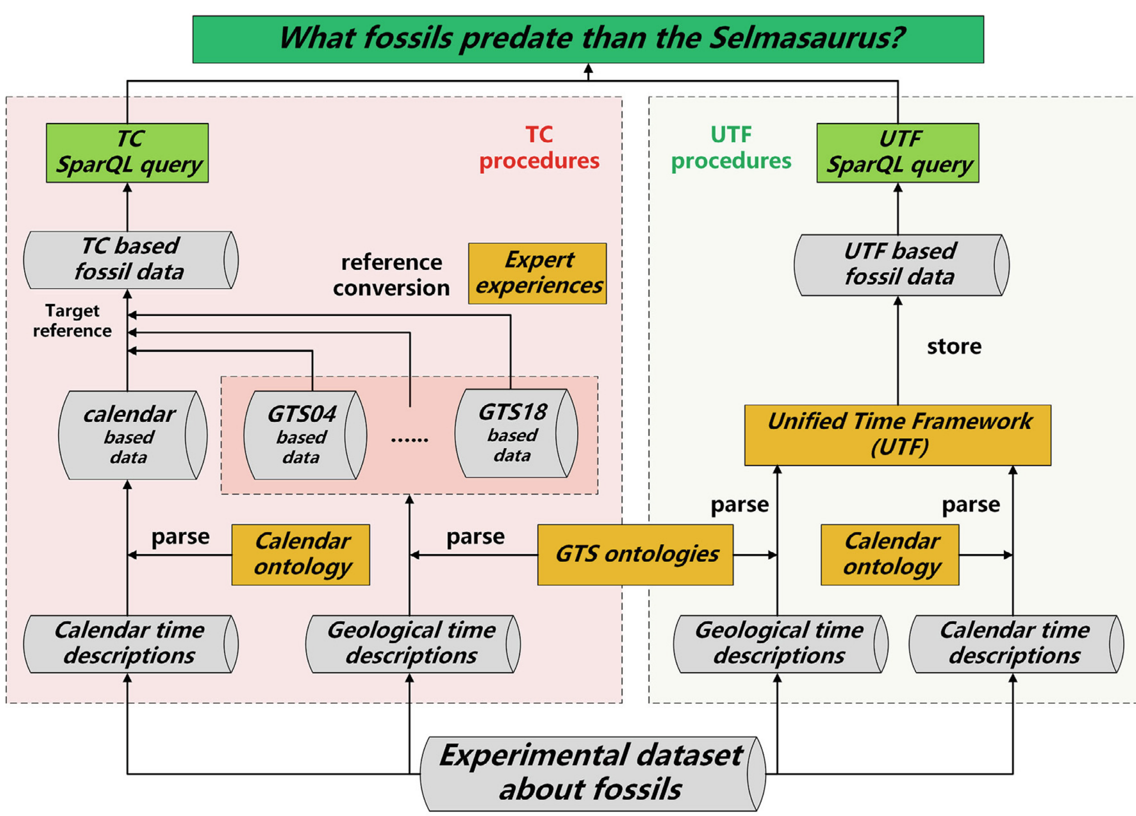

Fig. 5. The calculation procedures of time conversion (TC) method and unified time framework (UTF). Note: the GTS04 based data means the data uses the Geological Time Scale 2004/2018 (GTS04/18) ontology to express the temporal information (International Commission on Stratigraphy, 2022).

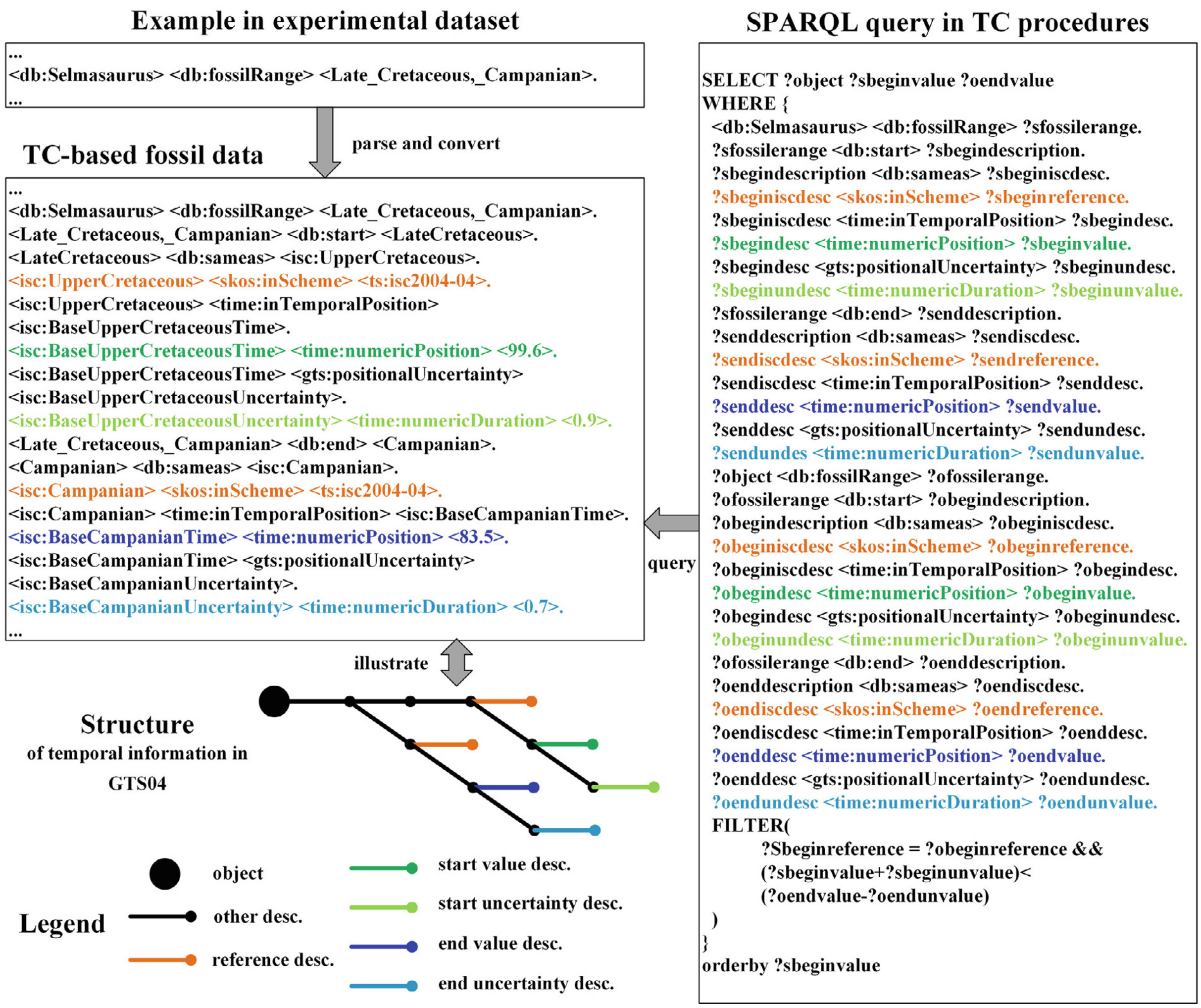

Fig. 6. An example of TC main procedures (typical corresponding abbreviations are as follows: "**s**beginvalue" indicates "**subject** begin value", "**o**endvalue" indicates "**object** end value", "sbeginisc**desc**" indicates "subject begin isc **description**", "sbegin**un**value" indicates "subject begin **uncertainty** value", and "reference **desc.**" indicates "reference **description**").

UTF structure using the relevant ontologies. Thus, fossil data has the same structure, which the UTF SPARQL query can directly calculate. Fig. 7 gives an example of UTF main procedures. As the UTF is fixed to the full-time references, the whole procedures and relevant SPARQL queries remain unchanged, making the UTF procedures easier to use.

3.3. Procedure analyses

Three major differences can be identified between the two methods by comparing the TC and UTF procedures: an external conversion process in TC calculation, expert experience dependence in TC calculation, and fixed query structure in UTF calculation.

Firstly, the TC calculation procedure has an external conversion process compared to the UTF procedure, which is due to the inability to calculate the parsed dataset directly. For example, the time descriptions of calendar data (<db:Polysacos> <db:fossilRange> <335 Ma>.) and GTS04 data (<db:Manatee> <db:fossilRange> <isc:Pleistocene>.) are not in a reference system. It is necessary to pay attention to the fact that the calculation of the different data for "isc:Pleistocene" and "335 Ma" requires a conversion process to transfer the former to its absolute geological time values of "0.781 Ma–0.126 Ma". Thus, the TC-based calculation has more complex processes than the UTF-based calculation.

Secondly, the reference conversion in the TC procedure depends on expert experiences. It is because target reference could be different from the original data changing. Thus, various reference conversion tools in the TC procedure need to be developed with expert experience on different reference conversions to fix the reference conversion problem.

Thirdly, the SPARQL queries for TC and UTF are different, indicating that the UTF SPARQL query has the fixed query structure while the TC SPARQL has not. The UTF defines the same structure as the temporal information about the fossils. In contrast, the TC-based dataset may have different storing structures depending on various target references.

It can be observed by comparing the TC and UTF procedures that the UTF-based method calculates temporal information more easily than the TC-based method. It has a fixed unified structure to represent temporal information, which leads data to have a unified structure and brings a fixed SPARQL query to calculate relevant temporal information. To put it simply, the UTF can be used to calculate temporal information automatically by replacing the manual experience depending time conversion method.

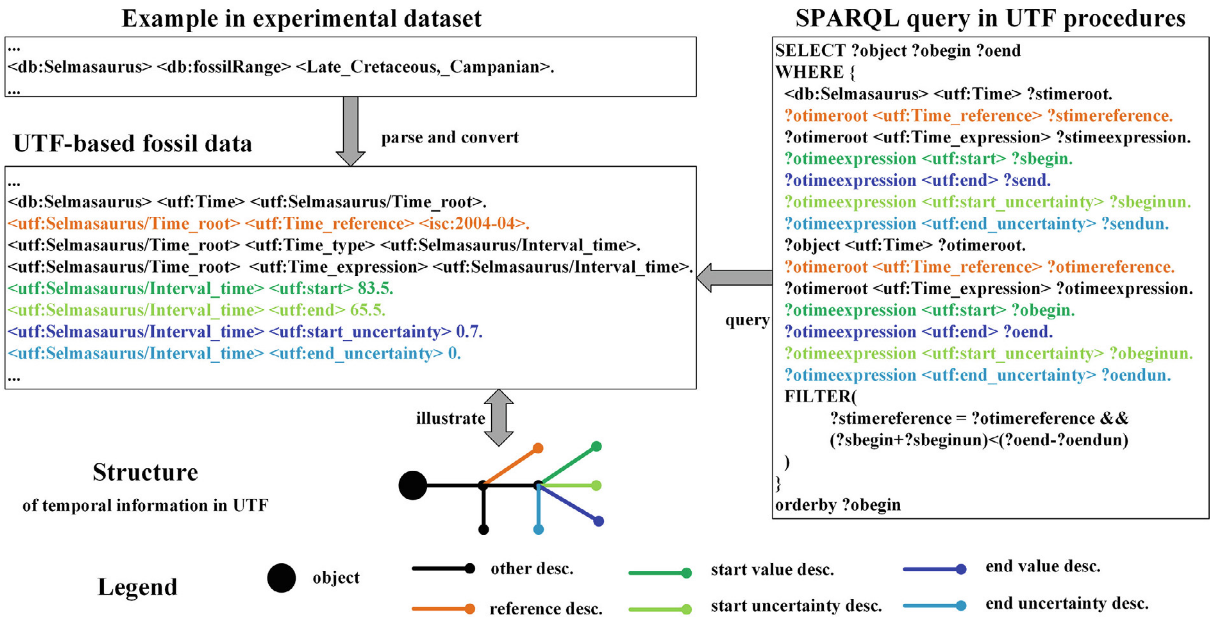

Fig. 7. An example of UTF main procedures (typical corresponding abbreviations are as follows: "**s**begin" indicates "**subject** begin", "**o**timeroot" indicates "**object** time root", "sbegin**un**" indicates "subject begin **uncertainty**", and "reference **desc.**" indicates "reference **description**"). Note that the Combination_time expression does not affect the SPARQL querys.

3.4. Performance analyses

To analyze the performance improvement of the UTF quantitatively, the time consumptions are compared between the UTF and TC procedures. Assuming that all the code including the TC and the UTF has been written well and the procedures can be worked automatically, the time consumption of two procedures are recorded and compared. The data is shown in Table 2.

The "Extra node number" column is set to compare the speed of performance change, which indicates the number of extra nodes added in each node of current experimental data. Results reveal that the average time cost of the UTF is lower than that of the TC and that the percentage reduction in time consumption rises as the number of extra nodes increases. To illustrate the performance improvement, the query efficiency (queries/ 100 ms) is calculated (Fig. 8).

Fig. 8 shows the performance difference between the UTF and the TC method. In general, the former performs at least 7% better than the latter. According to the trend in Fig. 8, the performance difference between the two methods will continuously increase with the extra node increasing. It is thus clear that the query efficiency of UTF is stable and much better than the current TC method. Therefore, the quantitative performance experiment indicates the UTF method has excellent performance compared with the TC method.

4. Discussion

By analyzing the different procedures of temporal information calculation, the experiment proves the UTF method can support temporal information calculation across different time references. Furthermore, a comparison of the two indicates that the UTF-based method is easier and more efficient than the TC-based method with its fixed data and query structure. This section presents a discussion of its dependencies for an in-depth understanding of the UTF.

4.1. The dependency on external ontologies

Although the UTF in geosciences knowledge system can support accurate and efficient temporal information calculation across dif-

Table 2
The performance comparison between the TC and UTF procedures.

Group id	Extra node number	TC		UTF		Reduced time consumption
		Time cost	Average time cost	Time cost	Average time cost	percentage
1	0	880.779	**88.08**±0.90	823.099	**82.31**±2.44	**6.55**
2	20	886.52	**88.65**±1.23	830.347	**83.03**±2.72	**6.34**
3	40	890.94	**89.09**±1.27	829.046	**82.90**±1.50	**6.95**
4	60	891.748	**89.17**±1.02	821.016	**82.10**±0.53	**7.93**
5	80	893.594	**89.36**±1.12	829.917	**82.99**±1.30	**7.13**
6	100	895.857	**89.59**±1.77	828.779	**82.88**±1.88	**7.49**
7	120	896.658	**89.67**±1.39	827.675	**82.77**±1.14	**7.69**
8	140	897.078	**89.71**±1.21	823.498	**82.35**±2.00	**8.20**
9	160	897.373	**89.74**±1.36	828.384	**82.84**±1.18	**7.69**
10	180	899.443	**89.94**±1.14	824.54	**82.45**±2.27	**8.33**
11	200	903.488	**90.35**±1.83	829.061	**82.91**±1.19	**8.24**

Note: the experiment uses the same query in section 3.2 experimental procedures "What fossils predate than the Selmasaurus?". The time unit is millisecond. Column of "Time cost" records the time consumption of the query in ten times.

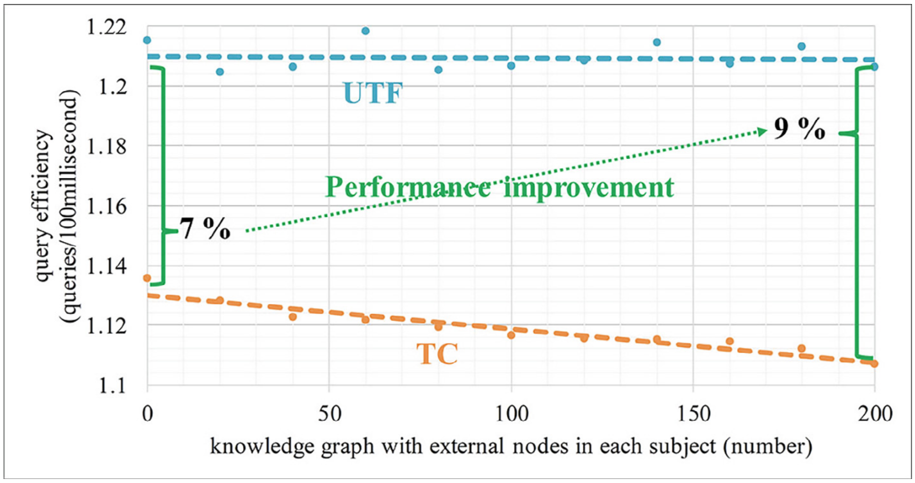

Fig. 8. The query efficiency comparison between the UTF the TC method.

ferent time references, it still has a limitation to be noticed, that is, the UTF is dependent on external ontologies. Specifically, the storing process needs domain ontologies to parse the domain knowledge. For example, without geological time ontology, the original description triple of "<db:Selmasaurus> <db:fossilRange> <Late_Cretaceous,_Campanian>" cannot parse the specific temporal information as "……<utf:start> <83.5>" and "……<utf:end> <65.5>". Thus, the UTF must work with the relevant domain ontologies to calculate temporal information across different time references.

At present, massive domain ontologies have been constructed as previously mentioned in Section 2.1, such as geological time scale ontologies in different versions, solar calendar ontology, and lunar calendar ontology. Current domain time ontologies can support the commonly used applications. All published existing time ontologies can be used directly in the UTF. For example, the unit triple "utf: ontology establish time utf:unit_symbol time: year" uses "time: <https://www.w3.org/2006/time#>". However, the partial lunar calendar, Julian calendar, Mayan calendar, and et al. are not constructed entirely. Moreover, part of them uses different languages to describe, such as Chinese, Latin, and Roman. Consequently, partial domain time ontologies cannot be directly applied to the UTF method, so the construction of domain ontologies can consistently contribute to the UTF-based temporal information calculation across different time references.

4.2. The dependency of the UTF popularity

Moreover, the UTF performance benefits will become more evident as its popularity increases. When all geosciences temporal information triples use the UTF, the time calculation process will radically change. For one thing, the process eliminates the need to convert temporal triples from different time references, thus greatly saving the time cost of translation; for another, it leads to a basic SPARQL query, the SPARQL query except for the FILTER section, to query all questions. It is because all the time elements are stored in the fixed position, which leads to the same SPARQL query to obtain. It will boost the performance of storage and calculation aspects.

Therefore, continuously increasing UTF popularity is an essential future task for UTF usage. At present, the UTF is set as the basic temporal framework in the Deep-time Digital Earth (DDE) program, which is an international big science program in the geoscience field (Stephenson et al., 2020; Wang et al., 2021a). DDE community is emerging with the UTF as the basic temporal framework. Currently, 62 existing time ontologies have been converted into a UTF version (https://doi.org/10.5281/zenodo.6734796). Moreover, The UTF is being tested in this program by increasing multidisciplinary knowledge with various time references. For example, the samples in different time references or different disciplines can jointly support the fine simulation of paleoclimate. With the DDE program being rolled out worldwide, the UTF will further exploit its strengths.

5. Conclusion

To achieve accurate and efficient temporal information calculation across different time references on a computer, this paper designs a Unified Time Framework (UTF) in the geosciences knowledge system with systematic time elements representation. The main conclusions are discovered as follows by comparing with the current Time Conversion (TC) method in the mechanism, procedure, data, and performance.

(i) The well-designed UTF can greatly support accurate and efficient temporal information calculation across different time references on a computer. It is possible to record geological time "Messinian", geographical time "CE 1950", geomagnetic time "C4A", paleontology time "CNM19", etc. on a time scale by using this structure, and also to calculate knowledge with different time references. Therefore, the UTF should be suggested to form a future geosciences infrastructure to apply in all geosciences knowledge systems as soon as possible.

(ii) The UTF shows a higher and stable performance of temporal information queries. It indicates that the UTF is a promising robust structure to massive geosciences knowledge calculations in the Big Data era. With the quantity and types of data increasing, the advantages of the UTF will be presented more comprehensively because the time query is the basic unit for all kinds of complex knowledge calculations.

(iii) The UTF still has a dependency on external domain time ontologies. While current domain time ontologies are able to support basic applications in both geological and geographical domains, other time ontologies are still required if it is necessary to calculate specific domain knowledge together (e.g., the relationship between Mayan culture and geological events). Thus, domain time ontology construction is a continuous task.

In short, the UTF has been well-designed to support accurate and efficient temporal information calculation across different time references, and it warrants further promotion in applications and domain ontology improvements. With the advent of the geoscience Big Data era, the UTF will support the calculation of cross-reference temporal information in various cross-geoscience projects to discover new geosciences knowledge.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge all suggestions from the scientists in the Deep-time Digital Earth (DDE) group. This work was funded by the National Natural Science Foundation of China (Grant Nos. 42050101 and 42101467) and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA23100101).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gsf.2022.101465.

References

Alekseev, A.S., 2015. The international stratigraphic chart: state of the art. Russ. Geol. Geophys. 56, 524–531.
Arboit, F., Min, M., Chew, D., Mitchell, A., Drost, K., Badenszki, E., Daly, J.S., 2021. Constraining the links between the Himalayan belt and the Central Myanmar Basins during the Cenozoic: An integrated multi-proxy detrital geochronology and trace-element geochemistry study. Geosci. Front. 12, 657.
Bergen, K.J., Johnson, P.A., De Hoop, M.V., Beroza, G.C., 2019. Machine learning for data-driven discovery in solid Earth geoscience. Science 363, eaau0323.
Dbpedia Community. 2021b. *Latest releases of core data from en.wikipedia.org* [Online]. Available: https://databus.dbpedia.org/dbpedia/collections/latest-core [Accessed 12.12.2021 2021].
Cox, S.J.D., 2016. Time ontology extended for non-Gregorian calendar applications. Semant. Web 7, 201–209.
Cox, S. J. D., Little, C. 2021. *Time Ontology in OWL* [Online]. Available: https://www.w3.org/TR/owl-time/ [Accessed 30th May 2021 2021].
Cox, S.J.D., Richard, S.M., 2015. A geologic timescale ontology and service. Earth Sci. Inform. 8, 5–19.
Dbpedia Community. 2021a. *DBpedia Ontology Archive* [Online]. Available: https://archivo.dbpedia.org/list [Accessed 2021].
Fan, J., Shen, S., Erwin, D.H., Sadler, P.M., et al., 2020. A high-resolution summary of Cambrian to Early Triassic marine invertebrate biodiversity. Science 367, 272.
Frasincar, F., Milea, V., Kaymak, U. 2010. tOWL: Integrating Time in OWL. *In:* De Virgilio, R., Giunchiglia, F. & Tanca, L. (eds.) *Semantic Web Information Management: A Model-Based Perspective.* Berlin, Heidelberg: Springer Berlin Heidelberg.
Fu, J., Pirajno, F., Yang, F., Shivute, E., Sun, Y., Ai, N., Qiu, K., 2021. Integration of zircon and apatite U-Pb geochronology and geochemical mapping of the Wude basalts (Emeishan large igneous province): A tool for a better understanding of the tectonothermal and geodynamic evolution of the Emeishan LIP. Geosci. Front. 12, 573.
Hobbs, J. R., Pan, F. 2006. *Time ontology in OWL* [Online]. Available: https://www.w3.org/TR/owl-time/ [Accessed 2021].
Hou, Z., Zhu, Y., Gao, Y., Song, J., Qin, C., 2018. Geologic time scale ontology and its applications in semantic retrieval. J. Geo-Info. Sci. 20, 17–27.
Jiang, J., Liu, R., 2021. Standardization of cartography and geographic information. In: Wang, J., Wu, F. (Eds.), Advances in Cartography and Geographic Information Engineering. Singapore, Springer Singapore.
Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., Van Kleff, P., Auer, S., Bizer, C., 2015. DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia. Semant. Web 6, 167–195.
Levine, J., 2016. The history of time and frequency from antiquity to the present day. Eur. Phys. J. H 41, 1–67.
Ma, X., Fox, P., 2013. Recent progress on geologic time ontologies and considerations for future works. Earth Sci. Inform. 6, 31–46.
Ma, X., Carranza, E.J.M., Wu, C., Van Der Meer, F.D., Liu, G., 2011. A SKOS-based multilingual thesaurus of geological time scale for interoperability of online geological maps. Comput. Geosci. 37, 1602–1615.
Ma, X., Carranza, E.J.M., Wu, C., Van Der Meer, F.D., 2012. Ontology-aided annotation, visualization, and generalization of geological time-scale information from online geological map services. Comput. Geosci. 40, 107–119.
Ma, X., Ma, C., Wang, C., 2020. A new structure for representing and tracking version information in a deep time knowledge graph. Comput. Geosci. 145, 104620.
Normile, D. 2019. Earth scientists plan a 'geological Google'. *Science,* 363**,** 917-917.
Pan, F., Hobbs, J.R., 2004. Time in owl-s. In: Proceedings of the AAAI Spring Symposium on Semantic Web Services, pp. 29–36.
Perrin, M., Mastella, L.S., Morel, O., Lorenzatti, A., 2011. Geological time formalization: an improved formal model for describing time successions and their correlation. Earth Sci. Inform. 4, 81–96.
Saïs, F., Malaverri, J.E.G., Quercini, G., 2020. MOMENT: Temporal Meta-fact Generation and Propagation in Knowledge Graphs. Proceedings of the 35th Annual ACM Symposium on Applied Computing. Association for Computing Machinery, Brno, Czech Republic.
Stephenson, M.H., Cheng, Q., Wang, C., Fan, J., Oberhänsli, R., 2020. Progress towards the establishment of the IUGS Deep-time Digital Earth (DDE) programme. Episodes 43, 1057–1062.
Tanon, P. T., Weikum, G., Suchanek, F. YAGO 4: A Reason-able Knowledge Base. *In:* Harth, A., Kirrane, S., Ngonga Ngomo, A.-C., Paulheim, H., Rula, A., Gentile, A. L., Hase, P., Cochez, M. (Eds.). Semant. Web, 2020//2020 Cham. Springer International Publishing, pp. 583-596.
The International Organization for Standardization, 2002. Geographic information-Temporal schema. Switzerland.
Wang, C., Hazen, R.M., Cheng, Q., Stephenson, M.H., Zhou, C., Fox, P., Shen, S.-Z., Oberhänsli, R., Hou, Z., Ma, X., Feng, Z., Fan, J., Ma, C., Hu, X., Luo, B., Wang, J., Schiffries, C.M., 2021a. The Deep-Time Digital Earth program: data-driven discovery in geosciences. Natl. Sci Rev.
Wang, C.B., Ma, X.G., Chen, J.G., 2018. The application of data pre-processing technology in the geoscience big data. Acta Petrol. Sin. 34, 303–313. in Chinese with English abstract.
Wang, S., Zhang, X., Ye, P., Du, M., Lu, Y., Xue, H., 2019. Geographic knowledge graph (GeoKG): a formalized geographic knowledge representation. ISPRS Int J. Geo-Inf., 8
Wang, S., Qian, L., Zhu, Y., Song, J., Lu, F., Zeng, H., Chen, P., Yuan, W., Li, W., Geng, W., 2021b. A web text mining approach for the evaluation of regional characteristics at the town level. Trans. GIS 25, 2074–2103.
Wong, J.H., Dillon, T.S., Wong, A.K., Lin, W.W., 2009. Text mining for real-time ontology evolution. Data Mining for Business Applications. Springer.
Zhang, C., Cao, C., Sui, Y., Wu, X., 2011. A Chinese time ontology for the Semantic Web. Knowl. Syst. 24, 1057–1074.
Zhang, X., Chen, N., Chen, Z., Wu, L., Li, X., Zhang, L., Di, L., Gong, J., Li, D., 2018. Geospatial sensor web: A cyber-physical infrastructure for geoscience research and application. Earth Sci. Rev. 185, 684–703.
Zhou, C., Wang, H., Wang, C., Hou, Z., Zheng, Z., Shen, S., Cheng, Q., Feng, Z., Wang, X., Lv, H., Fan, J., Hu, X., Hou, M., Zhu, Y., 2021. Geoscience knowledge graph in the big data era. Sci. China Earth Sci. 64, 1105–1114.
Zou, Q., Park, E.G., 2011. Modelling ancient Chinese time ontology. J. Inf. Sci. 37, 332–341.