Extending SUMO to Geological Times

Alexandre Rademaker[1,4], **Alexandre Tessarollo**[2,4], **Henrique Muniz**[1,4], **Adam Pease**[3]

[1]IBM Research, Brazil

[2]Petrobras, Brazil

[3]Infosys, USA

[4]FGV/EMAp, Brazil

alexrad@br.ibm.com, apease@articulatesoftware.com

{alexandretessarollo, hn.muniz.a}@gmail.com

Abstract. *One of the challenges of Natural Language Processing of Oil&Gas domain is reasoning with geological times. Although there are some initiatives for specifying the vocabulary of this information, they fall short on enforcement of expected properties, such as no overlapping between Ages (Epoch, Eras etc) and hierarchy compliance. We used the Suggested Upper Merged Ontology (SUMO) and its associated automated reasoning tools to tackle these matters and uncovered some inconsistencies on geological time International Chronostratigraphic Chart (ICC) official published material.*

1. Introduction

Oil&Gas Exploration and Production companies annually invest billions of dollars gathering documents, including reports, scientific articles, business intelligence articles and so on. These documents are the main base for major decisions such as whether to drill exploratory wells, bid or buy, production schedules and risk assessments. However most of the processing of this fundamental data is still done by human professionals actually reading it rather than a computational system. Considering that this unstructured data is growing exponentially, management of such data and finding relevant content quickly has become one of companies and professionals most critical challenges [Antoniak et al. 2016, Schoen et al. 2018]. Natural Language Processing on the specific domain of Oil&Gas has its own challenges, some of them presented in [Rademaker 2018].

Assessing geoscience papers one can notice that among the most common properties raised are usually geographic location [Palkowsky 2005] and geological time, e.g. '165 Million years ago (Ma)', 'during the Jurassic Period,' etc. Applications, such as `http://www.agenames.org/` attest the relevance of such information. It was implemented to perform (space and time) query and scan documents for stratigraphic terms, identifying the stratigraphic context of a publication. In this work we aim to set the ground for a 'deep' natural language processing pipeline capable of not only identifying references to terms but also reasoning about them, answering user questions. It is notable that even simple inferences are not yet available for users. Consider the application mentioned above, we would expect it to return not only utterances that explicitly mention the term 'Maastrichtian,' but also numeric expressions referring to the interval from 72.1 to 66 Million years ago if a user searches for the Maastrichtian Age. Some of us have previously presented in [Muniz et al. 2018] an extension of Princeton WordNet [Fellbaum 1998] for geological time terms, making the first step towards our intended pipeline. Here we continue working on modeling temporal aspects but focus on their definition in logic rather than vocabulary.

While a lexical resource can provide a computer an inventory of words, it cannot provide the information needed for computation about time periods and the facts that hold for those time periods. We are concerned with deductive reasoning that can compute answers to questions, rather than simply retrieving a document that may contain words similar to those in the question a user asked. We are also concerned with a software engineering model of capturing such information, so that it can have a long period of utility on a variety of applications. While it might appear quicker to develop an ontology from scratch, specific to our present domain and application, modern software development practices include reusing a majority of code from a library and building extensions compatible with that library. That is the approach we follow here, building on the Suggested Upper Merged Ontology (SUMO) [Pease 2011, Niles and Pease 2001].

Another decision to make is in what formal language to code the ontology. While much effort in the field today is done in taxonomies and semantic networks, or in semantic web languages like OWL and RDF, such approaches must grapple with the fact that many facts that are easily stated in human language cannot be formally stated in those languages. In particular, we need to be able to make statements about what is possible, or what may be true during a given period of time.[1] This requires a logic beyond first order logic (FOL) (and therefore well beyond Description Logic). This provides another motivating factor for adopting SUMO and its higher order logical language, SUO-KIF [Pease 2009]. Its associated translations to TPTP, TFF0 [Sutcliffe et al. 2012] and THF provide a range of options for use with the best modern theorem proving tools, such as Vampire [Kovács and Voronkov 2013] and LEO-III [Steen and Benzmüller 2018]. Since we need to perform expressive inference, this provides another motivation for this choice. We also can use the same automated theorem proving tools to check the consistency of our formalizations, which is an approach to software quality not available to procedural production systems (like CLIPS or SWRL). For this paper, given the focus on geologic time periods and arithmetic calculations with them, we will focus on the TFF0 translation of SUMO and proving within Vampire.

The paper is organized as follows. In Section 2 we outline the domain we are interested in modeling. We discuss related work and the currently available ontologies for Geological Time Periods in Section 3. Section 4 presents our formalization of the domain in SUMO. Briefly, the reason to use the knowledge rep-

[1] For example, consider sentences like "Regions marked by important erosion and truncation of pre-salt strata, uplifted and exposed sub-aerially before the deposition of Aptian salt, *can* form structural lows at present or be part of horsts uplifted after the Aptian." [Alves et al. 2017].

resentation language SUO-KIF is that a description logic doesn't allow us to capture the original natural language definitions from the domain, just the taxonomy of concepts and argument types. OWL does not allow for arities beyond binary, modal statements including temporal qualifications of formulas etc. Without expressive rules supported in SUO-KIF, most of the statements and terms would not be properly formalized, leaving the semantics to the imagination of the user (and each user is likely to have a slightly different intuition), rather than accessible through logical inference. Finally, we conclude and present some future work in Section 5.

2. Geological Time Periods

The geologic timescale is used by geologists, paleontologists, and other geoscientists to describe the timing and relationships of events in Earth's history. The table of geologic time spans set forth by the International Commission on Stratigraphy (ICS), a sub-committee of the International Union of Geological Sciences, is described in `http://www.stratigraphy.org`. The geologic timescale is organized in a hierarchical fashion. Eons (or aeons) are divided into eras. Eras contain periods that contain epochs, and finally epochs contain ages. The first three eons (Hadean, Archean, Proterozoic) are collectively referred as the Precambrian super-eon. The most recent eon, the Phanerozoic is subdivided into several periods.

The International Commission on Stratigraphy publishes regularly the International Chronostratigraphic Chart (ICC) [2] as the current standard of the organization of the geologic timescale of the Earth. In the current version, the chart contains 175 names of geological periods. One can read about the development of the chart in [Cohen et al. 2013].

As explained in that paper, geological time periods are not as well-established as one might expect. The committee was tasked with producing a chart that solved the issues of conflicting and overlapping regional strata. We assume the chart and its periods and boundaries represent the consensus between scientists working on this area. A fragment of the ICC is presented in Figure 1.

3. Related Work

Temporal Logic is a term broadly used to cover all approaches to representing and reasoning about time and temporal information within a logical framework. It can be more narrowly defined to refer the modal-logic introduced by Arthur Prior [Prior 1962] under the name of Tense Logic and subsequently developed further by many researchers. Over time, Temporal Logic has been used for many applications such as a formalism for clarifying philosophical issues about time, as a framework to precisely define the semantics of temporal expressions in natural language, as a language for encoding temporal knowledge in artificial intelligence and as a tool for specification and verification of computer programs [Goranko and Galton 2015].

[2] It was previous called International Stratigraphic Chart (ISC). It can be found at `http://www.stratigraphy.org/index.php/ics-chart-timescale`.

System/ Period	Series/ Epoch	Stage/ Age	Age (Ma)	
Paleogene	Paleocene	Danian	younger	
Cretaceous	Upper/ Late	**Maastrichtian**	66.0	72.1
		Campanian	72.1	83.6
		Santonian	83.6	86.3
		Coniacian	86.3	89.8
		Turonian	89.8	93.9
		Cenomanian	93.9	100.5
	Lower/ Early	Albian	100.5	~113.0
		Aptian	~113.0	~125.0
		Barremian	~125.0	~129.4
		Hauterivian	~129.4	~132.9
		Valanginian	~132.9	~139.8
		Berriasian	~139.8	~145.0
Jurassic	Upper/ Late	Tithonian	older	

Figure 1. A fragment of the ICC presenting the Maastrichtian age.

In a more practical point of view, one of the seminal works is Allen's interval algebra. It is a calculus for temporal reasoning that was introduced in [Allen 1983]. The calculus defines possible relations between time intervals and provides a composition table that can be used as a basis for reasoning about temporal descriptions of events.

Many vocabularies for time concepts where developed for the Semantic Web initiative. The most notable OWL/RDF vocabulary actively maintained for the time domain is the OWL-Time from W3C[3] but as noted above, lacks the language and reasoning frameworks needed to compute answers to numerical queries about times and dates. Interesting to note is that most of them are derived from the formalization presented in [Hobbs and Pan 2004], where the presentation is a mix of first order logic formulas and description logics (OWL) formulas and it is not easy to grasp the intended target formal logic language in the paper. For instance, the use of some ternary predicates, such as `timeBetween`, makes the presentation not directly entirely convertible to OWL. The authors say "This effort has been informed by temporal ontologies developed at a number of sites; it is intended to capture the essential features of all of them and make them easily available to a large group of Web developers and users, embedded in the ontology mark-up language OWL."

The geologic timescale represented in the chart described in the last section is a complex data structure composed of abstract elements, instants and time intervals, and their relationship with specific concrete representations of geologic records and the observations made of those concrete representations. The International Commission on Stratigraphy guidelines recommends a very precise usage of these components in order to establish a standard timescale for global correlations. However, this has been primarily described in text [Remane et al. 1996]. In [Cox and Richard 2005], a representation of the model using the Unified Modeling Language (UML) was presented. The model builds

[3] `https://www.w3.org/TR/owl-time/`

on existing components from standardization of geospatial information systems.

Later on, an OWL ontology for the geologic timescale, the ISC ontology, derived from the UML model was presented in [Cox and Richard 2014]. All versions of the International Stratigraphic Chart from 2004 to 2014 have been encoded using the ISC ontology. A particular aspect of the ISC ontology is that the elements of the timescale retain the same identifiers across the multiple versions, though the information describing each geochronologic unit evolves with the versions of the timescale. The ISC ontology contains many sub-ontologies including the Geologic Timescale [4] (GTS), the Temporal Hierarchical Ordinal Reference System model [5] (THORS), the Simple Knowledge Organization System (SKOS) [Isaac and Summers 2008] and the already mentioned OWL-Time.

It is worth noting that although ISC ontology makes use of the different vocabularies, because it is written in OWL[6], few logical axioms can be provided beyond the simple taxonomy of concepts. All geological periods are OWL individuals and properties on these instances are defined by 'annotation properties'. Annotation properties can not be used in property axioms. Thus, in OWL one cannot even define subproperties or domain/range constraints for annotation properties. The object of an annotation property must be either a data literal, a URI reference, or an individual.[7] As we will see in Section 4, this imposes a strong limitation in the modeling of the required constraints.

In the GTS ontology, *age, epoch, sub-period, period, era, eon*, and *super-eon* are sub-classes of `GeochronologicEra` (abbreviated as GE). However, there is no formally defined hierarchy between these concepts. Instead, greater emphasis is placed on the boundaries of the periods and, many times, only the approximate duration of the period is given in the chart. It is important to note that geologists qualify the units as "early", "mid", and "late" when referring to time, and "lower", "middle", and "upper" when referring to the corresponding rocks. For example, the lower Jurassic Series in chronostratigraphy corresponds to the early Jurassic Epoch in geochronology. The adjectives are capitalized when the subdivision is formally recognized, and lower case when not; thus "early Miocene" but "Early Jurassic".

While the commission was created exactly to unify and organize the classification of both strata and geochronological periods, it appears that the work is not finished and subject to disagreement. In [Cohen et al. 2013] the authors says "[...] disagreement often arises, because type sections that are favoured for historical reasons may be abandoned, previously established boundary levels may be greatly changed, and in some instances historical units are replaced by different new ones." Thus while the ontology might look very much a finished product, it seems that its contents are still subject to debate. Another evidence is that between 2012 and 2018 there were eleven different versions of the International

[4] `http://resource.geosciml.org/ontology/timescale/gts.html`

[5] `http://resource.geosciml.org/ontology/timescale/thors.html`

[6] The URI and namespaces are the standard instruments to vocabulary reuse in Semantic Web technologies.

[7] `https://www.w3.org/TR/owl-ref/`

Chronostratigraphic Chart.

The boundaries between periods used to be annotated using the THORS ontology, which is used to define the hierarchy between instances of GE. Fragments of the `ISO19108:2002` standard (Geographic information – temporal schema) are also used to specify the temporal position of geochronologic boundaries. [8]. In the more recent versions, THORS ontology properties are mapped to W3C OWL-Time properties. The time interval of a GE is given in terms of its boundaries to other GEs via `time:hasBegin` and `time:hasEnd`. Each boundary is an instance of `gts:GeochronologicBoundary` and it is temporally located via `time:inTemporalPosition` which specifies a `time:numericPosition` with a value, frame (e.g., "Ma"), and a numeric uncertainty when necessary. Nevertheless, the approximate numeric ages provided in the ICS Chart with the ( ) mark were not modeled in the ontology. The boundary modeling should be sufficient for representing the hierarchical relationship between GEs, but ISC ontology further defines a explicit set inclusion relationship between GEs via the `thors:member` property. Also, SKOS is also used to represent inclusion via `skos:narrower`, `skos:broader` along with theirs transitive versions, `skos:narrowerTransitive` and `skos:broaderTransitive`.

4. Expanding SUMO with Geochronological Eras

The Suggested Upper Merged Ontology (SUMO) [Niles and Pease 2001] is a formal ontology written in a higher order logic.[9] It is being used for research and applications in search, linguistics and reasoning. It consists of an upper level ontology, a mid-level and dozens of domains ontologies. Together they form roughly 20,000 terms and 80,000 human-authored logical statements. SUMO is the only formal ontology that has been mapped to all of the WordNet lexicon which provides a strong basis for natural language processing applications [Niles and Pease 2003]. There is an associated open source toolset for development, debugging and inference on the ontology [Pease and Benzmüller 2013].

SUMO contains most of the content we need for our application, including definitions for time points and intervals and relations between intervals (adapted from [Allen 1984]). For modeling the geochronological times, we have used two main SUMO classes `TimeInterval` and `TimePoint` and the functions and predicates associated to them.

While a tutorial on the SUO-KIF language is beyond scope here, the interested reader is referred to [Pease 2011]. In brief, the syntax is valid Lisp S-expressions,[10] a prefix notation in which predicates are followed by one or more arguments. Variables are denoted by an initial question mark.

Figure 2 presents the definition of the `GeochronologicTime` class and one of its sub-classes, the `GeochronologicSuperEon` class.[11] The remain sub-

[8] `https://www.iso.org/standard/26013.html`

[9] `http://www.ontologyportal.org`

[10] `https://en.wikipedia.org/wiki/S-expression`

[11] The current full version of Geochronologic Time as an extension of SUMO is found at `https:`

classes of GeochronologicTime are defined in a similar fashion. Note that all defined classes are sub-classes of the SUMO TimeInterval class, inheriting all its properties. Following the definition of the classes we have two important axioms that guarantee the consistency of the model, none of them encoded in the formalization of ISC presented in Section 3. The first axiom says that no two instances of GeochronologicTime in the same rank can overlap. That is, no two Epoch (Era, Eon, Period etc) can overlap temporally. The second axiom enforces the hierarchical system of time intervals. It says that an Age must occur during an Epoch. The remaining axioms for the other classes are similar.

```
(partition GeochronologicTime GeochronologicAge GeochronologicEpoch
           GeochronologicSubPeriod GeochronologicPeriod
           GeochronologicEra
           GeochronologicEon GeochronologicSuperEon)

(subclass GeochronologicTime TimeInterval)
(subclass GeochronologicSuperEon GeochronologicTime)
(termFormat EnglishLanguage GeochronologicSuperEon "supereon")
...

(=>
  (and
    (instance ?X GeochronologicTime)
    (instance ?Y GeochronologicTime)
    (instance ?X ?Class)
    (instance ?Y ?Class)
    (not (equal ?X ?Y))
    (subclass ?class GeochronologicTime))
  (not
    (overlapsTemporally ?X ?Y)))

(=>
  (instance ?X GeochronologicAge)
  (exists (?Y)
    (and
      (instance ?Y GeochronologicEpoch)
      (or
        (starts ?X ?Y)
        (during ?X ?Y)
        (finishes ?X ?Y)))))
...
```

Figure 2. GeochronologicTime classes

Next, in Figure 3, we define the time boundaries between geochronological times. Following the International Commission on Stratigraphy convention, we defined the class GeochronologicBase sub-class of the SUMO TimePoint class for representing a boundary between periods. The GeochronologicPresent constant represents the beginning of the year 1950, taken as the 'current time' by ISC [Cox and Richard 2005]. The function MillionYearsAgoFn basically defines the time unit 'Millions of year ago' (Ma). The boundaries between periods can be precisely or approximately defined. In the case of uncertainty, boundaries can be in a range (e.g. 182.7 ± 0.7) or approximations (e.g. 500.5). To represent all these cases we defined three predicates maBoundary, maApproxPoint and maPoint and associated GeochronologicBase instances and numbers.

//github.com/ontologyportal/sumo/blob/master/GeochronologicTimes.kif

```
(subclass GeochronologicBase TimePoint)

(instance GeochronologicPresent (BeginFn (YearFn 1950)))

(instance MillionYearsAgoFn UnaryFunction)
(domain MillionYearsAgoFn 1 Number)
(range MillionYearsAgoFn 1 TimePoint)

(equal (MillionYearsAgoFn ?X)
       (BeginFn (YearFn (AdditionFn 1950 (MultiplicationFn ?X -1000000)))))

(instance maBoundary TernaryPredicate)
(domain maBoundary 1 GeochronologicBase)
(domain maBoundary 2 RealNumber)
(domain maBoundary 3 RealNumber)

(=>
  (maBoundary ?Base ?X ?Y)
  (temporallyBetween
    (MillionYearsAgoFn (AdditionFn ?X ?Y))
    ?Base
    (MillionYearsAgoFn (SubtractionFn ?X ?Y))))

(instance maApproxPoint BinaryPredicate)
(domain maApproxPoint 1 GeochronologicBase)
(domain maApproxPoint 2 RealNumber)

(=>
  (maApproxPoint ?Base ?X)
  (exists (?Y)
    (and
      (approximateValue ?X ?Y)
      (equal ?Base (MillionYearsAgoFn ?Y)))))

(instance maPoint BinaryPredicate)
(domain maPoint 1 GeochronologicBase)
(domain maPoint 2 RealNumber)

(=>
  (maPoint ?Base ?X)
  (equal ?Base (MillionYearsAgoFn ?X)))
```

Figure 3. GeochronologicTime boundaries

We must emphasize that all predicates used in the previous code fragments, such as `overlapsTemporally`, `during`, `temporallyBetween` etc., are formally defined in SUMO.[12] They are not merely symbols as in the OWL Ontology presented in Section 3. Given all the above definitions, we can finally present in Figure 4 the SUMO encoding for the fragment of ICS Chart presented in Figure 1.

```
(instance Maastrichtian GeochronologicAge)
(termFormat EnglishLanguage Maastrichtian "Maastrichtian")
(termFormat PortugueseLanguage Maastrichtian "Maestrichtiano")
(meetsTemporally Campanian Maastrichtian)
(meetsTemporally Maastrichtian Danian)
(finishes Maastrichtian LateCretaceous)
(equal (BeginFn Maastrichtian) BaseMaastrichtian)
(equal (EndFn Maastrichtian) BaseCenozoic)

(instance Danian GeochronologicAge)
(termFormat EnglishLanguage Danian "Danian")
(termFormat PortugueseLanguage Danian "Daniano")
(equal (BeginFn Danian) BaseCenozoic)
(equal (EndFn Danian) BaseSelandian)

(instance BaseMaastrichtian GeochronologicBase)
(MaBoundary BaseMaastrichtian 72.1 0.2)

(instance BaseCenozoic GeochronologicBase)
(MaPoint BaseCenozoic 66.0)
```

Figure 4. The SUMO encoding of Maastrichtian Age, the SUMO version of the ISC Ontology fragment from Figure 1.

It is important to note that Figure 1 presents only a small fragment of the axioms added to SUMO. We have expanded SUMO with all the 175 names of geological periods presented in the current version of the International Chronostratigraphic Chart.

Given the definitions above, we can employ the SUMO to TFF0 language translation [Pease 2019] available in SigmaKEE [Pease and Schulz 2014], with Vampire (or another prover that implements TFF0) to query whether, for example, if 125 Ma is earlier than 113 Ma (as shown in Figure 5) or if all the 175 geological periods comply with our axioms. Note that in the proof shown here, the type definitions are removed and the proof only shows the axioms from the portion of SUMO needed for the proof. The TFF0 version of SUMO is produced automatically by the Sigma system, and the relevant axioms among the tens of thousands in SUMO are found automatically by Vampire 4.2.2. Axioms marked "axiom" are those from the human-authored SUMO. Axioms marked "plain" are those derived automatically by Vampire. This is a resolution proof, or proof by contradiction, so a successful conclusion is a proof of `$false`. The proof has been simplified to remove trivial steps and allow it to fit on one page.

5. Conclusion and Future Work

To set the foundations for an application that could reason over geological time, handle equally "Maastrichtian Age" and numeric expressions referring to the in-

[12]The definitions can be inspected at `http://ontologyportal.org`.

```
tff(f18028,axiom,(! [X0 : $int,X1 : $int,X2,X3] : (($less(X0,X1) &
  equal(X3,s__BeginFn(s__YearFn__1InFn(X1))) &
  equal(X2,s__BeginFn(s__YearFn__1InFn(X0)))) => s__before(X2,X3)))).
tff(f16133,axiom,(! [X0 : $real] : equal(s__MillionYearsAgoFn__1ReFn(X0),
  s__BeginFn(s__YearFn__1InFn(s__FloorFn__0In1ReFn(
    $sum(1950.0,$product(X0,-1000000.0)))))))).
tff(f16080,negated_conjecture,(
  ~s__before(s__MillionYearsAgoFn__1ReFn(125.0),
             s__MillionYearsAgoFn__1ReFn(113.0)))).
tff(f16090,axiom,(
  ! [X0 : $real] : s__FloorFn__0In1ReFn(X0) = $to_int(X0))).
tff(f21055,plain,(
  ! [X0 : $int,X1 : $int,X2,X3] : (s__before(X2,X3) | (~$less(X0,X1) |
  ~equal(X3,s__BeginFn(s__YearFn__1InFn(X1))) |
  ~equal(X2,s__BeginFn(s__YearFn__1InFn(X0)))))),ennf_trans,[f18028]).
tff(f22979,plain,(
  ~s__before(s__MillionYearsAgoFn__1ReFn(125.0),
             s__MillionYearsAgoFn__1ReFn(113.0))),cnf_trans,[f16080]).
tff(f36673,plain,(( ! [X0:$real] : (s__FloorFn__0In1ReFn(X0) = $to_int(X0)) )),
  cnf_trans,[],[f16090]).
tff(f36716,plain,(
  ( ! [X0:$real] : (equal(s__MillionYearsAgoFn__1ReFn(X0),
    s__BeginFn(s__YearFn__1InFn(s__FloorFn__0In1ReFn(
      $sum(1950.0,$product(X0,-1000000.0)))))),cnf_trans,[f16133])))).
tff(f40282,plain,(
  ( ! [X2,X0:$int,X3,X1:$int] : (s__before(X2,X3) | ~$less(X0,X1) |
  ~equal(X3,s__BeginFn(s__YearFn__1InFn(X1))) |
  ~equal(X2,s__BeginFn(s__YearFn__1InFn(X0)))) )),cnf_trans,[f21055]).
tff(f40348,plain,(
  ( ! [X0:$real] : (equal(s__MillionYearsAgoFn__1ReFn(X0),
    s__BeginFn(s__YearFn__1InFn($to_int(
      $sum(1950.0,$product(X0,-1000000.0))))))) )),
  definition_unfolding,[f36716,f36673]).
tff(f40413,plain,(
  ( ! [X4:$int,X5:$int] : (~equal(s__MillionYearsAgoFn__1ReFn(113.0),
    s__BeginFn(s__YearFn__1InFn(X5))) | ~$less(X4,X5) |
    ~equal(s__MillionYearsAgoFn__1ReFn(125.0),
           s__BeginFn(s__YearFn__1InFn(X4)))) )),
  resolution,[f22979,f40282]).
tff(f40594,plain,(
  ( ! [X0:$int] : (~$less(X0,$to_int(
    $sum(1950.0,$product(113.0,-1000000.0)))) |
    ~equal(s__MillionYearsAgoFn__1ReFn(125.0),
           s__BeginFn(s__YearFn__1InFn(X0)))) )),
  resolution,[f40413,f40348]).
tff(f40664,plain,(( ! [X0:$int] : (~equal(s__MillionYearsAgoFn__1ReFn(125.0),
          s__BeginFn(s__YearFn__1InFn(X0))) | ~$less(X0,-112998050)) )),
  evaluation,[f40594]).
tff(f40665,plain,(
  ~$less($to_int($sum(1950.0,$product(125.0,-1000000.0))),-112998050)),
  resolution,[f40664,f40348]).
tff(f40734,plain,(~$less($to_int(-124998050.0),-112998050)),
  evaluation,[f40665]).
tff(f40735,plain,($false),evaluation,[f40734]).
% Time elapsed: 0.119 s
```

Figure 5. A Simplified Proof in the TFF0 Version of SUMO with the Vampire Prover.

terval from 72.1 to 66 Million years ago, and represent complex statements involving time in the Oil&Gas domain, we chose to extend SUMO based on the International Chronostratigraphic Chart and the ISC ontology.

Considering geological time is sub-divided in intricate ways and its modeling is a work in progress, we believe this work can contribute to updates and improvements of the ISC ontology. With our SUMO extension we were able to clarify some points in the most recent published version of ISC Ontology such as Capitanian Age and Upper Mississippian Sub Period inconsistent endings and the missing information about the approximate numeric ages. It also provides a formal specification of constraints that can be employed in first order logical reasoning. Undoubtedly, the presented SUMO encoding of geological time opens the possibility of a broader effort on the formalization of other important domain specific information artifacts, such as a chronostratigraphic chart of a given area.

As future work, we still need to encode in SUMO the stratotype or type sections.[13] Stratotypes are physical locations or outcrop of a particular reference exposure of a stratigraphic sequence or stratigraphic boundary; they are represented in the ISC ontology. Next, we aim to implement some concrete use cases for the work presented here. It will probably involve the use of some additional facts, extracted from texts, that combined with the axioms presented in this article will turn possible the answer to questions formulated by technical users. It is worth to remember that this article is part of a long-term project for 'deep' processing technical documents from the Oil&Gas domain for extracting concepts, facts and answering user queries.

References

Allen, J. (1984). Towards a general theory of action and time. *Artificial Intelligence*, 23:123–154.

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Commun. ACM*, 26(11):832–843.

Alves, T. M., Fetter, M., Lima, C., Cartwright, J. A., Cosgrove, J., Gangá, A., Queiroz, C. L., and Strugale, M. (2017). An incomplete correlation between pre-salt topography, top reservoir erosion, and salt deformation in deep-water santos basin (se brazil). *Marine and Petroleum Geology*, 79:300 – 320.

Antoniak, M., Dalgliesh, J., Verkruyse, M., and Lo, J. (2016). Natural language processing techniques on oil and gas drilling data. In *Intelligent Energy International Conference*, pages 1–6.

Cohen, K., Finney, S., Gibbard, P., and Fan, J. (2013). The ICS International Chronostratigraphic Chart. *Episodes*, 36(3):199–204.

Cox, S. J. D. and Richard, S. M. (2005). A formal model for the geologic time scale and global stratotype section and point, compatible with geospatial information transfer standards. *Geosphere*, 1(3):119–19.

[13] `https://en.wikipedia.org/wiki/Stratotype`

Cox, S. J. D. and Richard, S. M. (2014). A geologic timescale ontology and service. *Earth Science Informatics*, 8(1):5–19.

Fellbaum, C., editor (1998). *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press.

Goranko, V. and Galton, A. (2015). Temporal logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2015 edition.

Hobbs, J. R. and Pan, F. (2004). An Ontology of Time for the Semantic Web. *ACM Transactions on Asian Language Information Processing*, 3(1):66–85.

Isaac, A. and Summers, E. (2008). Skos primer. Technical report, W3C. latest version available at http://www.w3.org/TR/skos-primer.

Kovács, L. and Voronkov, A. (2013). First-order theorem proving and vampire. In *Proceedings of the 25th International Conference on Computer Aided Verification*, volume 8044 of *CAV 2013*, pages 1–35, New York, NY, USA. Springer-Verlag New York, Inc.

Muniz, H., Chalub, F., Rademaker, A., and de Paiva, V. (2018). Extending wordnet to geological times. In *Global Wordnet Conference 2018*, Singapore.

Niles, I. and Pease, A. (2001). Toward a Standard Upper Ontology. In Welty, C. and Smith, B., editors, *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems*. FOIS-2001.

Niles, I. and Pease, A. (2003). Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In *Proceedings of the IEEE International Conference on Information and Knowledge Engineering*, pages 412–416.

Palkowsky, B. (2005). A New Approach to Information Discovery - Geography Really Does Matter. In *SPE Annual Technical Conference*, pages 9–12, Dallas.

Pease, A. (2009). Standard upper ontology knowledge interchange format. `https://github.com/ontologyportal/sigmakee`.

Pease, A. (2011). *Ontology: A Practical Guide*. Articulate Software Press, Angwin, CA.

Pease, A. (2019). Arithmetic and inference in a large theory (extended abstract). In *4th Conference on Artificial Intelligence and Theorem Proving*.

Pease, A. and Benzmüller, C. (2013). Sigma: An Integrated Development Environment for Logical Theories. *AI Communications*, 26:9–97.

Pease, A. and Schulz, S. (2014). Knowledge Engineering for Large Ontologies with Sigma KEE 3.0. In *The International Joint Conference on Automated Reasoning*.

Prior, A. (1962). Tense logic and the continuity of time. *Studia Logica*, 13:133–148.

Rademaker, A. (2018). Challenges for information extraction in the oil and gas domain. In Carbonera, J. L., Guizzardi, G., Fiorini, S. R., and Abel, M., editors, *Proceedings of the XI Seminar on Ontology Research in Brazil (ONTOBRAS)*, São Paulo, Brazil.

Remane, J., Bassett, M. G., Cowie, J. W., Gohrbandt, K. H., Lane, H. R., Michelsen, O., and Naiwen, W. (1996). Revised guidelines for the establishment of global chronostratigraphic standards by the international commission on stratigraphy (ics). *Episodes*, 19(3):77–81.

Schoen, E., Smith, R., and Boden, J. (2018). AI Supports Information Discovery and Analysis in an SPE Research Portal. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers.

Steen, A. and Benzmüller, C. (2018). The higher-order prover Leo-III. In Galmiche, D., Schulz, S., and Sebastiani, R., editors, *Automated Reasoning. IJCAR 2018*, volume 10900 of *LNCS*, pages 108–116. Springer, Cham. Url (preprint): https://tinyurl.com/y7wg5w5w.

Sutcliffe, G., Schulz, S., Claessen, K., and Baumgartner, P. (2012). The TPTP Typed First-order Form with Arithmetic. In *International Conference on Logic for Programming Artificial Intelligence and Reasoning (LPAR 2012)*, pages 406–419.