A novel approach for extracting well-founded ontology views

Jose Lozano and **Joel Carbonera** and **Mara Abel**
Institute of Informatics
Universidade Federal do Rio Grande do Sul
Porto Alegre Brazil

Abstract

When the size of an ontology increases, it becomes hard to be managed. Ontology view extraction is an approach that can be used for overcoming the challenges that arise in this scenario. In this context, an ontology view is a subset of an ontology tailored to a specific set of user requirements. Well-founded ontology views were recently proposed as ontology views that follow well-founded ontological principles, which ensures some desirable ontological properties. In this paper, we propose a novel approach for extracting well-founded ontology views, which is more flexible than the previous approach. We also present a method for evaluating the quality of approaches for extracting ontology views. We apply this method for demonstrating that our novel approach produces ontology views that are more accurate than those produced by the previous approach. We illustrate our approaches using a domain ontology for Petrography.

Introduction

Ontologies tend to evolve over time by incorporating new knowledge. The resulting ontology can lead to a scenario of *information overload*, where the information exceeds the cognitive capability of the users. Ontology views have been adopted as a solution for overcoming this scenario, since they are extracted from a base ontology according to specific user criteria, and provide only the knowledge that is relevant for a given task at hand.

The literature provides some approaches for extracting ontology views (Noy and Musen 2003; Bhatt et al. 2004; Lozano et al. 2014). Particularly, in (Lozano et al. 2014), the authors propose the notion of *well-founded ontology view* (WFOV), which is an ontology view that preserves some important *ontological meta-properties* (such as *identity* and *existential dependence*). The authors also define a set of *conservation principles* and apply them for guiding a sub-ontology extraction algorithm.

In this paper, we propose a new approach for extracting WFOVs, which modifies the basic approach defined in (Lozano et al. 2014). Our novel approach eliminates a source of information overload from the basic approach and provides more flexibility, since it allows the user to specify how some aspects of the ontology are considered during the extraction process. We also carried out an experiment for demonstrating that our approach produces WFOVs that are smaller and that fit better to their target conceptualizations than the WFOVs extracted by the original approach (Lozano et al. 2014). This experiment was based on a *data-driven* method for evaluating approaches for ontology view extraction. This method is based on comparisons of the *f-measures* of different ontology views, considering sets of terms extracted from the scientific literature related to different communities or tasks.

In Section , we provide an overview of the main approaches available in the literature for extracting portions of ontologies. In Section , we present a basic definition of the notion of well-founded ontology view and describe the basic approach for extracting WFOVs. Section presents our approach for extracting WFOVs. Section describes the method that we used for evaluating our approach. Section describes the application of the different approaches for extracting WFOVs in a case scenario with their corresponding evaluations. Finally, Section presents our conclusions.

Related Works

In general, the literature provides two main approaches that can be used for extracting manageable portions of ontologies. The extraction of *ontology modules* (Doran, Tamma, and Iannone 2007; d'Aquin, Sabou, and Motta 2006; Seidenberg and Rector 2006) fragments a given base ontology into a set of *smaller*, *non-overlapping* and possibly *interconnected* parts, or *modules*. The alternative approach, is the extraction of *ontology views* (Noy and Musen 2003; Bhatt et al. 2004), where smaller (and possibly overlapping) subsets of the base ontology are extracted according to the *user requirements*. Since they are tailored to specific tasks or interests, ontology views provide to the agent (users or computer applications) only the knowledge that is relevant for reaching some goal.

Some of these approaches (Seidenberg and Rector 2006; d'Aquin, Sabou, and Motta 2006; Noy and Musen 2003) are dependent on some representation language (such as OWL), while others (Doran, Tamma, and Iannone 2007; Bhatt et al. 2004), language-independent, adopt an abstract ontology representation that is based on graphs. Besides that, most of the approaches extract modules or views starting from some target concepts and include in the subset (module or view) only the ontology elements (concepts, relations and properties) that are directly related to the concepts that are already included in the subset.

Algorithm 1 The basic approach for WFOV extraction.

```
Require: Well-Founded Ontology
  procedure SEL(O_b, tConcepts, tRelations, S_o)
    S_o.C ← S_o.C ∪ tConcepts
    S_o.R ← S_o.R ∪ tRelations
    newC ← ∅
    newR ← ∅
    for all c ∈ tConcepts do
      conservesTAX(O_b, c, newC, newR)
      conservesQUA(O_b, c, newC, newR)
      conservesIP(O_b, c, newC, newR)
      conservesED(O_b, c, newC, newR)
      conservesRD(O_b, c, newC, newR)
      conservesFR(O_b, c, newC, newR)
      conservesPR(O_b, c, newC, newR)
      newC ← newC − S_o.C
      newR ← newR − S_o.R
    end for
    if newC ≠ ∅ then
      SEL(O_b, newC, newR, S_o)
    else
      if newR ≠ ∅ then
        S_o.R ← S_o.R ∪ newR
      end if
    end if
  end procedure
```

In (Lozano et al. 2014), the authors propose using ontological meta-properties (such as identity, rigidity and existential dependency) for guiding the extraction of ontology views. Their approach has the advantage of including in the views the ontology elements (concepts, relations and properties) that need to be included in the view due to their *ontological status*. For example, if the concept A is included in an ontology view and instances of A are *existentially dependent* on instances of a concept B, B should also be included in the view. This dimension of analysis is not considered by the other approaches discussed in this section. Since our work proposes an improvement of the approach proposed by (Lozano et al. 2014), in the Section we shall present this approach in more details.

In Table 1, we present a comparison of the approaches discussed in this section. The approaches are identified as: 1 (d'Aquin, Sabou, and Motta 2006), 2 (Doran, Tamma, and Iannone 2007), 3 (Noy and Musen 2009), 4 (Seidenberg and Rector 2006), 5 (Bhatt et al. 2004) and 6 (Lozano et al. 2014).

Table 1: Comparison of sub-ontology extraction approaches.

Approach	1	2	3	4	5	6
Language-independent	No	Yes	No	No	Yes	Yes
Use of Meta-properties	No	No	No	No	No	Yes
Context	Module	Module	View	Module	View	View

Well-founded Ontology Views

In this Section, we present the approach proposed by (Lozano et al. 2014), for extracting *well-founded ontology views* (WFOV). Since this approach relies on a set of ontological meta-properties, firstly we shall discuss them. After, we present the characterization of a WFOV. Finally, we present the basic approach for extracting WFOVs, proposed by the authors.

Ontological Meta-Properties

The approach proposed by (Lozano et al. 2014) uses the formal characterization of the ontological meta-properties provided by the *Unified Foundational Ontology* (UFO) (Guizzardi 2005). This ontology provides a set of categories of universals, which are characterized according to a set of meta-properties. The categories of universals can be viewed as *meta-types*, since they are *types of types*. Thus, they can be used for classifying classes in specific domain ontologies. When some class C is classified by some meta-type MT, this means that C has the meta-properties that characterize MT, and this entails some formal consequences, according to the UFO axiomatization. The UFO has been used for supporting the development of domain ontologies (Carbonera et al. 2011; 2013; Carbonera, Abel, and Scherer 2015; Abel, Perrin, and Carbonera 2015) in a well-founded basis. Here we will present the main meta-properties and meta-types provided by UFO and that are used by the approach of (Lozano et al. 2014). A detailed account of UFO can be found in (Guizzardi 2005).

One of the main categories of universals provided by UFO is *Substantial Universal*, whose instances are individuals that, in general, are *existentially independent* of all other individuals. Some of its instances can be *existentially dependent* when they are considered *inseparable parts* of their hosts. *Sortal Universals* are substantial universals that provide or carry some *principle of identity* (PI) for their instances. In this context, a PI is the principle that supports the judgment whether two instances of the universal are the same.

Another important ontological meta-property used by UFO is the *rigidity*. A certain universal is rigid when its extension (set of all particulars) is the same in all possible worlds. That is, an instance of a rigid universal cannot cease to be an instance of it without ceasing to exist. For example, *Person* can be viewed as a *rigid universal*, since persons cannot cease to be persons without ceasing to exist; meanwhile all instances of *Student* (which is an *anti-rigid universal*) can still exist (as persons) if they cease to be students.

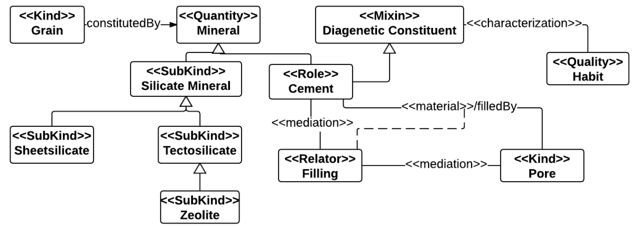
Within the sortal universals, UFO includes three distinct types of *substance sortals*, which are *rigid sortals that provide their own principle of identity*: *Kind*, which represents functional complexes (Person, Dog, Chair, etc); *Collective* (Swarm, Forest, etc), which represents collectives; and *Quantity*, which represents objectified portions of matter (Wine, Water, Gold, etc). Besides that, *Subkind* is a *rigid sortal* that does not provide its own PI, but *carries* a principle of identity that is supplied by a given substance sortal.

UFO also defines two *anti-rigid sortals*: *Roles* and *Phases*. Phases are universals that constitute possible stages in the history of a substance sortal. Phases are *relationally independent*, since they depend solely on intrinsic properties. For example, *Baby*, *Toddler*, *Kid*, *Teenager* and *Adult* are considered phases of *Human*. On the other hand, Roles are *relationally dependent*, since they depend on extrinsic (relational) properties. This is the case, for example, when we say that for an instance of *person* to be considered a *Student*, she must be *enrolled at* an *educational institution*.

Other substantial universals do not have the properties of sortals; they are *dispersive universals*. This is the case, for example, of *Categories*, which are *rigid* universals that do not provide or carry a PI for their instances. Categories represent essential properties that are common to all instances of many disjoint universals that provide distinct PIs. *Rational agent* is an example of Category, since it abstracts an essential property (namely, the rationality) of instances of *Person* and *Artificial Agent*, which are disjoint universals, with distinct PIs. *Role Mixins*, on the other hand, are *anti-rigid* universals that *do not provide and do not carry* a PI for their instances. They can be viewed as generalizations of roles of different substance sortals. For example, *Customer* is a role mixin that generalizes *Personal Customer*, which is a role of *Person*; and *Corporate Customer*, which is a role of *Organization*. Finally, *Mixins* are universals that *do not provide and do not carry* a PI for their instances and that are *semi-rigid*; that is, they have some instances that are *necessarily* their instances, but they also have some instances that are only *contingently* their instances. They usually generalize rigid and anti-rigid universals. For example, *Seatable Object* is a mixin that generalizes *Chair*, which is a rigid universal; and *Solid Crate*, which is an anti-rigid universal (actually, it is a phase of a *Crate*, which can also be a *Broken Crate*).

On the other hand, *Moment Universals* are Universals whose instances are *existentially dependent* individuals that *inhere* in other individuals. Some moment universals depend existentially on a single entity. This is the case of *Quality Universals* and *Modes*. *Quality Universals* represents the *properties* in the conceptual models. A Quality Universal characterizes other Universals and is related to *Quality Structures*, that is, a structure that represents a set of all values that a quality can assume. Thus, considering the property *Color* as a Quality Universal, a given instance of *Car* could be characterized by an instance of *Color*, which is associated with a value (called *quale*) in the *ColorStructure*, which represents all the possible values that the property *Color* can assume. On the other hand, *Modes* are universals whose instances are *existentially dependent* individuals, and that are not associated to *Quality Structures*. Examples of modes are *Skill*, *Belief*, *Headache*, etc. Both *Quality universals* and *Modes* are related to the entities that they characterize through a relation of *characterization*. Besides that, *Relators* are moments that depend existentially on two or more entities. Examples of relators are *Enrollment*, *Contract*, etc. Relators are related to entities that it relates through a relation of *mediation*. The relators also represent the relational dependency of roles and role mixins. Due to this, roles and role mixins must be related to some relator, through a relation of *mediation*.

Figure 1: A WFOV for the Diagenesis community, extracted from a domain ontology for Petrography



UFO proposes four types of parthood relations: *componentOf*, *memberOf*, *subCollectionOf* and *subQuantityOf*. Each parthood relation can only be established between individuals of specific UFO meta-types, respecting some ontological constraints embedded in UFO. These relations can be characterized by five *meronymic meta-properties* that indicate: *essential part*, *inseparable part*, *immutable part*, *immutable whole* and *shareable part*.

As important as the characterization of the meta-properties and meta-types are, UFO also provides some postulates that a model should follow:

- **Postulate 1:** Every individual in a conceptual model of the domain must be an instance of a *sortal*.
- **Postulate 2:** An individual represented in a conceptual model of the domain must instantiate *exactly one* ultimate *Substance Sortal* (*kind*, *quantity* or *collective*).
- **Postulate 3:** A rigid universal cannot specialize (restrict) an anti-rigid one.
- **Postulate 4:** A dispersive universal cannot specialize a Sortal.

Furthermore, it is important to notice that every sortal that does not provide its own principle of identity (Role, Phase and SubKind) must be subsumed by exactly one concept that provides its own identity (one of the Substance Sortals).

Basic Definitions

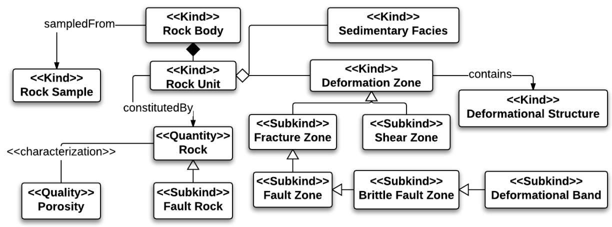
The notion of ontologically well-founded ontology view is defined considering certain *principles of conservation* proposed by (Lozano et al. 2014). These principles were built considering a set of philosophically well-founded ontological meta-properties. In this work, the selected meta-properties were obtained from the UFO ontology.

In order to illustrate the proposed conservation principles, we present portions of two WFOVs generated from a base ontology for the domain of Petrography (a field of Geology). These WFOVs were generated for meeting the interests of two different communities of users within the domain of Petrography: *Diagenesis* (Figure 1) and *Microstructural Analysis* (Figure 2).

The principles of conservation proposed by (Lozano et al. 2014) are:

Conservation of identity: If a view v includes a concept c that does not provide its own principle of identity, then v should also include all the supertypes of c from which

Figure 2: A WFOV for the community of Microstructural analysis, extracted from a domain ontology for Petrography



c inhered its principle of identity, as well as, all the subsumption relations that are held between these concepts. For example, if *zeolite* is included in v, *mineral* should also be included, since *mineral* provides the identity to *zeolite*.

Conservation of the existential dependence: If a concept c_1 is included in the view v, and instances of c_1 are *existentially dependent* on instances of c_2, then it is necessary to include in v also the concept c_2 and the relation held between c_1 and c_2. For example, if *Porosity* is included in v, the concept *Rock* must be included because the *porosity* is existentially dependent of *Rock*.

Conservation of relational dependence: If a concept c_1 is included in the view v, and c_1 is relationally dependent on a relation (materialized through a given *relator*) with the concepts in $\{c_2, ..., c_n\}$, then it is necessary to include in v also: the relator r, all the concepts in $\{c_2, ..., c_n\}$ and all relations that are held between the concepts in $\{c_2, ..., c_n\}$, r and c_1 that are necessary for the conservation of the relational dependence. For example, if the concept is *Cement* is included in v, the concepts *pore* and *filling* should also be included in v, because a *mineral* is considered *cement* when the *mineral* is filling *pore*. Thus, cement is relational dependent of *filling* and *pore*.

Conservation of taxonomy: If a view v includes the concept c_1, it should also include all the concepts that are subsumed by c_1. For example, if concept *Silicate Mineral* is included in v, all the concepts that it subsumes are included in the v.

Conservation of attributes: If a view v includes a concept c_1, every attribute[1] of c_1 must also be included in v. For example, if the concept *Diagenetic Constituent* is included in v, the concept *habit* should also be included, because it is a quality of *Diagenetic Constituent*.

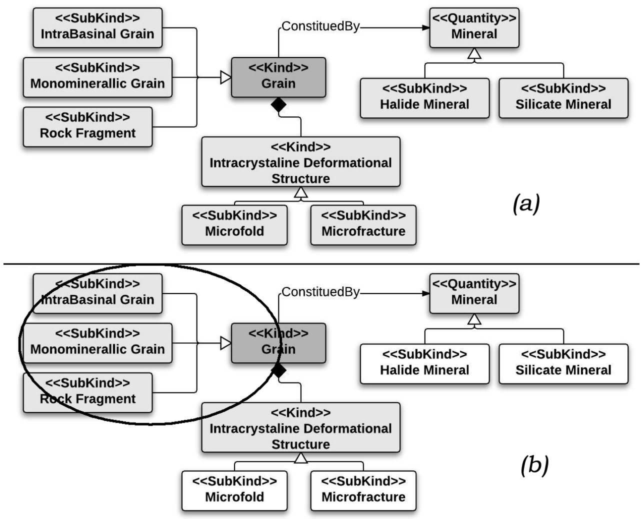
Conservation of formally related concepts: If a view includes a concept c_1, every concept that is related to c_1 in a formal relation is added. For example, if the concept *Rock Unit* is included in v, the concept *Rock* is also included because there is the formal relation *constituted by* between *Rock Unit* and *Rock*.

Conservation of partonomy: If a view includes a concept c_1, all the concepts whose instances are parts of instances of c_1 should be included. For example, if the concept *Rock Unit* is included in v, the concepts *deformation zone* and *sedimentary facies* should also be included.

[1] Adopting the UFO, attributes are considered *Quality Universals*

Figure 3: New version example

(a) If it were applied the approach 1 in the target concept, the concepts colored in light gray are included in the view. (b) If were applied the new version (approach 2)



Basic Approach for extracting WFOV

The basic approach for extracting WFOVs is formalized in the Algorithm 1. It takes as input the following parameters: the ontology base (O_b), a set of user required concepts ($targets$), a set of relations ($relations$), and the resulting extracted sub-ontology (S_o). At the beginning, $relations$ and S_o are empty. The algorithm analyses each concept in $targets$. For each concept, the conservation principles are applied for ensuring that the result will be an ontologically well-founded ontology view. The conservation principles are applied through the following functions: *conservesTAX*, for the *conservation of taxonomy*; *conservesQUA*, for the *conservation of attributes*; *conservesIP*, for *conservation of identity principle*; *conservesED*, for the *conservation of existential dependence*; *conservesRD*, for the *conservation of relational dependence*, *conservesFR*, for the *conservation of formally related concepts*; and *conservesPR*, for the *conservation of partonomy*. In the main loop, these functions accumulate concepts (in $newC$) and relations (in $newR$) that are necessary for ensuring the defined principles for a given concept c in $tConcepts$. More details regarding this approach can be found in (Lozano et al. 2014).

A novel approach for extracting well-founded ontology views

Our new approach aims at reducing the number of concepts included in the views by the basic approach, and covering more precisely the requirements of a task at hand. In an overview, our approach relaxes some criteria adopted by the basic approach and provides more flexibility to the user.

The main difference in our approach regarding the basic approach concerns the application of the *principle of conservation of taxonomy*. While the basic approach includes in the resulting view the taxonomy of every concept that is already included in the view, our novel approach includes only the immediate taxonomy of the original target concepts. This modification was motivated by the fact that, in general, only the taxonomies of the target concepts are useful for the task for which the view was built. Besides that, the inclusion of the taxonomies of every concept in the view leads to a rapid increase in the size of the view. In this way, the inclusion of irrelevant taxonomies can be considered as a source of information overload.

Moreover, our approach also provides more flexibility to the user, by allowing the setting of three parameters. These parameters are the variables *wP* (with Partonomy), *wRT* (only Rigid Taxonomy) and *wFR* (with formal relation). In this way, the ontology engineer can specify if the desired WFOV should include the partonomies of every concept or not; if it should include only the rigid concepts in the taxonomies or if non-rigid concepts should be included as well; and if it should include all the concepts that are related (through formal relations) to concepts already included in the view.

Algorithm 2 Novel algorithm for extracting well-founded ontology views

```
Require: Well-Founded Ontology
  procedure EXTRACTOR(O_b, tConcepts, tRelations, S_o, wP, wRT, wFR)
    S_o.C ← S_o.C ∪ tConcepts
    newC ← ∅
    newR ← ∅
    for all c ∈ tConcepts do
      if wRT then
        conservesTAXR(O_b, c, newC, newR)
      else
        conservesTAX(O_b, c, newC, newR)
      end if
    end for
    newC ← newC ∪ tConcepts
    newR ← newR ∪ tRelations
    selection(O_b, newC, newR, S_o, wP, wFR)
  end procedure
```

Our approach is formalized in the algorithm 2. Firstly, it applies the *conservation of taxonomy* only to the original target concepts. At this point, the parameter *wRT* controls if the taxonomy takes all the concepts or only the rigid ones (Algorithm 4 presents how to recover only rigid concepts in the taxonomy). Then, the algorithm calls the $selection$ algorithm (algorithm 3), which applies the other principles of conservation, according to the parameters. Notice that the algorithm 3 is a variation of the basic algorithm proposed by (Lozano et al. 2014) (presented in subsection), which does not apply the conservation of the taxonomy of every concept in the main loop, and which controls through parameters the application of some principles of conservation. For instance, in Figure 3 (a), if we consider *Grain* as the target concept, the basic approach would include in the resulting view all the concepts in gray. On the other hand, our novel approach applies the conservation of taxonomy only to the target concept (*Grain*), including the concepts surrounded by a circle in Figure 3 (b). As a consequence, the sub-ontology will not include the taxonomy of *Mineral* and *Intracrystalline Deformational Structure*, depicted in Figure 3 (b) (in white).

Algorithm 3 Parameterizable algorithm for selecting ontology elements for the view

```
Require: Well-Founded Ontology
  procedure SELECTION(O_b, tConcepts, tRelations, S_o, wP, wFR)
    S_o.C ← S_o.C ∪ tConcepts
    S_o.R ← S_o.R ∪ tRelations
    newC ← ∅
    newR ← ∅
    for all c ∈ tConcepts do
      conservesQUA(O_b, c, newC, newR)
      conservesIP(O_b, c, newC, newR)
      conservesED(O_b, c, newC, newR)
      conservesRD(O_b, c, newC, newR)
      if wFR then
        conservesFR(O_b, c, newC, newR)
      end if
      if wP then
        conservesPR(O_b, c, newC, newR)
      end if
      newC ← newC − S_o.C
      newR ← newR − S_o.R
    end for
    if newC ≠ ∅ then
      selection(O_b, newC, newR, S_o, wP, wFR)
    else
      if newR ≠ ∅ then
        S_o.R ← S_o.R ∪ newR
      end if
    end if
  end procedure
```

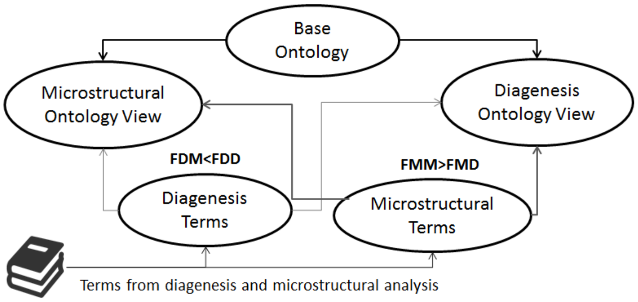
Evaluation Method

We assume that the quality of the ontology view extraction approach can be measured by the degree to which the extracted ontology views *fit* to the required conceptualizations. In our work, we adopt an approach for evaluating this *fitness* in an indirect way.

We assume that the required conceptualization (of a community, of some task) is properly represented, in natural language, in the relevant literature. Thus, our evaluation is based on measuring the correspondences between a given ontology view (built for some specific task or some community) and the set of terms extracted from the relevant literature (related the correspondent task or community for which the ontology view was built). We measure these correspondences through well-known measures used in Information Retrieval: *Precision*, *Recall* and *F-measure* (Powers 2011).

Considering this, we assume that in a useful approach for extracting ontology views, an ontology view generated for a community *A* should have a value of f-measure that is greater than the f-measure value of any well-founded ontology view generated for community *B*, when compared to

Figure 4: Evaluation Method applied in the Petrography domain.



a set of terms extracted from the literature of the community *A*. In other words, a view generated for the community *X* should fit better the conceptualization of the community *X* rather than the ontology view generated for another community.

Considering LT as the set of terms extracted from the literature, the *precision* (P) of the generated ontology view O is given by

$$P(OT, LT) = \frac{|OT \cap LT|}{|OT|} \quad (1)$$

, the *recall* (R) is given by

$$R(OT, LT) = \frac{|OT \cap LT|}{|LT|} \quad (2)$$

, and the *f-measure* (F) is given by

$$F(OT, LT) = 2 * \frac{P(OT, LT) * R(OT, LT)}{P(OT, LT) + R(OT, LT)} \quad (3)$$

where $|S|$ indicates the cardinality of the set S and OT is the set of terms that identify the set of ontology elements (concepts, relations and properties) of the ontology view O.

Evaluation Results

This section describes the application of the evaluation method described in Section for comparing the performance of our approach (A2) for extracting ontology views with the performance of the basic approach proposed by (Lozano et al. 2014). In our evaluation, we compared the f-measure of the ontology view generated by the basic approach A1 with the ontology views generated for each parameter combination of our approach (A2). For performing this comparison, we considered two WFOVs extracted from the domain ontology of *Petrography* proposed by (Lozano 2014): a WFOV for the community of *Diagenesis* and a WFOV for the community of *Microstructural analysis*. These two communities of users employ different sets of concepts from the ontology of *Petrography*. This base ontology of Petrography includes 366 concepts and 387 relations.

Our approach also requires the extraction of sets of terms that are representative for the considered communities. For this step, a domain expert selected six peer-reviewed papers about *Diagenesis*, such as (Worden and Burley 2003), and six papers about *Microstructural analysis*, such as (Haertel and Herwegh 2014). After the extraction of terms was performed, following the sequence of steps defined in (Abel 2001): (i) exclude all common words: prepositions, articles, adverbs and connection verbs and; (ii) mark all geological terms specific to the domain.

Algorithm 4 Conserve Taxonomy only Rigid

procedure CONSERVESTAXR($O_b, c, newC, newR$)
  for all $v \in O_b.C | \exists r = Rel(subsumption, c, v)$ **do**
    if $metaType(v) \in \{\text{SubKind, Collective, Kind, Quantity, Category}\}$ **then**
      $newR \leftarrow newR \cup r$
      $newC \leftarrow newC \cup v$
      $conservesTAXR(O_b, v, newC, newR)$
    end if
  end for
end procedure

This extraction was done manually, for ensuring the quality of the extraction. We excluded the terms that were not exclusive to the communities of *Diagenesis* and *Microstructural analysis*. We also excluded the terms that were common for both communities. The result was two lists of geological terms; one (DT) for the community of *Diagenesis* and other (MT) for the community of *Microstructural analysis*.

The next step consists of generating the two well-founded ontology views (one for each community), from a set of key terms that are representative of the community. These key terms were provided by domain experts. The WFOVs that were considered in this evaluation were extracted using *Detrital Constituent*, *Diagenetic Constituent* and *Pore* for the WFOV of Diagenesis and *Deformational Band*, *Fault*, *Breccia* and *Microfracture* for the community of *Microstructural analysis*.

The result of this step is an ontology view for *Diagenesis* (DO) and ontology view for *Microstructural analysis* (MO). For our proposed approach, we extracted one WFOV for each community, considering each combination of parameters.

The last step is to calculate and compare the f-measures, considering the WFOVs and the sets of selected terms. As depicted in Figure 4, we expect that the f-measure ($FDD = FMeasure(DO, DT)$) between ontology view for *Diagenesis* (*DO*) and the terms for the community of *Diageneis* (*DT*) is greater than the f-measure ($FDM = FMeasure(MO, DT)$) between ontology view for *Microstructural analysis MO* and the terms for the community of *Diageneis DT*. And, in the same way, it is also expected that the f-measure ($FMM = FMeasure(MO, MT)$) between *MO* and *MT* is greater than the f-measure ($FMD = FMeasure(DO, MT)$) between *DO* and *MT*. In the following subsections, we describe the evaluation of these two cases. In *Case 1*, we evaluate the ontology view for *Diagenesis*, comparing it with the ontology view for *Microstructural analysis*, considering the terms for *Diagenesis*. In *Case 2*, we evaluate the ontology view for *Microstructural analysis*, comparing it with the ontology view for *Diagenesis*, based on the microstructural terms.

In Table 2, we present the results of the evaluation process of approaches A1 (basic approach) and A2 (novel approach), in the two considered cases. Notice that the table presents the *ratios* between the considered measures of the two ontology views, evaluated according to a set of terms. Thus, in the row *Case 1*, the table presents, for both approaches A1 and A2, the ratios between the measures (precision, recall and f-measure) of the ontology view for *Diagenesis* and the measures of the ontology view of *Microstructural analysis*, considering the terms of *Diagenesis*. Notice that it is expected that the resulting ratios are greater than 1. In a similar way, the row *Case 2*, the table presents, for both approaches A1 and A2, the ratios between the measures of the ontology view for *Microstructural analysis* and the measures of the ontology view of *Diagenesis*, considering the terms of *Microstructural analysis*.

Evaluation of the Ontology View for Diagenesis

In the *approach A1* (basic approach), the ontology view generated for the *Diagenesis* community obtained a ratio of *f-measure* smaller than 1. This means that the basic approach does not satisfy the expectation. This happens because the approach A1 includes many ontology elements that are not necessary for the community. However, the ratio of *recall* is greater than 1, as expected. This means that the ontology view generated for the *Diagenesis* community contains more relevant terms than the ontology views obtained for the *Microstructural analysis* community.

For the *approach A2*, proposed in this paper, in all cases, the ratio of P, R and F was greater than 1. This occurs because it applies the principle of conservation of taxonomy just once to the target concepts, eliminating taxonomies that are useless for the community in focus. In general, *approach A2* results satisfy our expectations about the generated ontology view for all set of representative terms given in this case study.

In the column *Case 1* of Table 2, it is possible to see that the new approach (A2) achieve better results than the basic approach (A1), for the ontology views of the community of *Diagenesis*. The best results of approach A2 for *Case 1* were achieved with *wRT* and with all the parameters as false.

Evaluation of the Ontology View for Microstructural Analysis

In *Case 2*, the ratio of f-measure achieved by the approach A1 is greater than 1. Also, the precision obtained by this approach achieved the highest value, in comparison with the approach A2, considering all parameter combinations.

The approach A2 also achieved high quality results in *Case 2*. Each combination of parameters of *approach A2* satisfies our expectations about the generated ontology views, considering the set of representative terms given in this case study. However, the precision in all combination of parameters for *Microstructural analysis* ontology view has a ratio smaller than 1. This means that the ontology view generated for the *Microstructural analysis* community contains few terms to cover the terminology used in the literature of *Microstructural analysis*, than the ontology view for *Diagenesis*.

The row *Case2* of Table 2 presents the ratio between the measures of the ontology view of *Microstructural analysis* and the measures of the ontology view of *Diagenesis*. In this row, it can be seen that approach *A2* achieved its best result using the parameter *wP*.

The approach *A2* achieved results of low quality in some settings because, for this domain, the key terms could be related by *formal relations* with other concepts that do not belong to the *Microstructural analysis* ontology view.

Table 2: Evaluation results. In this Talbe, *P*, *R* and *F* mean *Precision*, *Recall* and *F-measure*, respectively

Case	**Measure**	**Approach**								
		A1	**A2**							
			-	wP	wFR	wRT	wP wRT	wRT wFR	wP wFR	wP wRT wFR
Case 1	**P**	0.69	**22.0**	**22.0**	15.67	**22.0**	**22.0**	15.67	15.67	15.67
	R	1.15	3.88	**4.43**	3.20	1.94	2.21	1.68	3.56	1.88
	F	0.97	**9.25**	9.0	7.60	**9.25**	9.0	7.60	7.60	7.60
Case 2	**p**	**2.39**	0.75	0.84	0.72	0.54	0.64	0.52	0.81	0.62
	R	1.46	3.38	3.69	3.15	5.23	**5.62**	4.77	3.46	5.15
	F	1.61	1.56	**1.71**	1.47	1.38	1.59	1.29	1.56	1.44

Conclusion

In this work, we propose a novel approach for extracting well-founded ontology views, by improving the approach proposed in (Lozano et al. 2014). This approach eliminates a source of information overload that is present in the previous approach. Moreover, the proposed approach also provides more flexibility to the user, by allowing the control of important aspects of the process of extracting ontology views.

It is important to notice that, although this work adopts the meta-properties defined by UFO, it can be viewed as a specific implementation of a more general idea. The notion of WFOV is defined according to a set of *principles of conservation* that should be followed by the extraction algorithm. The set of principles of conservation can be changed, by including, excluding or modifying the principles proposed in this work. In this way, the general approach proposed in (Lozano et al. 2014), and extended in this work, can be considered as independent of UFO.

In this work, we also propose a method for evaluating approaches for extracting ontology views. This method uses well-known measures used in information retrieval (*precision*, *recall* and *f-measure*), for evaluating the fitness of the resulting ontology views to the target conceptualization (of a community or task). According to this method, our novel approach outperforms the basic approach (Lozano et al. 2014) for extracting ontology views in most of the considered cases. We hypothesize that this method can inspire methods that can be applied for evaluating ontology modules. This hypothesis should be investigated in future works.

In future works, we also plan to improve the proposed approach by identifying and eliminating other sources of information overload in the resulting ontology views. Besides that, we also intend to investigate if the ontological meta-properties considered in this work can also be applied for guiding the extraction of ontology modules.

References

Abel, M.; Perrin, M.; and Carbonera, J. L. 2015. Ontological analysis for information integration in geomodeling. *Earth Science Informatics* 8(1):21–36.

Abel, M. 2001. *Study of Expertise in Sedimentary Petrography and their importance for Knowledge Engineering*. Ph.D. Dissertation, Federal University of Rio Grande do Sul.

Bhatt, M.; Flahive, A.; Wouters, C.; Rahayu, W.; Taniar, D.; and Dillon, T. 2004. A distributed approach to sub-ontology extraction. In *Proceedings...*, volume 1, 636–641 Vol.1. Advanced Information Networking and Applications.

Carbonera, J. L.; Abel, M.; and Scherer, C. M. 2015. Visual interpretation of events in petroleum exploration: An approach supported by well-founded ontologies. *Expert Systems with Applications* 42:27492763.

Carbonera, J. L.; Abel, M.; Scherer, C. M.; and Bernardes, A. K. 2011. Reasoning over visual knowledge. In *Joint IV Seminar on Ontology Research in Brazil and VI International Workshop on Metamodels, Ontologies and Semantic Technologies*, 49–60.

Carbonera, J. L.; Abel, M.; Scherer, C. M.; and Bernardes, A. K. 2013. Visual interpretation of events in petroleum geology. In *Proceedings of ICTAI 2013*.

d'Aquin, M.; Sabou, M.; and Motta, E. 2006. Modularization: a key for the dynamic selection of relevant knowledge components. *WoMo*.

Doran, P.; Tamma, V.; and Iannone, L. 2007. Ontology module extraction for ontology reuse: An ontology engineering perspective. In *Proceedings CIKM*, 61–70.

Guizzardi, G. 2005. *Ontological Foundations for Structural Conceptual Models*. Ph.D. Dissertation, University of Twente, The Netherlands.

Haertel, M., and Herwegh, M. 2014. Microfabric memory of vein quartz for strain localization in detachment faults: A case study on the simplon fault zone. *Journal of Structural Geology* (0):–.

Lozano, J.; Carbonera, J. L.; Abel, M.; and Pimenta, M. 2014. Ontology view extraction: an approach based on ontological meta-properties. In *Proceedings of ICTAI*.

Lozano, J. 2014. Ontology view: a new sub-ontology extraction method. Master's thesis, Federal University of Rio Grande do Sul.

Noy, N. F., and Musen, M. A. 2003. The prompt suite: interactive tools for ontology merging and mapping. *International Journal of Human-Computer Studies* 59(6):983 – 1024.

Noy, N., and Musen, M. 2009. Traversing ontologies to extract views. In Stuckenschmidt, H.; Parent, C.; and Spaccapietra, S., eds., *Modular Ontologies*, volume 5445 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg. 245–260.

Powers, D. M. 2011. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *International Journal of Machine Learning Technology* 37–63.

Seidenberg, J., and Rector, A. 2006. Web ontology segmentation: Analysis, classification and use. In *Proceedings of the 15th*, WWW '06, 13–22. New York, NY, USA: International Conference on World Wide Web.

Worden, R., and Burley, S. 2003. Sandstone diagenesis: the evolution of sand to stone. *Inernational Association of Sedimentologists*.