Extended version: Categorization Power of Ontologies with Respect to Focus Classes

Vojtěch Svátek, Ondřej Zamazal, and Miroslav Vacura

Department of Information and Knowledge Engineering,
University of Economics, W. Churchill Sq.4, 130 67 Prague 3, Czech Republic,
{svatek|ondrej.zamazal|vacuram}@vse.cz

Abstract. When reusing existing ontologies, preference might be given to those providing extensive subcategorization for the classes deemed important in the new ontology (focus classes). The reused set of categories may not only consist of named classes but also of some compound concept expressions that could be viewed as meaningful categories by human ontologist. We define the general notion of focused ontologistic categorization power; for the sake of tractable experiments we then choose a restricted concept expression language and a map it to syntactic axiom patterns. The occurrence of the patterns has been verified in two ontology collections, and for a sample of pattern instances their ontologistic status has been assessed by different groups of users.

1 Introduction

Reusing parts of existing semantic web ontologies when designing a new one, or when merely proposing the schema for an RDF dataset to be published, is commonly understood as best practice. With the growing number of ontologies on the semantic web it also becomes more likely to find multiple ontologies covering the given topic. However, mere thematic relevance may not be enough: since the target ontology/schema is to be used in a certain application context, it should exhibit features required in this context. For example, if a reasoner is to be applied on the ontology, its expressiveness should not exceed that expected by the reasoner. In this paper we investigate one more structural feature of ontologies to be potentially reused: their *categorization power*, i.e. its suitability for assigning meaningful categories – not necessarily expressed as *named* classes but possibly in the form of *compound concept expressions* – to individual domain objects (instances). Namely, many tasks related to the management of ontologically described data refer to detailed categorization of objects: companies may provide specific offers to different categories of customers, buyers may only be interested in specific categories of products, and the like. Reusing a categorization structure pre-existing in a widespread vocabulary (or one with potential for future widespread, e.g., cataloged in a respected collection such as LOV[1] [9]) may

[1] http://lov.okfn.org/

not only save a part of the design effort but also allow to better interface with other applications, e.g., in federated querying or concerted recommendation. We therefore hypothesize that ontologies providing more subcategories for classes important in the given use case – to be called *focus classes* in our approach – would be a more desirable subject of reuse, as whole or in (relevant) part.

To informally introduce the key concepts of our approach (more formally grounded in Section 2), let us start with a toy ontology O[2] as motivating example:

```
Class: Person
Class: Man       SubClassOf: Person
Class: Woman     SubClassOf: Person
Class: MarriedMan         EquivalentTo: Man and hasSpouse some Thing
Class: ProductivePerson
  EquivalentTo: Person and insuranceCategory some {Enterpreneur,Employed}
Class: Country
ObjectProperty: hasSpouse   Domain: Person    Range: Person
ObjectProperty: bornIn      Domain: Person    Range: Country
ObjectProperty: insuranceCategory     Domain: Person
  Range: {Enterpreneur,Employed,Child,Retired}
DataProperty: zipCode       Domain: Person    Range: string
Individual: UK     Types: Country
Individual: Italy  Types: Country
```

Let us assume we want to build a rich ontology for categorizing persons and need to assess if O is a good reuse candidate. Class `Person` (possibly discovered by lexical search via an ontology search engine) thus becomes our focus class, FC, in O. A simple quantification of the categorization power of O wrt. `Person` could be 4, i.e. the sum of its asserted and inferred subclasses. However, the entities from O can be assembled to many compound expressions containing a subset of `Person` instances, such as: `insuranceCategory value Retired`, `hasSpouse some Thing`, or `Woman and bornIn value Italy`. We can imagine that some of these have only been 'refused entry' to the named class 'elite' (the concept signature of O) due to stringent parsimony or even sloppy modeling. On the other hand, some structurally similar compound expressions are unsuitable for categorizing persons. For example, `bornIn some Thing` does not refine `Person` in any way, while `insuranceCategory value Child and insuranceCategory value Retired` is void. Furthermore, complex conjunctions and especially disjunctions, although possibly containing adequately large subsets of the extent of the focus class, might be mentally too complex to grasp.

For both named subclasses of FC and compound expressions from the former group (for which it would not surprise us to see them transformed to named classes) we propose the term *ontologistic category*. We use this adjective to make distinction from the notion of 'ontological category': while 'ontological' would refer to 'category of beings that exists' (i.e. we cannot deny the existence of categories with complex, unintuitive descriptions or with very small sets of instances), 'ontologistic' refers to a category plausible as reusable domain concept

[2] In Manchester OWL syntax, `http://www.w3.org/TR/owl2-manchester-syntax/`.

to a human ontologist. Intuitively, we should primarily derive the categorization power of an ontology with respect to FC from the set of ontologistic categories rather than from that of all possible concept expressions.

The approach taken in this paper and reflected in its structure is: to create the overall framing of the focused categorization task (Section 2); to choose a restricted (finite and easily manageable) concept expression language and a map it to syntactic axiom patterns (Section 3); to verify the occurrence of the patterns in ontology collection/s (Section 4); to check on a sample of pattern instances if and under what conditions their respective concept expressions are 'ontologistically' plausible (Section 5). We also provide an overview of related research (Section 6) and summary conclusions with future work prospects.

2 General Model of Ontologistic Categorization Power

Let $PS(FC, O)$ be the set of concept expressions (CEs) that are proper specializations of named class FC with respect to ontology O:

$$PS(FC, O) = \{CE; O \models (CE \sqsubset FC)\}$$

Then *focused ontologistic categorization power* (FOCP) of O with respect to FC could theoretically be defined as

$$FOCP(FC, O) = |CE; CE \in PS(FC, O) \wedge oc(CE)|$$

where the binary function *oc* returns *true* if the CE in its argument is an *ontologistic category* (OC). Obviously, such a definition would be anything but rigorous and operational. First, $PS(FC, O)$ will be infinite in common OWL DL dialects, e.g., considering concept expressions nesting with unlimited depth. Second, the concept of ontologistic category, as outlined in the introduction (and exemplified later in this paper) is fuzzy, context-dependent and subjective. For practical purposes we thus need to 1) restrict the language $\mathcal{L}$ of the CEs, to assure $PS(FC, O)$ finiteness, and 2) approximate the 'typical' result of *oc* (as returned by human oracles in various contexts) by a formula based on measurable features of the CEs. In this paper we only consider boolean features, namely, the presence of *axiom patterns*, mapped on the CEs, in O.

Let an *FC-matching axiom pattern*[3] be a set of OWL axioms with placeholder variables (for concepts, roles and individuals), such that one of them, in concept position, is the *FC-variable* (to be substituted by FC in pattern matching). For example, a simple axiom pattern could be C `rdfs:subClassOf` FC.

Let a *pattern-CE mapping function* m be a function that takes an n-tuple of entities substituted for variables (other than the FC-variable) in one instantiation of a pattern p and returns the corresponding CE of a certain *type* t, provided (optional) pruning constraints $prun_t$ associated with this type are satisfied. The

[3] From now on simply 'axiom pattern', since no other kind of axiom pattern is referred to in the paper.

CE types are understood in the context of the categorization task (having 1-to-1 mapping to patterns) and conform to the CE language $\mathcal{L}$.

The set of patterns used for FOCP computation, together with their mapping functions, should satisfy some requirements: 1) they should assure that the resulting CE is a subconcept of FC (e.g., the 'subclass' pattern above satisfies this trivially), and, 2) the occurrences of two different patterns should not yield the same CE (to assure unique counting). We should further, using a description logic (DL) reasoner and/or syntactic pruning constraints, omit CEs that are either unsatisfiable or equivalent to FC. An open question is whether we should only count logically equivalent (though syntactically different) expressions once. While trivially derivable equivalent concepts such as double negations should be avoided, dissimilar expressions only alignable via complex derivations might have descriptive value of their own. Since the CE language chosen in this paper only covers a small subset of OWL (e.g., without negation and Boolean connectives in general), we tolerate multiple equivalent expressions entering the FOCP computation.

If the above pattern were instantiated as `Man rdfs:subClassOf Person`, an adequate m would take the substitution (`Man`/C) and return the CE `Man`, with type t corresponding to 'named subclass'. No pruning constraints would apply.

Let the *occurrence function* for pattern p wrt. FC, denoted as $Occ(p, FC, O)$, return the number of matches of p in O such that FC is substituted to all occurrences of the FC-variable in p. In the body of Occ the axioms from p are conjunctively interpreted, together with the associated pruning constraints. The matches correspond to the n-tuples submitted to the mapping function m.

The *approximate FOCP* of O with respect to a pattern set $P = \{p_1, ..., p_n\}$ can then be defined as the weighted sum of the Occ function results in O,

$$\widehat{FOCP}(FC, O, \mathcal{L}, P) = Occ(p_1, FC, O) * w_1 + ... + Occ(p_n, FC, O) * w_n$$

where $w_i \in [0, 1]$ is the weight of pattern p_i indicating the likelihood that its occurrence would produce an OC.

3 Concept Expression Language and Axiom Patterns

For the remainder of this paper we restrict the language of CEs as defined by the following, extremely simple grammar, guaranteeing finiteness for any finite ontology signature:

```
CE := namedClass | simpleExistentialRestriction | valueRestriction
simpleExistentialRestriction := objectProperty 'some' namedClass
valueRestriction := objectProperty 'value' individual
```

When considering the types of CEs corresponding to this ad hoc language (further called $\mathcal{L}_0$), we specifically cater for the top concept (`Thing`) in the role of existential restriction filler. Therefore the CE types are eventually four, as shown in Tab. 3. The second column displays the CE structure in DL notation; C stands

Type.	CE in DL	Subst.	Abox path length	Axiom pattern size
t1	C	C	3	1
t2	$\exists R.\top$	R	2	1
t3	$\exists R.C$	R, C	5	3
t4	$\exists R.\{i\}$	R, i	3	4

Table 1. Summary of CE types in $\mathcal{L}_0$

for named concept (class), R for role (object property) and i for individual. We see that t2 and t4 are mere refinements of t3 (generic existential restriction) in DL terms. The third column indicates which symbols from the CE correspond to variables substituted in the associated axiom pattern. The set of variables in t2 is subset of those of t3 and t4; if we consider one or more CEs for t3 or t4 with some R then we should also consider the CE for t2 with this R. The fourth column measures the length of Abox path (as sum of resource nodes and predicate edges) connecting the categorized individual with entities ('responsible' for the categorization) substituted for variables from the third column: it is smallest for t2 where only the adjacent edge is applied; t1 and t4 require a whole triple (instantiation or property assertion, respectively), while t3 needs two triples (both assertion and the ensued instantiation to the 'filler' class). The order of the patterns in the table however reflect the increased complexity of their detection in the Tbox using the proposed axiom patterns (fifth column), which we detail in the next subsection.

3.1 Syntactic axiom patterns in the ontology schema

The CEs in $\mathcal{L}_0$ could in principle be mapped to diverse constellations of axiom patterns with varying expressiveness. However, we implement the patterns primarily in RDFS terms, namely, over `rdfs:subClassOf`, `rdf:type`, `rdfs:domain` and `rdfs:range` axioms, such that all of their arguments are either atomic expressions or variables for which they can be substituted; we so far avoided `rdfs:subPropertyOf` to keep the pattern structure simpler (we will consider adding it in the future). Besides the patterns also address the *pruning* of CEs whose ineligibility follows from the ontology structure. The patterns (p1,..,p4)[4] are pairwise mapped on the previously defined CE types (t1,..,t4), except that for t4 we also supply an additional pattern p5 in which the individual i is not part of the ontology itself but of an associated SKOS codelist. Since the pattern occurrence is to be computed specifically with respect to the entities responsible for categorization (third column in Tab. 3), we present the patterns in terms of their *occurrence function* $Occ(p\#, FC)$.[5]

[4] We denote these specific patterns using normal font, to differentiate with the superscript notation (p_i) of abstract symbols in Section 2.

[5] For simplicity we omit O in the formula; the identity of the ontology follows from the FC it contains. We also avoid the use of DL notation and express the OWL

Pattern p1 CEs of type t1 are simply *subclasses* of FC, i.e. they are matched by the previously mentioned axiom pattern; the occurrence function is the number of these subclasses:

$$Occ(p1, FC) = |\{C;\ C\ \texttt{rdfs:subClassOf}\ FC\}| \tag{1}$$

The subclasses can be both direct or indirect, and possibly even inferred using other kinds of axioms, i.e. they are subclasses of FC in the deductive closure of the ontology computed by a reasoner.

Pattern p2 Next we will consider properties having FC in their *domain*:

$$Occ(p2, FC) = |\{P;\ P\ \texttt{rdfs:domain}\ FC \wedge P \notin prun_2(FC)\}| \tag{2}$$

where $prun_2(FC)$ is the set of properties that have to be pruned as ineligible for this pattern. Again, even cases when FC is *inferred* as domain of P are considered. Since we do not take into account the right-hand side of P, each corresponding CE (of type t2) would contain all instances of FC that appear in the subject of a triple with P as predicate. The corresponding mapping function m thus maps the pattern on the DL expression $\exists P.\top$ (t2 in Table 3, with P substituted for R). $prun_2(FC)$ essentially contains the properties P that appear in an existential restriction[6] $FC \sqsubseteq \exists P.C$; for such properties the CE would contain *all* instances of FC.

Pattern p3 Now we proceed to the *range* of properties with FC in domain and then to the *subclasses of this range*:

$$\begin{aligned} Occ(p3, FC) = |\{(P, C);\ \exists D\ \ P\ \texttt{rdfs:domain}\ FC \wedge P\ \texttt{rdfs:range}_a\ D \wedge \\ \wedge\ C\ \texttt{rdfs:subClassOf}\ D\ \wedge\ (P, C) \notin prun_3(FC)\}| \end{aligned} \tag{3}$$

where $prun_3(FC)$ is, again, the set of properties that have to be pruned as ineligible for this pattern. The CE (of type t3) would include all instances of FC that appear in the subject of a triple with P as predicate and some i as object such that i is instance of C. The inferential closure is again used, however, with the exception of the range axiom, which is only considered as asserted (therefore the 'a' index in $\texttt{rdfs:range}_a$) – otherwise not only subclasses of D but also classes having a common superclass[7] with D would be returned as C. The pattern maps on the DL expression $\exists P.C$ (t3). $prun_3(FC)$ contains the pairs (P, C) that appear in an existential restriction $FC \sqsubseteq \exists P.E$ such that $E \sqsubseteq C$; for such properties the CE would contain *all* instances of FC.

---

axioms using predicate URIs, to avoid collision with general math notation. The only reference to beyond-RDFS construct is the value restriction for p5.

[6] The restriction can also be inherited from a superclass or part of a complete definition, or can have the form of a `value` or `self` restriction or of a cardinality restriction that specializes the existential one; analogously for other $prun_n$'s below.

[7] This superclass would become an inferred range of P.

Pattern p4 This pattern extends the previous one with an individual that is instance of C:

$$Occ(p4, FC) = |\{(P, i);\ \exists C, D\ \ P\ \texttt{rdfs:domain}\ FC \wedge P\ \texttt{rdfs:range}_a\ D \wedge \\ \wedge\ C\ \texttt{rdfs:subClassOf}\ D\ \wedge\ i\ \texttt{rdf:type}\ C\ \wedge\ (P, i) \notin prun_4(FC)\}| \quad (4)$$

where $prun_4(FC)$ is analogous to the previous variants. The CE (of type t4) would include all instances of FC that appear in the subject of a triple with P as predicate and (the specific individual) i as object. The inferential closure is used as before. The pattern maps on the DL expression $\exists P.\{i\}$ (t4). $prun_4(FC)$ contains the pairs (P, i) that appear in an existential restriction $FC \sqsubseteq \exists P.E$ such that $i \in E$; for such properties the OC would contain *all* instances of FC.

Pattern p5 SKOS is the most widespread alternative to OWL to consider when specifying simpler 'ontological' taxonomies. This variant thus extends the previous one for a specific source of instance i – a SKOS code list (concept scheme):

$$Occ(p5, FC) = |\{(P, i);\ \exists s\ \ P\ \texttt{rdfs:domain}\ FC \wedge \\ \wedge\ P\ \texttt{rdfs:range}_a\ (\texttt{skos:Concept} \sqcap \textbf{value}(\texttt{skos:inScheme}, s)) \wedge \\ \wedge\ i\ \texttt{skos:inScheme}\ s\ \wedge\ i\ \texttt{rdf:type}\ \texttt{skos:Concept}\ \wedge\ (P, i) \notin prun_4(FC)\}| \quad (5)$$

where $\textbf{value}(\texttt{skos:inScheme}, s)$ shortcuts the DL concept expression $\exists Q.\{s\}$ such that $Q =$`skos:inScheme`. The CE (again of type t4) is defined as in Pattern 4. The difference is merely in the selection method for i – rather than instance of a class from the current ontology, it has to be a `skos:Concept` linked to concept scheme s that is in the range of P. The inferential closure is used as before.

With respect to the requirements on axiom patterns from Section 2, all patterns assure (p2–p5 via the domain axiom) that the mapped CE is a *specialization* of the FC. It is also easy to see that the patterns, except p4 vs. p5, are *mutually exclusive* since they produce structurally different CEs. (Formally, to assure exclusivity of p2 and p3, D in p3 should not be `owl:Thing`. We however do not anticipate that explicit range axioms would have the default value `Thing`.)

4 Survey on Syntactic Pattern Occurrence

The *research questions* to be answered by the analysis were:

1. How many ontologies, and for how many FCs, provide a decent number of 'categorizing' CEs mapped on the patterns from Section 3.
2. What are the differences in the occurrence of the individual patterns overall and across different collections.

For our experiments we used two collections of ontologies. First is a small collection from the domain of conference organization, called *OntoFarm*,[8] and

[8] `http://owl.vse.cz:8080/ontofarm/`

the second is the collection from *Linked Open Vocabularies* (LOV) portal.[9] While the former is rather an experimental collection of ontologies (used, among other, in the Ontology Alignment Evaluation Initiative[10]) with heterogeneous styles and relatively rich in axioms, the latter contains real-world (mostly) light-weight ontologies with connection to the Linked Open Data Cloud. In the analysis we made use of our *Online Ontology Set Picker framework*[11] to process ontologies from two collections. OntoFarm has 16 ontologies, and for LOV we used January 2016 snapshot where 529 ontologies were available of which we could process successfully 509 at syntactical level.

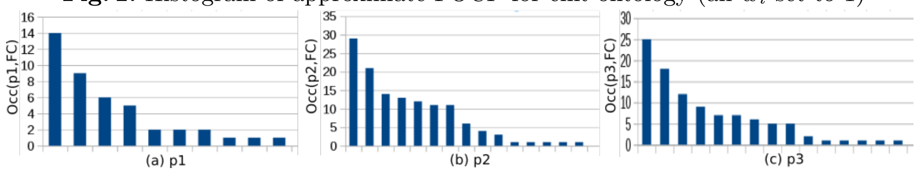
In the rest of the section we report on the summary as well as selected detailed results of the axiom pattern occurrence analysis; separately for the four patterns and for the fifth ('SKOSsy') pattern.[12]

Patterns p1, p2, p3 and p4 We matched the five patterns described in Sec. 3, including the pruning operation, on each ontology (together with its imports). Fig. 1 depicts the histogram of $\widehat{FOCP}$ for weights set to 1, i.e. as a plain sum of pattern occurrences,[13] for p1, p2 and p3 in the *cmt* ontology from OntoFarm (neither p4 nor p5 are matched in this ontology). Cmt has 29 classes. We see that there are 10, 15 and 15 focus classes (FCs) having at least one match of p1, p2 or p3, respectively. This ontology has four levels. As expected, FCs with higher categorization power are placed closer to the root (even for p2 and p3 where this is not entailed straightforwardly). For p1 there is the following order of FCs: Person, ConferenceMember, User, Document, Paper, Author, Decision, ProgramCommitteeMember, Review, Reviewer. For p2 there is the following order of FCs: Person, User, ConferenceMember, Document, Administrator, Conference, Paper, Reviewer, ProgramCommitteeMember, Author, Co-author, ProgramCommittee, Review, ProgramCommitteeChair, Bid. For p3 there are the same FCs as for p2 but their order is slightly changed. In all, six of the FCs with nonzero $\widehat{FOCP}$ are top-level classes: Person, Document, Decision, Conference, ProgramCommittee and Bid. There are four FCs from third level: Author, ProgramCommitteeMember, Administrator and Reviewer Only one FCs, Co-author, is a leaf class. All other FCs with nonzero $\widehat{FOCP}$ are from the second level of the taxonomy.

In order to provide aggregate results we counted the pattern occurrences across the FCs for all classes of all ontologies. We summed up these results at ontology level by identifying 'categorizable FCs' for which the $\widehat{FOCP}$ reached some threshold τ (1, 3 and 5) for patterns p1,..,p4;[14] see Table 2 for OntoFarm and Table 3 for LOV. They show the percentage of ontologies for which more

[9] `http://lov.okfn.org/`
[10] `http://oaei.ontologymatching.org/`
[11] `http://owl.vse.cz:8080/OOSP/`
[12] Supplementary material is at `http://owl.vse.cz:8080/EKAW2016/`
[13] This is the (degenerated) version of $\widehat{FOCP}$ we will use throughout this section. It is only in the end of the paper that we formulate a 'smarter' one.
[14] Pattern p5 was only matched rarely; it is thus analyzed separately below.

Fig. 1. Histogram of approximate FOCP for cmt ontology (all w_i set to 1)



than a n portion (quantized by 0.1) of FCs are 'categorizable FCs' according to the given pattern. For example, from Fig. 1 we can see that *cmt* would be included in the OntoFarm percentage, for $\tau = 5$, up to $n > 0.1$ in V1 (as 4 of the 29 classes are 'categorizable') and up to $n > 0.2$ in V2 and V3.

Table 2. OntoFarm: ratio of ontologies with over n portion of classes 'categorizable'

Pattern	τ	> 0.0	> 0.1	> 0.2	> 0.3	> 0.4	> 0.5	> 0.6	> 0.7	> 0.8	> 0.9
p1	1	100.0 %	100.0 %	81.3 %	25.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
	3	100.0 %	100.0 %	18.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
	5	93.8 %	31.3 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
p2	1	100.0 %	93.8 %	56.3 %	31.3 %	25.0 %	6.3 %	6.3 %	6.3 %	6.3 %	0.0 %
	3	100.0 %	43.8 %	25.0 %	12.5 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
	5	93.8 %	37.5 %	6.3 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
p3	1	93.8 %	50.0 %	18.8 %	6.3 %	6.3 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
	3	93.8 %	50.0 %	12.5 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
	5	93.8 %	31.3 %	6.3 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
p4	1	25.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
	3	18.8 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
	5	6.3 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %

Comparing the results in Table 2 and Table 3 we can see that proportionally more OntoFarm ontologies provide some categorization options (i.e. CEs) for at least some focus classes. In the case of pattern p4 considering $\tau = 3$ and $\tau = 5$ LOV ontologies provide proportionally slightly more ontologies with categorization by instances. Since by nature LOV ontologies are 'data schemes', instances are more typical there than in OntoFarm ontologies. Interestingly, OntoFarm ontologies fail to provide categorization options for more than 40% of their classes, considering $\tau = 3$. On the other side, there are always some LOV ontologies providing categorization options for almost all of their classes regarding pattern p2, and for at least 60% or 40% of their classes regarding patterns p1, p3 and p4.

Table 4 shows the average $\widehat{FOCP}$ across FCs separately for each OntoFarm ontology. The highest average per pattern is in bold. The *OpenConf* ontology has the highest average for categorization by subclasses, i.e. pattern p1; this can be attributed to its rich multi-level subsumption hiearchy. The highest average

Table 3. LOV: ratio of ontologies with over n portion of classes 'categorizable'

Pattern	τ	> 0.0	> 0.1	> 0.2	> 0.3	> 0.4	> 0.5	> 0.6	> 0.7	> 0.8	> 0.9
p1	1	80.9%	74.5%	48.7%	24.2%	10.2%	2.8%	0.6%	0.2%	0.0%	0.0%
	3	63.7%	44.4%	10.2%	1.2%	0.2%	0.2%	0.2%	0.0%	0.0%	0.0%
	5	53.2%	22.8%	1.0%	0.2%	0.2%	0.2%	0.2%	0.0%	0.0%	0.0%
p2	1	85.7%	75.0%	63.7%	48.3%	35.0%	21.0%	14.9%	9.0%	4.3%	1.8%
	3	77.0%	56.4%	31.6%	15.7%	7.9%	3.9%	1.6%	0.6%	0.6%	0.4%
	5	64.2%	35.6%	15.9%	6.7%	2.9%	1.6%	0.8%	0.4%	0.4%	0.2%
p3	1	50.5%	24.4%	11.6%	5.1%	2.8%	1.6%	1.2%	0.8%	0.2%	0.0%
	3	43.4%	17.9%	7.7%	2.6%	1.6%	0.8%	0.4%	0.2%	0.0%	0.0%
	5	35.8%	12.8%	4.7%	1.4%	1.2%	0.6%	0.4%	0.0%	0.0%	0.0%
p4	1	23.0%	8.1%	3.3%	1.6%	1.0%	0.4%	0.0%	0.0%	0.0%	0.0%
	3	18.9%	5.1%	2.6%	0.8%	0.6%	0.2%	0.0%	0.0%	0.0%	0.0%
	5	16.5%	4.7%	2.6%	0.8%	0.6%	0.0%	0.0%	0.0%	0.0%	0.0%

for categorization by properties, patter p2, is in *cmt*. This can be explained by the fact *cmt* has more properties than classes (59 vs. 29) and there are few complete or partial definitions. *Iasted* features the highest average of categorization by classes in range, pattern p3. It has a rich multi-level hierarchy and there are often general classes in range of its properties. Finally, *OpenConf* also has the highest average of categorization by instances. The designer of this ontology applied a specific modeling style with enumerated classes whose instances serve for categorization, e.g., of papers or reviews. Only three other ontologies feature nonzero average for pattern p4. On the supplementary webpage we further provide a similar table for LOV, the complete list of FCs for each ontology from OntoFarm and LOV, and finally, we provide the list of top 3 FCs for each pattern.

Table 4. Average of axiom pattern matches

ontology	p1	p2	p3	p4	ontology	p1	p2	p3	p4
cmt	1.43	**4.30**	3.37	0	confOf	1.15	0.51	0.92	0.00
MyReview	0.77	2.36	1.85	0.1	sigkdd	1.08	0.56	0.86	0.00
edas	1.22	0.57	1.3	0	Conference	1.67	1.62	1.82	0.00
ekaw	2	0.38	3.08	0	confious	1.11	0.51	2.19	0.12
crs_dr	0.71	2.14	0	0	PCS	1	2.08	1.79	0.00
paperdyne	0.93	1.33	0.87	0	iasted	1.95	0.26	**4.09**	0.00
linklings	0.78	0.32	0.22	0	OpenConf	**2.19**	1.55	1.11	**0.19**
MICRO	0.84	0.63	0.56	0.13	Cocus	1.16	0.55	2.45	0.00

Pattern p5 Finally, we performed an analysis of ontologies which use SKOS concepts for entity categorization (pattern *p5*). OntoFarm ontologies do not feature this pattern. Regarding LOV, there are 7 ontologies out of 509 which enable users to apply this categorization. Table 5 contains information about those 7 ontologies regarding number of all SKOS concepts that can be used for categorization and number of all schemes from which those concepts come. Further, there is information about the date of last modification. In two cases, concept schemes are not available. For other ontologies the number of SKOS concepts (SKOS schemes) usable for categorization varies from 11 to 274 (from 1 to 16, resp.). Although this is not present in many ontologies, we can assume that this practice will appear more often in future as those seven ontologies were modified on average in 2014 while average of modified date for the whole LOV is 2012.

Table 5. Entity categorization power via SKOS concepts (pattern p5)

vocabulary	nr. of schemes	nr. of all concepts	modified
http://www.loc.gov/premis/rdf/v1	16	79	2012-09-14
http://purl.org/procurement/public-contracts	4	86	2012-10-10
http://data.ign.fr/def/geofla	1	NA	2015-01-11
http://datos.gob.es/def/sector-publico/organizacion#	3	NA	2015-02-19
http://data.ign.fr/def/topo	13	274	2014-04-12
http://contsem.unizar.es/def/sector-publico/pproc	3	11	2015-01-11
http://rdf.insee.fr/def/geo	3	30	2015-01-11

Generally, the *research questions* from the start of the session have been answered in the sense that: 1) The proportion of ontologies exhibiting 'categorizable' FCs is neatly displayed in Tab. 3 (for LOV as large and representative collection). Overall, the majority of ontologies have 'interesting' categorization power for *some* of their classes as FCs; however, if we require more than 20-30 % of classes to be 'categorizable', the proportion of ontologies satisfying this requirement is rapidly dropping. 2) There are important differences in the proportion of pattern occurrence, with p2 being most widespread, followed by p1. This also holds for the smaller OntoFarm collection, which is structurally richer than (on average) the LOV; however, as appears, the richer axiomatization only allows for a smaller proportion of ontology classes to be categorized, possibly by the effect of CE pruning.

5 Ontologistic Categorization Experiment

The CE sets on which the pattern occurrence counts are obtained by automated analysis are mere rough approximations of the true OC sets for the respective FCs. In order to get finer insights, we proceeded to detailed investigation of sample CEs by human 'ontologists', both experts and relative novices (students of relevant subjects). Since we take named (sub)classes as OCs by default, we only examined the CEs of t2, t3 and t4 (we did not further distinguish between the t4 variant returned by pattern p4 and by the 'SKOSsy' p5). The general research questions, this time, were:

1. Is the OC status of CEs correlated with the CE type and/or the background of the human assessor?
2. What is the proportion of clear vs. borderline cases?
3. Which deeper semantic distinctions either lead to negative assessment (even in absence of logical causes for such assessment) or make the decision tricky?

We report on the most interesting results of this effort below.[15]

Initial sampling As regards the CE sampling for both threads of analysis (expert/novice), we used the same collections as in Section 4, i.e. OntoFarm and LOV. From each collection 10 CEs per type have originally been randomly sampled, yielding 80 CEs. After manual removal of duplicities (for OntoFarm as smaller collection) and CEs containing entities with cryptic names, 59 CEs remained (28 from OntoFarm and 31 from LOV); there were 17 CEs of t2 (existential restriction with 'filler `Thing`'), 20 of t3 (existential restriction to specific class), and 22 of t4 (value restriction).

Expert ontologist assessment and insights. The analysis has been done by the three authors of the paper, all with 10–20 years of experience in ontological engineering. They first examined the sample of 59 CEs independently and assessed it on the 5-point Likert scale: for each CE X the question "Is X an OC?" was answered as either 'certainly', 'perhaps', 'borderline', 'perhaps not' or 'certainly not'. Then a consensus was sought in a F2F session. The independent assessment had 76% agreement: in 45 out of 59 cases there was no contradictory assessment (certainly/perhaps yes vs. certainly/perhaps no);[16] we will call these cases *clear positives* (43 cases, incl. the one used twice) and *clear negatives* (3 cases), respectively. The consensus session then yielded a complete consensus on the remaining cases; in 12 out of the 14 'clash' cases the final result was 'yes' (namely, a conceivable situation was formulated in which the CE would be a plausible OC), one case was found dubious due to implausible inference (see the second 'insight' below) and in one case the CE was assumed semantically equivalent to its FC, both resulting into 'no'. Of the seven ultimately negative results, five were of t2, one of t3 and one of t4. Selected general insights into less obvious decisions, with examples, follow (see the supplementary page for complete assessments with commentaries):

– Ontologies tied to software applications, such as some OntoFarm ones (capturing the processes supported by conference software) use object properties to capture relationships that are only relevant within a *short time frame*, e.g., `cmt:finalizePaperAssignment`; a meaningful category of persons would rather refer to their long-term responsibility for paper assignment rather than to the instantaneous action of 'finalizing' it.

[15] More detail can be found, again, at `http://owl.vse.cz:8080/EKAW2016/`.

[16] For all these cases also held that out of the three assessments at least two had the same polarity while the third was sometimes 'borderline'.

– In some cases the use of inferential closure for the filler class in t3 leads to linking relatively *thematically unrelated* entities (especially in the DBpedia ontology), such as in `dbo:beatifiedPlace some dbo:WineRegion` for instances of `dbo:Person`. While this case was found marginally acceptable (there could be some correlation between religiosity and wine production), a similar one, `dbo:headChef some dbo:BaseballPlayer` was rejected not only due to thematic leap but also due odd inference result: the FC was `dbo:Village`, the declared domain of `dbo:headChef` is `dbo:Restaurant`, but the ontology (actually, the 2014 version from the LOV endpoint) enables to infer the axiom `dbo:Restaurant rdfs:subClassOf dbo:Village`.
– Some CEs of t4 are plausible but less useful due to their *inherently limited extent*: for instance, categorizing instances of `geopolitical:area` as `geopolitical:isSuccessorOf value` X, where X is another (geopolitical) area.

Novice ontologist assessment. There were two groups of students involved: Bc-level students in a course on Artificial Intelligence (AI) and MSc-level students in a specialized course on Ontological Engineering (OE). Both courses provided a certain degree of OWL modeling experience (in Protégé and Manchester syntax) prior to this exercise, although OE went into more depth as regards the underlying DL and reasoning. There were 17 AI students and 10 OE students altogether. In both courses the students were first provided with a 30' overview of the notions of CE (in $\mathcal{L}_0$), OC and FC roughly as presented in Section 1 of this paper. Then they completed an assignment consisting of 20 atomic tasks, all available in a single sheet of a web questionnaire.[17] In each atomic task the student was required to provide an answer to the question "Is the class CE a meaningful category for categorizing objects of class FC", where FC was a named focus class and CE was a concept expression in Manchester syntax. The answer was again from the 5-point Likert scale, with an additional option '*no judgment, since I don't understand the example*'.

The 59 CEs from the initial sample were randomly divided into three questionnaire versions (one value restriction CE was used twice) to eliminate cribbing; the numbers of returned questionnaires per version were 7, 9 and 11, respectively, with balanced proportion of AI vs. OE students. To avoid protracting and biasing the experiment, the students were instructed to only judge the CEs by the expression itself, i.e. without consulting the respective ontology specification or other external resources. However, specifically for the 'conference' domain of OntoFarm, they were provided with a brief domain glossary (since as students they were not expected to have experience with conference organization matters). In both sessions, 30' sufficed to all students for completing the (20-task) assignment.

We aggregated the results by questionnaire task, and then both by the course and by CE type. The aggregation was carried out by simple summation over the

[17] The questionnaire was in Czech. Its English translation is available from the paper web page `http://owl.vse.cz:8080/EKAW2016/`.

values rescaled to the $[-1;1]$ interval (i.e. 'certainly' turned to 1, 'perhaps' to 0.5, both 'borderline' and 'no judgment' to 0, etc.), and then normalized by dividing by the number of students on the task. This way, for example, a task assigned to eight students, with the responses 'certainly', 'perhaps', 'borderline' and 'perhaps not', all present twice, yields the normalized sum (NS) of $(2*1+2*0.5+2*0+2*-0.5)/8 = 0.25$. Follows a short digest of the results:

- The average NS over all 60 tasks was 0.07, i.e. rather low, although positive. Of the 60 NS values, 28 were positive, 5 zero and 27 negative. The values strictly below 0.25 and above -0.25, possibly viewed as 'borderline aggregates', were 34 (57%).
- The cases[18] with highest positive and lowest negative values are in Table 6; the type is listed in the third column. We see that cases with highest positive polarity tend to achieve higher absolute values than cases with highest negative polarity, and that t4 dominates the upper end of the spectrum. Interestingly, the negative cases correspond each to a different type and also have different semantic roots: the village with baseball player head-chef was already discussed before (distant and dubious inference), the 'conference in city' one deals with a seemingly mandatory property leading to $OC \equiv FC$ (here, however, the experts' consensual opinion diverged: how about future editions not yet having a location, or virtual conferences?), and the 'day followed by Friday' only holds for one individual, in turn.
- The average NS was higher for the OE students (0.12) than for the AI students (0.04), which might be attributed to more developed 'ontologistic thinking' of the latter. The inter-task variance,[19] indicating the tendency towards giving uneven values (averaged over the students filling the same task) across the questionnaire, was about the same (0.16) for both courses.
- The average NS was highest for t4 (0.21, with 15 positives, 1 zero and 7 negatives), lower for t3 (0.02, with with 9 positives, 3 zero and 8 negatives) and lowest for t2 (-0.05, with with 4 positives, 1 zero and 12 negatives).

In comparison with the 'expert ontologist' assessment:

- The students gave a significantly lower score: only about a half of CEs are viewed as OCs, compared to 88% (52/59) by the final consensus of experts. This can be explained by their lower ability to figure out specific situations in which less obvious CEs might become plausible.

[18] Most namespace prefixes used can be expanded using the `prefix.cc` service. Prefixes unlisted by this service follow: p-act=`http://purl.org/procurement/public-contracts-activities#`, p-aut=`http://purl.org/procurement/public-contracts-authority-kinds#`, p1=`http://www.loc.gov/premis/rdf/v1`, p1-sm=`http://id.loc.gov/vocabulary/preservation/storageMedium#`, sigkdd=`http://oaei.ontologymatching.org/2016/conference/data/sigkdd.owl`

[19] The *intra-task* variance, indicating the degree of agreement within the students filling the same task, was computed and it is available from supplementary web page. It will be included in the CR version.

Table 6. CEs with highest and lowest average NS of student scores

FC	Expression	Type	Avg.NS
ofrd:FridgeFreezer	ofrd:styleOfUnit value ofrd:SingleDoor	4	0.91
gr:BusinessEntity	pco:mainActivity value p-act:GeneralServices	4	0.86
gr:BusinessEntity	pco:authorityKind value p-aut:LocalAuthority	4	0.61
akt:Generalized-Transfer	akt:information-transfer-medium-used value akt:Email-Medium	4	0.59
p1:Storage	p1:hasStorageMedium value p1-sm:mag	4	0.59
fabio:Item	fabio:isStoredOn value fabio:web	4	0.50
...	...	...	...
dbo:Village	dbo:headChef some dbo:BaseballPlayer	3	-0.50
sigkdd:Conference	sigkdd:City_of_conference some Thing	2	-0.56
gr:DayOfWeek	gr:hasNext value gr:Friday	4	-0.56

– If we apply the same method of average NS computation on the initial assessment of experts the proportion of 'borderline aggregates' between -0,25 and 0,25 is only 14% (in contrast to 57% for the students' values).

– There is agreement on less frequent OC status of t2 (i.e. lower reliability of pattern p2). Out of the 17 respective CEs, as mentioned above, only 4 were viewed as OCs by students and 12 by the experts (who in turned judged all CEs of other types, except two, as OCs).

– As regards the case-by-case comparison between students and experts, there is also correlation in the sense that the 43 experts' clear positives obtained a positive average NS from students (0.14), while the 14 initially 'clash' cases obtained a slightly negative average NS (-0.07) and the 3 negative cases obtained a clearly negative average NS (-0.24).

As regards the *research questions* from the start of this section: ad 1) the OC status assessment strongly depends both on the CE type and the expert/novice distinction; ad 2) most cases are clear for experts but not for novices; 3) semantic distinctions leading to negative or inconclusive assessment might be related to temporality, inference over distant paths or inherently small cardinality of some concepts (referring to the exemplified insights above).

A conclusion to be made with respect to FOCP computation is that t4 (value restriction) and to some degree t3 (existential restriction) might successfully complement t1 (named class) in the role of OC. As regards the relatively poor performance of p2, it might be premature to completely abandon it at this phase, since for some models it might yield the only unnamed CEs, as the analysis from Section 4 indicates, and its contribution to the overall categorization power should still be considered. A naïve weighted approximate FOCP formula for $\mathcal{L}_0$

and $P = \{p1, p2, p3, p4\}$, derived from the students' assessment, could be

$$\widehat{FOCP}(FC, O, \mathcal{L}_0, P) = Occ(p1, FC, O) * 1.0 + Occ(p2, FC, O) * 0.3$$
$$+ Occ(p3, FC, O) * 0.5 + (Occ(p4, FC, O) + Occ(p5, FC, O)) * 0.7$$

where the numerical weight coefficients roughly correspond to the ratio of CEs with positive average NS, per type.

6 Related Work

Since we are unaware of prior work on the same topic, we reference related research that overlaps with ours at the abstract level (the notion of 'classification power'), systematically applies other kinds of metrics on ontologies, or addresses similar application-level goals (ontology reuse or transformation) by other means.

The term classification/categorization power previously appeared in many scientific texts, however, rarely as a rigorously defined notion. For example, on many occasions, automated classifiers are reported to have certain 'classification power' with respect to classes from an ontology, which is merely an informal circumscription of measures such as accuracy or error rate. The 'power' also clearly pertains to the classifier and not to the ontology. Partially relevant is the analysis made by Giunchiglia & Zaihrayeu [3], who categorized 'lightweight' ontologies with respect to two dimensions: complexity of labels (simple noun phrases vs. use of connectives and prepositions) and use of 'intersection' operator allowing to combine atomic entities of different nature (e.g., the atomic concepts 'Italy' and 'vacation' implicitly combine into 'vacation in Italy'). Maximal 'classification power' is obtained when both explicitly complex labels and implicit concept combinations are allowed. This however only applies to classifying documents extrinsic to the ontology, since 'intersection' of concepts of different nature is not coherent with the set-theoretic semantics of DL. Overall, their 'classification power' is a global property of the method by which the ontology has been built. In contrast, our notion of FOCP applies to individuals intrinsic to the DL world of the ontology and is calculated with respect to a focus class.

As regards the analysis of ontology repositories in terms of various aggregated features and metrics (logical, graph, lexical etc.), there has recently been renewed interest, following up with the early work of Tempich et al. [8] (aiming to build a benchmark for testing ontology tools). A large scale study of OWL ontology metrics has been carried out by Matentzoglu et al. [5]. However, the categorization power of ontologies has not been, to our knowledge, studied, never mind with the flavor presented here.

Our own ongoing work on the PURO modeling language [7] deals with various options how the same 'background' state of affairs can be expressed in OWL. PURO structurally resembles OWL but relaxes some of its modeling constraints. A library of transformation patterns allows to proceed from one PURO model to alternative OWL ontologies in different encoding styles. An example relevant to our case is the notion of *enterpreneur*, which is likely to be expressed as type in PURO, but could be translated to relationship (`insuranceCategory`) restricted to the `Entrepreneur` individual in OWL (i.e. a compound concept expression). Analogously, *born in* may possibly be a relation in PURO but can be translated not only to OWL property restrictions but also to named classes such as `PersonBornInUK`, assuming we prefer an 'encapsulating' encoding style used, e.g., in the DBpedia Ontology.[20] Modeling in PURO and applying the transformation patterns may thus make hidden OCs explicit in the domain. A similar account of alternative 'typecasting' (but with smaller coverage) has been given by Krisnadhi [4].

The broad context of our research, the task of ontology reuse, has been studied by Schaible et al. [6]: the users expressed their preferences on reuse strategy in a survey. The results indicate that reusing multiple entities from the same vocabulary may often be preferred; this corroborates the relevance of our approach to measuring the categorization power of ontologies with respect to focus classes. Reuse support [2] is also systematically sought by the maintainers of LOV [9], primarily at keyword relevance level; we are in contact with them and will seek to integrate our complementary approaches.

7 Conclusions and Future Work

Ontologies are an important means of subcategorizing entities already known to belong to a general focus class, and the scope of subcategories need not be confined to named classes, especially in the linked data world, which is relatively 'property-centric'. High categorization power of an ontology for certain classes might serve as argument for *reusing* this ontology when building a new one, or preparing a dataset scheme, if the notions corresponding to these classes are important in the modeled domain. Additionally, plausible compound concept expressions can be *transformed* to named ones if needed. Studying the role of compound class expressions as meaningful (focused) categorization options thus appears of high importance. In this respect we provided, to our knowledge, the first systematic study on the categorization power of OWL ontologies covering several compound concept expression patterns. The main contributions of the paper (including oversize material swapped to an accompanying website) are: 1) formulation of the problem and description of the pattern set; 2) empirical analysis of two ontology collections for (syntactical) pattern occurrence; 3) in-depth ontologistic analysis of a sample of pattern occurrences, carried out both by ontologistic experts (paper authors) and two groups of slightly trained students.

While the presented research focused on the general principles and empirical analysis, aspects of it have also been implemented into our OOSP tool, allowing to recommend ontologies with respect to their estimated FOCP.[21] The estimation currently does not distinguish between the CE types; it is however going

[20] `http://wiki.dbpedia.org/services-resources/ontology`

[21] This work has been published separately as a demo paper [11]. The demo paper only contains a minimalist informal explanation of the notions of FC, CE and OC; apart from that the content of the demo paper is disjoint with the current submission.

to be tuned by lowering the weight for instances of patterns exhibiting lower proportion of plausible OCs, as indicated in Section 5.

As future work we plan to investigate if inclusion of additional types of concept constructors into the $\mathcal{L}$ language could still yield relevant results (true OCs) while keeping the computational complexity tractable. One candidate is the *inverseOf* predicate: in some 'modeling styles' only one of the pair of mutually inverse properties is included in the ontology and some categorization tasks might then be carried out against the direction of such a relationship.

In addition to the structural aspect of the CEs, we may also leverage on their *lexical* aspects. Presence of suffixes such as `-Category`, `-Type` or `-Kind` in either property or filler class names may indicate that the filler individual is actually a type (note the highly scoring case with `pco:authorityKind` in Table 6).

We also envisage to align this schema-oriented analysis to *data-oriented* one. If a representative sample of data is available for an ontology, we can of course derive the OC status of CEs from the numbers of FC instances satisfying them (or not), which can be obtained by a simple SPARQL query. This empirical analysis will also allow us to refine the reliability estimate of the patterns (in combination with ontologistic assessment as in this paper) and clues used in the schema-oriented approach, which is usable even if instance data are not available.

In longer term, the goal is to embed a properly tuned ontology/fragment recommender into our OBOWLMorph OE tool [1] and also make available via an API to third-party tools such as the ProtégéLOV Plugin [2].

Ondřej Zamazal has been supported by the CSF grant no. 14-14076P, "COSOL – Categorization of Ontologies in Support of Ontology Life Cycle". Vojtěch Svátek has been supported by the the UEP IGA F4/28/2016 project.

References

1. Dudáš, M., Hanzal, T., Svátek, V., Zamazal, O.: OBOWLMorph: Starting Ontology Development from PURO Background Models. In: OWLED'15.
2. García-Santa, N., Atemezing, G. A., Villazón-Terrazas, B.: The ProtégéLOV Plugin: Ontology Access and Reuse for Everyone. In: ESWC (Satellite Events) 2015: 41-45.
3. Giunchiglia F., Zaihrayeu I.: Lightweight Ontologies. In: Encyclopedia of Database Systems, Springer, 2009, 1613–1619.
4. Krisnadhi, A. A., Hitzler P., Janowicz K. On capabilities and limitations of OWL regarding typecasting and ontology design pattern views. In: OWLED'15.
5. Matentzoglu, N., Bail, S., Parsia, B.: A Snapshot of the OWL Web. In: 12th International Semantic Web Conference, 2013.
6. Schaible J., Gottron T., Scherp A.: Survey on Common Strategies of Vocabulary Reuse in Linked Open Data Modeling. In: ESWC 2014: 457–472.
7. Svátek, V., Homola, M., Kluka, J., Vacura, M.: Metamodeling-based coherence checking of OWL vocabulary background models. In: OWLED'13, Montpellier.
8. Tempich C., Volz R.: Towards a Benchmark for Semantic Web Reasoners - An Analysis of the DAML Ontology Library. In: EON-2003.
9. Vandenbussche, P.-Y., Vatant, B.: Linked Open Vocabularies. *ERCIM News*, 2014.
10. Zamazal O., Svátek V.: OOSP: Ontological Benchmarks Made on the Fly. In: Workshop SumPre'15 at ESWC 2015.

11. Zamazal O., Svátek V.: Ontology Search by Categorization Power. In: Workshop SumPre'16 at ESWC 2016.