

Focused Categorization Power of Ontologies: General Framework and Study on Simple Existential Concept Expressions

Vojtěch Svátek^{a,*}, Ondřej Zamazal^a, Viet Bach Nguyen^a, Miroslav Vacura^{a,b}, Jiří Ivánek^a and
Ján Klúka^c

^a Department of Information and Knowledge Engineering, Prague University of Economics and Business, Nám. W. Churchillia 4, 130 67 Praha 3, Czech Republic

E-mail: {svatek, ondrej.zamazal, nguv03, ivanek}@vse.cz

^b Department of Philosophy, Prague University of Economics and Business, Nám. W. Churchillia 4, 130 67 Praha 3, Czech Republic

E-mail: vacuram@vse.cz

^c Department of Applied Informatics, Comenius University in Bratislava, Mlynská dolina, 842 48 Bratislava, Slovakia

E-mail: kluka@fmph.uniba.sk

Abstract. When reusing existing ontologies for publishing a dataset in RDF or developing a new ontology, preference may be given to those providing extensive subcategorization for the classes deemed important in the new dataset schema or ontology (focus classes). The reused set of categories may not only consist of named classes but also of some compound concept expressions viewed as meaningful categories by the knowledge engineer and possibly later transformed to a named class, too, in a local setting. We define the general notion of focused categorization power of a given ontology, with respect to a focus class and a concept expression language, as the (estimated) weighted count of the categories that can be built from the ontology's signature, conform to the language, and are subsumed by the focus class. For the sake of tractable experiments we then formulate a restricted class expression language based on existential restrictions, and heuristically map it to syntactic patterns over ontology axioms (so-called FCE patterns). The characteristics of the chosen concept expression language and associated FCE patterns are investigated using three different empirical sources derived from ontology collections: first, the class expression pattern frequency in class definitions; second, the occurrence of FCE patterns (mapped on the class expression patterns) in the Tbox of ontologies; and last, for two different samples of class expressions generated from the Tbox of ontologies (through the FCE patterns) their ‘meaningfulness’ was assessed by different groups of users, yielding a ‘quality ordering’ of the concept expression patterns. The different types of complementary analyses / experiments are then compared and summarized. Aside the various quantitative findings, we also come up with qualitative insights into the meaning of either explicit or implicit compound concept expressions appearing in the semantic web realms. To allow for further experimentation, a web-based prototype (named OReCaP) was also implemented, which covers the whole process of ontology reuse from keyword-based ontology search through the actual FCP computation to the selection of ontologies to be reused and their enrichment with new named concepts defined through the chosen compound expressions.

Keywords: ontology reuse, concept expression, focus class, categorization, OWL, ontology pattern

1. Introduction

*Corresponding author. E-mail: svatek@vse.cz

The main motivation of providing machine-readable semantics to data on the web in the form of ontolo-

gies is that of achieving interoperability of independently built data sources and applications. For example, if the same kind of product offered by different e-shops is semantically described using the same web ontology, comparison and automatic recommendation of these offers can be provided to customers; similarly, the spending structure of different municipalities can be jointly analyzed if the spending categories are linked to common concepts.

Obviously, interoperability depends not only on the existence of ontologies but also on their *reuse*. Rather than coining new entities in isolation, a dataset publisher should invest into finding relevant ontologies and integrating their entities into the schema of the given dataset. Similarly, the designers of a new ontology should consider reusing parts of existing ontologies from the same domain. Besides the interoperability benefit, reusing a categorization structure already existing in another ontology may also save a part of the design effort.

Since the majority of ontologies is nowadays published in the same standard language, OWL [2], the reuse is easy from the technological point of view, whether the method is the direct reuse of existing ontology entities or their subsumption/equivalence mapping from the dataset schema or from the new standalone ontology. However, despite the general agreement on the benefits of ontology reuse, this best practice is not massively adhered to yet [5]. One reason might be that selecting an ontology, or a fragment of it, suitable for being reused, from a larger pool of ontologies (typically pre-selected via keyword-based search in ontology repositories) is a non-trivial task for which only recently formal methods have emerged. They mostly rely upon

- ontology/entity popularity metrics, such as how many instances in how many different datasets already refer to them, and
- their a priori credibility [22], especially, whether the ontologies are listed in catalogs such as LOV – Linked Open Vocabularies¹ [28].

However, these approaches face the ‘cold start’ problem. Due to rapid growth of the ontological ‘ecosystem’ on the web, many emerging ontologies relevant for a certain dataset (or, generally, a reuse case) might not yet have achieved significant popularity ratings. Furthermore, the fact that an ontology as whole is the-

matically related to the reuse case does not mean that it structurally fits well the data that is to be semantically described. All this calls for complementing such ‘extrinsic’ sources of evidence with ‘intrinsic ones’, reflecting what is in the ontology itself, beyond mere unstructured keyword matching; the ontology’s axiom structures should be examined.

1.1. Working assumptions

Our proposed approach to enhancing web ontology reuse is based on the assumptions that

1. A large part of the use cases of ontologies on the web consists in *assigning* data objects to certain *categories* (with some consequences following from this assignment). Furthermore, prior to the assignment, the objects are already known to be instances of some (more general) class, to which we will refer as the *focus class* (FC).
2. The categories being assigned may not necessarily have the form of an *atomic concept* (named class) explicitly present in the ontology, but can also be *compound concept expressions* (anonymous classes) that are built from atomic entities of the ontology (classes, properties and sometimes individuals) using elements of the underlying description logic (DL) language.
3. The number and ‘quality’ of the categories an ontology provides for a focus class, summarized by a metric (or, a family thereof) we will denote as *focused categorization power* (FCP), can serve as indicator of the reusability of the ontology for a dataset containing objects semantically corresponding to this class as its instances.
4. The ‘quality’ of the *compound* categories is strongly correlated with the degree to which users would consider *creating a named class* (equivalent to this category) as meaningful.

The first two assumptions are supported, among other, by the findings of a study on competency questions by Ren et al., from 2014 [20]. They collected 168 competency questions (CQ) from two ontology projects, and clustered them into twelve archetypes. Of them, at least three correspond to tasks that can be characterized as sub-categorization of instances of a relatively generic class (here, pizza or software), namely:²

¹<https://lov.linkeddata.es/>

²CEx# stands for a concept expression (in fact, rather a named class), OPE for an object property expression, and I for an individual (object).

- #1: “Which [CE1] [OPE] [CE2]?” (e.g., “Which pizzas contain pork?”)
- #3: “What type of [CE] is [I]?” (e.g., “What type of software (API, Desk-top application etc.) is it?”)
- #4: “Is the [CE1] [CE2]?” (e.g., “Is the software open source development?”)

(For completeness: other archetypes have to do with quantitative features of objects or with relationships between specific objects. Interpreting them as sub-categorization would be less natural.)

Note that while the second assumption gives a value to compound concept expressions, we can realistically expect that for certain reuse cases (e.g., when the categories have to be imported into a static hierarchy such as a thesaurus or product catalog on a web page) the compound concept expressions have to be structurally and lexically *transformed* to named classes.

The validity of the third assumption obviously depends on the meaning of “(ontology) provides (categories)”. Namely, an ontology is, structurally, a collection of *axioms* rather than of categories (concept expressions). The generation of categories from ontology axioms will essentially be a heuristic process.

The fourth assumption brings a bottleneck: the respective notion of quality escapes an automatic empirical evaluation, as it depends on the subjective perception of users. We however believe that the direct human perception factor is indispensable for assuring the versatility of the reused schema with respect to different use cases. We could of course imagine some alternative, indirect quality measures, in particular, the distinction whether an expression would fit the *expressiveness of the query engine* to be used for querying the dataset described with the ontology. There could be however many mutually overlapping queries, and the expressions corresponding to each of them would only mildly contribute to the semantic underpinning of the dataset. Imagine for example two ontologies, one with two relevant atomic concepts (subclasses of the focus class) and one with four. Without caring for human perception, we might be tempted to construct the compound categories using the conjunction operator, since conjunctive queries are the cornerstone of ontology-based data access. The second ontology would then be assigned (set aside other aspects) a five times higher categorization power (15 vs. 3 possible non-empty conjunctions) than the first ontology. It could however easily happen that none of the conjunctions would not have any special meaning beyond that

trivially following from its constituents, which would make this strong degree of preference for the second ontology rather inadequate. For these reasons, and also due to the fact that the computational use of the reused ontologies may be rather diverse and unpredictable a priori, we view the FCP computation as rather agnostic of this usage.

Having said that, we still acknowledge that the reuse process in whole should be aware of the subsequent application context. Namely, the FCP computation will, aside its numerical side, also imply the *categories themselves* being collected into a list. This list can then be processed in the reuse task in the narrow sense: concept picking, within their associated axioms, from the ontologies. It is in this later phase that the subsequent use of the structures will come into play, including the expressiveness of the tools such as query engines and reasoners. Most notably, if their expressiveness is high and DL reasoning is to be applied over the data, as many axioms as possible should be picked from the source ontologies. On the other hand, if the published data is only to be processed by simple tools, transferring the complete axiomatization might not be necessary.

1.2. Motivating example

Let us now present a concrete motivation example, which we will use (among other examples) throughout the paper to illustrate various components of our approach.

Example 1 Let us consider a used vehicle retailer website that is to be enhanced with structured descriptions of the offered items (vehicles) in RDF. The descriptions should reuse ontology/ies whenever possible, in view of achieving interoperability with (presently unknown) applications of partners, search engines and aggregators. How could the computation of FCP be useful when assessing the potential of existing ontologies for being reused in this case? And how would the reuse then take place?

- Various customers are interested in different *categories* of vehicles (in terms of technical parameters, make, operation history, etc.); the reused ontologies thus should allow to represent as many such user-demanded categories as possible. The categories would then be assigned to data items representing the corresponding vehicles within the website, this way made easier to search and

browse. Note that the number of vehicle categories may not be correlated with the number of classes in the ontology. For the task of recommending a particular vehicle, categories refining the ‘vehicle’ concept are more relevant than others. For example, if a considered ontology O were a broader one covering transport in general, its capability to express categories of, e.g., traffic signs, would not be an argument in favor of its reuse in the given case.

- The (re)usability analysis of a given ontology O should first seek in O (using lexical methods, possibly including thesauri etc.) a class expressing the general notion of ‘vehicle’, make it the *focus class*, and then perform the FCP calculation with respect to it. A high value of the FCP will indicate the suitability of O for becoming a part of the schema of the website’s data.
- The whole or part of the reused ontology/ies (after a merging step, if more ontologies are reused for the ‘vehicle’ concept) would possibly become a product taxonomy published as a navigation structure of the shop website. In the navigation structure, the compound concepts then should not be served in the form of logical formulas, but would first be *transformed* (ideally, in a partially automated way) into regular noun phrases that would, explicitly or implicitly, correspond to *named OWL classes*.

To name some hypothetical applications (in the context of this particular domain) where the category-centric processing would make sense:

- within the retailer website, the user could *navigate* over a (multi-)tree consisting of both original and constructed categories;
- the designed categories would be considered as dimensional values in an internal *sales reporting* dashboard;
- meaningful sales segments to be coherently *marketed* would be identified.

1.3. Research contributions and paper structure

The research presented in this paper aims to bring the following contributions:

1. Introduce the notion of *focused categorization power* of ontologies as a fundamentally new concept in analyzing ontologies (particularly relevant in ontology reuse scenarios, but possibly

even beyond), position it with respect to *related research*, and provide it with a *formal underpinning* and an *algorithmic solution*.

2. Explore a particular *concept expression language* (for compound expressions) over which the FCP can be computed, and both empirically and analytically demonstrate that this language, based on simple existential restrictions, is indeed a relevant one for this first approximation.
3. For this language, provide a collection of *syntactic patterns* (called FCE patterns) through which the expressions can be constructed from the ontology axioms (primarily RDFS-level ones: subclassing, domain and range), thus making the FCP calculation operational. The FCE patterns are then verified both *empirically* (how prevalent they are in ontology collections) and through *cognitive experiments* (how reliable they are in producing ‘sensible’ categories); their tentative *weighting* for the FCP computation is thus induced.
4. Finally, further experimentation is made easy for other researchers by making available a prototype *implementation* of the whole approach, including a convenient GUI. (Notably, the tool also provides additional functionality not having a theoretical underpinning in the paper: the initial ontology and focus class search, and the ultimate selective reuse of concepts from the chosen ontologies.)

The present paper is an evolution of a previous conference paper [24], which provided a brief explanation of the notion of FCP, proposed an initial concept expression language, equipped the individual concept expression types (conforming to this language) with patterns for their construction based on the ontology Tbox, and provided the results of a cognitive experiment on compound category assessment. The present paper however extends the previous one along numerous axes, and the majority of its volume consists of entirely new content:

- The *core FCP model* has been completely reworked and extended by a number of notions, thus gaining more formal rigor. Among other, the notion of concept expression language (CEL) is now formally defined, and several instances of it introduced, including the CEL of compound expressions used further in the experiments.
- The relationship between concept expressions and *syntactic patterns* in ontologies that trigger

the consideration of these expressions for the FCP is now more elaborated.

- The survey on syntactic pattern *occurrence in ontology collections*, only present in an online addendum to the previous paper [24], is now an integral part of the paper, and has been re-run to provide fresher statistics covering the ontologies meanwhile contributed to the LOV catalog. Its results are now also more thoroughly discussed.
- For the existing ontologies, beside the occurrence of heuristic patterns, the presence of the actual *compound concept expressions* (inside axioms) has been explored, covering three ontology collections.
- The *cognitive experiment* has been repeated on a different data sample, with more guidance provided to the experimental subjects, and with additional meta-data collection.
- The different analyses / experiments in the paper are now framed by a *comparative analysis*.
- The prototype *implementation* (under the name OReCaP), made available both as a running instance and in source code, is an entirely new result.³

From the methodological point of view, the research presented in the paper is a mix of conceptual reflections, (lightweight) formalization, empirical analysis of existing data sources, experimentation involving human subjects, and prototype implementation allowing for further experiments even by third parties. As consequence, the different ‘facets’ of the research are only elaborated to a limited degree. In particular, a thorough study spanning across many candidate concept expression languages was left to future work. Similarly, the implemented prototype is merely provisional; due to its confinement to the single initial concept expression language, its main goal is to support initial experiments by fellow researchers, demonstrating our approach on whatever material they choose, rather than to provide guaranteed performance in real-life ontology reuse scenarios. Therefore, its systematic evaluation was also deferred until a more advanced version supporting a wider scope of concept expression languages. Finally, the formal apparatus introduced is merely descriptive and does not aim to elaborate on non-obvious formal properties of the elements of the

approach. These simplifications were needed in order to integrate the multiple complementary views of the focused categorization problem within the limited space of a single paper.

The rest of the paper is structured as follows. Section 2 presents the formal framework of focused categorization power in general terms, and provides its formalization. Section 3 introduces a particular concept expression language having suitable computational properties for an initial study. Section 4 complements this language with syntactic patterns (FCE patterns) allowing to generate concept expressions conforming to it from the ontology Tbox. Section 5 surveys the occurrence of class expression patterns in ontology axioms. Section 6, analogously, surveys the occurrence of the previously defined syntactic patterns (FCE patterns) in the Tbox of ontologies. Section 7 describes a series of cognitive experiments in which humans provided an assessment of ‘meaningfulness’ of concept expressions belonging to different patterns. Section 8 provides a comparison of those complementary surveys and experiments and a summarizing discussion. Section 9 describes a tentative operationalization of the results from the analyses, and explains the functionality of the implemented OReCaP prototype. Section 10 reviews some related methods and projects. Finally, Section 11 wraps up the paper and outlines the directions for future work.

2. General framework of focused categorization power

The aim of this section is to formally underpin the whole approach as well as to motivate the empirical analyses to which most of the remainder of the paper is devoted. We will proceed from the notion of *concept expression language* to the notion of *L-category*, as a concept (class) expression refining a (focus) class and conforming to some concept expression language, and then to the central notion of *focused categorization power* (FCP). Finally, we analyze what sources may inform the computation of the FCP through a *weighting* mechanism and how the FCP can be approximated by matching FCE patterns in the ontology.

2.1. Concept expression language

OWL [2], being a primarily *concept-based* formal ontology language, allows to create compound class expressions of many kinds using a collection of con-

³Its very early version was presented as a short demo paper [17]; the current version however has substantially enhanced functionality.

structors. The use of the syntactic constructors can be restricted in different ways, producing a formal system in logic often called a fragment or sublanguage. There is a number of decidable fragments of description logics [11], only some of them being useful for semantic web tasks. The so-called OWL Profiles as sublanguages of OWL defined in the current OWL 2 standard are examples of such fragments [1]; however, there may be other non-standard restricted sublanguages of OWL that may be useful for other semantic web community needs.

Therefore we define the notion of *concept expression language*⁴ that generalizes the notion of a restricted formal sublanguage of OWL to that only usable for the creation of class expressions; we use the description logic (DL) syntax instead of the textual functional-style OWL syntax, for brevity.

The notion of concept expression language, in our terms, refers essentially a set of class expression *patterns* constructed over a set of variables together with a set of allowed *substitutions* for those variables. Given a particular ontology signature, a concept expression language generates a set of concept expressions by applying all allowed substitutions to all patterns.

Before formally defining the notion of concept expression language proper, let us first recall or introduce a few preliminary notions. A *signature* is a triple $\mathcal{S} = (\mathcal{C}_\mathcal{S}, \mathcal{P}_\mathcal{S}, \mathcal{I}_\mathcal{S})$ of sets of classes $\mathcal{C}_\mathcal{S}$, properties $\mathcal{P}_\mathcal{S}$, and individuals $\mathcal{I}_\mathcal{S}$, collectively called *entities*. The *signature of an ontology O* is the signature $\text{sig}(O) = (\mathcal{C}_O, \mathcal{P}_O, \mathcal{I}_O)$ where \mathcal{C}_O , \mathcal{P}_O , \mathcal{I}_O are respectively the sets of all classes, properties and individuals occurring in O (cf. [2, Sect. 5]). Note that the punning feature of OWL 2 allows these sets to be non-disjoint. By $\mathcal{PE}_\mathcal{S}$ and $\mathcal{CE}_\mathcal{S}$, we denote, respectively, the sets of property expressions and class expressions (CEs) over the signature \mathcal{S} , defined in the standard way [2, Sects. 6 and 8].

In order to bring in the notions of variable and substitutions, let us introduce a special signature $\mathcal{V} = (\mathcal{C}_\mathcal{V}, \mathcal{P}_\mathcal{V}, \mathcal{I}_\mathcal{V})$ such that the sets $\mathcal{C}_\mathcal{V} = \{C, C_1, C_2, \dots\}$, $\mathcal{P}_\mathcal{V} = \{P, P_1, P_2, \dots\}$, $\mathcal{I}_\mathcal{V} = \{i, i_1, i_2, \dots\}$ are countable, infinite, and mutually disjoint. The entities in $\mathcal{C}_\mathcal{V}$, $\mathcal{P}_\mathcal{V}$, $\mathcal{I}_\mathcal{V}$ will be, respectively, called *class variables*, *property variables*, and *individual variables*, and collectively we will call them *variables*. Class expressions $\mathcal{CE}_\mathcal{V}$ and property expressions $\mathcal{PE}_\mathcal{V}$ over \mathcal{V} are

⁴Formally we should use the term ‘OWL concept expression language’; we omit the ‘OWL’ attribute for brevity, since we do not go beyond the limits of OWL in any part of this research.

then called *class expression patterns* (CE patterns, for brevity) and *property expression patterns*, respectively.

Further, given a signature \mathcal{S} , any function σ that maps class variables to class expressions over \mathcal{S} (i.e., $\sigma(C) \in \mathcal{CE}_\mathcal{S}$ for each $C \in \mathcal{C}_\mathcal{V}$), property variables to property expressions over \mathcal{S} ($\sigma(P) \in \mathcal{PE}_\mathcal{S}$ for each $P \in \mathcal{P}_\mathcal{V}$), and individual variables to individuals in \mathcal{S} ($\sigma(i) \in \mathcal{I}_\mathcal{S}$ for each $i \in \mathcal{I}_\mathcal{V}$) is called a *substitution*. Let $\Sigma_\mathcal{S}$ denote the set of all substitutions for a signature \mathcal{S} . Substitutions can easily be inductively extended to property and class expression patterns in the obvious way, e.g., $\sigma(p^-) = \sigma(p)^-$, $\sigma(c_1 \sqcup c_2) = \sigma(c_1) \sqcup \sigma(c_2)$, $\sigma(\exists p.c) = \exists \sigma(p).\sigma(c)$, $\sigma(\{i_1, \dots, i_n\}) = \{\sigma(i_1), \dots, \sigma(i_n)\}$, etc. (cf., Baader & Narendran [4]). With preliminaries in place, we may proceed to our definitions.

Definition 1 A concept expression language (CEL) is a pair $\mathcal{L} = (\mathcal{CP}_\mathcal{L}, \mathcal{R}_\mathcal{L})$, where

- $\mathcal{CP}_\mathcal{L}$ is a set of class expression patterns, and
- $\mathcal{R}_\mathcal{L}$ is a substitution-restricting function mapping every signature \mathcal{S} to a subset of $\Sigma_\mathcal{S}$, the set of allowed substitutions.

One example of a CEL may be the trivial language of named concepts, let us call it \mathcal{L}_{nam} for future reference, where $\mathcal{CP}_{\mathcal{L}_{nam}} = \{C\}$, and $\mathcal{R}_{\mathcal{L}_{nam}}$ only allows a named (atomic) class to be substituted for C . Thus for every signature \mathcal{S} , $\mathcal{R}_{\mathcal{L}_{nam}}(\mathcal{S})$ only contains substitutions that map the class variable C to a member of $\mathcal{C}_\mathcal{S}$.

Another example would be a CEL, let us call it \mathcal{L}_{extop} (language of existential restrictions to the top concept), such that $\mathcal{CP}_{\mathcal{L}_{extop}} = \{\exists P.T\}$ and for each signature \mathcal{S} , $\mathcal{R}_{\mathcal{L}_{extop}}(\mathcal{S})$ contains only substitutions that map the property variable P to an atomic property of \mathcal{S} .

For clarity, we explicitly define the notion of class expressions generated by a CEL:

Definition 2 Given a CEL $\mathcal{L} = (\mathcal{CP}_\mathcal{L}, \mathcal{R}_\mathcal{L})$ and a signature \mathcal{S} , any class expression $\sigma(c)$ obtained by applying a substitution $\sigma \in \mathcal{R}_\mathcal{L}(\mathcal{S})$ to a class expression pattern $c \in \mathcal{CP}_\mathcal{L}$ is an \mathcal{L} -expression in \mathcal{S} . We say that the expression $\sigma(c)$ conforms to (also matches) the CE pattern c .

2.2. \mathcal{L} -categories and focused categorization power

The notion of CEL allows us to define the notion of *\mathcal{L} -category* for a given OWL class, as a specific kind of \mathcal{L} -expression, and the associated notion of *differentiating concept expression*.

Definition 3 Let $\mathcal{L} = (\mathcal{CP}_{\mathcal{L}}, \mathcal{R}_{\mathcal{L}})$ be a CEL. Let us assume an OWL ontology O with signature $\text{sig}(O) = (\mathcal{C}_O, \mathcal{P}_O, \mathcal{I}_O)$ and a named class $FC \in \mathcal{C}_O$. An \mathcal{L} -category⁵ for FC in O is then any concept expression in the form $FC \sqcap D$, where D is an \mathcal{L} -expression in $\text{sig}(O)$. D is then called a differentiating concept expression (DCE) of the \mathcal{L} -category.

Within this paper we will sometimes informally refer to a DCE as to the *subcategory* of the class FC . Note, however, that DCE is a mere syntactic notion, as for some FC and D it would not hold that $FC \sqcap D$ would ‘differentiate’ among the instances of FC – i.e., as we will discuss later, $FC \sqcap D$ may not be extensionally a proper subconcept (subcategory) of FC .

The capability of O to finer subcategorize the resources that are already known to belong to FC can be, at an abstract level, expressed in terms of the number as well as ‘quality’ of \mathcal{L} -categories for FC in O , for a suitably selected \mathcal{L} . We will denote the measure of this capability as *focused categorization power* (FCP) of O with respect to FC over \mathcal{L} ; the class FC plays the role of a *focus class*. Formally, at an abstract level:

$$\text{FCP}(O, FC, \mathcal{L}) = \sum_{Cat \in Cats}^k w(Cat, O, FC) \quad (1)$$

where $Cats$ is the set of all \mathcal{L} -categories for FC in O unique under the inferential closure (i.e., one representative of each equivalence class of categories is present in $Cats$) and w is a *weight function* assigning every Cat a weight from $[0, 1]$, expressing its ‘quality’ as a category of FC .

2.3. Weight function desiderata

Intuitively, leveraging on some of the assumptions from Section 1.1, an ideal weight function $w(Cat, O, FC)$ should take into account the following four characteristics of an \mathcal{L} -category Cat :

Absolute size The weight w must be 0 if the extension of Cat is an empty set (Cat is an unsatisfiable concept). w should also be low if the extension of Cat is a singleton (or, very small) set.

⁵For simplicity, we will sometimes simply write ‘category’ rather than \mathcal{L} -category throughout the paper, provided the language \mathcal{L} is either clear from the context or irrelevant.

Relative size A hard constraint is that Cat must not be equivalent to its focus class FC (if it is, w must be 0). On the other hand, w should be highest if Cat contains a ‘reasonable’ proportion of the instances of FC , neither extremely high nor extremely low. For example, a category with around 10 instances may be quite meaningful for a focus class with a few dozen of instances (say, the DCE EuropeanMonarchy wrt. the focus class EuropeanCountry), but less so for a focus class with millions of instances (say, a DCE expressing the notion of ‘village next to a spacecraft launch location’ wrt. the focus class Village).

Expression complexity The weight w should increase with the decreasing complexity of Cat (namely, of its simplest equivalent form) as DL expression. The rationale is that a simpler category can be better decoded by people and more conveniently displayed.

Semantic coherence It might happen, especially for ontologies with broad coverage and when applying the inferential closure, that a category would be assembled from thematically unrelated entities. The weight w should be lowered in such cases. A real example derived from the DBpedia ontology (we also refer to it in Section 7.1) is the category of persons ‘beatified in a wine region’; it is unlikely that such a concept spanning across the religious and agricultural domain would be of high interest in any of them. However, this characteristic is obviously most subtle and least eligible for automatic assessment of all four.⁶

For the unrestricted OWL concept expression language (say, with $\mathcal{CP}_{\mathcal{L}}$ including all concept constructors of OWL 2), the size of $Cats$ would typically be very large (actually, infinite if $\mathcal{R}_{\mathcal{L}}$ allowed arbitrary substitutions) for any realistic ontology. Proportionally few categories would however satisfy the above characteristics: they would be unsatisfiable, overly complex, or the like. The extent of the set of categories should thus be constrained by some finer CEL chosen for constructing the DCEs, ideally, one contributing to meeting the desired characteristics. Obviously, the choice of the language straightforwardly addresses the *expression complexity* aspect. On the other hand, the *size of the category extension*, whether absolute

⁶Presumably, graph-based metrics relying on the number of different paths connecting the constituent entities might be applied. However, we do not elaborate on this topic in the current paper.

or relative, is only influenced indirectly by the language (and the semantic coherence is unlikely to be significantly correlated with the CEL choice). For the sake of this paper, we will anticipate a rather restricted CEL, whose expression complexity will not be a concern. The subsequent analyses thus primarily address the category quality in terms of the absolute or relative size only.

The least complex CEL allowing to construct meaningful DCEs is the previously introduced language of *named classes*, \mathcal{L}_{nam} . Each \mathcal{L}_{nam} -category is then a conjunction of the given focus class FC and another named class D . Intuitively, $w(FC \sqcap D, O, FC)$ should be 0 for such D that either $FC \sqcap D = \emptyset$ or $FC \sqsubseteq D$ (due to violating the hard constraint on absolute/relative size). On the other hand, w should be high (presumably even equal to 1) for the asserted (even transitive) subclasses of FC , since these are the concepts viewed as proper subcategories of FC by the ontology designer/s. Note however that w might deserve to be non-zero even for some other non-disjoint classes. For example, $Student \sqcap Teacher$ can be a valid category of the FC $Student$ in an academic ontology, since some graduate students may also have a teacher contract; obviously, none of the classes would be a subclass of the other.

2.4. Focused category pattern

While the CE patterns defining a given CEL allow for substitutions producing class expressions to be ‘counted’ in the FCP computation, such expressions cannot be directly extracted from the OWL code of the ontology, which consists of (TBox, and sometimes ABox) *axioms* and not of standalone expressions. We will thus introduce the general notion of *focused category pattern* – a syntactic pattern for transforming OWL axioms occurring in ontologies (primarily, in their Tbox) to class expressions that are to be used as the DCE of categories of a focus class. Intuitively, the definition of a focused category pattern relies on OWL axiom templates that can be matched with an OWL ontology.

Definition 4 A focused category extraction pattern (*an FCE pattern, for short*) is a 5-tuple $p = (c, Vars, Tpl, Cons, mf)$ where

- c is a CE pattern.
- $Vars$ is a sequence of mapping variables.

- Tpl is a set of conjunctively⁷ interpreted axiom templates, namely, OWL axioms (possibly) containing variables in the position of entities. One distinguished variable is (usually, though not mandatorily) the FC-variable, which is to be substituted by the focus class FC when matching the pattern. Every variable from $Vars$ has to appear in Tpl at least once, and there can also be other (non-mapping) variables in Tpl not present in $Vars$.
- $Cons$ is a set of FCE pattern validity constraints.
- mf is a mapping function, which has as input a tuple of entities (from the matched ontology) instantiating $Vars$, and returns a CE conforming to c , provided all the constraints from $Cons$ are satisfied by the tuple.

The constraints are, in fact, a ‘local’ method of assuring that some of those CEs for which w_{01}^I (in Equation 3) would be 0 are not generated through the pattern at all.

Considering we have two kinds of patterns (CE and FC), which both abstract over some OWL structures, a question naturally arises why these two could not be more integrated or based one upon the other: obviously, the OWL axioms have concept expressions as their left- and right-hand sides, and the same could hold at the pattern level in principle. There are however two factors that would make such an integration cumbersome or impossible:

- First, while the variables in CE patterns are strictly *typed* (in the sense that the abstract signature \mathcal{V} consists of sets of variables for classes, properties and individuals that are pairwise disjoint), the FCE patterns should be versatile wrt. possible use of *punning* in the ontology: across multiple axioms, the same identifier might, e.g., once refer to an individual and once to a class.
- Second, the role of the FCE patterns is not merely in matching the OWL structures but also in constructing concept expressions based on the match. The CE pattern of the constructed expressions will not always (or, probably, just rarely) correspond to the structure of expressions explicitly present in the axioms matched by the FCE pattern.

⁷For simplicity we only consider conjunctions of axioms. Disjunctions could be expressed by means of multiple FCE patterns for the same CE pattern, if needed.

For these reasons, we keep the two notions as completely separate. In the rest of the paper we will carefully distinguish which type of pattern is being discussed (using the ‘CE’ and ‘FC’ acronyms) unless the type of pattern is entirely clear from the context.

2.5. Hypothetical sources of category weight

Automatic assignment of w to a certain \mathcal{L} -category $Cat = FC \sqcap D$ for the sake of $FCP(O, FC, \mathcal{L})$ computation in (1) could be based on a number of different sources, possibly in combination:

1. *Inferential relationships* between the focus class FC , the entities from which D consists, and possibly other entities in O . As pointed above for the specific case of D being a named class, but valid generally, if $FC \sqcap D = \emptyset$ or $FC \sqsubseteq D$ then w must be 0. Moreover, if $|FC \sqcap D| = 1$ then $FC \sqcap D$ is not a very useful category either. This happens, e.g., with D being $\exists P.\{i\}$ such that i is a particular individual and P is an inverse-functional property (inferentially yielding $|D| \leq 1$).
2. Heuristics based on *FCE patterns*, applied on the context of FC and (constituent entities of) D in O (Tbox).
3. Frequency of instantiation of Cat in some *RDF dataset/s* (Abox).
4. General likelihood that the *class expression pattern* $c \in \mathcal{CP}_{\mathcal{L}}$ to which D conforms produces a meaningful category. This may relate
 - to the complexity of c , and,
 - to the overall distribution function of absolute/relative extension size of the CEs conforming to c , i.e., how many conforming CEs have a ‘reasonable’ extension, on average.

In one sentence, to judge the quality of a category Cat in ontology O , we can look at the Tbox of O (sources 1 and 2), at the Abox of O (source 3), and at the quality of the CE pattern that corresponds to Cat , independent of O and the particular Cat (source 4). Hence, while the first three sources consider Cat individually, the fourth one only considers its CE pattern – which is universally applicable but necessarily rough.

Let us demonstrate these factors on a simple (artificial) example from the vehicle domain of our motivating example.

Example 2 A considered ontology contains in its signature a focus class Car ⁸ and a property $hadAccident$. In a CEL For the existential restriction \exists and the top concept \top (such as \mathcal{L}_{extop} mentioned in Section 2.1) the category $Car \sqcap \exists hadAccident. \top$ can be constructed. The DCE of this category conforms to the general CE pattern $\exists P. \top$. Through the lenses of the four different sources mentioned at the beginning of this subsection:

1. *Inference*: If the axioms of the ontology are consistent with reality, the *reasoner* would not infer any prohibitive statement, since the set of cars that have had an accident is neither equivalent to the set of all cars, nor is it empty (and not even a singleton). The source thus does not ‘forbid’ this category (does not shrink its weight).
2. *FCE patterns*: Let the property $hadAccident$ have Car as its domain ($rdfs:domain$). An *FCE pattern* based on the $rdfs:domain$ property (see Section 4 for details) is then matched in the ontology TBox. The rationale of heuristically linking this FCE pattern to the CE pattern $\exists P. \top$ can be illustrated as follows: since everything having had an accident (more literally wrt. the expression structure, ‘having had *something* as its accident’) is inferred to be a car thanks to the domain axiom, it is intuitive to assume that there might exist both cars that have had an accident and those that have not.⁹ Thus, again, the source suggests the category as meaningful.
3. *Abox*: Let us assume there is already an external, publicly available car *dataset* published using this ontology, which contains 2500 instances of Car , of which 1000 appear in the subject position in at least one triple with predicate $hadAccident$. The category, if translated to a SPARQL¹⁰ query to this dataset,¹¹ is instantiated

⁸For easy readability of the DL formulas, we will mostly use the simple DL notation without IRI prefixes in the examples. The namespace will be irrelevant for artificial examples and clear from the context in real-world examples.

⁹Hypothetically, if all cars in the universe have had an accident, $Car \sqcap \exists hadAccident. \top$ would not have been a meaningful subcategory of Car . However, a carefully modeled ontology should then have contained the respective existential axiom as a necessary condition on the Car class ($Car \sqsubseteq \exists hadAccident. \top$). The presence of such conditions, disabling the FCE patterns for use in the particular case, is actually accounted for in our approach, see again Section 4.

¹⁰<https://www.w3.org/TR/sparql11-overview/>

¹¹SELECT DISTINCT ?c {?c a :Car ; :hadAccident []}

1000 times, which yields a solid absolute size figure, and also the relative size wrt. the focus class frequency ($1000/2500 = 0.4$) is fine. Even this third source should thus have a positive effect on the weight.

4. *CE pattern:* As we partially verify in Section 7, the CE pattern $\exists P.T$ is a priori not as likely to produce good categories as are some other CE patterns. This would have led to a weight decrease; however, in this case, the support of the *specific* CE by Abox data should overrule the negative impact of the mere *CE pattern* this CE conforms to.

2.6. Category-specific vs. CE-pattern-inferred weight

A fully informed category weight computation (for the sake of FCP computation, see Equation 1) could be possibly decomposed as

$$w(FC \sqcap D, O, FC) = \\ w^I(FC \sqcap D, O, FC) \otimes w^P(pattern(D)) \quad (2)$$

where

- w^I is an individual weight function for the given category (presumably based on the first three sources from the list in Section 2.5: inferential relationships, FCE patterns, and possibly Abox occurrence),
- $pattern(D)$ is the CE pattern that the (differentiating) concept expression matches,
- w^P is the CE pattern weight function (based on the fourth source from the list in Section 2.5: likelihood of the CE pattern producing a meaningful category), and
- \otimes is a function for combining the two partial weights into a single weight of the category.¹²

As regards the CE pattern weight w^P , we are aware of three empirical sources, in turn, which might provide a *quality estimate* of a certain CE pattern (and

¹²As we exemplified (on ‘cars with an accident’) in Sect. 2.5, the presence of (trustworthy) Abox data should overrule the assumptions made based on the specific Tbox information (within $w^I(FC \sqcap D, FC)$) or by the CE pattern (within $w^P(pattern(D))$). A formula truly considering the Abox should thus grant it a more prominent position. However, since we resign on the impact of the Abox in the next elaboration of the same formula (Eq. 3), we may afford such a simplification.

thus, indirectly, the individual categories in which the DCE would match this CE pattern):

- *Humans*, who can assess concrete expressions conforming to this CE pattern as less or more meaningful subcategories of a focus class.
- *RDF datasets* instantiating expressions of this pattern (under the inferential closure) with smaller or bigger numbers of Abox statements.
- Existing ontology *axioms* in whose right-hand sides¹³ (RHS) an expression matching the given CE pattern appear less or more often.

Data collected from these sources then have to be *aggregated* across the different expressions matching the same CE pattern; the human feedback also has to be aggregated over the participating human assessors. All three sources also have their biases and limitations: *human assessment* is generally subjective, *RDF data* may not yet be available for newer ontologies, and the presence of *axioms with a compound RHS* is biased by the higher/lower tendency of different communities to axiomatize ontologies; the design rationale for the axioms (intended for DL reasoning) may also differ from the usefulness criteria of categories meant to subcategorize a focus class.

Orthogonally, we can also aggregate the occurrence counts of indicative *FCE patterns* (heuristically mapped to CE patterns) in ontologies. This source, in turn, will indicate the upper bound of the *quantity* of categories that can be obtained for a focus class using such patterns (whatever the degree of *quality* of these categories would be).

To avoid any mismatch of the presented ‘weight sources’ list with the ‘weight sources’ list from Section 2.5, note that the sources from Section 2.5 are applied ‘deductively’, to estimate the weight of a particular category, while the sources in this section serve for ‘inductive’ derivation of the (mean) weight pertaining to a whole CE pattern.

In different sections of the paper we will empirically investigate these sources: the CEs in axioms in Section 5, the FCE patterns in Section 6, and the human assessment of FC+subcategory pairs in Section 7 (thus only deferring the Abox analysis to later research), and discuss their potential and limitations in more detail.

¹³Although the OWL grammar allows for general concept inclusions, having a compound concept on the left-hand side (see <http://ontogenesis.knowledgeblog.org/1288/>), its usage is currently marginal. For example, the 2017 snapshot of the LOV catalog only contained three ontologies featuring them.

The goals of empirical analysis are both to identify interesting *insights* on these sources in general and to suggest tentative *CE pattern weights* for the purpose of FCP computation.

2.7. Tbox-based approximation of FCP via patterns

We will now specifically consider the situation when the Abox information on the usage of a specific category is either unavailable or unreliable, and we thus have to derive the category weight from the *ontology* alone. Under such conditions, the individual weight function w^I is likely to be close to Boolean, since relevant deductive inference (or some highly reliable heuristics over the axioms' syntax, expressed as an FCE pattern) can, in most cases, only tell us that a category will be unsatisfiable, with a singleton (or otherwise very small) extension, or equivalent to the FC, all of which will make its quality inferior. Under the assumption of a *Boolean individual weight function*, let us denote it as w_{01}^I , we can then express the overall weight of a category as a simple product:

$$w(FC \sqcap D, O, FC) = w_{01}^I(FC \sqcap D, O, FC) \cdot w^P(pattern(D)) \quad (3)$$

which will yield

- 0 if $w_{01}^I = 0$: this happens when the reasoner arrives at some specific conclusion ('unsatisfiable', 'singleton' or 'FC-equivalent', see the beginning of Section 2.5) about the extension of the category according to O , and thus judges the category as useless, and
- $w^P(pattern(D))$ if $w_{01}^I = 1$: this happens when the reasoner does not arrive at such a kind of conclusion.

Let us then assume that we distinguish among k CE patterns allowed in the CEL \mathcal{L} , $\mathcal{CP}_{\mathcal{L}} = \{c_1, \dots, c_k\}$, each associated with a weight via the CE pattern weight function w^P . Let each CE pattern c_i be also associated with a set of FCE patterns $PS_i = \{p_{i1}, \dots, p_{ij}\}$ that allow to construct the corresponding categories from the axioms of ontologies. The validity constraints $Cons_{i1}, \dots, Cons_{ij}$ contained in these FCE patterns are, in fact, a 'local' method of assuring that some of those CEs for which w_{01}^I (in Equation 3) would be 0 are not generated through the pattern at all.

Then we can express the *approximate FCP* of an ontology O for a focus class FC over \mathcal{L} with respect to

the (set of) FCE pattern sets $PS = \{PS_1, \dots, PS_k\}$ as the weighted sum of the FCE pattern occurrence counts in O :

$$\widehat{FCP}(O, FC, \mathcal{L}, PS) = \sum_{i=1}^k Occ(PS_i, FC, O) \cdot w^P(c_i) \quad (4)$$

where $Occ(PS_i, FC, O)$ is the sum of FCE pattern occurrence counts in O for all FCE patterns $p_{ij} \in PS_i$. Note that with respect to (3), the multiplication by $w^P(c_i)$ (the CE-pattern-specific weight) is performed over the *counts* of FCE pattern matches (thus, properly constructed categories) and not over the individual category weights, but the final weighted sum will be the same as if the weights of all individual categories produced by (3) were summed according to (1).

We will introduce usable examples of FCE patterns in Section 4 for a CEL that we will meanwhile introduce in Section 3.

2.8. Pattern-based CE extraction and approximate FCP computation algorithm

The computational process leading to extraction of class expressions using FCE patterns is rather straightforward with respect to the general description of the approach in Section 2.7, see Listing 1. Each call of the function **match**(O, FC, p_{ij}) returns a class expression generated by a new match of p_{ij} in O respective to the given focus class FC , and taking into account the constraints $Cons$ in the pattern; when no new CE can be generated, the function returns *false*. As a result, a set CE of pairs "CE pattern, CE" is returned. The only additional step is the pruning of mutually equivalent CEs, which should be only counted once (lines 14–31). First, for each CE from CE , an auxiliary named class is generated (by function **newIRI**) and stored, and an axiom defining this auxiliary class as equivalent to the given CE is added to an auxiliary version of the ontology (line 19). In the deductive closure of this auxiliary ontology, equivalence axioms between pairs of auxiliary classes are then sought (line 24), and for each such pair, only one of the CEs is retained. The choice of the CE to be discarded could be based on the weight of the CE patterns:

$$\begin{aligned} \textbf{discard}(c_i, ce_i, c_j, ce_j) &= ce_i \text{ if } w^P(c_i) < w^P(c_j) \\ &= ce_j \text{ otherwise} \end{aligned}$$

For completeness, we also include the approximate FCP computation algorithm, as Listing 2. It simply amounts to summation of CE pattern weights w^P over the collected class expressions.

```

1 Input:  $O$ ,  $FC$ ,  $PS = \{PS_1, \dots, PS_k\}$ 
2 Output:  $CE$ 
3 Procedure:
4  $CE := \emptyset$ 
5 for each  $PS_i$ 
6   for each  $p_{ij} \in PS_i$ ,
7      $p_{ij} = (c, Vars, Tpl, Cons, mf)$ 
8     do while  $\text{match}(O, FC, p_{ij}) \neq \text{false}$ 
9        $ce := \text{match}(O, FC, p_{ij})$ 
10       $CE := CE \cup \{(c, ce)\}$ 
11      enddo
12    endfor
13  endfor
14  $Aux := \emptyset$ 
15  $O_{aux} := O$ 
16 for each  $(c, ce) \in CE$ 
17    $aux := \text{newIRI}$ 
18    $Aux := Aux \cup \{aux\}$ 
19    $O_{aux} := O_{aux} \cup \{aux \equiv ce\}$ 
20 endfor
21  $O_{auxc} := \text{closure}(O_{aux})$ 
22 for each  $aux_i \in Aux$ 
23   for each  $aux_j \in Aux \setminus \{aux_i\}$ 
24     if  $\{aux_i \equiv aux_j\} \sqsubseteq O_{auxc}$  then
25       find  $aux_i \equiv ce_i$ ,  $aux_j \equiv ce_j$  in  $O_{aux}$ 
26       find  $(c_i, ce_i)$ ,  $(c_j, ce_j)$  in  $CE$ 
27        $ce_d := \text{discard}(c_i, ce_i, c_j, ce_j)$ 
28        $CE := CE \setminus \{(c_d, ce_d)\}$ 
29        $Aux := Aux \setminus \{aux_d\}$ 
30     endif
31   endfor
32 endfor

```

Listing 1: Pattern-based CE extraction algorithm

```

1 Input:  $CE$ 
2 Output:  $\widehat{FCP}$ 
3 Procedure:
4  $\widehat{FCP} := 0$ 
5 for each  $(c, ce) \in CE$ 
6    $\widehat{FCP} := \widehat{FCP} + w^P(c)$ 
7 endfor

```

Listing 2: Computation of approximate FCP

3. Simple existential CEL: \mathcal{L}_{SE}

Having explained the whole proposed general framework of FCP computation, let us now return to its most essential element: the language that determines what kinds of CEs are considered from the onset.

As set up by the second assumption from the introduction (importance of both atomic and compound CEs for the FCP), for meaningful analysis of the FCP computation landscape we have to combine some variant of the language of named concepts (\mathcal{L}_{nam} from Section 2) with a concept expression language of differentiating concept expressions (DCEs) involving some DL concept constructor. For the purposes of the first proof of concept we sought a ‘small’ CEL well tractable, both by humans and machines, in the empirical investigations. Compared to other common types of concept constructors, the *simple existential restriction*, $\exists P.C$ where only atomic classes can be substituted for C , has some nice features in this respect – though, admittedly, subjective (we are aware that similar arguments could also be found for other candidate ‘starter’ CELs):

- The *interplay* between properties and classes, which are prone to be viewed separately in some ontology design environments, e.g., Protégé, may intuitively give rise to concepts ‘interesting to look at’, whatever their ultimately judged contribution to the FCP will be.
- The number of matching CEs only grows linearly with the number of classes $|C_O|$ and number of properties $|\mathcal{P}_O|$ in the ontology (at most it can reach $|\mathcal{P}_O| \cdot |C_O|$).
- The existence of a property assertion (witnessing the validity of the existential restriction) can be easily *checked in the Abox*, even complying to the open-world assumption (OWA). This favors¹⁴ the existential restriction over the universal restriction or max-cardinality restrictions, whose validation has to rely on the closed-world assumption (CWA).

OWL concept constructors’ usability for focused categorization is analyzed in more detail in Section 5.

When considering the CE pattern $\exists P.C$, it may be practical to also consider some of its special cases

¹⁴The use of the Abox in the FCP computation process is a work in progress.

or extensions that do not significantly worsen the tractability:

- One of them is $\exists P.\top$, the *unqualified existential restriction*, i.e., the restriction to the class `owl:Thing` in OWL terms, indicating that there is a property link but its filler does not matter when matching Abox data. The upper bound of the number of matching CEs is $|\mathcal{P}_O|$. (In Section 2.1 we already introduced the CEL \mathcal{L}_{extop} , which only allows this CE pattern.)
- The other is $\exists P.\{i\}$, i.e., the *individual value restriction*. The upper bound of the number of matching CEs is $|\mathcal{P}_O| \cdot |\mathcal{I}_O|$, but $|\mathcal{I}_O|$ is rarely high, since ontologies normally contain few individuals;¹⁵ in addition, such ‘internal codelist’ individuals have their overarching class already assigned within the ontology, which may lead to pruning the number of relevant properties (only leaving those with the particular class in the range of the property).

Along with the CE patterns based on existential restriction, we will also consider the sole CE pattern of \mathcal{L}_{nam} (the language of named classes) mentioned in Section 2, since we believe it should be by default a part of any practical CEL setting.

Based on these consideration (and, additionally, eliminating some uninteresting edge cases) we defined a suitable CEL, for the sake of the empirical research described further in this paper, as follows.

Definition 5 *The simple existential CEL \mathcal{L}_{SE} is the CEL with the set of CE patterns $\mathcal{CE}_{\mathcal{L}_{SE}} = \{C, \exists P.\top, \exists P.C, \exists P.\{i\}\}$ and with $\mathcal{R}_{\mathcal{L}_{SE}}$ allowing for each signature $S = (\mathcal{C}, \mathcal{P}, \mathcal{I})$ only those substitutions that assign a named class from \mathcal{C} to the class variable C , and a named property from \mathcal{P} to the property variable P .*

Therefore, the CE patterns considered for the sake of the presented research are: a named class; an unqualified property restriction (i.e., one with \top as the filler); a property restriction with a named class as the filler; and, an individual value restriction (a property restriction with a singleton class as the filler).

¹⁵In long term we may be also interested in ontologies coupled with (especially, SKOS) codelists, for which the size of \mathcal{I}_O may grow much larger. To date, however, a clear coupling with codelists is uncommon at least for those ontologies listed in LOV. See Section 6, where we experimented with a tentative FCE pattern for such a setting.

Since \mathcal{L}_{SE} does not allow for concept nesting, it guarantees *finiteness* of the set $\{Cat_i\}$ of categories for any O and FC . With respect to the three *OWL 2 profiles*, OWL 2 EL, OWL 2 QL, and OWL 2 RL, \mathcal{L}_{SE} , though simple, is only within the scope of OWL 2 EL, since the individual value restriction is only possible in this profile. Since the role of the CEL in our context is however not to form a new ontology on which reasoners would be systematically applied, but merely to provide a ground on which the FCP calculation could be performed (the reasoner only being used for one-shot computation of the deductive closure of the ontology), this expressiveness aspect does not present any serious problem.

Table 1 gives an overview of the CE patterns:

- The first column assigns the CE patterns a numbering local to the given CEL setting, for convenience.¹⁶
- The second column indicates the structure of the CE pattern itself in DL notation; the CE patterns substitutions are already restricted according to $R_{\mathcal{L}_{SE}}$: C stands for a named concept (class), P for a role (object property), and i for an individual.
- The third column indicates which variables from the CE structure are to be substituted by corresponding $sig(O)$ elements. The set of variables in **c2** is a subset of those of **c3** and **c4**; therefore, if we consider one or more CEs for **c3** or **c4** with some property P then we should also consider the CE for **c2** with this P .
- The fourth column measures the length of the Abox path (as number of triples) connecting the individual j to be assigned to the CE with entities (‘responsible’ for the assignment) substituted for variables from the third column, in other words, the minimal size of the SPARQL graph pattern to be used for the CE instance detection in the Abox. The SPARQL pattern will rely on the `rdf:type` predicate (**c1**), on the specific predicate P (**c2**, **c4**), or on both (**c3**). For example, for **c3** the templated SPARQL graph pattern (where ‘a’ stands for `rdf:type` and the blank node [] for the existentially quantified individual of the restriction) is

$$\text{?j } P \text{ [a } C \text{] .}$$

¹⁶We denote these specific CE patterns in bold face, to differentiate them from CE pattern meta-variables using the math font with subscript notation (c_i). We apply this distinction also to focused category patterns (**p1**, **p2**, ... vs. p_i) in the next section.

Table 1
Summary of CE patterns in \mathcal{L}_{SE}

	Structure	Sign.	Abox	FCE pattern
c1	C	C	1	p1
c2	$\exists P.T$	P	1	p2
c3	$\exists P.C$	P, C	2	p3
c4	$\exists P.\{i\}$	P, i	1	p4

- The order of the CE patterns in the table reflects the increased complexity of their detection in the Tbox using FCE patterns specifically designed for \mathcal{L}_{SE} . Those FCE patterns (presented next in Section 4), identified by their numbers, are in the fifth column.

Looking back at the study of competency questions by Ren et al. [20] referenced in Section 1, we immediately see that while the archetypes #3 (“What type of [CE] is [I]?”) and #4 (“Is the [CE1] [CE2]?”) assume named subclasses of the focus class, i.e. are covered by the CE pattern **c1**, archetype #1 (“Which [CE1] [OPE] [CE2]?”) assumes categorization via an object property and a target class (or individual), i.e. is covered by the CE pattern **c3** (or **c4**).

4. Focused category extraction patterns for \mathcal{L}_{SE}

Having defined \mathcal{L}_{SE} as a particular CEL, we can now proceed to formulating FCE patterns suitable for this CEL, through which categories can be constructed from the ontology axioms. We will then also briefly review the algorithmic approach for pattern-based CE generation, and outline the usage of patterns in the (optional) named class generation step. While these two aspects can be possibly generalized to any CEL, we stick to their variant relevant in the \mathcal{L}_{SE} setting, for simplicity.

4.1. FCE pattern collection for \mathcal{L}_{SE}

We present below five FCE patterns for \mathcal{L}_{SE} and illustrate them on our running example of used cars.

Pattern p1 Let us first consider the FCE pattern for the CE pattern **c1** of \mathcal{L}_{nam} ; the axiom templates are serialized as triples in the Turtle¹⁷ RDF notation:

$$\mathbf{p1} = (\mathbf{c1}, (C), \{ C \text{ rdfs:subClassOf } FC . \}, \emptyset, \mathbf{m1}) \quad (5)$$

where FC is the FC-variable and **m1** is a function returning a *named class* for a variable sequence only containing a single variable instantiated by this class. For example, if the FCP of an ontology O with respect to class `Car` is being computed, and the sole axiom template of pattern **p1** is instantiated by

$$\text{Truck rdfs:subClassOf Car .}$$

m1 will simply return the CE `Truck`, and the created category will thus be `Car ⊓ Truck`. Note however, that in general the subclass relationship (to satisfy **p1**) can be both direct or indirect, and possibly even inferred using other kinds of axioms, i.e., the class matched to C can be any subclass of FC in the deductive closure of the ontology computed by a reasoner.

Also note that even such a simple FCE pattern might not be the only possible for the given CE pattern. If we accept the possibility that not all classes in different hierarchical paths are pairwise disjoint, we could also apply to **c1** a pattern for possibly unrelated C and FC , with the axiom template set being, e.g.,

$$\{ C \text{ rdfs:subClassOf owl:Thing} .$$

$$FC \text{ rdfs:subClassOf owl:Thing} . \} \quad (6)$$

and with the constraints in *Cons* checking that none of $C \sqsubseteq FC$, $FC \sqsubseteq C$, $C \sqsubseteq \neg FC$ and $FC \sqsubseteq \neg C$ holds. Since the weight of the expressions built using such an FCE pattern should be rather tiny on average, we omitted it in our study, for simplicity.

As we move from \mathcal{L}_{nam} to the rest of \mathcal{L}_{SE} , i.e., to the realms of compound CEs, the FCE patterns are somewhat less obvious. We chose to implement them primarily in RDFS terms, namely, over `rdfs:subClassOf`, `rdf:type`, `rdfs:domain` and `rdfs:range` axioms, since these prevail in most linked vocabularies that are the prime subject in our ontology reuse setting. We however so far avoided `rdfs:subPropertyOf` to keep the pattern structure simpler (we will consider adding it in the future).

The relationship between `rdfs:domain` and the existential restriction will be central in our conception of FCE patterns for \mathcal{L}_{SE} . Note that for any C , FC and P holds that from the validity of $P \text{ rdfs:domain } FC$

¹⁷<https://www.w3.org/TR/turtle/>

we can trivially infer $\exists P.C \sqsubseteq FC$. The existential CE will thus necessarily be a subclass of the focus class FC in such a case; the only issue remaining is whether this subclass will have an ‘interesting size’ (neither too large nor too small) as subcategory of FC , in the spirit of the discussion from Section 2.3.

Pattern p2 The FCE pattern for CE pattern **c2** is still rather concise. It assumes that if the focus class FC is the domain of a property P , the instances of FC can be categorized according to whether they actually appear in the subject position in a triple with this property, whatever the object is (which truly corresponds to the existential restriction with filler \top , at the CE level). Again, even cases when FC is *inferred* as domain of P are considered.

$$\mathbf{p2} = (\mathbf{c2}, (P), \{ P \text{ rdfs:domain } FC . \}, \mathbf{Cons2}, \mathbf{m2}) \quad (7)$$

Here, notably, the constraint set, **Cons2**, is not empty. Currently we only consider one constraint in **Cons2**: such that P must not appear in an existential restriction¹⁸ $FC \sqsubseteq \exists P.C$; for such properties the CE would contain *all* instances of FC .

Note that the boundary between the axiom template and the constraints might not be strict, since the constraints could also be expressed as axiom templates in principle. However, they will serve for reducing the number of matches, so their internal structure would contain negations and thus would not be expressible in RDFS, which we propose as the core language for Tpl , for simplicity.

As regards **m2**, it is, again trivially, a function that generates the existential restriction with filler \top , taking a sequence only containing a single variable instantiated by this property.

In order to illustrate the application of the two FCE patterns introduced so far together, in the calculation of the approximate FCP, see formula (4) from page 11, we will return to our running example of choosing an ontology for categorizing the used cars for sale. To the difference of the previous, artificial, examples, we will now refer to a real-world ontology.

¹⁸The 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 entails the existential one; analogously for similar constraints for other FCE patterns.

Example 3 The *schema.org* ontology contains a class **Vehicle**¹⁹ having four subclasses: **Car**, **BusOrCoach**, **Motorcycle** and **MotorizedBicycle**. *Schema.org* also proposes 82 properties having **Vehicle** in their domain.²⁰

The instantiated \widehat{FCP} formula (cf. (4)) for $\mathcal{CP}_{\mathcal{L}} = \{\mathbf{c1}, \mathbf{c2}\}$ (the respective \mathcal{L} thus only covering a fraction of the expressiveness of \mathcal{L}_{SE}), $PS = \{\{\mathbf{p1}\}, \{\mathbf{p2}\}\}$, $FC = \text{Vehicle}$, $O = \text{schema.org}$, then becomes

$$\begin{aligned} \widehat{FCP}(FC, O, \mathcal{L}) &= \\ Occ(\{\mathbf{p1}\}, FC, O) \cdot w^P(\mathbf{c1}) + \\ + Occ(\{\mathbf{p2}\}, FC, O) \cdot w^P(\mathbf{c2}) &= \\ 4 \cdot w^P(\mathbf{c1}) + 82 \cdot w^P(\mathbf{c2}) \end{aligned} \quad (8)$$

Intuitively, while $w^P(\mathbf{c1})$ should be close to the theoretical maximum (named subclasses are by default meaningful subcategories), $w^P(\mathbf{c2})$ should be much smaller. Positive examples of *schema.org* properties wrt. **p2** are **review** (by substituting this property for P in $\exists P.\top$ we yield the category of cars for which at least one review exists), **knownVehicleDamages** (category of cars that suffered from at least one damage) or **vehicleSpecialUsage** (category of cars that have been used in a special mode, e.g., as a cab). Properties such as **fuelType** or **numberOfDoors** would, in turn, only yield meaningful categories when filled with a certain value – since *all* cars are supposed to consume *some* fuel (or another source of energy) and have *some* number of doors. The inclusion of their **p2**-matches in the computation will thus make \widehat{FCP} slightly overstated.

Let the weights be, for example, $w^P(\mathbf{c1}) = 1$ and $w^P(\mathbf{c2}) = 0.25$. This would yield $\widehat{FCP} = 4 \cdot 1 + 82 \cdot 0.25 = 24.5$. The categorization power of *schema.org* with respect to the focus class **Vehicle** would thus be

¹⁹Here and in the rest of the running example we omit the <http://schema.org/> and other prefixes, for brevity.

²⁰There is actually a tweak in this example. *schema.org* uses **schema:domainIncludes** instead of **rdfs:domain**, and analogously for the range axioms (**schema:rangeIncludes**). These properties do not have inferential semantics, and merely relate a given property to a class that “is (one of) the type(s) the property is expected to be used on” (<https://schema.org/domainIncludes>), and “constitutes (one of) the expected type(s) for values of the property” (<https://schema.org/rangeIncludes>), respectively. However, if a property only has a single **schema:domainIncludes** axiom, which is the case for most properties linked to **schema:Vehicle**, the heuristic use of the pattern is similarly plausible as for **rdfs:domain**.

for example greater than that of a purely taxonomic ontology in which a class (semantically corresponding to) Vehicle would have 20 transitive subclasses, but smaller than that of a purely taxonomic ontology with 30 such subclasses.

Let us now proceed to the FCE patterns for the remaining CE patterns included in \mathcal{L}_{SE} .

Pattern p3 The CE pattern **c3** includes 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 an instance of class C . The FCE pattern for **c3** then extends **p2** such that we also consider the range of P , which is an auxiliary class D . Since D itself would not make a proper distinction (all instances satisfying **c2** will by the definition of the `rdfs:range` predicate be connected by P to an instance of D), we further need to involve a ‘refining’ class C as subclass of D , which is also reflected by the second mapping variable.

$$\begin{aligned} \mathbf{p3} = (\mathbf{c3}, (P, C), \{ & P \text{ rdfs:domain } FC ; \\ & \text{rdfs:range}_a D . \quad C \text{ rdfs:subClassOf } D . \}, \\ & \mathbf{Cons3, m3}) \quad (9) \end{aligned}$$

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 `rdfs:rangea`) – otherwise not only subclasses of D but also classes having a common superclass with D would be returned as C (since the superclass is also inferred as a range of P).

The constraint in **Cons3** states that P must not appear (in the inferential closure of O) 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 .

The mapping function **m3** is, again, trivial, except that its input is now a pair (P, C) , from which the qualified existential restriction $\exists P.C$ is built.

As this FCE pattern is already a bit more complex, we will also demonstrate its application (and the added value compared to **c2**) graphically, in Fig. 1. The upper part of the diagram displays the structure (Tbox) of **p3**; the bold labels indicate asserted statements, and the italic label indicates an inferred statement (here we show a case where FC only belongs to the domain of P by ‘inheritance’ from its subclass). The bottom part shows example constellations of an instance and non-instance of $\exists P.C$ in the Abox. While both x_1 and x_2

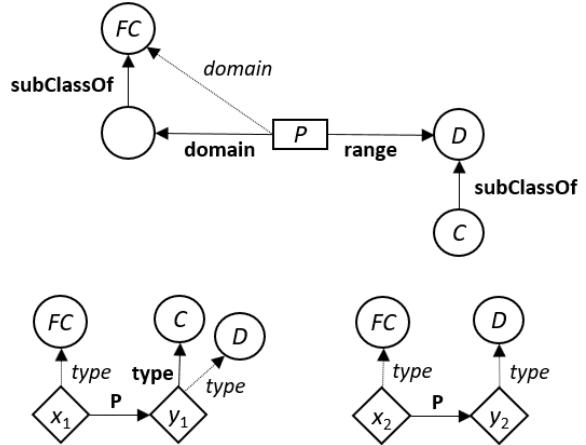


Fig. 1. Tbox of **p3** and Abox of instance and non-instance of $\exists P.C$

belong, via inference over the domain axiom, to FC , and both also belong to $\exists P.\top$ and (via the inference over the range axiom) to $\exists P.D$, it is only x_1 that also belongs to $\exists P.C$ as the considered category.

Pattern p4 The CE pattern **c4** includes all instances of FC that appear in the subject of a triple with P as predicate and (the specific individual) i as object. The FCE pattern for **c4** then extends **p3** with an individual that is instance of C and is also used in the mapping variable sequence (replacing there the class C):

$$\begin{aligned} \mathbf{p4} = (\mathbf{c4}, (P, i), \{ & P \text{ rdfs:domain } FC ; \\ & \text{rdfs:range}_a D . \quad C \text{ rdfs:subClassOf } D . \\ & i \text{ rdf:type } C . \}, \mathbf{Cons4, m4}) \quad (10) \end{aligned}$$

The inferential closure is used as in **p3**, i.e., with the exception of the range axiom.

The constraint in **Cons4** states that there must not be (in the inferential closure of O) an existential restriction $FC \sqsubseteq \exists P.\{i\}$; for such properties the CE would contain *all* instances of FC .

The mapping function **m4** is analogous to **m3**, with merely wrapping i into a singleton enumeration.

Pattern p5: SKOS concept as value For the CE pattern **c4** we include one more FCE pattern, which goes beyond the RDFS expressiveness and relies on the connection of the ontology to an external resource – a *SKOS²¹ codelist*. This variant thus extends the previ-

²¹<http://www.w3.org/2008/05/skos>

ous one for a specific source of instance i , which is a code list (also called concept scheme in SKOS):

$$\begin{aligned} \mathbf{p5} = (\mathbf{c4}, (P, i), \{ & P \text{ rdfs:domain } FC ; \\ & \text{ rdfs:range}_a \text{ skos:Concept ,} \\ & [\text{ a owl:Restriction ; owl:hasValue } s ; \\ & \text{ owl:onProperty skos:inScheme }] . \\ & i \text{ a skos:Concept ; skos:inScheme } s . \}) \end{aligned} \quad \text{Cons4, m4)} \quad (11)$$

The difference is merely in the selection method for i – rather than an instance of a class from the current ontology, it has to be a `skos:Concept` linked to a concept scheme s that is in the range of P . The inferential closure is used as before, and the constraint and the mapping function are the same as in **p4**.

We can now illustrate the application of all considered FCE patterns on our used cars example.

Example 4 The property `vehicleSpecialUsage` of `schema.org` has the class `CarUsageType` as its range. `CarUsageType`, in turn, has no subclass but three enumerated instances: `DrivingSchoolVehicleUsage`, `RentalVehicleUsage` and `TaxiVehicleUsage`. It would thus yield three categories via pattern **p4** (and none for **p3**), in addition to the one for **p2** from Example 3. In contrast, the properties `fuelType` or `numberOfDoors` only have a disjunction of generic classes (`QualitativeValue`, `Text` and `URL`) in their range.²² They therefore do not yield any category via **p4**, within `schema.org`. However, if the range of such properties were later associated with relevant SKOS code lists containing fuel types and possible numbers of doors, respectively, **p5** could be applied.

All five FCE patterns assure (**p2–p5** via the domain axiom) that the mapped CE is a *specialization* of the FC.²³ Therefore, any \mathcal{L} -category $Cat = FC \sqcap D$ generated from an ontology using these patterns is equivalent to its differentiating concept expression D . It is also easy to see that the patterns, except **p4** vs. **p5**, are *mutually exclusive* since they produce structurally dif-

ferent CEs. (Formally, to assure the 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`.)

As regards the complexity of the FCE patterns in terms of (templated) RDF triples in the template axioms, it is as listed, for the respective CE patterns, in the fifth column of Table 1. **p4** is the most complex one of the ‘straightforward’ FCE patterns, with four triples; however, **p5** (for the same CE pattern, **c4**) is even more complex, having eight triples overall, due to the overheads of representing OWL restrictions in RDF.

We have previously stated that the proposed FCE patterns are merely heuristic (approximate) with respect to the optimally chosen set of \mathcal{L}_{SE} categories. Let us now briefly elaborate on this statement and summarize its different aspects:

- First of all, the fourth assumption from Section 1, i.e., that the quality of compound categories should be correlated with the degree to which users would consider a corresponding named class as meaningful, rules out any kind of optimality guarantee, since human judgment is always subjective. The designer wishing to reuse an ontology may even question the *asserted subclass axioms* from the source ontology, thus lowering the *precision* of the **p1**. The (human-judged) precision of FCE patterns for *compound CEs* will naturally be even less stable.
- The situation is analogous for the FCE patterns’ *recall*. Even **p1** will omit any category with a differentiating concept expression (DCE) beyond those subsumed by the FC. Typical cases of ‘lost’ categories are those with the FC and the DCE being ‘close neighbors’ having a common ‘natural’ superclass (e.g., `Student` and `Teacher` being subsumed by `Person`).
- As regards the FCE patterns **p2–p5**, which rely on domain/range axioms in their axiom templates and on (negated) existential restrictions in their validity constraints, their accuracy depends on the degree in which the ontology is equipped by these kinds of axioms; see Section 8.2 for a more detailed discussion.

4.2. FCE pattern usage for named category generation

The generic algorithm in Listing 1 ends by generating a list of CEs (plus their CE patterns) for a given

²²Such special-purpose classes would have to be filtered out; typically they could be automatically detected by appearing in the range of a huge proportion of properties.

²³In contrast, this would for example not be the case for the alternative FCE pattern for **c1** (cf. Equation 6) allowing hierarchically unrelated C and FC .

focus class. However, to finalize our *schema.org* example narrative, we will also consider the subsequent step of *named category generation* (already mentioned in Example 1) as an additional use case of the FCE patterns.

Example 5 Let us assume that the categories discovered for the purpose of the FCP computation are retained for future use. The vehicle retailer might then decide to rebuild the product catalog so as to cover the categories, even the compound ones, that are populated by a significant number of vehicle items. The categories derived from the same property, say, `vehicleSpecialUsage`, would then generate a branch of the catalog structure, which, after some (possibly, NLP-assisted) tweaks could look like this:

```
Vehicles
  Vehicles with special usage
    Vehicles with driving school usage
    Vehicles with rental usage
    Vehicles with taxi usage
```

This structure might either become materialized in the underlying ontology as such or might only be generated at the web engineering level. The taxonomy would naturally follow the specialization of FCE patterns: the categories yielded by **p2** would be direct subcategories of the FC (and siblings of those yielded by **p1**, i.e. of the original named subclasses); the categories yielded by **p3** would be subcategories of that yielded by **p2** (for the same property P), and the categories yielded by **p4** would be either at the same level as those of **p3**, or some might be a subcategory of a category yielded by **p3** if there were an `rdf:type` link between the instance i in the match of **p4** and a subclass of the class in the range of P in the match of **p4**. (Since the derivation of named categories is not the main topic of this paper, we omit the formalization of this derivation for space reasons, however simple it is.)

5. CE patterns in ontology axioms

As the first empirical source for estimating the frequency of occurrence of the CE patterns of the CEL \mathcal{L}_{SE} we chose the axioms of publicly available ontologies.

5.1. Ontology axiom sources

For our analysis we used three collections of ontologies:

1. The collection indexed by the *Linked Open Vocabularies* (LOV) portal,²⁴ containing real-world (but mostly light-weight) ontologies typically used for linked data publishing.
2. The BioPortal collection,²⁵ the well-known repository of biomedical ontologies.
3. A small experimental collection of ontologies having heterogeneous styles and relatively rich in axioms, from the domain of conference organization, called *OntoFarm*.²⁶

The analysis took place in November 2017.

5.2. Motivation of the analysis

The impetus for this empirical analysis was the close relationship between the central motivating task of the research, that of using CEs (either named classes, or anonymous CEs that can possibly be transformed to named classes) for *categorizing individuals*, on the one hand, and the task of *defining new named*²⁷ *classes* (using anonymous CEs) within ontologies on the other hand. We conjecture that the distribution of anonymous CE structures in both tasks would be to some degree similar. The differences in the relevance of constructs wrt. these tasks are discussed below.

Note that the subsequent use of the constructed CEs is, in some aspects, similar in both tasks, too. Notably, in all cases, an aspect of *categorization* is present:

- An anonymous CE in the RHS of an *equivalence axiom* allows to *intrinsically* (i.e. using the apparatus of DL tightly coupled with OWL as representation language) *infer* individuals to be instances of the named class in the left-hand side (LHS) of the axiom.
- An anonymous CE in the RHS of a *subsumption axiom* (which represents a mere necessary, and not sufficient, condition) allows, analogously, to

²⁴<https://lov.linkeddata.es/dataset/lov/>

²⁵<https://bioportal.bioontology.org/>

²⁶<https://owl.vse.cz/ontofarm/>; the collection is used, among other, in the Ontology Alignment Evaluation Initiative, <http://oaei.ontologymatching.org/>.

²⁷We can for now ignore general concept inclusions (having an anonymous CE as their left-hand side), which are allowed in some dialects of OWL but only used sparingly in real-world ontologies.

intrinsically rule out individuals from being instances of the named class in the LHS of the axiom.

- An anonymous CE that can be merely constructed from the *ontology signature* under some CEL and *for some FC* still allows to subcategorize individuals, but merely *extrinsically*, i.e., either using manual assignment or some other external source (e.g., a machine-learning-based classifier).

Technically, the expected outcomes of the analysis were the following:

- Findings about the frequency of various class expression patterns in the RHS of axioms in existing ontologies (possibly interesting for the community even beyond the focused categorization setting)
- Positioning of the CEL \mathcal{L}_{SE} with respect to the rank of its CE patterns in a CE pattern frequency table.

The considered list of CE patterns is, essentially, that of first-level constructors in the axiom RHS. However, since we were particularly interested in the compound CE patterns of \mathcal{L}_{SE} , we singled them out under the labels ‘ExistThing’ (**c2**), ‘ExistNamed’ (**c3**) and ‘ExistIndiv’ (**c4**). All other variants of existential restrictions not falling under these three are counted under ‘ExistAnon’, since the filler then is an anonymous CE rather than a named CE.²⁸

The frequency of CE pattern occurrence, in absolute counts, was classified by three dimensions:

- By the analyzed ontology collection.
- By the distinction of equivalence or subclass axioms (and the sum of both).
- By the level of nesting: outermost constructor vs. further levels.

5.3. Results

The results are in Table 2, however omitting the last dimension (nesting level) for brevity. The most frequent constructors are listed: 10 for LOV and BioPortal, and 5 for OntoFarm (where only the top of the ranking is relevant, due to very low counts). The constructors corresponding to \mathcal{L}_{SE} are in italics.

²⁸Strictly speaking, the singleton enumeration $\{i\}$ in **c4** is also an anonymous CE, and the notion of ‘individual value restriction’ is rather a syntactic sugar in DL terms. For our purposes it is however distinguished from the restrictions having some other compound filler.

Table 2
CE patterns in axiom RHS

LOV			
Equivalence		Subclass	
Construct	#	Construct	#
Conjunction	1134	<i>ExistNamed</i>	3187
<i>ExistNamed</i>	126	Universal	1331
Disjunction	108	Conjunction	553
Enumeration	54	Exactly	545
ExistAnon	31	Minimally	483
<i>ExistIndiv</i>	30	ExistAnon	467
Exactly	12	Maximally	237
Negation	4	<i>ExistIndiv</i>	77
Minimally	3	Disjunction	56
Maximally	1	Negation	23

BioPortal			
Equivalence		Subclass	
Construct	#	Construct	#
Conjunction	122725	<i>ExistNamed</i>	1977805
ExistAnon	24239	ExistAnon	64046
<i>ExistNamed</i>	2338	Conjunction	18409
Enumeration	854	Universal	6587
Disjunction	658	<i>ExistIndiv</i>	4698
Universal	22	Exactly	1179
<i>ExistThing</i>	11	Disjunction	856
Minimally	10	Minimally	579
Exactly	7	Maximally	577
Negation	6	Negation	96

OntoFarm			
Equivalence		Subclass	
Construct	#	Construct	#
<i>ExistNamed</i>	40	<i>ExistNamed</i>	317
Conjunction	27	Universal	47
Disjunction	15	Exactly	23
Universal	2	Minimally	22
Enumeration	1	ExistAnon	15

As regards the \mathcal{L}_{SE} patterns, we can see that ‘ExistNamed’ (**c3**) appears very often at top positions, throughout the collections and axiom types; it dominates in subclass axioms, most notably in BioPortal. ‘ExistIndiv’ (**c4**, i.e. the value restriction) is less frequent but still plays some role in both larger collections. ‘ExistThing’ (**c2**) appears even less often.

The other constructors appearing at the first level of the axiom RHS nesting are:

- *Conjunction* (\sqcap). From the point of view of FCP proper, considering logical conjunctions backed by plausible axiom patterns, analogous to those

from Section 4, could be interesting and deserves a further study. For example, for the *Car* focus class, *conjunctions of property restrictions* such as

$$\exists \text{hasSeller}.\text{Corporate} \\ \sqcap \exists \text{usage}.\text{TaxiUsage}$$

(where the two properties are required to share their domain) could provide further weight to an ontology in the FCP computation, atop the weight of their constituent property restrictions. This would however lead to a significantly more complex CEL. In contrast, considering mere *conjunctions of named classes* (which would still fit into a ‘simple’ CEL) would be problematic due to the combinatorial explosion of class pairs, in which only a tiny fraction would be mutually compatible (considering that the absence of explicit disjointness axioms does not guarantee class compatibility in real-world ontologies). Also note that from the point of view of data description and querying as such, emulating conjunctive concepts is simple: multiple categories are just applied at the same time, either to categorize a given individual or to form a simple query (e.g., by ticking multiple categories in the product catalog form). With respect to the above example, the two newly constructed named classes, as categories assigned/ticked in combination, might be called *CarWithCorporationSeller* and *CarWithTaxiUsage*, respectively.

– *Disjunction* (\sqcup). Unlike for conjunctive expressions, any pair of expressions is mutually compatible in a disjunction. As mentioned in Section 3, the number of disjunctive combinations will thus be very high for larger ontologies, and the question of which combinations should be taken into account for FCP or not would not be easy to decide at the axiom pattern level. Intuitively, concepts ‘closer to each other’ in the ontology might be more adequately combined (e.g., the customer might be more likely to seek a “Sedan or Hatchback” than a “Bus or Bicycle”), and similarly in combining property expressions for properties with the same domain. However, such a heuristic would favour shallow ontologies over ontologies with a more elaborate class taxonomy. The utility of disjunctively defined categories in practical contexts of RDF data publishing (e.g.,

within e-shop catalogs) deserves a further study, in any case.

- *Existential restriction to an anonymous CE* (‘*ExistAnon*’). This could be one of the meaningful extensions to \mathcal{L}_{SE} . Instead of only allowing a names class as property filler (as in **c3**), we would allow a certain subset of anonymous CEs. The weight of such expressions in the FCP computation would of course have to be adequately decreased compared to less complex expressions. We will study this option in our future research; one promising candidate CE pattern would be the embedding of two existential restrictions (corresponding of a two-piece property chain).
- *Universal restriction* (\forall). Universal restrictions rarely appear in equivalence axioms, since classifying individuals by them (using the ‘necessary and sufficient conditions’ on the RHS) is impossible on the open semantic web due to the OWA (we can never be certain that a to-date unknown statement would not contradict the universal restriction). Yet, universally restricted concepts may be valid categories in the FCP context, since the usage of the ontology involving the ‘extrinsic’ categorization may already be performed under the CWA. For example, the used vehicle retailer may have trustful evidence that *all* previous owners of the car have been corporations. In terms of heuristically deriving a universal restriction category (with a named class as filler) from the Tbox, there would be probably very little difference from the existential restriction (to the same named class): the same Equation 9 (based on a domain and range axiom) could be applied, except that the pattern constraint should be changed to a universal restriction (ruling out the possibility that the category would coincide with the focus class). Additionally, we would assume that a universal property restriction category would be implicitly applied in conjunction with the existential one. For example, based solely on the universal restriction, the CE $\forall \text{hadAccident}.\text{FatalAccident}$ would include all cars not having any accident recorded in their history at all. To avoid this effect, we should extend it to $\exists \text{hadAccident}.\top \sqcap \forall \text{hadAccident}.\text{FatalAccident}$.
- *Cardinality restrictions* ($\leq_n, \geq_n, =_n$, i.e., ‘Minimally’, ‘Maximally’ and ‘Exactly’ in Table 2). They appear relatively frequently in subclass restrictions, much less so in equivalence ones. While some of them may be valid categories in

- the FCP context, their heuristic derivation from the Tbox would require to specify the numerical parameter. As in the case of the universal restriction, the heuristic pattern for the generation of cardinality restrictions would basically be the same as for the existential restriction (Equation 9), only the pattern constraints would differ.
- *Enumeration* ($\{\dots\}$). Using enumerations in practical focused categorization scenarios is mostly irrelevant. Enumerations presumably appear as the outermost level in the RHS of (equivalence) axioms whose LHS is a class serving as an embedded (thus typically small), intentionally closed ‘code list’ inside the ontology. The ‘code list class’ is then not used on its own, but rather helps define new concepts. The broadly usable ‘code’ individuals embedded in the ontology are then unlikely to coincide with the ad hoc instances of a focus class that we would like to subcategorize.
 - *Negation* (\neg). In some contexts, a negated concept might perhaps be a sensible subcategory of a focus class; however, such contexts would be hard to guess a priori. Considering the negation categories in a generic way would simply double the number of categories for each CE pattern to be negated, and thus would not have other impact on the FCP than would have the doubling of the weight of those CE patterns.

In all, the analysis suggested that the \mathcal{L}_{SE} patterns play a significant role in the family of all anonymous expressions commonly used in OWL. The design of an FCP formula restricted to this CEL is thus meaningful, considering that we want to investigate the landscape of the whole focused categorization problem within a simple language first. The analysis however also indicated directions for extending the used language in the future – most notably, in relation to the expected complexity of the *queries* to be applied on data published using the reused ontologies.

6. FCE pattern occurrence in the Tbox of ontologies

The questions to be answered by this analysis were:

1. How many ontologies, and for how many FCs, provide a decent number of ‘categorizing’ CEs through heuristic mapping from the patterns from Section 4.

2. What are the differences in the occurrence of the individual FCE patterns overall and across different ontology collections.

The answers to these questions would help us estimate how likely it is that the particular FCE pattern would yield potential categories when a focus class is provided as an input. (They will not, though, provide an estimate of the quality of such categories; this question will be addressed in Section 7.)

6.1. Ontology sources and data aggregation method

In the analysis we made use of our *Online Ontology Set Picker framework*²⁹ (a tool allowing to quickly select ontologies according to various metrics) to process ontologies from two collections. One is LOV; we again used its November 2017 snapshot where 617 ontologies were available, of which we could process successfully 568 at syntactical level. The other is OntoFarm, which contains 16 ontologies. We omitted Bioportal in this analysis, since it is on average rather poor in properties, which are a necessary ingredient for generating categories within \mathcal{L}_{SE} .

In order to provide aggregate results, we counted the occurrences of FCE patterns from Section 4 across all classes of all ontologies evaluated in the role of FC. We summed up these results at ontology level by identifying ‘categorizable’ classes, i.e. classes for which the pattern occurrence reached some threshold τ (1, 3, and 5) for patterns **p1**, ..., **p4**,³⁰ see Table 3. The tables show the percentage of ontologies for which more than a c portion of classes is ‘categorizable’ as FCs using the given pattern. The portion of ‘categorizable’ classes is quantized by 0.1, from >0.0 to >0.4 ; higher values are not so relevant, since, intuitively, not all classes of an ontology but only some proportion of more general ones would likely serve as FCs.

For illustration of the bottom-up calculation steps, let us take the example of an OntoFarm ontology called *cmt*. This ontology satisfies the parameter thresholds for $\tau = 5$, $c > 0.1$, for **p1**, as 4 of its 29 classes ($n \sim 14\% > 0.1$) are ‘categorizable’: each has five or more named subclasses (i.e. its τ is at least 5). As we see in the respective cell of Table 3, *cmt* thus belongs to the 31.3% of OntoFarm ontologies having such a proportion of ‘categorizable’ FCs.

²⁹<https://owl.vse.cz/00SP/>

³⁰Pattern **p5** was only matched rarely; it is thus analyzed separately.

Table 3

Ratio of ontologies with a $> c$ portion of classes ‘categorizable’

OntoFarm						
Pat.	τ	> 0.0	> 0.1	> 0.2	> 0.3	> 0.4
p1	1	100.0%	100.0%	81.3%	25.0%	0.0%
	3	100.0%	100.0%	18.8%	0.0%	0.0%
	5	93.8%	31.3%	0.0%	0.0%	0.0%
p2	1	100.0%	93.8%	56.3%	31.3%	25.0%
	3	100.0%	43.8%	25.0%	12.5%	0.0%
	5	93.8%	37.5%	6.3%	0.0%	0.0%
p3	1	93.8%	50.0%	18.8%	6.3%	6.3%
	3	93.8%	50.0%	12.5%	0.0%	0.0%
	5	93.8%	31.3%	6.3%	0.0%	0.0%
p4	1	25.0%	0.0%	0.0%	0.0%	0.0%
	3	18.8%	0.0%	0.0%	0.0%	0.0%
	5	6.3%	0.0%	0.0%	0.0%	0.0%

LOV						
Pat.	τ	> 0.0	> 0.1	> 0.2	> 0.3	> 0.4
p1	1	80.1%	73.1%	47.7%	25.4%	11.6%
	3	60.4%	43.0%	13.0%	3.2%	1.4%
	5	49.3%	22.4%	3.2%	1.4%	1.1%
p2	1	83.8%	72.4%	60.7%	46.8%	34.5%
	3	71.1%	51.6%	29.8%	16.2%	8.6%
	5	56.2%	32.7%	14.4%	7.2%	3.9%
p3	1	49.3%	29.0%	14.1%	6.2%	3.0%
	3	42.1%	20.6%	8.8%	3.2%	1.8%
	5	34.2%	14.1%	5.5%	1.9%	1.4%
p4	1	22.0%	9.9%	4.6%	2.1%	1.2%
	3	18.1%	7.2%	3.3%	1.2%	0.9%
	5	16.2%	6.9%	3.2%	1.2%	0.9%

6.2. Results for patterns **p1** to **p4**

Let us now try to answer the questions posed at the start of the section, by examining the result table.

Unsurprisingly, the percentage of ‘categorizable’ classes is in most cases highest for patterns **p1** and **p2**, which reflects the simplicity of these patterns: many classes have some named subclass (**p1**), and many also appear in the domain of some property (**p2**). With increasing c , the percentage drops more slowly for **p2** than for **p1**, reflecting the fact that even classes lowest in the subclass hierarchy (i.e. not having named subclasses) do appear in the domain of properties.

For **p3** the percentage is generally lower; however, it ‘beats’ **p1** when it comes to the setting with higher thresholds for both c and τ , i.e., when we want to include even less general (focus) classes and require them to have a higher number of categories. This

is apparent even in the much smaller OntoFarm (for $c > 0.2, \tau = 5$), but particularly in LOV; we indicate such values using boldface in the table. Presumably, focus classes down in the hierarchy may be connected (through some properties’ domain and range) to more general classes, thus ‘borrowing’ their subclass hierarchy (note the structure of **p3** in Equation 9, where class C from the range of property P ‘lends’ its subclass D for the filler of the existential restriction category).

p4 is least frequent, however, for higher c and $\tau = 3$ the percentage also gets closer to that of **p1** (presumably, through the same mechanism as we discussed for **p3**).

The findings about **p3** and **p4** are potentially interesting. If we managed to prove that the categories generated by these patterns are often meaningful, these two patterns could help yield useful categories even for focus classes positioned lower in the subclass hierarchy, for which **p1** does not work well.

6.3. Results for pattern **p5**

For completeness, we performed an analysis of ontologies which use SKOS concepts for entity categorization (pattern **p5**). While OntoFarm ontologies do not feature this pattern at all, in LOV there were 7 ontologies out of 509 containing its instantiation, to date.

Table 4 presents information about those 7 ontologies in terms of the number of SKOS concepts that can be used as the ‘categorization individual’ and the number of SKOS concept schemes from which those concepts come. Further, we include information about the date of the last modification of the ontology. In two cases, no concept schemes are available. For the other ontologies the number of SKOS concepts (and SKOS schemes, respectively) usable for categorization varied from 11 to 274 (from 1 to 16, respectively). Although the phenomenon captured by **p5** (reference to a SKOS code list in an ontology) is currently marginal, we assume that it may have an increasing trend, since those seven ontologies had been last modified, on average, in 2014, while the average of the last update dates for the whole of LOV was in 2012.

7. Cognitive experiments: CE assessment

The previous analyses carried out over the ontology Tbox (axioms RHS in Section 5 and FCE pattern occurrence in Section 6) only indirectly contributed to the central question: whether the compound CEs from

Table 4
Categorization via SKOS concepts (pattern p5)

Ontology	# of schemes	# of concepts	Last 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://consem.unizar.es/def/sector-publico/pproc	3	11	2015-01-11
http://rdf.insee.fr/def/geo	3	30	2015-01-11

\mathcal{L}_{SE} , and to what degree, can serve as meaningful subcategories of their focus classes.

In order to get finer insights, we proceeded to a detailed investigation of sample CEs by human ‘ontologists’, both experts and relative novices (students of relevant subjects). We performed two campaigns of experiments, the first in Spring 2016³¹ and the second, with a slightly modified and extended setup, in Spring 2018.

The overall scheme of the experiments was as follows:

- Provide the human assessors with a set of ‘focus class – subcategory’ pairs, such that the subcategories correspond to *different CE patterns*. Each pair constituted an elementary *assessment task*.
- Collect the *assessment* for each task, reflecting the degree to which the assessor perceived the category as ‘meaningful’, or, better, ‘meaningful and reusable’ (or also, for short, what is the perceived quality of the category), via a *questionnaire*.
- Aggregate the collected data across the assessors.
- Seek *correlations* between the category quality and the CE pattern, possibly corrected through *metadata* (also collected in the questionnaire) such as the level of written English of the assessors or their general comprehension of the meaning of the constituent entities of the category.
- Examine the *qualitative feedback* also collected through the questionnaire.

A summary of the experimental setting in both campaigns is in Table 5.

7.1. Summer 2016 campaign

Initial sampling As regards the FC+subcategory (from now on, ‘task’) sampling for both threads of

Table 5
Overview of cognitive experiments with students

Campaign	Summer 2016	Summer 2018
# of students	27	31
# of different tasks	59	40
# of tasks/person	20	8
Avg. time/task	90 sec	5 min
Tasks source	LOV, OntoFarm	LOV
CE patterns	c2, c3, c4	c1, c2, c3, c4
Judgment support	Glossary of conference terms	Entity definitions from the ontologies; CE structure verbalization
Collected metadata and other info	Comprehension of entity semantics	English skills; comprehension of entity semantics; justification for negative assessment

analysis (expert/novice), we used the same collections as in the survey from Section 6, i.e. LOV (a 2016 snapshot) and OntoFarm (this collection has remained unaltered for nearly a decade). From each collection 40 tasks were randomly generated while maintaining an approximately even proportion of subcategories for patterns c2, c3 and c4, eventually yielding 80 tasks. After a manual removal of duplicates (for OntoFarm as smaller collection) and tasks involving entities with cryptic names without meaning in natural language, 59 tasks remained (28 from OntoFarm and 31 from LOV); there were 17 tasks for c2, 20 for c3 and 22 for c4.

Expert ontologist assessment and insights. In the first stage, the assessment was made by three researchers with 10–20 years of experience in ontology engineering (authors of this paper: VS, OZ and MV). They first examined the 59 tasks independently and assessed them on the 5-point Likert scale: for each task, with a focus class Y and a subcategory X, the question “Is X a meaningful subcategory of Y?” was answered as either ‘certainly’, ‘perhaps’, ‘borderline’, ‘perhaps not’ or ‘certainly not’. Then a consensus was sought in

³¹Already described in our early publication [24]. Here we provide a synoptic view of both campaigns.

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 not); we will call these cases *clear positives* (42 cases) 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 meaningful subcategory of the FC), one case was found dubious due to an implausible inference (see the second ‘insight’ below, on the ‘village head chef’) and in one case the subcategory was assumed semantically equivalent to its FC, both resulting into ‘no’. Of the five ultimately negative results, four were of **c2** and one of **c3**. Complete assessments and commentaries are in a supplementary page. Selected general insights into less obvious decisions, with examples, follow:³²

- 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 than to the instantaneous action of ‘finalizing’ it.
- In some cases the use of inferential closure for the filler class in **c3** leads to linking relatively *thematically unrelated* entities (especially in the DBpedia ontology), such as in `dbo:beatifiedPlace some dbo:WineRegion` for instances of the FC `dbo:Person`. This case was still found marginally acceptable, as there could be some correlation between religiosity and wine production. However, a similar one, with subcategory `dbo:headChef some dbo:BaseballPlayer` proposed for the FC `dbo:Village`, was rejected not only due to the thematic leap but also due to an odd inference result: the declared domain of `dbo:headChef` is `dbo:Restaurant`, but the ontology (actually, the 2014 version from the LOV endpoint) enabled to infer the axiom `dbo:Restaurant rdfs:subClassOf dbo:Village`.

- Some CEs of **c4** 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 Ontology 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-minute overview of the notions of CE (in \mathcal{L}_{SE}) and FC roughly as presented in Sections 2 and 3 of this paper. Then they completed an assignment consisting of 20 atomic tasks, all available in a single sheet of a web questionnaire.³³ 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 collected on the 5-point Likert scale, with an additional option ‘*no judgment, since I don’t understand the example*’.

The 59 tasks from the initial sample were randomly divided into three questionnaire versions (one **c4** task was used twice) to eliminate cribbing; the numbers of returned questionnaires per version were 7, 9 and 11, respectively, with a 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 pattern. The aggregation was carried out by simple summation over the answer values rescaled to the $[-1; 1]$ interval (i.e. ‘certainly’ turned to 1, ‘perhaps’ to 0.5, both ‘borderline’

³²In this section we write the example CEs in Manchester syntax, with the keyword `some` indicating an existential restriction in **c2** and **c3** CEs and the keyword `value` indicating a value restriction in **c4** CEs.

³³The questionnaire was in Czech. The English translation of a sample task is available in Appendix A.

and ‘no judgment’ to 0, etc.), and then normalized by dividing by the number of students on the task. The normalized sum of the answer values of the i -th task is then

$$NS_i = \frac{\sum_{j=1}^k v_{i,j}}{k} \quad (12)$$

where $v_{i,j}$ is the rescaled answer value of respondent j for task i , and k is the total number of respondents of task i . For example, a task assigned to eight students, with the answers ‘certainly’, ‘perhaps’, ‘borderline’ and ‘perhaps not’, all present twice, yields

$$NS = (2 * 1 + 2 * 0.5 + 2 * 0 + 2 * -0.5) / 8 = 0.25$$

A short digest of the results follows:

- 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%).
- Perhaps most important, the average NS was highest for **c4** (0.21, with 15 positives, 1 zero and 7 negatives), lower for **c3** (0.02, with 9 positives, 3 zero and 8 negatives), and lowest for **c2** (-0.05, with 4 positives, 1 zero and 12 negatives).
- The cases³⁴ with highest positive and lowest negative values are in Table 6; the CE pattern is listed in the third column. We see that the cases with highest positive polarity tend to achieve higher absolute values than the cases with highest negative polarity, i.e., it was more likely for the students to agree on the positive than on the negative cases. We also see that **c4** dominates the upper end of the spectrum. Interestingly, the negative cases correspond each to a different CE pattern and also have different semantic roots: the ‘village with baseball player as head-chef’ one has

³⁴Most 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>

already been discussed before (distant and dubious inference), the ‘conference in city’ one deals with a seemingly mandatory property leading to $Cat \equiv FC$ (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 ‘ontological thinking’ of the former. The *inter-task* variance, 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. However, the *intra-task* variance, indicating the degree of disagreement among the students filling the same task, was higher for the AI students (2.51) than for the OE students (2.12), i.e. the rating of the latter was more coherent.

In comparison with the ‘expert ontologist’ assessment:

- The students gave a significantly lower score: only about a half of the tasks had a positive NS, compared to 92% (54/59) in the final consensus of experts. This can be explained by their lower ability to figure out specific situations in which less obvious categories might become meaningful.
- 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 the less frequent ‘meaningfulness’ of **c2** CEs (i.e., on the lower reliability of the FCE pattern **p2**). Out of the 17 respective tasks, as mentioned above, only 4 were viewed as ‘meaningful’ by the students and 13 by the experts (who in turned judged all tasks with other patterns, except one, as ‘meaningful’).
- As regards the case-by-case comparison between the students and the experts, there is also a 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).

Table 6
CEs with highest and lowest average NS of student scores, 2016 campaign

FC	Subcategory	CE pattern	Avg. NS
ofrd:FridgeFreezer	ofrd:styleOfUnit value ofrd:SingleDoor	c4	0.91
gr:BusinessEntity	pco:mainActivity value p-act:GeneralServices	c4	0.86
gr:BusinessEntity	pco:authorityKind value p-aut:LocalAuthority	c4	0.61
akt:Generalized-Transfer	akt:information-transfer-medium-used value akt:Email-Medium	c4	0.59
p1:Storage	p1:hasStorageMedium value p1-sm:mag	c4	0.59
fabio:Item	fabio:isStoredOn value fabio:web	c4	0.50
...
dbo:Village	dbo:headChef some dbo:BaseballPlayer	c3	-0.50
sigkdd:Conference	sigkdd:City_of_conference some Thing	c2	-0.56
gr:DayOfWeek	gr:hasNext value gr:Friday	c4	-0.56

7.2. Summer 2018 campaign

In the second campaing we tried to modify the setting so as to avoid some biases and gaps appearing in the first campaign, in particular:

- The even distribution of tasks between LOV and OntoFarm was judged inadequate, as OntoFarm is by an order of magnitude smaller, addresses one domain only, and its ontologies have been created artificially, even if based on real-world non-ontological resources.
- Assessing the CEs solely based on their formal representation risked of suffering from a comprehension bottleneck.

There was no a priori expert assessment this time (assuming that the correlation of the expert and novice assessment had been adequately studied in the first campaign).

Initial sampling and task preparation This time, all tasks were generated from the LOV. In contrast to the 2016 campaing, we also added CEs matching the pattern **c1**, to verify the assumption that named subclasses are by default meaningful categories. We first pre-selected LOV ontologies that satisfied the following two conditions:

- they had more than 90% of their classes equipped with the `rdfs:label` value (since we would need these values for the questionnaire)
- they had at least 10 classes (to eliminate the long tail of very small ontologies).

The actual sampling was then performed on approx. 130 thousand CEs generated from the 72 ontologies that satisfied the above conditions. From this pool we randomly sampled ten tasks for every CE pattern (**c1**, **c2**, **c3**, **c4**). No manual filtering of tasks was applied this time, thus entities with cryptic IRIs were allowed (relying on the availability of the `rdfs:label` values). In addition to the labels, we also collected from the ontologies:

- *verbal definitions* of the involved entities, as the `rdfs:comment` values or even as definitions manually found in the documentation of the ontologies
- selected *axioms* in which the entities appeared in the RHS.

Unfortunately, the sampling results exhibited some potentially undesirable features, and we did not have time to redesign the sampling because of the planned experiment dates (within the schedule of both courses) that we were unable to shift. Namely:

- Some domain ontologies contained links to upper-level ontologies. If the FC was then picked from an upper-level ontology, it was highly abstract (e.g. ‘Feature’, ‘Object’ or ‘Endeavor’), and its relationship to domain-specific concepts of the CE was hard to figure out. The assessment then had ‘strong philosophical flavor’, and the setting was unrealistic wrt. our target use case, since upper-level entities would not typically be sought as focus classes when publishing linked datasets.
- One of the tasks referred to an ontology in a language different from English (namely, Spanish).

By the results and the students' feedback it however does not seem that these infelicities would have seriously biased the experiment.

The structure of the CEs was *verbalized*, using simple NLP patterns plus occasional manual tweaking to assure grammatical correctness of the generated sentences. For example, if the property label P in a CE matching the pattern **c2** was a noun phrase, the CE was verbalized as "A specific kind of **FC**: such that has a **P**". On the other hand, if the property was labeled by a verb phrase, the verbalization of the CE became "A specific kind of **FC**: such that **P** something." The verbalization for the other anonymous CE patterns was analogous, e.g., for **c3**, "... such that has a **P** that is a **C**" (for a noun phrase P) and "... such that **P** a **C**" (for a verb phrase P), respectively.

Novice ontologist assessment. There were, again, two groups of students involved, from the same courses as in 2016 (Bc-level AI and MSc-level OE). The amount of prior training in OWL was also similar as in the 2016 campaign. There were 15 AI students and 16 OE students altogether. In both courses the students were first provided with a 30-minute overview of the notions of CE (in \mathcal{L}_{SE}) and FC roughly as presented in sections 2 and 3 of this paper. Then they completed an assignment consisting of 8 atomic tasks (two for each of the four CE patterns), all available in a single sheet of a web questionnaire.³⁵ The time allocated to completing the questionnaire was 40 minutes, i.e. 5 minutes per task.

The task question was slightly modified: it explored to what the degree the category is meaningful and *reusable*. The rationale was that possibly even subcategories with very small absolute or relative frequency might be viewed as meaningful (this term being rather vague and subjective), but undoubtedly their reusability should be perceived as low. Besides, the novelty of the assessment task setting compared to the 2016 campaign was in the following:

- The questionnaire separated the meta-question on *comprehension* from the actual 'meaningfulness' assessment, for each task. A separate question now inquired to what degree the student understood the meaning of the *individual entities* (assuming that the lack of familiarity with the enti-

ties strongly impacts the competency to assess a compound CE), with possible values that can be shortened as: 'quite familiar', 'roughly', 'pretty vague idea' and 'no clue'.

- The students could textually *justify a negative value*.
- For the anonymous CEs, the students could provide a *noun phrase* to which the verbalization of the category could be compressed.

We will however not discuss the last two types of metadata elements (the unstructured ones) in the current paper, to avoid thematic dilution.

In addition to the FCP tasks assessment, the questionnaire also examined the students' assessment of their own level of *written English*.

We computed the normalized sum (NS) of the task assessment values, as in the 2016 campaign (using Equation 12). The core results are as follows:

- The average NS over all 40 tasks was 0.27, i.e. much higher than in the first campaign (0.07). Of the 40 NS values, 34 were positive, 2 zero, and only 4 negative. This can possibly be attributed to the longer time available for each task, to the higher amount of available documentation, and/or to the verbalization of the CEs.
- The relative position of the anonymous CE patterns did not change from the 2016 campaign. The average NS was 0.33 for **c4** (all ten NS values being positive), 0.30 for **c3** (8 positives, 1 zero and 1 negative) and 0.16 for **c2** (7 positives, 1 zero and 2 negatives). The NS for **c1** was the highest, 0.39. This value is still surprisingly low, it is however strongly influenced by a single task with a negative NS. Upon removing this outlier, the NS would be 0.47.
- The cases with highest positive and lowest negative values are in Table 7; both the FC and the subcategory are now shown at the level of labels, just as presented to the students. The property and its filler, whether a class or an individual, are separated with a colon. The underlying ontology is referenced in the third column, through its nickname, which can be resolved against the LOV portal by appending it to <https://lov.linkeddata.es/dataset/lov/vocabs/>; to give a rough idea on the ontology size, the number of its classes is shown in parentheses. The CE pattern is listed in the fourth column, and the average NS value in the fifth column. Interestingly, the students were positive to-

³⁵The questionnaire was again in Czech, except the CEs (verbalized in English, to avoid issues with the inflection grammar of Czech). The English translation of a sample task is available in Appendix B.

wards some highly abstract focus classes (Quantity, and even Thing). As regards the negative cases, the first three rather seem to do with too-far-reaching connections between the involved entities; the last case suggests some odd modeling in the ontology itself (the students commented that a subcontinent should rather contain a country than be contained in it).

We also computed the relative frequencies reflecting the impact of the (declared) *English writing skills* and of the *comprehension of entities* on the assessment value:

- Of the 168 assessments by students with excellent or very good English skills, 102 (61%) were positive ('certainly' or 'perhaps'); in contrast, of the 80 assessments by students with fair or basic English skills, only 37 (46%) were positive.
- Of the 176 assessments where the students comprehended the meaning of the CE entities ('quite familiar' or 'roughly'), 126 (72%) were positive ('certainly' or 'perhaps'). In contrast, only 13 assessments of 72 (18%) where the students did not comprehend the meaning of the CE entities ('pretty vague idea' or 'no clue') were positive.

In order to reflect the degree of entity semantics comprehension in the assessment (with the assumption that more weight should be given to 'more informed' assessment), we also applied simple numerical weighting: the formula from Equation 12 was changed to

$$NS_i = \frac{\sum_{j=1}^k c_{i,j} \cdot v_{i,j}}{k} \quad (13)$$

where $c_{i,j}$ was set to 1 if the comprehension of the entities of task i by respondent j was 'quite familiar', to 0.66 for 'roughly', to 0.33 for 'pretty vague idea', and to 0 for 'no clue'. The results did not dramatically change. The average NS per CE pattern became as follows: 0.35 for **c1**, 0.32 for **c4**, 0.25 for **c3** and 0.14 for **c2**. Compared to the non-weighted normalized sum, **c4** marked the smallest decrease of all CE patterns.

Summary of the experiment Across the different campaigns and settings, the order of the CE patterns according to the average 'meaningfulness' of the categories remains stable: (**c1**) > **c4** > **c3** > **c2**. We can also conjecture that the categories are more likely to be viewed as 'meaningful' if the assessor is an expert, or if some of the following holds: more information is available on the entities and context; the categories are

better presented (e.g., verbalized); more time is available per category (task).

7.3. Data availability

The data from both experiments is available on the web through https://bit.ly/FCP_cogex. The page contains, for each of the campaigns, the links to:

- The description of each task (consisting in evaluation of the meaningfulness of a class expression) in each questionnaire variant
- The table with calculation of aggregated results.

8. Discussion

Since the empirical part of the paper may appear a bit fragmented and the results hard to align, in this section we first provide an integrative meta-view of the surveys / experiments settings and results.

From this we depart to a discussion of limitations and open questions of the analysis.

8.1. Meta-view of the empirical analyses

In Table 8 we synoptically summarize the three empirical pillars of our research so far, as elaborated in Sections 5, 6 and 7. We see that the surveys/experiments are to a large degree complementary, differing in their features: in the *structural type of the data source*; in the focus on either *direct* analysis of CEs or on their underlying patterns that are only *indirectly* tied to the CEs; in the *objectivity/subjectivity* of the obtained data; and, finally, in the actual *scope of the CE and FCE patterns*, and of the *focus classes* examined. Notably, all data sources currently refer to the Tbox. As mentioned before (esp. in Section 2.1), the fourth pillar of the empirical analysis of the FCP problem should be the analysis of the CEs occurrence in the *RDF datasets Abox* (which is ongoing but did not fit into the current paper).

The last row in the table attempts to summarize the core findings of each analysis. At the first sight the arrangement of the CE/FCE patterns might look incoherent. Especially, **c2** is (by Sec. 5) very rarely used in the axioms RHS, which contrasts with the fact that we can generate a huge number of matching CEs (by Sec. 6) through FCE pattern **p2** and the proportion of these CEs that are 'meaningful' is not dramatically lower than for the other CE patterns (by Sec. 7). This paradox can however be explained as follows:

Table 7
CEs with highest and lowest average NS of student scores, 2018 campaign

FC	Subcategory	Onto (# cl.)	CE pattern	Avg. NS
Security mechanism	BlockCipher	stac (211)	c1	0.79
Object	Machine	pext (488)	c1	0.75
Sensor	Gyro sensor	ha (88)	c1	0.75
Social position	has title: Prime minister	pext (488)	c4	0.75
Quantity	Mass	schema (625)	c1	0.67
Document	cites: Protocol	vivo (146)	c3	0.67
Thing	has dimension: Thing	ecrm (84)	c2	0.64
...
Service	has forecast: Thing	km4c (582)	c2	-0.21
CRM entity	is separated from: Beginning of existence	ecrm (84)	c3	-0.29
Feature	Camper service	km4c (582)	c1	-0.33
Subcontinent	in country: Thing	swpo (57)	c2	-0.50

- Sec. 5 considers the CEs within the RHS of axioms, where they primarily serve as a means for inferring the subordination of *arbitrary* instances (or classes) to the named class appearing in the LHS of the axiom; it is thus desirable that the RHS would specify some restrictive filler (as in **c3**) and not just *owl:Thing* (as in **c2**), to avoid ‘false positives’.
- On the other hand, Sections 6 and 7 already study the categorization in the setting with a known *focus class*, i.e. not for arbitrary instances. Then even the categorization of individuals based solely on the property they appear together with, irrespective the filler (as in **c2**), might work.

As regards the high ranking of **c4** in the cognitive experiments, in contrast with its low ranking in the other two studies, there is no discrepancy at all, since the cognitive experiments analyzed the meaningfulness of the categories (and thus, indirectly, their CE patterns) and not the frequency of occurrence of CE patterns at all (they were sampled evenly). The findings are thus completely orthogonal: there will typically be few **c4** CEs at hand, but they should be counted with a high weight in the FCP computation. Recalling the results from Section 6.2, for focus classes lower in the taxonomy, even the quantity of **c3** and **c4** CEs may proportionally increase.

8.2. Limitations and open questions

Since the paper explores a substantially novel problem space, the coverage of its different corners is still rather limited. In this section, we discuss three limitations and/or open questions, in turn: the omission

of Abox data in the whole process, the simplification made in the empirical analysis of CE pattern frequency, and the reliance on particular Tbox design principles in the application of FCE patterns.

The presented research currently ignores the role of *Abox analysis*, both in the *actual FCP computation* (within a concrete reuse problem) for a specific ontology as well as in the *inductive process* of CE pattern weight estimation. The former means that the presented approach is tuned towards the ‘cold-start’ setting (new ontologies not yet referenced in data); otherwise the relevance of concrete CEs could be estimated (though with some computational cost) from existing datasets instantiating them. The omission of the latter (which would consist in generating CEs for a given CEL over a representative set of ‘training’ ontologies, measuring their frequency of instantiation in data, and aggregating those frequencies by CE pattern) to date does not have any principled reason, and is only owing to limited human resources on the project.³⁶

The computation of the *CE pattern frequency* in axioms from Sec. 5 targeted the top-level structure in the axiom RHS (with additional distinction of the CE patterns from \mathcal{L}_{SE}). How complex are the embedded expressions within this top structure remains undistinguished. For example, for the conjunction, it would make sense to distinguish between a conjunction of two named classes, a conjunction of two property restrictions, a conjunction of a named class and a restriction, and several others. Characterizing the utility of a

³⁶While some early experiments are under way, their inclusion (in a low-maturity state) would have made the current paper lengthy and the contribution diluted.

Table 8
Summary of the complementary surveys / experiments

Survey / experiment	CE pattern frequency in axioms (Sec. 5)	FCE pattern occurrence in the Tbox (Sec. 6)	CE assessment via the cognitive experiments (Sec. 7)
Data source type	RHS of axioms (OWL)	Axioms (primarily RDFS)	Axioms (primarily RDFS); human assessors
Source ontologies	LOV, BioPortal, OntoFarm	LOV, OntoFarm	LOV, OntoFarm (2016 campaign); LOV (2018 campaign)
CEs examined	Directly	Indirectly (through FCE patterns)	Directly (by humans), but generated using FCE patterns
Objectiveness	Objective	Objective	Subjective (human judgment)
CE/FCE patterns	c2, c3, c4	p1, p2, p3, p4, p5	c1, c2, c3, c4
Focus classes	(Not relevant)	All classes	59+40 randomly selected classes
Gist of the quantitative findings	CE frequency: c3 most frequent (by an order of magnitude)	FCE pattern occurrence: p2 > p1 > p3 > p4 > p5 with p1 and p3 being swapped for less general FCs	Proportion of CEs meaningful for human assessors: c1 > c4 > c3 > c2 (but not differing dramatically)

CE pattern based on the top-level structure alone is obviously superficial, as it would vary according to the embedded levels.

The design of our *FCE patterns*, where *global* restrictions (domain+range) are used in the pattern template and *local* (existential) restrictions in the validity constraints, puts a strong assumption on the engineering approach used in the ontology analyzed. If the ontology lacks global restrictions³⁷ then the FCP will be under-estimated, since the FCE patterns will not match it at all. On the other hand, if the ontology lacks local restrictions, the FCE patterns will possibly generate many useless categories (those equivalent to the FC, or with empty or singleton extension), which would then have to be manually filtered out. This caveat could be compensated by computing the overall proportions of the number of classes plus local and global restrictions in the examined ontology and then adjusting the FCP accordingly.

Another potentially arguable assumption under which we gave preference to the *existential property restriction* as the cornerstone of our initially chosen CEL, over the *Boolean conjunction* (which would have been another simple candidate for a CE pattern), is that we *cannot* rely on proper equipment of ontologies with *disjointness axioms*. As mentioned in Section 5, in cases where we can (through pruning the ontology using those axioms) single out truly compatible pairs of classes, the consideration of conjunctive categories would be promising; in other cases it would however

lead to explosion of output CEs, which we wanted to avoid.

9. Operationalization and prototype implementation

While most of this paper is devoted to the description and formalization of the FCP ‘view’ of ontologies and to the associated empirical analyses, we also provide a tentative (or rather, illustrative) operationalization of the empirical findings into FCP computation weights, and describe an early prototype of an ontology search (for reuse) tool leveraging on FCP.

9.1. Tentative operationalization of the empirical results

We understand the obtained insights into the usage of OWL concept expressions and their perception by humans in general as a research achievement per se. However, the starting point for the overarching empirical study was an ‘engineering’ goal (possibly modest compared to the extent of the performed surveys and experiments): to propose *adequate weights* for the CE patterns to be used when computing the FCP in the context of an ontology reuse scenario.

In these terms, based on the cognitive experiments in particular, we can see that **c4** (individual value restriction) and to some degree **c3** (existential restriction) may successfully complement **c1** (named class) when harvesting categories from a reused ontology. The survey of the CEs in axioms, in turn, suggests that the chosen CE language (\mathcal{L}_{SE}) covers a decent part of the

³⁷Note that some best practices, e.g., in biomedicine, used to argue against domain/range axioms, see e.g. [19]

Table 9

Alternative CE pattern weights derived from the cognitive experiments

	c1	c2	c3	c4
Average NS mapped to [0, 1]				
Campaign 2016		0.48	0.51	0.61
Campaign 2018	0.70	0.58	0.65	0.67
Campaign 2018 weighted	0.68	0.57	0.63	0.66
Ratio of positive NS values				
Campaign 2016		0.24	0.45	0.65
Campaign 2018	0.90	0.70	0.80	1.00
Campaign 2018 weighted	0.90	0.80	0.90	1.00

space of relevant languages and could be applied in the first, albeit rough approximation of an ideal FCP formula. Thus, if a requirement came to design an ontology reuse metric leveraging on the FCP, based on the cognitive experiments we could tentatively set the weights in Equation 4 according to the NS of the assessments in the cognitive experiments. Two simple options of generating weights from the assessment NS are quantified in Table 9: by transforming the average NS from the $[-1; 1]$ interval to the $[0; 1]$ interval, and by computing the ratio of positive NS values of all NS values for the given CE pattern.

Considering that the lowered average NS values of **c1** in the 2018 campaign were presumably caused by randomly sampled highly general classes, which would not normally appear in the focus class role (as in the Feature class in Table 7), we should probably set $w^P(\mathbf{c1})$ to 1. The remaining weights can be set, heuristically reflecting both calculation methods and the results of both campaigns in combination, for example as $w^P(\mathbf{c2}) = 0.5$, $w^P(\mathbf{c3}) = 0.7$, and $w^P(\mathbf{c4}) = 0.8$. This would yield the following FCP estimate formula, to be applied on an ontology O and focus class FC , the language \mathcal{L} being \mathcal{L}_{SE} :

$$\widehat{FCP}(O, FC, \mathcal{L}, \{\{\mathbf{p1}\}, \{\mathbf{p2}\}, \{\mathbf{p3}\}, \{\mathbf{p4}\}\}) = \\ Occ(\{\mathbf{p1}\}, FC, O) \cdot 1 + Occ(\{\mathbf{p2}\}, FC, O) \cdot 0.5 + \\ + Occ(\{\mathbf{p3}\}, FC, O) \cdot 0.7 + Occ(\{\mathbf{p4}\}, FC, O) \cdot 0.8$$

The current version of the OReCaP tool (see the next subsection), when launched, proposes these values, i.e. $w^P(\mathbf{c2}) = 0.5$, $w^P(\mathbf{c3}) = 0.7$, and $w^P(\mathbf{c4}) = 0.8$. The formula should however only be taken as a very first, tentative initial setting (to be adapted by the user as needed) and not as a granted recipe.

9.2. OReCaP prototype

To demonstrate the whole focused categorization framework, specifically for \mathcal{L}_{SE} , patterns **p1–p4** and in the context of the ontology reuse scenario, as well to make it possible for both us and any interested subjects to perform experiments on various input ontologies and focus classes, we implemented a web-based application named OReCaP.³⁸ The application is available online at <https://fcp.vse.cz/orecap>. A demo of its earlier, limited version was already presented at the ISWC 2020 conference [17]. What is novel in the currently available version compared to that in the demo paper is, in particular, the support of *actual ontology reuse*, mediated by the user's interaction with different categories constructed using different FCE patterns.

The interaction workflow with the tool consists of several, possibly iterative, phases:

1. The process starts with a *keyword-based search* where the input consists of at least one *focused class keyword* and of optional *additional keywords*. The intuition behind the search interface is that the focused class keyword/s denotes the high- or medium-level type/s of entities whose instances are to be further sub-categorized using concepts from the ontology; the additional keywords, on the other hand, correspond to whatever domain terms. Imagine, for example, that the data is currently stored in a relational database. The focused class keyword might then often be the name of the top-level table (which can be, e.g., ‘Client’, ‘Patient’, ‘Vehicle’, ‘Account’, or the like); the additional keywords can be taken, e.g., from the names of subordinate tables, table columns, or predefined values for the fields.
2. The search returns a *sorted list of ontologies* whose classes match one or more of the provided keywords by their IRI, name or description; classes with a match of focused class keyword are listed first. The matched classes are listed for each ontology. Classes that match the *focused class keywords* are preselected (i.e., checked) by default; classes that matches the *additional keywords* are not preselected but can be selected (checked) manually by the user.
3. The next step is to execute the *FCP calculation* for a chosen ontology, given the selected

³⁸The acronym refers to the terms ‘ontology reuse’ and ‘categorization power’.

classes as focus classes, by clicking on the ‘Calculate FCP’ button. In a pop-up window, metadata about the ontology including its IRI and namespace is displayed, along with the total FCP score, which is calculated based on the FCP weight values and the categorizations (i.e., class expressions) listed at the bottom. This score is the sum of all partial scores for each focus class. The weight values can be adjusted for each calculated ontology according to the user’s assessment of each CE pattern, and the resulting FCP score will change accordingly. The global FCP weights can be changed in the settings section, so that every new FCP calculation would use them as the default weight values.

4. The calculated FCP score is then displayed below the ontology overview in the search results, and also saved to a *comparison list*, which shows the FCP-based ranking of the ontologies.
5. From the comparison list it is possible to proceed to the *reuse summary* view of a given ontology. There the user can inspect all generated categories in the form of a pop-up tree, and can check those s/he eventually wants to reuse.
6. The actual *reuse* is currently performed two ways: first, a particular ontology (chosen by the ranking) can be downloaded as is. Second, the axioms defining the named classes newly constructed from compound expressions can be downloaded separately from the reuse summary view, to be later manually added to the ontology.

Let us assume the user wanting to publish data about business contracts and their payment, and seeking a suitable ontology for their subcategorization. A partial screenshot showing the overall search results together with FCP scores for two ontologies is in Fig. 2. The focus class keywords (here, just one) are entered in the top-left field; their matches are therefore proposed (highlighted in blue) as focus classes, by the tool. The user has also provided additional keywords, which may improve the result ranking but do not produce further focus classes (unless the user pro-actively highlights them, too).

For the first ontology, PPROC, a snippet of the reuse summary window is then shown in Fig. 3; the numbers (e.g., “(1/5)”) indicate how many categories were chosen in the given sub-tree. The user has chosen two categories, corresponding to ‘contract that has been anyhow modified’ (CE pattern **c2**) and ‘contract that has been ended by procedure resignation’ (CE pattern **c3**).

The set of axioms defining new named classes based on such compound CEs, with shortnames currently created by trivial concatenation, can then be downloaded; for example:

$$\begin{aligned} \text{ContractHavingContractModification} &\equiv \\ \text{Contract} \sqcap \text{ContractModification}. \top \end{aligned}$$

A custom namespace can be set for those new classes.

OReCaP makes use of the Linked Open Vocabulary API³⁹ for the keyword-based search and for retrieving the ontology metadata. The FCP calculation itself is implemented on top of OWL API, in combination with the jFact reasoner.⁴⁰ OWL API is used to load and parse the ontology source codes, and jFact is used to infer class expressions. Both the front-end and back-end implementation of OReCaP are open-source and available on GitHub⁴¹ under the MIT license.

10. Related work

Since the research described in this paper addresses the focused categorization power problem from various angles, multiple areas of related research can be identified. In this section we report on the following, in turn: abstract notions similar to our notion of FCP; empirical studies on presence of concept expressions and structural patterns in ontology repositories; cognitive experiments on assessing ontological structures; concept learning in DL; ontology reuse metrics and methods.

10.1. Abstract notion of focusing or categorization (power) in ontologies

We are unaware of prior work on the same topic of FCP as we coin it in the current paper. We will however reference some related research that overlaps with ours at the abstract level.

The notion of focusing recently appeared in the work of Gogacz et al. [9]. The so-called focusing solution pairs a set of predicate symbols that describe a database schema (that is, a set of predicates) with a set

³⁹<https://lov.linkeddata.es/dataset/lov/api>

⁴⁰<https://github.com/owlcs/owlapi>, <https://github.com/owlcs/jfact>

⁴¹<https://github.com/nvbach91/orecap>, <https://github.com/nvbach91/fcp-api>

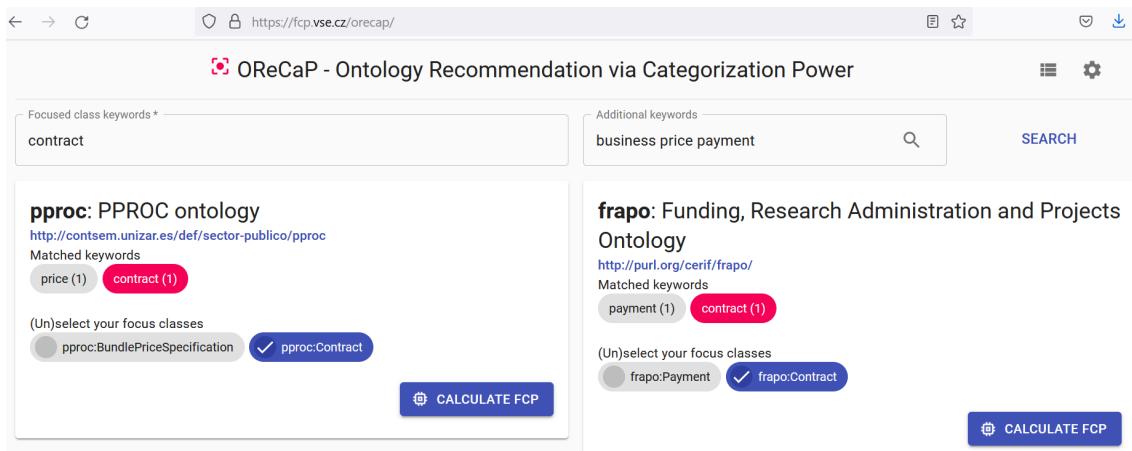


Fig. 2. OReCaP interface: two ontologies found via keyword search, with focus class and additional match

Re-use summary

- > **pproc:additionalDocumentReference** (0/1) <expressions>
- > **pproc:complementaryContract** (0/30) <expressions>
- ✓ **pproc:contractModification** (1/1) <expressions>
- ✓ **pproc:contractModification** owl:Thing
- > **pproc:contractObject** (0/1) <expressions>
- > **pproc:contractOrProcedureExtinction** (1/5) <expressions>
- pproc:contractOrProcedureExtinction** owl:Thing
- ✓ **pproc:contractOrProcedureExtinction** — **pproc:ProcedureResignation** <classes>
- pproc:contractOrProcedureExtinction** — **pproc:ProcedureWaive**

Fig. 3. OReCaP interface: the reuse summary, with user's choices of two CEs

of assumptions on the partial completeness on the data and the ontology (closed and fixed queries). In their approach, focusing is about choosing which parts of the data and ontology are to be declared complete, to allow for efficient reasoning. In our approach, focusing is about choosing ontology classes for whose instances in data we would like to obtain many meaningful categorization options; the categorization itself however need not rely on logical reasoning (never mind using the ontology), but can be based on whatever kind of classification model or even made by humans.

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 (typically, machine-learning-based) 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. The association of the notions of ‘categorization’ is thus merely verbal.

Partially relevant is the analysis made by Giunchiglia & Zaihrayeu [8], 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 FCP applies to individuals intrinsic to the DL world of the ontology and is calculated with respect to a focus class.

A. Rector’s work on entangling hierarchies (normalization) [18] addressed a different problem than us, but to some degree analogously considered the compound concepts as an alternative to named ones. This applies to ‘partitioning’ or ‘refining’ concepts, that only modify the ‘self-standing’ concepts; secondary partitioning aspects should not be expressed through subclassing (yielding a multi-hierarchy) but through existential restrictions filled with classes from separate ‘codelist’ taxonomies. For example, a class *SteroidHormone* would not have two parents, *Steroid* and *Hormone*, but only the first one, while the latter subordination would be made through an existential restriction: $\text{SteroidHormone} \equiv \text{Steroid} \sqcap \exists \text{playsRole}.\text{Hormone}$.

Our own ongoing work on the PURO modeling language [25] 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 *entrepreneur*, 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 CE), assuming we prefer an encoding style using object properties with ‘codelist individuals’. 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.⁴² Modeling in PURO and applying the transformation patterns may thus make hidden (compound) relevant concepts explicit in the domain as atomic concepts. A similar account of alternative ‘typecasting’ beyond named classes (but with smaller coverage) has been given by Krisnadhi [12].

Certain research in cognitive psychology might also be relevant wrt. the notion of FCP. In particular, the notion of *graded structure* of categories [6] can be applied to our concept pattern weighting. The authors suggest that there is “a basic level of abstraction (e.g. CHAIR, DOG, …)” which is “further discriminated at the subordinate level (e.g. KITCHEN CHAIR, SPANIEL) and abstracted at the superordinate level (FURNITURE, ANIMAL)” [6]. Presumably, this basic level might often coincide with the focus classes to be subcategorized in practical settings. Furthermore, the graded structure may include “ad hoc, goal derived categories such as GOOD PLACES TO HIDE FROM THE MAFIA” [6], analogous to our CEs defined via an existential restriction.

10.2. Empirical studies of ontology features

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. [26] (aiming to build a benchmark for testing ontology tools). A large scale study of OWL ontology metrics was carried out by Matentzoglu et al. [16]. However, the categorization power of ontologies has not been, to our knowledge, studied, never mind with the flavor presented here.

Our study on concept expression frequency in axioms from Section 5 looks similar to the recent study carried out in the MontoloStats project [15]. Both studies essentially analyze the same ontology repositories (primarily, LOV and Bioportal), and refer to the suitability of ontologies for reuse. There is however a difference in the restrictions coverage. For an unclear reason, the MontoloStats study does not cover existential restrictions (which are central for our study, and also shown as empirically very frequent) at all, nor the conjunctive concepts. On the other hand, it covers (sub-class axioms with) named class in the RHS, and also property axioms such as domain/range or functional

⁴²<http://wiki.dbpedia.org/services-resources/ontology>

property. Notably, even over the common subset of CE patterns in restrictions (such as disjoint, universal and cardinality restriction CEs) that some differences in the computed ranking appear between MontoloStats and our research; these may be due to additional distinct features in the methodology used.

Yet another stream of empirical research aims to study ontologies not on their own but from the point of view of LOD datasets in which they are used. This was the subject of the project by Asprino et al. [3], which produced a condensed representation of the global, virtual, ‘LOD ontology’ in the form of so called equivalent set graphs. Various metrics related to the connectedness and extensional size of ontology entities were computed; while this research does only addresses compound concepts, it is in line with our ongoing activity in analyzing the Abox imprint of (named as well as compound) concept expressions.

10.3. Cognitive experiments on assessing ontological structures

Several cognitive studies using ontologies as material have been published. There are several cognitive studies present in the ontological engineering research. However, they primarily address the capability of humans themselves to carry out the categorization of objects to a set of classes or to understand the structure of OWL expressions. A recent example of the former is a study on classifying domain entities to upper-level ontology classes [23]. An example of the latter is an earlier study on the human capability of deriving useful information from differently verbalized OWL statements [27]. Our research in Section 7 of this paper differs in that the humans were to assess the automatically build concepts as more or less plausible, thus generating ground truth. (Semi-automated) verbalization was present, too, but only played an auxiliary role, the actual subject of assessment being the formal CEs themselves.

10.4. Concept learning in DL

The heuristic construction of compound CEs from the ontology axioms, triggered by the identification of a focus class, bears some resemblance to *concept learning* in DL [14], which is triggered by the specification of the target concept. However, (supervised) concept learning aims to identify CEs that best approximate a concrete goal concept as whole. In contrast, CEs logically equivalent to the focus class are uninter-

esting as its subcategories, while categories that only cover a few per cent of the FC instances may be valid means for partitioning its instances. Moreover, concept learning fully depends on presence of instance data for the *particular concept* to be learned. Yet, for a newly designed ontology such data might not be available (the well-known cold start problem). Finally, the overall goal is different: while concept learning primarily aims at enriching the ontology Tbox with new axioms [13], in our approach the prime task is to reuse the ontology (or its fragments) ‘as it is’, and the formation of new equivalence axioms is an optional, secondary step carried out with human interaction.

10.5. Ontology reuse metrics and methods

The broad context of our research, the task of ontology reuse, was studied by Schaible et al. [21]: 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.

Vocabulary reuse techniques similar to the use of FCP-based metrics also appeared in a recent project on combining popularity metrics with the credibility of the vocabulary designers [22]. As regards the designer credibility, this is a feature of the ontology itself similarly to FCP, but it is completely orthogonal.

Reuse support [7] is also systematically sought by the maintainers of LOV [28], primarily at keyword relevance level; we are in contact with them and will seek to integrate our complementary approaches.

11. Conclusions and future work

Ontologies are an important means of subcategorizing entities already known to belong to a general focus class. Ontologies with the best *categorization power*, in terms of the number and quality of available subcategories of the *focus class/ies*, should have preference in major *ontology reuse* scenarios, both for publishing data in RDF or building a new ontology. We demonstrated that the scope of subcategories need not be confined to named classes but can also cover *compound ones*. In the first approximation, we treated the computation of the focused categorization power of ontologies, beyond named subclasses, in terms of *simple existential restrictions over properties*. This appears par-

ticularly relevant for publishing datasets on the linked data cloud, which is relatively ‘property-centric’.

We provided, to our knowledge, a pioneering study on the focused categorization power problem, starting from its *formalization*, through the *operationalization* of the FCP computation (for a particular concept expression language) via heuristic Tbox patterns, to *empirical analyses* carried out both fully automatically over ontology repositories and via cognitive experiments with human users; a software demonstrator is also provided.

Ongoing research addresses the analysis of the CE *instantiation in linked datasets* (Abox). This information can be employed two-ways. First, directly for the CEs generated from the Tbox of the given ontology considered for reuse. Second, the occurrence of instantiation of a compound CEs can be compared to the occurrence of its constituent entities, in the same dataset. The ratio of the compound vs. individual instantiation can then be aggregated per the pattern of compound CE, and provide an empirical grounding for the ‘meaningfulness’ of the CE pattern, complementary to the human feedback from the cognitive experiments. A dedicated thread of the Abox analysis will also study the availability of external, deferefencable *SKOS codelists*, which would enter the FCP computation through FCE patterns such as **p5**.

Another area of ongoing research concerns the techniques of *pattern-based lexical transformation* of compound CEs to named ones (now merely as a trivial concatenation with *having* in OReCaP). Based on an already performed empirical analysis of naming patterns in the LOV ontologies, the transformation method will be finalized and implemented; a naming pattern catalog will also appear, as a side-product.

In middle term, we plan to extend the *concept expression language* considered. A simple extension to \mathcal{L}_{SE} could be the inclusion of the *inverseOf* predicate: in some modeling styles only one of the pair of mutually inverse properties is present in the ontology, and the entity categorization should then be carried out even against the direction of the respective property. Next, we plan to move towards some other *class expression patterns* identified as common in OWL axioms, by our study, such as universal restrictions, conjunctions (for ontologies equipped with disjointness axioms), disjunctions (relevant for group categorization) and some nested expressions, especially concepts based on *property paths* of length 2. Besides we also want to consider *data property restrictions*; they are rare in ontology axioms, but possibly highly

relevant for generating subcategories, since data properties abound, in particular, in markup-oriented vocabularies such as *schema.org*. Relevance for the FCP is quite obvious for Boolean-valued properties, but even numerical properties, where the categories would correspond to value intervals, should be taken into account.

Within the scope of the different CE patterns, the syntactic *FCE patterns* providing the CEs will also be extended (e.g., by considering the *subPropertyOf* relationship) and new heuristic alternatives possibly added. The algorithm for *pattern-based generation* of CEs will also be enhanced, primarily by applying post-pruning according to the pattern applicability constraints. Overall, the transition to more expressive CELs and thus more expressive patterns will of course mandate deeper *computational complexity analyses*.

In addition to the logical-structural side of the CEs, we may also leverage on their lexical, foundational-semantic and graph-based aspects. For the *lexical aspects*, e.g., the 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 ‘code’ of a type (note the highly scoring case with *pco:authorityKind* in Table 6). The mapping (if available – in rare cases) of the analyzed ontology to *foundational ontologies*, or ontology *background models* such as OntoClean metaproPERTIES [10] or PURO models [25], will provide a similar kind of information – e.g., that the individual in the property filler is inherently a type to be assigned to the instances of the class in the axiom LHS. Finally, the *graph-based* view may lead to penalization of compound CEs for which the graph distance of their constituents in the ontology is too high. All this, in turn, would require an adaptation of the formal definition of a CEL, in which the placeholder expressions would need to have linguistic patterns (or just regexps, in the first approximation), graph distance parameters or foundational meta-features assigned.

The pool of testing subjects in *cognitive experiments* should be extended to external ontology engineering experts (on the top of students and internal experts, engaged in the previous experiments). The new experiments should also feature more carefully pre-selected (rather than random) assessment tasks, so as to provide targeted feedback to so far unclear constellations of focus class and subcategory.

Since the main foreseen practical application of the whole approach is the improvement of *ontology reuse*, we also plan to explore this task in full, considering

the FCP-based approach merely as a single element to be combined with other state-of-the-art supporting techniques for ontology reuse. Studies in different domains – e-commerce, academia, etc. – will be carried out. Especially in *e-commerce*, where product catalogs abound, manual *re-engineering* of the product type textual labels (intuitively designed for e-shop visitors' easy navigation), to the underlying CE patterns should be performed, to gain additional empirical material both for the FCP computation itself and for the lexical transformation of compound CEs to named categories. The OReCaP tool itself, with its two-level distinction of category quality (mere reuse of an anonymous CE, and its transformation to a new named class), could serve as collector of empirical material from volunteers willing to experiment with it.

As regards ontology reuse for describing a new RDF dataset, we should put more attention to the possible availability of *textual resources* describing the source data in the given ontology reuse case. For example, a used car retailer may employ data sheets filled with values of various textually expressed car parameters. On the one hand, the initial keyword search for relevant ontologies should better only employ the focus concepts, since the labels of parameters might dilute the search results (they would often be generic, e.g., most attributes describing a car could also apply to other machines, products, or even physical objects in general). On the other hand, as soon as the candidate ontologies have been pre-selected, the parameter labels can influence the FCP computation itself, boosting the individual weight w^I of categories that lexically match them.

We should also take into account not only the reuse of an ontology for publishing a dataset, but also the reuse of an ontology by *another ontology*. The latter can then be either ‘hard’ (by ontology import) or ‘soft’ (by merely referencing its entities) [5].

To conclude, it is worth noting that while the reuse scenario was our primary motivation in studying the FCP problem, we foresee that the scope of usage for the FCP models and techniques might be wider, including, e.g., the general evaluation of ontology design quality, detection of ontology patterns or effective visualization. In this sense, we believe that our study may have brought to light a substantially innovative perception of the ontologies published (not only) on the semantic web.

Acknowledgments

The research had been supported by CSF 18-23964S, “Focused categorization power of web ontologies”, by projects ORBIS, funded by Slovak SRDA agency under contract No. APVV-19-0220, KATO, funded by Slovak VEGA agency under contract No. 1/0778/18, and by TAILOR, funded by EU Horizon 2020 research and innovation programme under GA No. 952215.

References

- [1] OWL 2 web ontology language profiles (second edition). W3C Recommendation, December 2012. URL <https://www.w3.org/TR/owl2-profiles/>.
- [2] OWL 2 web ontology language structural specification and functional-style syntax (second edition). W3C Recommendation, December 2012. URL <https://www.w3.org/TR/owl2-syntax/>.
- [3] Luigi Asprino, Wouter Beek, Paolo Ciancarini, Frank van Harmelen, and Valentina Presutti. Observing LOD using equivalent set graphs: It is mostly flat and sparsely linked. In Chiara Ghidini, Olaf Hartig, Maria Maleshkova, Vojtech Svátek, Isabel F. Cruz, Aidan Hogan, Jie Song, Maxime Lefrançois, and Fabien Gandon, editors, *The Semantic Web - ISWC 2019 - 18th International Semantic Web Conference, Auckland, New Zealand, October 26-30, 2019, Proceedings, Part I*, volume 11778 of *Lecture Notes in Computer Science*, pages 57–74. Springer, 2019. . URL https://doi.org/10.1007/978-3-030-30793-6_4.
- [4] Franz Baader and Paliath Narendran. Unification of concept terms in description logics. *J. Symbolic Computation*, 21:277–305, 2001.
- [5] Mariano Fernández-López, María Poveda-Villalón, Mari Carmen Suárez-Figueroa, and Asunción Gómez-Pérez. Why are ontologies not reused across the same domain? *J. Web Semant.*, 57, 2019.
- [6] Liane Gabora, Eleanor Rosch, and Diederik Aerts. Toward an ecological theory of concepts. *Ecol. Psychology*, 20, 2008.
- [7] Nuria García-Santa, Ghislain Auguste Atemezing, and Boris Villazón-Terrazas. The ProtégéLOV plugin: Ontology access and reuse for everyone. In *The Semantic Web: ESWC 2015 Satellite Events*, pages 41–45, 2015. .
- [8] Fausto Giunchiglia and Ilya Zaihrayeu. Lightweight ontologies. In Ling Liu and M. Tamer Özsu, editors, *Encyclopedia of Database Systems*, pages 1613–1619. Springer US, 2009. ISBN 978-0-387-39940-9.
- [9] Tomasz Gogacz, Víctor Gutiérrez-Basulto, Yazmín Ibáñez-García, Filip Murlak, Magdalena Ortiz, and Mantas Simkus. Ontology focusing: Knowledge-enriched databases on demand. In Giuseppe De Giacomo, Alejandro Catalá, Bistra Dilkina, Michela Milano, Senén Barro, Alberto Bugarín, and Jérôme Lang, editors, *ECAI 2020 - 24th European Conference on Artificial Intelligence, 29 August-8 September 2020, Santiago de Compostela, Spain, August 29 - September 8, 2020 - Including 10th Conference on Prestigious Applications of Artifi-*

- cial Intelligence (PAIS 2020)*, volume 325 of *Frontiers in Artificial Intelligence and Applications*, pages 745–752. IOS Press, 2020. .
- [10] Nicola Guarino and Christopher A. Welty. An overview of OntoClean. In Steffen Staab and Rudi Studer, editors, *Handbook on Ontologies*, pages 201–220. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-540-92673-3. . URL http://dx.doi.org/10.1007/978-3-540-92673-3_9.
- [11] Ullrich Hustadt, Renate A Schmidt, and Lilia Georgieva. A survey of decidable first-order fragments and description logics. *Journal of Relational Methods in Computer Science*, 1(251–276):3, 2004.
- [12] Adila Alfa Krisnadhi, Pascal Hitzler, and Krzysztof Janowicz. On the capabilities and limitations of OWL regarding typecasting and ontology design pattern views. In Valentina A. M. Tamia, Mauro Dragoni, Rafael Gonçalves, and Agnieszka Lawrynowicz, editors, *OWLED*, volume 9557 of *Lecture Notes in Computer Science*, pages 105–116. Springer, 2015. ISBN 978-3-319-33244-4.
- [13] Jens Lehmann and Lorenz Bühmann. ORE - a tool for repairing and enriching knowledge bases. In *Proceedings of the 9th International Semantic Web Conference (ISWC2010)*, volume 6497 of *Lecture Notes in Computer Science*, pages 177–193, Berlin/Heidelberg, 2010. Springer. .
- [14] Jens Lehmann and Pascal Hitzler. Concept learning in description logics using refinement operators. *Machine Learning*, 78(1-2):203–250, 2010.
- [15] Sven Lieber, Ben De Meester, Anastasia Dimou, and Ruben Verborgh. MontoloStats - ontology modeling statistics. In Mayank Kejriwal, Pedro A. Szekely, and Raphaël Troncy, editors, *K-CAP*, pages 69–76. ACM, 2019.
- [16] Nicolas Matentzoglu, Samantha Bail, and Bijan Parsia. A snapshot of the OWL web. In Harith Alani, Lalana Kagal, Achille Fokoue, Paul T. Groth, Chris Biemann, Josiane Xavier Parreira, Lora Aroyo, Natasha F. Noy, Chris Welty, and Krzysztof Janowicz, editors, *International Semantic Web Conference (1)*, volume 8218 of *Lecture Notes in Computer Science*, pages 331–346. Springer, 2013. ISBN 978-3-642-41334-6.
- [17] Viet Bach Nguyen, Vojtech Svátek, Gollam Rabby, and Ondrej Zamazal. OReCaP - towards ontology reuse via focused categorization power. In Kerry L. Taylor, Rafael S. Gonçalves, Freddy Lécué, and Jun Yan, editors, *Proceedings of the ISWC 2020 Demos and Industry Tracks: From Novel Ideas to Industrial Practice co-located with 19th International Semantic Web Conference (ISWC 2020), Globally online, November 1-6, 2020 (UTC)*, volume 2721 of *CEUR Workshop Proceedings*, pages 295–300. CEUR-WS.org, 2020. URL <http://ceur-ws.org/Vol-2721/paper575.pdf>.
- [18] Alan L. Rector. Modularisation of domain ontologies implemented in description logics and related formalisms including OWL. In John H. Gennari, Bruce W. Porter, and Yolanda Gil, editors, *K-CAP*, pages 121–128. ACM, 2003. ISBN 1-58113-583-1.
- [19] Alan L. Rector, Nick Drummond, Matthew Horridge, Jeremy Rogers, Holger Knublauch, Robert Stevens, Hai Wang, and Chris Wroe. Owl Pizzas: Practical experience of teaching OWL-DL: Common errors & common patterns. In Enrico Motta, Nigel Shadbolt, Arthur Stutt, and Nicholas Gibbins, editors, *EKAW*, volume 3257 of *Lecture Notes in Computer Science*, pages 63–81. Springer, 2004.
- [20] Yuan Ren, Artemis Parvizi, Chris Mellish, Jeff Z. Pan, Kees van Deemter, and Robert Stevens. Towards competency question-driven ontology authoring. In Valentina Presutti, Claudia d’Amato, Fabien Gandon, Mathieu d’Aquin, Steffen Staab, and Anna Tordai, editors, *The Semantic Web: Trends and Challenges - 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings*, volume 8465 of *Lecture Notes in Computer Science*, pages 752–767. Springer, 2014. .
- [21] Johann Schaible, Thomas Gottron, and Ansgar Scherp. Survey on common strategies of vocabulary reuse in linked open data modeling. In Valentina Presutti, Claudia d’Amato, Fabien Gandon, Mathieu d’Aquin, Steffen Staab, and Anna Tordai, editors, *ESWC*, volume 8465 of *Lecture Notes in Computer Science*, pages 457–472. Springer, 2014. ISBN 978-3-319-07442-9.
- [22] Ioannis Stavrakantonakis, Anna Fensel, and Dieter Fensel. Linked open vocabulary ranking and terms discovery. In Anna Fensel, Amrapali Zaveri, Sebastian Hellmann, and Tassilo Pellegiani, editors, *SEMANTICS*, pages 1–8. ACM, 2016. ISBN 978-1-4503-4752-5.
- [23] Robert Stevens, Phillip Lord, James Malone, and Nicolas Mantzaoglou. Measuring expert performance at manually classifying domain entities under upper ontology classes. *J. Web Semant.*, 57, 2019.
- [24] Vojtech Svátek, Ondrej Zamazal, and Miroslav Vacura. Categorization power of ontologies with respect to focus classes. In *Knowledge Engineering and Knowledge Management - 20th International Conference, EKAW 2016, Bologna, Italy, November 19-23, 2016, Proceedings*, pages 636–650, 2016. . URL https://doi.org/10.1007/978-3-319-49004-5_41.
- [25] Vojtech Svátek, Martin Homola, Jan Kluka, and Miroslav Vacura. Metamodeling-based coherence checking of OWL vocabulary background models. In Mariano Rodriguez-Muro, Simon Jupp, and Kavitha Srinivas, editors, *OWLED*, volume 1080 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2013.
- [26] Christoph Tempich and Raphael Volz. Towards a benchmark for semantic web reasoners - an analysis of the DAML ontology library. In York Sure and Óscar Corcho, editors, *EON*, volume 87 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2003.
- [27] Allan Third. "Hidden semantics": what can we learn from the names in an ontology? In *INLG*, pages 67–75, 2012.
- [28] Pierre-Yves Vandenbussche, Ghislain Atemezing, María Poveda-Villalón, and Bernard Vatant. Linked open vocabularies (LOV): A gateway to reusable semantic vocabularies on the web. *Semantic Web*, 8(3):437–452, 2017.

Appendix A: Questionnaire task description from the 2016 cognitive experiment campaign

FC and category in Manchester syntax:

- Individual: ?i Types: Person
- Individual: ?i Types: bornIn some Thing

?i is an instance of a class including all objects that are in the relationship ‘bornIn’ to at least one object.

Is the class (bornIn some Thing) a meaningful category for categorizing objects of class Person?

Hint: the expression, e.g. (bornIn some Thing), suitable for categorization of the FC should satisfy the following conditions:

1. there may theoretically exist a higher number (i.e. not 0, 1 or just a few) FC instances satisfying the expression (absolute number of instances)
2. there may theoretically exist a higher number (i.e. not 0, 1 or just a few) FC instances *not* satisfying the expression, in turn
3. ideally (this is not a strict condition) the expression is not apparently tied to some specific subclass of the given FC, such that it would not make any sense to consider it for other subclasses of this FC.

Possible values:

- certainly (2)
- perhaps (1)
- borderline (0)
- perhaps not (-1)
- certainly not (-2)
- no judgment, since I don't understand the example (N)

Appendix B: Questionnaire task description from the 2018 cognitive experiment campaign

The focus class is from the ontology no. 162945, <http://www.ontotext.com/proton/protonext>

Ontology name: Proton Ontology

Ontology description:

"PROTON (PROTo ONtology) was developed in the SEKT project as a lightweight upper-level ontology, serving as a modeling basis for a number of tasks in different domains. To mention just a few applications: PROTON is meant to serve as a seed for ontology generation (new ontologies constructed by extending PROTON); it can be used for automatic entity recognition and more generally Information Extraction (IE) from text, for the sake of semantic annotation (metadata generation). PROTON was extended to cover the conceptual knowledge encoded within the most popular datasets from Linked Open Data like DBpedia,

GeoNames, etc."

Involved entities:

Focus class: **Object**

Property: **owner**

Target class: **Parliament**

Proposed category:

A specific kind of Object: such that has an owner that is a Parliament

Description of entities:

Object: Objects are entities that could be claimed to exist - in some sense of existence. An object can play a certain role in some happenings. Objects could be substantially real - as the Buckingham Palace or a hard-copy book - or substantially imperceptible - for instance, an electronic document that exists only virtually, one cannot touch it.

owner: The relationship between an object and an agent who owns it.

Parliament: A legislative assembly representative at national or regional level. It can also be called Senate, etc.

It holds:

Parliament SubClassOf: PoliticalEntity

Do you understand the meaning of these entities, making use of the description above?

Possible values:

- Yes, I am quite familiar with them (3)
- Yes, roughly (2)
- I only have a pretty vague idea (1)
- Absolutely no clue (0)

Is "Object that has an owner that is a Parliament" a meaningful and reusable category for instances of the class Object?

Possible values:

- Certainly (2)
- Perhaps (1)
- Borderline (0)
- Perhaps not (-1)
- Certainly not (-2)
- No judgment, since I don't understand the example (N)