

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

Vojtěch Svátek [a,*], Ondřej Zamazal [a], Miroslav Vacura [a,b] and Jiří Ivánek [a]
[a] *Department of Information and Knowledge Engineering, University of Economics, Prague, Nám. W. Churchilla 4, 130 67 Praha 3, Czech Republic*
E-mail: {svatek, ondrej.zamazal, ivanek}@vse.cz
[b] *Department of Philosophy, University of Economics, Prague, Nám. W. Churchilla 4, 130 67 Praha 3, Czech Republic*
E-mail: vacuram@vse.cz

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 concept expression language based on existential restrictions, and heuristically map it to syntactic patterns over ontology axioms. The characteristics of the chosen concept expression language and associated patterns are investigated using three different empirical sources derived from ontology collections: first, the concept expression type frequency in class definitions; second, the occurrence of the heuristic patterns (mapped on the expression types) in the Tbox of ontologies; and last, for two different samples of concept expressions generated from the Tbox of ontologies (through the heuristic patterns) their 'meaningfulness' was assessed by different groups of users, yielding a 'quality ordering' of the concept expression types. 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.

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

1. Introduction

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

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

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 [3]. 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* [14], especially, whether the ontologies are listed in catalogs such as LOV – Linked Open Vocabularies[1] [20].

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 thematically related to the reuse case does not mean that it structurally fits well the data that is to be semantically described.

Our proposed approach to enhancing web ontology reuse is based on the intuition 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), is an indicator of the reusability of the ontology for a dataset containing objects known to belong to this class.

Additionally, we hypothesize 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. This aspect is however mainly left to future work and is only elaborated in the current paper at a high level.

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 ontologi/es whenever possible, in view of achieving interoperability with (presently unknown) applications of partners, search engines and aggregators. How could the computation of focused categorization power be useful when assessing the potential of existing ontologies for being reused in this case?

– 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

[1] `https://lov.linkeddata.es/`

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 ontologi/es (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 will first be *transformed* (ideally, in a partially automated way) into regular noun phrases (that would, explicitly or implicitly, correspond to named OWL classes).

The present paper is an evolution of a previous conference paper [16], 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 extends the previous one along numerous axes:

- 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 formally defined, and several instances of it introduced, including the CEL of compound expressions used further in the experiments.
- The survey on *syntactic pattern occurrence* in ontology collections, only present in an online addendum to the previous paper [16], 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 mere 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 rest of the paper is structured as follows. Section 2 presents the formal framework of focused categorization power in general terms. Section 3 introduces a particular concept expression language having suitable computational properties for an initial study. Section 4 complements this language with syntactic patterns allowing to generate concept expressions conforming to it from the ontology Tbox. Section 5 surveys the occurrence of concept expression types in ontology axioms. Section 6, analogously, surveys the occurrence of the previously defined syntactic 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 of different types. Section 8 provides a comparison of those complementary surveys and experiments, and discusses a possible operationalization of their results. Section 9 reviews some related methods and projects. Finally, Section 10 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 covered by the rest of the paper. We will proceed from the notion of *concept expression language* to the notion of *$\mathcal{L}$-category*, as a concept 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 patterns in the ontology.

2.1. Concept expression language

OWL [2], being a primarily *concept-based* formal ontology language, allows to create complex class expressions of many kinds using a collection of constructors. 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 [8], 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*[2] that generalizes the notion of a restricted formal sublanguage of OWL to that only usable for the creation of concept expressions; we use the description logic (DL) syntax instead of the textual functional-style OWL syntax, for brevity.

Definition 1 *A* concept expression language *(CEL) is a formal language for constructing concept expressions (CE) corresponding to class expressions in OWL 2 [2, section 8]. Every CEL is a pair* $\mathcal{L} = (T_\mathcal{L}, R_\mathcal{L})$, *where*

- $T_\mathcal{L}$ *is the set of concept expressions* types *(CE types) allowed in* $\mathcal{L}$,
- $R_\mathcal{L}$ *is the set of* restrictions *placed on the concept expressions instantiating the CE types from* $T_\mathcal{L}$.

The possible elements of $T_\mathcal{L}$ *are:* C, $C \sqcap D$, $C \sqcup D$, $\neg C$, $\exists P.C$, $\forall P.C$, $\leq_n C$, $\geq_n C$, $=_n C$ *and* $\{i_1, i_2, \ldots, i_k\}$ *(enumeration), where C and D are placeholder variables for concept expressions, P is a placeholder variable for a role expression corresponding to a property expression in OWL 2*[3], $i_1, i_2, \ldots, i_k$ *are placeholder variables for individuals, and n is a placeholder variable for natural numbers.*

The restrictions in $R_\mathcal{L}$ *may limit the substitutions for variables in concept and role expressions in an arbitrary way.*

One example of a CEL may be the trivial language of named concepts, let us call it $\mathcal{L}_{nam}$ for future reference. $T_{\mathcal{L}_{nam}} = \{C\}$, and $R_{\mathcal{L}_{nam}}$ only allows a named (atomic) class to be substituted for C.

Another example would be a formal language, let us call it $\mathcal{L}_{extop}$ (language of existential restrictions to the top concept), such that $T_{\mathcal{L}_{extop}} = \{\exists P.C\}$ and $R_{\mathcal{L}_{extop}}$ contains two restrictions: a restriction on C only allowing it to be substituted with the top concept $\top$ and a restriction on P only allowing it to be substituted with an atomic property.

For conciseness, we will also explicitly define the notion of expression corresponding to a CEL, called $\mathcal{L}$-expression:

Definition 2 *Given a CEL* $\mathcal{L} = (T_\mathcal{L}, R_\mathcal{L})$, *an* $\mathcal{L}$-expression *is a concept expression created from a concept expression type* $t \in T_\mathcal{L}$ *by substituting all of its placeholder variables with expressions in conformance to all of the restrictions* $r_i \in R_\mathcal{L}$.

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* $\mathcal{L} = (T_\mathcal{L}, R_\mathcal{L})$ *be a CEL. Let us assume an OWL ontology O with signature* $sig(O) = \mathcal{C}_O \cup \mathcal{P}_O \cup \mathcal{I}_O$ *(i.e., consisting of the set of all classes, properties and individuals of the ontology), and a named class* $FC \in \mathcal{C}_O$. *An* $\mathcal{L}$-category[4] *for FC in O is then any concept expression in the form* $FC \sqcap D$, *where D is an* $\mathcal{L}$*-expression created by substituting entities from* $sig(O)$ *for placeholder variables in a CE type* $t \in T_\mathcal{L}$. *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 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 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

[2] Formally 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.

[3] I.e., it can be either an atomic property or a compound role expression such as that built using the inverse property constructor.

[4] For 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.

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:

$$FCP(O, FC, \mathcal{L}) = \sum_{i=1}^{k} w(Cat_i, FC) \quad (1)$$

where $\{Cat_1, ..., Cat_k\}$ is the set of $\mathcal{L}$-categories for FC in O (under inferential closure, i.e. equivalent Cat_i are only counted once) and w is a *weight function* assigning every Cat_i a weight from $[0, 1]$, expressing its 'quality' as category of FC.

2.3. Weight function desiderata

Intuitively, an ideal weight function $w(Cat_i, FC)$ should take into account the following four characteristics of an $\mathcal{L}$-category Cat:

Absolute size 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.

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 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. 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.[5]

For an unrestricted version of the OWL concept expression grammar (say, $\mathcal{L}$ corresponding to the whole of OWL 2), the size of $\{Cat_1, ..., Cat_k\}$ would typically be very large – actually, infinite (esp. due to the possibility of recursively composing property restrictions), for any realistic O. 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 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 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, in a well-designed ontology, $w(FC \sqcap D, FC)$ 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 such D that $D \sqsubset FC$, since these are the concepts viewed as proper subcategories by the ontology designer. However, w could also be accidentally (especially, in imperfectly normalized ontologies, which abound in practice) nonzero 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).

[5] Presumably, 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.

2.4. 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 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 as 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 *syntactic 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 *concept expression type* of D, $t_i \in T_{\mathcal{L}}$, produces a meaningful category. This may relate
   - to its complexity, and,
   - to the overall distribution function of absolute/relative extension size of the CEs conforming to it, i.e., how many CEs of this type have a 'reasonable' extension, on average.

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`[6] and a property `hadAccident`. In a CEL supporting 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.`$\top$ will be derived. The DCE of this category would be an instantiation of the general CE type $\exists P.\top$. Through the lenses of the different sources mentioned:

1. 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 the weight).
2. Assuming the property `hadAccident` has `Car` in its domain (`rdfs:domain`), a meaningful *syntactic pattern* (see Section 4) is matched in the ontology TBox. Thus, again, the source thus suggests the category as meaningful.
3. 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 SPARQL[7] query to this dataset,[8] is instatiated 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. As we will reveal in Section 7, the CE type $\exists P.\top$ is a priori not as likely to produce good categories as are some other CE types. 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 *type* of this CE.

Several parts of the remainder of the paper investigates the mentioned sources of category weight. The whole of Section 4 is devoted to *syntactic patterns*, also indirectly addressing the *inferential relationships* (reflected in the patterns' matching constraints), i.e. the sources 1 and 2. The *type-level* weight assignment (source 4) is discussed immediately below (Sections 2.5 and 2.6), and experimentally derived, via aggregation over the individual categories, through human users engaged in the cognitive experiments (Section 7). Eventually, there are also empirical experiments underway focusing on *Abox data* as source (no. 3) of category weight; the results are however bound to future publications due to the obstacles related to the sparsity of public Abox data with respect to most existing ontologies, and also to avoid an excessive length of the current paper.

[6]For 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.

[7]`https://www.w3.org/TR/sparql11-overview/`

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

2.5. Category-specific vs. type-inferred weight

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

$$w(FC \sqcap D, FC) = w^I(FC \sqcap D, FC) \otimes w^T(type(D)) \quad (2)$$

where

- w^I is an individual weight function for the given category (presumably based on the first three sources from the list (Section 2.4): inferential relationships, Tbox patterns, and possibly Abox occurrence),
- $type(D)$ is the type of the (differentiating) concept expression,
- w^T is the CE type weight function (based on the fourth source from the list above: likelihood of the type producing a meaningful category), and,
- $\otimes$ is a function for combining the two partial weights into a single weight of the category.[9]

As regards the CE *type* weight w^T, we are aware of three empirical sources, in turn, which might provide a *quality estimate* of a certain CE type (and thus, indirectly, the individual categories in which the DCE would match this CE type):

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

Data collected from these sources then have to be *aggregated* across the different expressions of the same CE type; 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 *syntactic patterns* (heuristically mapped to CE types) 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 the 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 2.4, note that the sources from Section 2.4 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 type.

In different sections of the paper we will empirically investigate these sources: the CEs in axioms in Section 5, the syntactic patterns in Section 6, and the human assessment of FC+subcategory pairs in Section 7 (thus only deferring the Abox analysis to later research, as we already noted in Section 2.4), and discuss their potential and limitations in more detail. The goals are both to identify interesting *insights* on these sources in general and to suggest tentative *CE type weights* for the purpose of the FCP computation.

2.6. 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) 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^I_{01}, we can then

[9] As we exemplified (on 'cars with an accident') in Sect. 2.4, 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 category type (within $w^T(type(D))$). 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.

express the overall weight of a category as a simple product:

$$w(FC \sqcap D, FC) = w^I_{01}(FC \sqcap D, FC) \cdot w^T(type(D)) \quad (3)$$

which will yield

- 0 if $w^I_{01} = 0$: this happens when the reasoner arrives at some specific conclusion (see above) about the extension of the category, and thus judges the category as useless, and
- $w^T(type(D))$ if $w^I_{01} = 1$: this happens when the reasoner does not arrive at such a kind of conclusion.

Let us then assume that we distinguish between k CE types allowed in the CEL $\mathcal{L}$, $T_{\mathcal{L}} = \{t_1, \ldots, t_k\}$, each associated with a weight via the type weight function w^T. Let each type t_i be also associated with a set of patterns $PS_i = \{p_{i1}, \ldots, p_{ij_i}\}$ that allow to construct the corresponding categories from the axioms of ontologies. Let us further assume that each pattern comprises conditions (constraints) allowing to evaluate w^I for its particular instantiation – the instantiations that would yield useless categories (e.g., equivalent to the FC) are then not returned as matches of 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) pattern sets $PS = \{PS_1, ..., PS_k\}$ as the weighted sum of the pattern occurrence counts in O:

$$\widehat{FCP}(O, FC, \mathcal{L}, PS) = Occ(PS_1, FC, O) \cdot w^T(t_1) + \ldots + Occ(PS_k, FC, O) \cdot w^T(t_k) \quad (4)$$

where $Occ(PS_i, FC, O)$ is the sum of pattern occurrence counts in O for all patterns $p_{ij} \in PS_i$.

We will define the notion of such category patterns, and bring usable example of patterns, in Section 4.

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

For meaningful analysis of the FCP computation landscape we have to combine $\mathcal{L}_{nam}$ (the trivial language of named concepts) with a concept expression language (CEL) of differentiating concept expressions (DCEs) involving some DL concept *constructor*. Yet, for the purposes of our initial study we sought a 'small' CEL well tractable, both by humans and machines, in the empirical investigations. As we show later in our survey of CE types in ontology axioms (Section 5), *existential restriction* is the most frequent[10] type of axiom right-hand side (RHS) in the Tbox of linked data ontologies. Compared to other common types of concept constructors, a *simple existential restriction*, $\exists P.C$, (where the filler C is an atomic class) has important advantages in this respect:

- The number of (syntactically enumerated) CEs only grows linearly with the number of classes $|\mathcal{C}_O|$ and number of properties $|\mathcal{P}_O|$ in the ontology (at most it can reach $|\mathcal{P}_O|.|\mathcal{C}_O|$). It has been observed that ontologies are often huge either in terms of classes or in terms of properties but rarely in terms of *both*. In contrast, e.g., for the *Boolean connectives* ($\sqcup$, $\sqcap$) the number of CEs grows quadratically (for the simplest expressions) with $|C|_O$, which makes the space large for some ontologies.
- 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 somewhat favorizes the existential restriction to the *universal restriction*, whose validation has to rely on the closed-world assumption (CWA). *Cardinality restrictions* except for the 'minimum' ones also require the CWA, and their checking may rely on more than one property assertion.

More detailed analysis of OWL concept constructors usability for focused categorization is in Section 5.

When considering $\exists P.C$, it may be practical to also consider some of its special cases or extensions that do not significantly worsen the tractability:

- One of them is $\exists P.\top$, i.e. the restriction to 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 such CEs is $|\mathcal{P}_O|$. (In Section 2.1 we already introduced the CEL $\mathcal{L}_{extop}$, which only allows this CE type.)

[10] Unless we consider the conjunction operator $\sqcap$, which however merely joins two CEs that can be considered independently in the FCP context, see more discussion in Section 5.

– The other is $\exists P.\{i\}$, i.e. the *value restriction*. The upper bound of the number of such CEs is $|\mathcal{P}_O|.|\mathcal{I}_O|$, but $|\mathcal{I}_O|$ (the number of instances in the ontology) is rarely high, since ontologies normally contain few instances;[11] in addition, such instances have their 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).

Based on this 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 4 *The* simple existential CEL $\mathcal{L}_{SE}$ *is a CEL with the set of types of concept expressions* $T_{\mathcal{L}_{SE}} = \{\exists P.C\}$ *and the restriction set* $R_{\mathcal{L}_{SE}}$ *consisting of the following ones:*

– *Variable C in $\exists P.C$ can only be substituted with $\top$, with a named concept, or with $\{i\}$ (a singleton enumeration concept)*
– *Variable P in $\exists P.C$ can only be substituted with a named property.*

Since $\mathcal{L}_{SE}$ does not allow for concept nesting, it guarantees *finiteness* of $\{Cat_i\}$ 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 value restriction (namely, the enumeration class in the filler) 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.

From the extremely restricted nature of $\mathcal{L}_{SE}$ we can trivially derive its allowed CE types. Along, we will also consider the sole CE type 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. Therefore, the CE types considered for the sake of the presented research are: a named class; a property restriction with $\top$ as filler; a property restriction with a named class as filler; and, a property restriction with a singleton class as filler (which is often called a 'value restriction'). In Table 1 we have an overview of the types:

Table 1
Summary of CE types in $\mathcal{L}_{nam} + \mathcal{L}_{SE}$

	Structure	Sign.	Abox	Pattern
t1	C	C	1	1
t2	$\exists P.\top$	P	1	1
t3	$\exists P.C$	P, C	2	3
t4	$\exists P.\{i\}$	P, i	1	4

– The first column assigns the types a numbering local to the given CEL setting, for convenience.[12] Obviously, **t1** belongs to $\mathcal{L}_{nam}$, while **t2**, **t3** and **t4** to $\mathcal{L}_{SE}$.
– The second column indicates the structure of the CE type itself in DL notation; the CE types substitutions are already restricted according to $R_{\mathcal{L}_{nam}}$ and $R_{\mathcal{L}_{SE}}$: 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 **t2** is a subset of those of **t3** and **t4**; therefore, if we consider one or more CEs for **t3** or **t4** with some property P then we should also consider the CE for **t2** 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 pattern will rely on the `rdf:type` predicate (**t1**), on the specific predicate P (**t2**, **t4**), or on both (**t3**).
– The order of the CE types in the table reflects the increased complexity of their detection in the Tbox using so-called focused category patterns (fifth column), which we detail in the next section.

[11] In long term we may be also interested in ontologies coupled with (especially, SKOS) codelists, for which the size of $|i|_O$ 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 pattern for such a setting.

[12] We denote these specific CE types using boldface, to differentiate them from the abstract symbols using the math font with subscript notation (t_i). Similarly for patterns (**p1**, **p2**, ... vs. p_i).

4. Focused category patterns for $\mathcal{L}_{nam} + \mathcal{L}_{SE}$

Having defined $\mathcal{L}_{SE}$ (as well as $\mathcal{L}_{nam}$) as a particular CEL, and identified the types of CEs allowed by this CEL, we can now proceed to formulating syntactic patterns through which such CEs can be constructed from the ontologies.

We will first introduce the general notion of *focused category pattern* – a syntactic pattern for mapping OWL axioms (in ontologies – primarily, in their Tbox) to concept expressions that are to be used as the DCE of categories of a focus class. Then we will present patterns specifically designed for the CE types allowed by the CELs $\mathcal{L}_{SE}$ and $\mathcal{L}_{nam}$. We will also briefly review the algorithmic approach for pattern-based CE generation, and outline the usage of patterns in the (optional) named category generation step.

4.1. Focused category pattern

Intuitively, the definition of a focused category pattern relies on OWL axiom templates that can be matched with an OWL ontology.

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

- *t is a* CE type.
- *Vars is a sequence of* mapping variables.
- *Tpl is a set of conjunctively*[13] *interpreted* axiom templates*, namely, OWL axioms (possibly) containing placeholder 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 placeholder 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 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 of type t, provided all the constraints from Cons are satisfied by the tuple.*

[13]For simplicity we only consider conjunctions of axioms. Disjunctions could be expressed by means of multiple FC patterns for the same CE type, if needed.

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

4.2. FC Pattern collection for $\mathcal{L}_{SE} + \mathcal{L}_{nam}$

We will subsequently present five patterns for $\mathcal{L}_{SE} + \mathcal{L}_{nam}$ and illustrate them on our running example of used cars.

Pattern **p1** Let us first consider the FC pattern for creating CEs of type **t1**, i.e. conforming to $\mathcal{L}_{nam}$; the axiom templates are serialized as triples in the Turtle[14] RDF notation:

$$\mathbf{p1} = (\mathbf{t1}, (C), \{C\ \texttt{rdfs:subClassOf}\ FC\ .\}, \emptyset, \mathbf{m1}) \quad (5)$$

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

`Truck rdfs:subClassOf Car .`

m1 will simply return the CE `Truck`, and the created category will thus be `Car` $\sqcap$ `Truck`. Note however, that in general the subclass relationship (to satisfy the pattern) 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 pattern might not be unique for the given CE type. If we accept the possibility that not all classes in different hierarchical paths are pairwise disjoint, we could also apply to **t1** a pattern for possibly unrelated C and FC, with the axiom template set being, e.g.,

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

and with the constraints in *Cons* checking that neither $C \sqsubseteq FC$ nor $FC \sqsubseteq C$ holds. Since the weight of the

[14]`https://www.w3.org/TR/turtle/`

expressions built using such a pattern should be rather tiny on average, we omitted it in our study, for simplicity.

As we move from $\mathcal{L}_{nam}$ to $\mathcal{L}_{SE}$, i.e. to the realms of compound CEs, the 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).

Pattern **p2** The pattern, for type **t2**, is still rather concise. It assumes that if the focus class *FC* is in 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.

$$\mathbf{p2} = (\mathbf{t2}, (P), \{\ \texttt{P rdfs:domain FC .}\}, \mathbf{Cons2}, \mathbf{m2}) \quad (7)$$

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 restriction[15] $FC \sqsubseteq \exists P.C$; 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 patterns introduced so far together, in the calculation of the approximate FCP, see Formula 3, 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.

Example 3 The *schema.org* ontology contains a class `Vehicle`[16] having four subclasses: `Car`, `BusOrCoach`, `Motorcycle` and `MotorizedBicycle`. *Schema.org* also proposes 82 properties having `Vehicle` in their domain.

The instantiated $\widehat{FCP}$ formula (cf. Equation 4) for $T_{\mathcal{L}} = \{\mathbf{t1}, \mathbf{t2}\}$ (the respective $\mathcal{L}$ thus only covering a fraction of the expressiveness of $\mathcal{L}_{SE} + \mathcal{L}_{nam}$), $PS = \{\{\mathbf{p1}\}, \{\mathbf{p2}\}\}$, *FC*=`Vehicle`, *O*=*schema.org*, then becomes

$$\begin{aligned}
&\widehat{FCP}(FC, O, \mathcal{L}) = \\
&Occ(\{\mathbf{p1}\}, FC, O) \cdot w^T(\mathbf{t1}) + \\
&+ Occ(\{\mathbf{p2}\}, FC, O) \cdot w^T(\mathbf{t2}) = \\
&\qquad 4 \cdot w^T(\mathbf{t1}) + 82 \cdot w^T(\mathbf{t2}) \quad (8)
\end{aligned}$$

Intuitively, while $w^T(\mathbf{t1})$ should be close to the theoretical maximum (named subclasses are by default meaningful subcategories), $w^T(\mathbf{t2})$ 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 us now progress to the patterns for the remaining two CE types of $\mathcal{L}_{SE}$.

[15] The restriction can also be inherited from a superclass or part of a complete definition, or can have the form of a `value` or `self` restriction or of a cardinality restriction that entails the existential one; analogously for similar constraints for other patterns.

[16] Here and in the rest of the running example we omit the `http://schema.org/` and other prefixes, for brevity.

Pattern **p3** A CE of type **t3** 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 C. The pattern for **t3** 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 **t2** 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.

$$\mathbf{p3} = (\mathbf{t3}, (P, C), \{\ P\ \texttt{rdfs:domain}\ FC\ ;\ \texttt{rdfs:range}_a\ D\ .\ C\ \texttt{rdfs:subClassOf}\ D\ .\}, \mathbf{Cons3}, \mathbf{m3}) \quad (9)$$

The inferential closure is again used, however, with the exception of the range axiom, which is only considered as asserted (therefore the 'a' index in $\texttt{rdfs:range}_a$) – otherwise not only subclasses of D but also classes having a common superclass with D would be returned as C (since the superclass would become an inferred range of P).

The constraint in **Cons3** states that P must not appear in an existential restriction $FC \sqsubseteq \exists P.E$ such that $E \sqsubseteq C$; for such properties the CE would contain *all* instances of FC.

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

Pattern **p4** The CE of type **t4** 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 pattern for **t4** 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):

$$\mathbf{p4} = (\mathbf{t4}, (P, i), \{\ P\ \texttt{rdfs:domain}\ FC\ ;\ \texttt{rdfs:range}_a\ D\ .\ C\ \texttt{rdfs:subClassOf}\ D\ .\ i\ \texttt{rdf:type}\ C\ .\ \}, \mathbf{Cons4}, \mathbf{m4}) \quad (10)$$

The inferential closure is used as before.

The constraint in **Cons4** states that there must not be an existential restriction $FC \sqsubseteq \exists P.E$ such that $i \in E$; 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 type **t4** we include one more pattern, which goes beyond the RDFS expressiveness and relies on the connection of the ontology to an external resource – a *SKOS*[17] *codelist*. This variant thus extends the previous one for a specific source of instance i, which is a code list (also called concept scheme in SKOS):

$$\mathbf{p5} = (\mathbf{t4}, (P, i), \{\ P\ \texttt{rdfs:domain}\ FC\ ;\ \texttt{rdfs:range}_a\ \texttt{skos:Concept}\ ,\ [\ \texttt{a owl:Restriction}\ ,\ \texttt{owl:hasValue}\ s\ ,\ \texttt{owl:onProperty}\ \texttt{skos:inScheme}\ ]\ .\ i\ \texttt{a skos:Concept}\ ;\ \texttt{skos:inScheme}\ s\ .\ \} \mathbf{Cons4}, \mathbf{m4}) \quad (11)$$

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 $p4$.

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.[18] 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 patterns assure (**p2–p5** via the domain axiom) that the mapped CE is a *specialization* of the

[17] `http://www.w3.org/2008/05/skos`

[18] Such 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.

FC.[19] 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 different 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 patterns in terms of (templated) RDF triples in the template axioms, it is as listed, for the respective CE types, in the fifth column of Table 1. **p4** is the most complex one of the 'straightforward' patterns, with four triples; however, **p5** (for the same type, **t4**) is even more complex, having eight triples overall, due to the overheads of representing OWL restrictions in RDF.

4.3. Pattern-based CE generation algorithm

The provisionally implemented algorithm for matching the presented FC patterns in the Tbox and generating concept expressions from their occurrences in ontologies (used both for the pattern occurrence survey described in Section 6 and for the generation of human assessment tasks described in Section 7) performs straightforward generation of categories that are immediately pruned with respect to a 'CE blacklist' previously constructed by an auxiliary (pre-processing) algorithm. The 'blacklist', built once for every FC, contains the CEs inferred to subsume (or be equivalent to) this FC, i.e. violating the constraints **Cons**# from the patterns listed in this section of the paper. The implementation of the 'blacklist' construction relies on a structural reasoner that is part of the OWL API.[20]

We do not list the algorithms to save space; they essentially amount to a literal translation of the concrete patterns and constraints described in this section to program code.

4.4. FC pattern usage for named category generation

We will, eventually, conclude the *schema.org* example narrative by demonstrating an additional use case of the patterns.

[19] In contrast, this would for example not be the case for the alternative pattern for **t1** (cf. Equation 6) allowing hierarchically unrelated C and FC.

[20] `https://github.com/owlcs/owlapi/`

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 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 types in ontology axioms

As the first empirical source for estimating the frequency of occurrence of the CE types 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,[21] containing real-world (but mostly light-weight) ontologies typically used for linked data publishing.

[21] `https://lov.linkeddata.es/dataset/lov/`

2. The BioPortal collection,[22] 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*.[23]

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*[24] *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, too, since 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 *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 concept expression types 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 types in a CE type frequency table.

The considered list of CE types is, essentially, that of first-level constructors in the axiom RHS. However, since we were particularly interested in the CE types from $\mathcal{L}_{SE}$, we singled them out under the labels 'ExistThing' (**t2**), 'ExistNamed' (**t3**) and 'ExistIndiv' (**t4**). 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.[25]

The frequence of CE type 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.

As regards the $\mathcal{L}_{SE}$ types, we can see that 'ExistNamed' (**t3**) appears very often at top positions, throughout the collections and axiom types; it dominates in subclass axioms, most notably in BioPortal. 'ExistIndiv' (**t4**, i.e. the value restriction) is less frequent but still plays some role in both larger collections. 'ExistThing' (**t2**) appears even less often.

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

[22] `https://bioportal.bioontology.org/`

[23] `https://owl.vse.cz/ontofarm/`; the collection is used, among other, in the Ontology Alignment Evaluation Initiative, `http://oaei.ontologymatching.org/`.

[24] We 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.

[25] Strictly speaking, the singleton enumeration $\{i\}$ in **t4** is also an anomymous CE, and the notion of '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 types in axiom RHS

LOV			
Equivalence		Subclass	
Construct	#	Construct	#
Conjunction	1134	*ExistNamed*	3187
ExistNamed	126	Universal	1331
Disjunction	108	Conjunction	553
Enumeration	54	Exactly	545
ExistAnon	31	Minimally	483
ExistIndiv	30	ExistAnon	467
Exactly	12	Maximally	237
Negation	4	*ExistIndiv*	77
Minimally	3	Disjunction	56
Maximally	1	Negation	23
BioPortal			
Equivalence		Subclass	
Construct	#	Construct	#
Conjunction	122725	*ExistNamed*	1977805
ExistAnon	24239	ExistAnon	64046
ExistNamed	2338	Conjunction	18409
Enumeration	854	Universal	6587
Disjunction	658	*ExistIndiv*	4698
Universal	22	Exactly	1179
ExistThing	11	Disjunction	856
Minimally	10	Minimally	579
Exactly	7	Maximally	577
Negation	6	Negation	96
OntoFarm			
Equivalence		Subclass	
Construct	#	Construct	#
ExistNamed	40	*ExistNamed*	317
Conjunction	27	Universal	47
Disjunction	15	Exactly	23
Universal	2	Minimally	22
Enumeration	1	ExistAnon	15

– *Conjunction* ($\sqcap$). From the point of view of focused categorization, logical conjunctions are actually not very interesting, since the conjunction can be simply achieved by applying multiple categories on the categorized individual. For example, for the `Car` focus class, the DCE `hasSeller.Corporate` $\sqcap$ `usage.TaxiUsage` can be distributed to two categories: one of them is `Car` $\sqcap$ `hasSeller.Corporation` and the other `Car` $\sqcap$ `usage.TaxiUsage`.

– *Disjunction* ($\sqcup$). When categorizing a single individual, the situation is analogous to that of the conjunction: it suffices to pick the particular disjunct (or multiple disjuncts) that evaluates true for that individual. Disjunctions may be potentially relevant for categorizing a set of individuals: for example, `Car` $\sqcap$ (`hasSeller.Corporation` $\sqcup$ `hasSeller.Government`) would contain cars of both kinds. As mentioned in Section 3, the number of disjunctive combinations may however be very high in some cases, and the question of which combinations should be allowed is not easy to decide at the axiom pattern level. The utility of disjunctively defined categories in practical contexts of RDF data publishing (e.g., within e-shop catalogs) deserves a further study.

– *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 **t3**), 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.

– *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 $\forall$`hadAccident.FatalAccident` would include all cars not having any accident recorded in their history at all. To avoid this effect, we should extend it to $\exists$`hadAccident`.$\top \sqcap$ $\forall$`hadAccident.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* ($\{\ldots\}$). 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 type to be negated, and thus would not have other impact on the FCP than would have the doubling of the weight of those CE types.

In all, the analysis suggested that the $\mathcal{L}_{SE}$ types play a significant role in the family of all anonymous expressions commonly used in OWL, and that the design of an FCP formula restricted to this simple CEL is thus meaningful. It however also indicated directions for extending the used language in the future.

6. FC 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 FC patterns overall and across different ontology collections.

The answers to these questions would help us estimate how likely it is that the particular FC 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*[26] (a tool allowing to quickly compute various metrics on ontologies) 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 FC 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 the patterns **p1**,..,**p4**;[27] see Table 3. The tables show the percentage of ontologies for which more than an *n* 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 $\tau = 5$, $n > 0.1$, 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.

[26] `https://owl.vse.cz/OOSP/`

[27] Pattern **p5** was only matched rarely; it is thus analyzed separately.

Table 3
Ratio of ontologies with an $> n$ 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 n, 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 n 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 $n > 0.2, \tau = 5$), 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 n 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 FC 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://contsem.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 2016[28] and the second, with a slightly modified and extended setup, in Spring 2018.

The overall scheme of the experiments was as follows:

1. Provide the human assessors with a set of 'focus class – subcategory' pairs, such that the subcategories correspond to *different CE types*. Each pair constituted an elementary *assessment task*.
2. 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*.
3. *Aggregate* the collected data across the assessors.
4. Seek *correlations* between the category quality and the CE type, 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.
5. Also examine *qualitative feedback* also collected through the questionnaire.

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

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 types	**t2**, **t3**, **t4**	**t1**, **t2**, **t3**, **t4**
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

7.1. Summer 2016 campaign

Initial sampling As regards the FC+subcategory (from now on, 'task') sampling for both threads of 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 of type **t2**, **t3** and **t4**, 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 **t2**, 20 for **t3** and 22 for **t4**.

Expert ontologist assessment and insights. In the first stage, the assessment was made by three researchers with 10–20 years of experience in ontological engineering (the first three authors of this paper). 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

[28] Already described in our early publication [16]. Here we provide a synoptic view of both campaigns.

sought in a F2F session. The independent assessment had 76% agreement: in 45 out of 59 cases there was no contradictory assessment (certainly/perhaps yes vs. certainly/perhaps 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 **t2** and one of **t3**. Selected general insights into less obvious decisions, with examples, follow:[29]

- 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 **t3** 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 **t4** are plausible but less useful due to their *inherently limited extent*: for instance, categorizing instances of `geopolitical:area` as `geopolitical:isSuccessorOf value` X, where X is another (geopolitical) area.

Novice ontologist assessment. There were two groups of students involved: Bc-level students in a course on Artificial Intelligence (AI) and MSc-level students in a specialized course on Ontological Engineering (OE). Both courses provided a certain degree of OWL modeling experience (in Protégé and Manchester syntax) prior to this exercise, although OE went into more depth as regards the underlying DL and reasoning. There were 17 AI students and 10 OE students altogether. In both courses the students were first provided with a 30-minute overview of the notions of CE (in $\mathcal{L}_{nam}$ and $\mathcal{L}_{SE}$) 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.[30] 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 **t4** 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 type. 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'

[29] In this section we write the example CEs in Manchester syntax, with the keyword `some` indicating an existential restriction in **t2** and **t3** CEs and the keyword `value` indicating a value restriction in **t4** CEs.

[30] The 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

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

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 **t4** (0.21, with 15 positives, 1 zero and 7 negatives), lower for **t3** (0.02, with with 9 positives, 3 zero and 8 negatives), and lowest for **t2** (-0.05, with with 4 positives, 1 zero and 12 negatives).
- The cases[31] with highest positive and lowest negative values are in Table 6; the CE type 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 **t4** dominates the upper end of the spectrum. Interestingly, the negative cases correspond each to a different type and also have different semantic roots: the 'village with baseball player as head-chef' one has 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 'ontologistic thinking' of the latter. 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 **t2** CEs (i.e., on the lower reliability of the 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 types, 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).

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

Table 6

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

FC	Subcategory	Type	Avg.NS
ofrd:FridgeFreezer	ofrd:styleOfUnit value ofrd:SingleDoor	**t4**	0.91
gr:BusinessEntity	pco:mainActivity value p-act:GeneralServices	**t4**	0.86
gr:BusinessEntity	pco:authorityKind value p-aut:LocalAuthority	**t4**	0.61
akt:Generalized-Transfer	akt:information-transfer-medium-used value akt:Email-Medium	**t4**	0.59
p1:Storage	p1:hasStorageMedium value p1-sm:mag	**t4**	0.59
fabio:Item	fabio:isStoredOn value fabio:web	**t4**	0.50
...	...	...	...
dbo:Village	dbo:headChef some dbo:BaseballPlayer	**t3**	-0.50
sigkdd:Conference	sigkdd:City_of_conference some Thing	**t2**	-0.56
gr:DayOfWeek	gr:hasNext value gr:Friday	**t4**	-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 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 of type **t1**, 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 type (**t1**, **t2**, **t3**, **t4**). 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 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 of type **t2** 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 types was analogous, e.g., for **t3**, "... 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 $\mathcal{L}_{nam}$ and $\mathcal{L}_{SE}$) 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 types), all available in a single sheet of a web questionnaire.[32] 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 entities 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 types did not change from the 2016 campaign. The average NS was 0.33 for **t4** (all ten NS values being positive), 0.30 for **t3** (8 positives, 1 zero and 1 negative) and 0.16 for **t2** (7 positives, 1 zero and 2 negatives). The NS for **t1** 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 type is listed in the fourth column, and the average NS value in the fifth column. Interestingly, the students were positive to-

[32] The 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

$$NS_i = \frac{\sum_{j=1}^{k} c_{i,j} \; . \; v_{i,j}}{k} \tag{13}$$

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 type became as follows: 0.35 for **t1**, 0.32 for **t4**, 0.25 for **t3** and 0.14 for **t2**. Compared to the non-weighted normalized sum, **t4** marked the smallest decrease of all CE types.

Summary of the experiment Across the different campaigns and settings, the order of the CE types according to the average 'meaningfulness' of the categories remains stable: (**t1** >) **t4** > **t3** > **t2**. 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).

8. Discussion

Since the empirical part of the paper may appear fragmented and the results hard to align, in this section we provide an integrative meta-view of the surveys / experiments settings and results. We also outline an approach to operationalizing these empirical results to directly applicable FCP calculation.

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*; [33] 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 types or 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 to 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 types / patterns might look incoherent. Especially, **t2** is (by Sec. 5) very rarely used in the axioms RHS, which contrasts with the fact that we can generate a huge number of CEs of this type (by Sec. 6) through pattern **p2** and the proportion of these CEs that are 'meaningful' is not dramatically lower than for the other CE types (by Sec. 7). This paradox can however be explained as follows:

- 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

[33] Partly also in the source ontologies used, this is however only caused by effectiveness concerns.

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

FC	Subcategory	Onto (# cl.)	Type	Avg.NS
Security mechanism	BlockCipher	stac (211)	**t1**	0.79
Object	Machine	pext (488)	**t1**	0.75
Sensor	Gyro sensor	ha (88)	**t1**	0.75
Social position	has title: Prime minister	pext (488)	**t4**	0.75
Quantity	Mass	schema (625)	**t1**	0.67
Document	cites: Protocol	vivo (146)	**t3**	0.67
Thing	has dimension: Thing	ecrm (84)	**t2**	0.64
...	...	...	...	...
Service	has forecast: Thing	km4c (582)	**t2**	-0.21
CRM entity	is separated from: Beginning of existence	ecrm (84)	**t3**	-0.29
Feature	Camper service	km4c (582)	**t1**	-0.33
Subcontinent	in country: Thing	swpo (57)	**t2**	-0.50

t3) and not just `owl:Thing` (as in **t2**), 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 **t2**), might work.

As regards the high ranking of **t4** 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 types) and not the frequency of occurrence of CE types at all (they were sampled evenly) wrt. type. The findings are thus completely orthogonal: there will typically be few **t4** 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 **t3** and **t4** CEs may proportionally increase.

8.2. *Tentative operationalization of the 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 types 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 **t4** (value restriction) and to some degree **t3** (existential restriction) may successfully complement **t1** (named class) when harvesting categories from a reused ontology. The survey of the CEs in axioms, in turn, suggests that the chosen CE language ($\mathcal{L}_{nam}+\mathcal{L}_{SE}$) covers a decent part of the 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 type.

Considering that the lowered average NS values of **t1** 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 $w^T(\mathbf{t_1})$ to 1. The remaining weights can be set, heuristically reflecting both calculation methods and the results of both campaigns in combination, for example as $w^T(\mathbf{t_2}) = 0.5$, $w^T(\mathbf{t_3}) = 0.7$, and $w^T(\mathbf{t_4}) = 0.8$. This would yield the following FCP estimate formula, to be applied on an ontology O and focus class FC, the language $\mathcal{L}$ being $\mathcal{L}_{nam} + \mathcal{L}_{SE}$:

$$\widehat{FCP}(O, FC, \mathcal{L}, \{\{\mathbf{p1}\}, \{\mathbf{p2}\}, \{\mathbf{p3}\}, \{\mathbf{p4}\}\}) =$$

Table 8
Summary of the complementary surveys / experiments

Survey / experiment	CE type frequency in axioms (Sec. 5)	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 patterns)	Directly (by humans), but generated using patterns
Objectiveness	Objective	Objective	Subjective (human judgment)
Types/patterns	**t2**, **t3**, **t4**	**p1**, **p2**, **p3**, **p4**, **p5**	**t1**, **t2**, **t3**, **t4**
Focus classes	(Not relevant)	All classes	59+40 randomly selected classes
Gist of the quantitative findings	CE frequency: **t3** most frequent (by an order of magnitude)	Pattern occurrence: **p2 > p1 > p3 > p4 > p5** with **p1** and **p3** being swapped for less general FCs	Proportion of CEs meaningful for human assessors: **t1 > t4 > t3 > t2** (but not differing dramatically)

Table 9
Alternative CE type weights derived from the cognitive experiments

	t1	**t2**	**t3**	**t4**
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

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

The formula should of course only be taken as an illustration of the possible operationalization of the research undertaken, and not as a granted recipe.

9. 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; ontology reuse metrics and methods.

9.1. Abstract notion of categorization (power)

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 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 [6], 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) [12] 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: $\texttt{SteroidHormone} \equiv \texttt{Steroid} \sqcap \exists\, \texttt{playsRole.Hormone}$.

Our own ongoing work on the PURO modeling language [17] deals with various options how the same 'background' state of affairs can be expressed in OWL. PURO structurally resembles OWL but relaxes some of its modeling constraints. A library of transformation patterns allows to proceed from one PURO model to alternative OWL ontologies in different encoding styles. An example relevant to our case is the notion of *enterpreneur*, which is likely to be expressed as type in PURO, but could be translated to relationship (`insuranceCategory`) restricted to the `Entrepreneur` individual in OWL (i.e. a compound 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.[34] 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 [9].

Certain research in cognitive psychology might also be relevant wrt. the notion of FCP. In particular, the notion of *graded structure* of categories [4] can be applied to our concept type 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)" [4]. 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" [4], analogous to our CEs defined via an existential restriction.

9.2. 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. [18] (aiming to build a benchmark for testing ontology tools). A large scale study of OWL ontology metrics was carried out by Matentzoglu et al. [11]. 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 [10]. 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 (subclass axioms with) named class in the RHS, and also property axioms such as domain/range or functional property. It is even over the common subset of CE types in restrictions, of MontoloStats and other research (such as disjoint, universal and cardinality restriction CEs) that some differences in the computed ranking appear; these may be due to additional distinct features in the methodology used.

9.3. 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

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

of OWL expressions. A recent example of the former is a study on classifying domain entities to upper-level ontology classes [15]. An example of the latter is an earlier study on the human capability of deriving useful information from differently verbalized OWL statements [19]. 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.

9.4. *Ontology reuse metrics and methods*

The broad context of our research, the task of ontology reuse, was studied by Schaible et al. [13]: 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 [14]. As regards the designer credibility, this is a feature of the ontology itself similarly to FCP, but it is completely orthogonal.

Reuse support [5] is also systematically sought by the maintainers of LOV [20], primarily at keyword relevance level; we are in contact with them and will seek to integrate our complementary approaches.

10. 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/es*, 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 particularly 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.

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 type of compound CE, and provide an empirical grounding for the 'meaningfulness' of the CE type, 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 FC-patterns such as **t5**.

Another area of ongoing research concerns the techniques of *pattern-based lexical transformation* of compound CEs to named ones. 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 *concept expression types* identified as common in OWL axioms, by our study, such as universal restrictions, disjunctions (relevant for group categorization) and some nested expressions. 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 types, the syntactic *FC-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 [7] or PURO models [17], 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 at to provide targeted feedback to so far unclear constellations of focus class and subcategory constellations.

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. From the engineering perspective, an ontology reuse GUI could also be powered, beyond the display of the plain FCP value, by a sampling engine providing a *representative overview of the kinds of CEs* (incl. compound ones) available through the considered ontology; this will make the reuse decision even more informed. 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 types 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.

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) [3].

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 belive 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 has been supported by CSF 18-23964S, "Focused categorization power of web ontologies".

OZ has also been supported by the long term institutional support of research activities by Faculty of Informatics and Statistics, University of Economics, Prague.

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] 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.

[4] Liane Gabora, Eleanor Rosch, and Diederik Aerts. Toward an ecological theory of concepts. *Ecol. Psychology*, 20, 2008.

[5] 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. .

[6] 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.

[7] 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`.

[8] 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.

[9] 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. Tamma, 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] Ioannis Stavrakantonakis, Anna Fensel, and Dieter Fensel. Linked open vocabulary ranking and terms discovery. In Anna Fensel, Amrapali Zaveri, Sebastian Hellmann, and Tassilo Pellegrini, editors, *SEMANTICS*, pages 1–8. ACM, 2016. ISBN 978-1-4503-4752-5.

[15] Robert Stevens, Phillip Lord, James Malone, and Nicolas Matentzoglu. Measuring expert performance at manually classifying domain entities under upper ontology classes. *J. Web Semant.*, 57, 2019.

[16] 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`.

[17] 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.

[18] 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.

[19] Allan Third. "Hidden semantics": what can we learn from the names in an ontology? In *INLG*, pages 67–75, 2012.

[20] 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 hardcopy 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 meaninful 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)