Type or Individual? Evidence of Large-Scale Conceptual Disarray in Wikidata

Atílio A. Dadalto[1], João Paulo A. Almeida[1],
Claudenir M. Fonseca[2], and Giancarlo Guizzardi[1,2]

[1] Ontology & Conceptual Modeling Research Group (NEMO),
Federal University of Espírito Santo (UFES), Brazil
atilio.dadalto@aluno.ufes.br, jpalmeida@ieee.org
[2] Conceptual and Cognitive Modeling Research Group (CORE),
Free University of Bozen-Bolzano, Italy
{cmoraisfonseca,giancarlo.guizzardi}@unibz.it

Abstract. The distinction between types and individuals is key to most conceptual modeling techniques. Despite that, there are a number of situations in which modelers navigate this distinction inadequately, leading to problematic models. We show evidence of a large number of modeling mistakes associated with the failure to employ this distinction in the Wikidata knowledge graph, which can be identified with the incorrect use of *instantiation*, which is a relation between an individual and a type, and *specialization* (or *subtyping*), which is a relation between two types.

Keywords: Types and instances, Taxonomies, Wikidata, Multi-Level Modeling

1 Introduction

Types are predicative entities, whose instances share some general characteristics, i.e., they are said to be repeatable invariances across multiple individuals. Individuals (or tokens), in their turn, are not general sorts of things, they are not repeatable; instead, they are particular entities, like Paul McCartney and John Lennon (instances of "person") or Jupiter and Mars (instances of "planet"). While we seem to be able to grasp this distinction intuitively, the boundaries between types and individuals are not always sharply drawn in everyday discourse. Consider, for instance, the paradigmatic case of "word" [11]. How many words are there in the sentence "the book is on the table"? The answer is *six* if we count the two occurrences of "the" as distinct words (or word tokens), or *five* if we count the word *types* used in the sentence. When we say "they drive the same car", do we mean the same *type of car* of the same *individual car*?

Given its occurrence in natural language, it is not surprising that this kind of ambiguity can arise also in knowledge representation and conceptual modeling. For instance, if we are capturing invariants about the domain of cars, what kinds of properties will characterize an entity named "car"? An *individual car* has a license plate and a production date, while a *car model* (a type) can be characterized by the tag sales price, the available colors, etc. Distinguishing between these two interpretations is key to grasp to what notion of "car" we refer to and what relations it can establish with other entities.

An instance of *car model* can specialize another type of car, in the way that "Porsche Speedster 23F" specializes "Four-Wheeled Car". An instance of *individual car* can instantiate "Porsche Speedster 23F", in the way that James Dean's Porsche did.

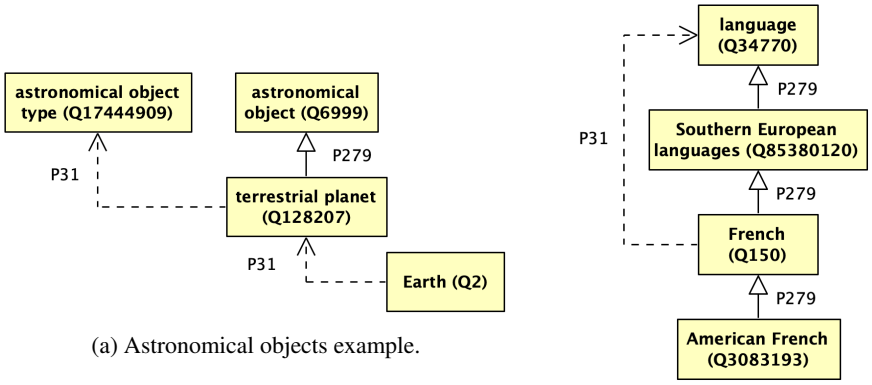
This paper examines the use of this distinction in practice, by employing Wikidata as a source of empirical data. Wikidata is structured as a graph with millions of nodes called *items*, which may represent a type (class) (e.g., the item for `planet (Q634)`) or an individual (e.g., the item for `Earth (Q2)`). The edges of this graph represent relations between items including specialization and instantiation. We here uncover a large number of items whose relations to other items indicate that their interpretation as a type or as an individual may be ambiguous. We investigate possible reasons behind these problems and, by using logical, ontological and semantic considerations, we propose some possible interpretation solutions for eliminating them. Finally, we demonstrate how we can leverage on an anti-pattern underlying the problems to build automated procedures that can proactively detect them before they are introduced to Wikidata.

This paper is further organized as follows: Section 2 introduces Wikidata's primitives for (multi-level) taxonomies. It shows some problems that occur when instantiation and specialization are combined in the platform. Section 3 identifies these problems at scale, updating some of the statistics collected in 2016 for Wikidata [2]. Section 4 examines these results in an attempt to identify a conceptual basis for explaining the identified problems, as well as proposing possible interpretation solutions for rectifying them. Section 5 presents a web application that can detect occurrences of the anti-pattern before they are introduced in Wikidata. Finally, Section 6 presents final considerations, including related work.

2 Taxonomies in Wikidata

Knowledge in Wikidata consists of *statements* that capture relations between *items*, which are "*are used to represent all the things in human knowledge*" [12]. A statement has the form of a "<subject> <property> <object>" triple. Examples of widely-used properties include `instance of (P31)` and `subclass of (P279)`. The property `instance of (P31)` represents a relation between an instance and a class (i.e., type), where the latter is predicated of the former. For example, `Earth (Q2)` is an `instance of terrestrial planet (Q128207)`, therefore exhibiting the properties of that class, in this case, being a planet of mostly rocky and metallic composition. The property `subclass of (P279)`, on the other hand, holds between two classes where the subclass has as instances a subset of the instances of the superclass. For example, `terrestrial planet (Q128207)` is a subclass of `planet (Q634)` meaning that every instance of the former is also an instance of the latter.

Wikidata also allows the declaration of classes of classes (or meta-classes). For example, `terrestrial planet` is `instance of` the class `astronomical object type (Q17444909)`, whose instances are specializations of `astronomical object (Q6999)`. See Figure 1(a), where boxes represent items; dashed and solid arrows represent `instance of (P31)` and `subclass of (P279)` respectively. We retain the capitalization of labels from Wikidata.



(a) Astronomical objects example.

(b) Languages example.

Fig. 1: Wikidata examples: (a) terrestrial planet as instance of astronomical object type and subclass of astronomical object; (b) French as instance and subclass of language.

The work of [3] clarifies this scheme of classes stratified in meta-levels (i.e., class, meta-class, meta-meta-class), using the concept of order, where individuals (entities that cannot have instances, like `Earth (Q2)` and `Alpha Centauri (Q12176)`) instantiate first-order classes (like `terrestrial planet (Q128207)`, `star (Q523)`), which in turn instantiate second-order classes (like `astronomical object type`), and so on into orders above (e.g., third-order, fourth-order).

This same clear stratification into orders is not present in other taxonomic structures of Wikidata, however. Consider, for instance, the following fragment concerning the French language, depicted in Figure 1(b). `French (Q150)` is both *instance of* and *subclass of* `language (Q34770)`. This opens up multiple interpretations: is French meant to be referring to a type of language or a specific, particular language? Of course, it is known that the French language is a particular language that has a certain number of speakers at a given point of time; however, variants of that language have spawned over the years, which can be considered instances of a class of French languages. The same applies to variants such as `American French (Q3083193)`, which denote the "varieties of the French language that are spoken in North America". The two facets (language as a class and language as a particular) are confounded in Wikidata.

3 Assessment of Taxonomic Structures in Wikidata

The fragment involving the French language identified in the previous section is an instance of a recurring pattern involving instantiation and specialization originally identified in [2]. More precisely, it is an occurrence of an *anti-pattern*, since it is a recurrent error-prone structure. The fragment exemplified by the French language is called here anti-pattern 1 (AP1 for short), and occurs whenever an item is instance of and subclass of another item (direct or indirectly) at the same time. AP1 prevents stratification into orders since, at the one hand, instantiation forces related items to be at different adja-

cent orders, and, at the other hand, a specialization of a class at a certain order must be in that same order (for formalization of the underlying theory and proofs, see [1, 3]). In this section, we discuss how we detect this pattern at scale in Wikidata and summarize the data we collected in the platform.

3.1 Data Collection

In order to deal with the size of Wikidata, we used a filtered dump of the Wikidata database[1]. Because our interest is only on taxonomic structures, only statements with the `subclass of (P279)` property were selected. The dump was created using wdumper[2] and processed using Stardog 7.4 and Jena 4.0.0. It has 2,452,006 entities, 26,264,034 statements and 38,224,283 triples, roughly 2.5% of the now almost 100,000,000 entities present in the complete Wikidata database as of April 2021.

3.2 Anti-Pattern Occurrences

To assess the occurrence of the anti-pattern, we have executed SPARQL queries in the filtered dump. Listing 1.1 shows the SPARQL query used to find AP1 occurrences considering transitiveness for *subclass of* statements. We have found 2,035,434 `?subject ?class` pairs involved in AP1, covering domains such as biology, gastronomy, awards, professions, sports, among others. Transitivity of subclassing is important as it reveals a large number of anti-pattern occurrences, which could indicate that it is harder to identify the specialization paths to indirect superclasses. The AP1 query without transitivity yields 1,279,629 results, while a query considering P279 transitivity returns 2,035,434 results.

Listing 1.1: SPARQL query for AP1.

```
SELECT DISTINCT ?subject ?class WHERE {
  ?subject wdt:P31 ?class .
  ?subject wdt:P279+ ?class .
}
```

3.3 Entities Most Frequently Involved in Anti-Pattern AP1

We have produced a ranking of the entities most frequently involved in the anti-pattern so that they could be further analyzed. The 10 top-ranked entities involved in AP1 are listed in Table 1 along with the number of times it participates in the anti-pattern. A comprehensive ranking with 200 entities and all scripts used in this paper are available at `https://purl.org/nemo/wapa`.

There is a clear overlap of subdomains in the ranking, especially but not limited to entities related to biology and biochemistry, e.g., `gene` as a "*basic physical and functional unit of heredity*" and `pseudogene (Q277338)` as a "*functionless relative of a gene*". For example, gene is a well-known complex concept frequently referring

[1] Wikidata dump generated in 14 September 2020, `https://zenodo.org/record/4046102`

[2] Further dump details and mirrors at `https://wdumps.toolforge.org/dump/749`.

Table 1: Ranking of occurrences of entities involved in AP1.

Place	Wikidata QID	English label	AP1 occurrences
1	Q7187	gene	971,982
2	Q8054	protein	757,360
3	Q4164871	position	103,545
4	Q277338	pseudogene	49,404
5	Q427087	non-coding RNA	49,132
6	Q2996394	biological process	30,315
7	Q12136	disease	12,293
8	Q14860489	molecular function	11,204
9	Q34770	language	6,795
10	Q5058355	cellular component	4,287

to a particular gene type repeatable in chromosomes of every cell (gene instances, i.e., particular biochemical structures composed of particular nucleotides) but also to the representation of a gene type (a data object) that results from genome sequencing operations. Multiple anti-pattern occurrences involving gene and pseudogene are introduced from batch adding or merging statements from external datasets such as UniProt and NCBI Gene, without proper consideration of imported entities as types or individuals. Hundreds of thousands of genes are directly related to `gene (Q7187)` in immediate instantiation and specialization relations! This pattern repeats for instances of `protein`, `protein domain`, `disease`, `rare disease`, `development defect during embryogenesis`, `head and neck disease`, `non-coding RNA`, `transfer RNA`. Users and softbots alike leverage databases such as GeneDB (genes), UniProt (proteins) Disease Ontology (diseases), InterPro, PubMed, NCBI Gene (RNAs), Gene Ontology (biological processes, cellular components), thus, introducing these violations. Other domains often present in AP1's top 20 entities are social roles and titles (e.g., `award (Q618779)`, `grade of an order (Q60754876)`, `position (Q4164871)`, and `public office (Q294414)`), language classification (e.g., `language (Q34770)`), and products of controlled origin denomination (e.g., `wine (Q282)`).

We inspected some of these top entities in the ranking to identify in which exact revision in the history of the Wikidata updates a violation was introduced. For example, take `language (Q34770)`. Originally, the item `Guarani (Q35876)` was simply represented as being an *instance of* `language`. However, revision 174811757 introduced the statement that `Guarani (Q35876)` is a *subclass of* `indigenous language of the Americas (Q51739)`—which is an indirect *subclass of* `language (Q34770)`. Together these statements configure a case of anti-pattern AP1. An anti-pattern checker could play a role in this context by detecting revisions that introduce inconsistencies prior to the inclusion of new statements.

4 Analysis and Discussion

The top-ranking entity involved in the anti-pattern we investigated is `gene`, which is described in Wikidata as a "basic physical and functional unit of heredity" with in-

stances such as TP53 (Q14818098), a "protein-coding gene in the species Homo sapiens". Inspecting their use in Wikidata, instances of gene like TP53 are most likely not "a particular gene from one cell from one person" but instead a *type* of which "many of us have tokens of — in fact many tokens of in each cell of our bodies" [11]. There is evidence for this in the properties ascribed to TP53, such as "found in taxon Homo Sapiens" and "encodes Tumor protein p53". This is consistent with an interpretation of gene as a second-order class, and its instances (e.g., TP53) as first-order classes. However, TP53, besides being declared as an instance of gene, is declared a subclass of protein-coding gene (Q20747295), which is itself a subclass of gene. Therefore, TP53 (and most of the other instances of gene) is also a subclass of gene. How should instances of TP53 be interpreted then, as they are also instances of gene like TP53 itself? We hypothesize that the subclassing statement is incorrect. TP53 is not a subclass, but an instance of the protein-coding gene subclass of gene. This issue may have never been flagged in Wikidata as instances of instances of gene are never instantiated explicitly in the platform (as it is not tracking "a particular gene from one cell from one person", but types of these). In fact, most gene talk is quantifying over types as discussed by Wetzel [11]. The same observation can be made for the other entities in the ranking related to biology and biochemistry such as: protein, pseudogene, non-coding RNA, and cellular component. These are all second-order types whose instances are first-order types classifying individual entities not recorded in the platform. Hence, there is a mismatch between ontological considerations (TP53 is a class instantiated by structures inside individual cells) and knowledge representation considerations (instances of TP53 are never recorded in Wikidata, suggesting it to be an individual).

Further in the ranking, we have the entity position (Q4164871), carrying the notion of "*social role [...] within an [...] organization*"). An instance of position is mayor (Q30185), "head of municipal government such as a town or city", instantiated by Frank Hilker (Q104772317). Clearly, he is an individual! Hence, mayor is a first-order class, suggesting position is a second-order class. However, mayor is declared as a subclass of public office (Q294414) which is a subclass of position. As a consequence, we come to the absurd inference that Frank Hilker is an instance of position (and consequently an instance of its superclasses, like artificial entity (Q16686448))! We hypothesize the declaration of mayor as a subclass of position is incorrect. The former being a first-order class and the latter a second-order class. As discussed in [3], order-crossing specialization is logically incorrect. Differently from the case of gene, the platform includes instances of instances of position (such as Frank Hilker); similarly, though, gene and position are second-order classes (meta-classes). It is important to note here that Wikidata has a specialized property to declare occupation of a position by a person (position held (P39)) and this is used instead of instantiation for most declarations of occupation. In any case, one needs to settle whether mayor and other entities like this are instances or specializations of position irrespective of the use of position held.

The case of biological process (Q2996394) also reveals confusion regarding the entity's order. It is a subclass of process (Q3249551), which in turn is a subclass of occurrence (Q1190554), the latter described as "occurrence of a fact or object in space-time". An occurrence may be qualified by point in time (Q186408), indicat-

ing that its instances are individual occurrences. Hence, biological process should be considered a first-order class. However, biological process includes among its instances entities such as birth (Q14819852) and death (Q4), entities bearing their own instances (e.g., the instance of death, death of James Dean (Q15213260)). Hence, death is a class of biological processes, which leads to an interpretation of biological process as a second-order class contradicting the earlier conclusion. Although present and declared as an instance of second-order class (Q24017414), the entity biological process type (Q47989961) has no instance of properties connecting it to any other items, not even the subclasses of biological process such as birth and death.

The case of language (Q34770), which we have raised earlier, involves the representation of extremely rich phenomena with much variation and diversity (a spectrum including macrolanguages, language families, and dialects). In this case, the criteria for individuation for a language is difficult to establish, and, as discussed earlier, items such as French can be regarded as a particular language or as a class of similar languages (given that each of its variations may be considered itself a language). We should note that language is an instance of languoid class (Q28923954), a "*[...] dialect, language, macrolanguage, language subfamily, family, or superfamily; each instance of these is a subclass of languoid*" according to its English description. Moreover, languoid class is explicitly marked as second-order class in Wikidata (it is an instance of Wikidata metaclass (Q19361238) which is an instance of third-order class (Q24017465)). This makes language a first-order class, and its instances individuals. As individuals, instances of language must not be involved in *subclass of* statements. To separate the two facets of a language, we need two items: one representing the language (say French of France (Q3083196)) as an instance of language (or dialect), and another as a subclass of language (or dialect) (referring to the class of French variants, whose instances include Quebec French (Q979914), Swiss French (Q1480152), and French of France).

Note that the ranking we have presented in this paper has been filtered to remove entities that are marked as instances of variable-order class (Q23958852), since these are explicitly flagged as not being stratified into a particular order. Variable-order [5] (or orderless [1]) classes have instances at different orders. Thus, being an orderless class can justify its *bona fide* occurrence in the (anti-)pattern, with no error incurred.

5 Automated Support

By leveraging on the type of analysis conducted in the previous section and the anti-pattern that can be identified with it, one can implement automated procedures for proactively identifying occurrences of this anti-pattern before it is introduced in Wikidata. In this section, we illustrate that by implementing such a procedure for the case of AP1 as a web application termed the Wikidata Anti-Pattern Analyzer[3] (WAPA). WAPA allows the user to input any entity from Wikidata to check for existing occurrences of AP1, or input a hypothetical statement to verify whether it would introduce new violations. Since it retrieves data directly from Wikidata's SPARQL endpoint, the results

[3] Available at https://atilioa.github.io/WikidataAntiPatternAnalyzer/.

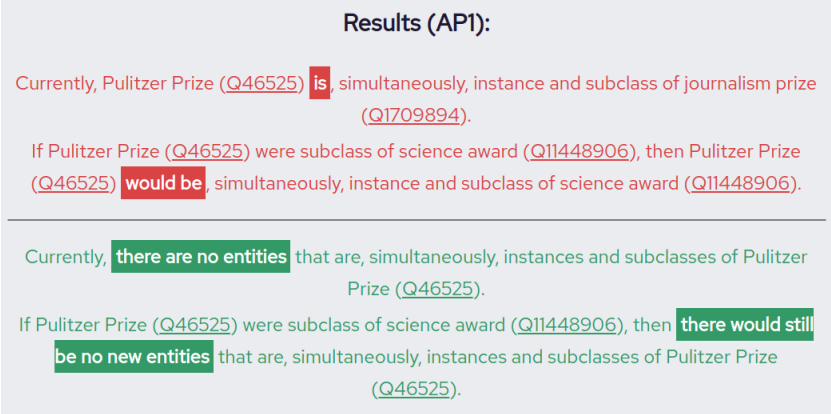


Fig. 2: WAPA results regarding hypothetical statement about `Pulitzer Prize (Q46525)`, reflecting the state of Wikidata as of April 2021.

reflect the current state of Wikidata (in the screenshots below, they reflect the state of Wikidata during the writing of this paper, April 2021).

As presented in Figure 2, one may check for AP1 violations that could be introduced to Wikidata with the addition of a statement "`Pulitzer Prize (Q46525) subclass of science award (Q11448906)`". Indeed, in this case, `Pulitzer Prize` would be, simultaneously instance and subclass of `science award`. Since WAPA always checks for existing violations before testing the hypothetical scenario, it would also inform that `Pulitzer Prize` is, simultaneously, instance and subclass of `journalism prize (Q1709894)` beyond the results for the hypothetical statement.

6 Final Considerations

In this paper, we conduct an empirical analysis of the Wikidata platform. We do that as a way to demonstrate how recurrent are anti-patterns exemplifying problems related to modeling of types and instances in large multi-level knowledge models. As this empirical data corroborates, this is a widespread problem with thousands and even million of occurrences in Wikidata. We also identify the items in Wikidata appearing in the highest number of occurrences of AP1. By conducting a conceptual analysis of these cases, we manage to venture an explanation for their occurrence, and propose interpretation solutions that would eliminate them. (Due to space limitations, the analysis conducted in Section 4 was limited to a subset of the top-ranking notions.) Finally, we show how this anti-pattern can inform the construction of automated procedures that can proactively detect this anti-pattern before it is introduced in such a knowledge model. In an earlier work, some of us explored the role of a multi-level modeling language (ML2) in detecting the occurrence of the anti-patterns discussed here [4]. Differently from that work, here we proposed a web application that can be used by Wikidata users to detect the problems in a language-independent manner.

We should note that the concepts of *order* and the stratification of taxonomies into consistent multi-level structures are concerns present in Wikidata since revisions introduced in mid 2016. Since then, to support stratified taxonomies, the platform includes at the top of its specialization hierarchy a set of classes representing different orders, namely `first-order class (Q104086571)`, `second-order class (Q24017414)`, `third-order class (Q24017465)`, `fourth-order class (Q24027474)`, `fifth-order class (Q24027515)`, and `fixed order metaclass of higher order (Q24027526)`. These classes are declared as equivalent to their counterparts in the OpenCyc ontology [5]. However, they are underused in the platform, and, as we show here and in [4], their mere inclusion in the platform without adequate computational aid has been insufficient to prevent the introduction of anti-patterns in new revisions. This motivated us to provide some automated support as shown here.

The dual facet of entities that are both types and instances is a phenomenon that is well-documented in (multi-level) conceptual modeling [3], in formal ontology [5, 8], and in linguistics [10]. In particular, the phenomenon of *systematic polysemy* in language accounts for many cases of this problem. For example, when we say "these ducks in the backyard are common around Europe", we are making a polysemic reference that overloads the term duck with particular duck instances (those in the backyard) with a duck type (that which is repeatable in a population of ducks and, hence, which is common around Europe). This polysemy that is present in natural language, we conjecture, is also manifested in the construction of lightweight representation structures such as Wikidata. This is specially the case when such a structure is collectively constructed in an asynchronous manner by millions of users, many of which are not expert modelers. This is made worse when these naive modeling strategies (oblivious to these problems) are codified in computer programs (e.g., softbots) that automatically transfer knowledge snippets from other existing data sources.

As we show here, by conducting an analysis of the logical and ontological reasons behind the phenomena causing these semantic confusions, we can proactive devise methodological (e.g., anti-patterns) and computational tools that can assist users in avoiding these mistakes. In this sense, the work presented here is in line with a number of successful initiatives of employing ontological principles to evaluate and rectify large-scale knowledge structures. These include, for example: (i) [6] and [7], which respectively use the DOLCE foundational ontology and the OntoClean methodology for analyzing and proposing correction to the Wordnet top-level; (ii) [9], which uses a lightweight version of DOLCE (termed DOLCE-Zero) for detecting anti-patterns in DBPedia. The works in (i) focus on detecting taxonomic problems related to ontological notions such as identity, unity, and dependence. In contrast, in (ii), the most common patterns detected are related to logical conflicts between disjoint types that are expected by and asserted to given properties. These are related to confusions between objects and events, agents and places, physical and social objects, etc. For example, `dbpedia#AlfonsoXIIofSpain dbo#birthPlace dbpedia#Madrid`, where `dbpedia#Madrid` is erroneously typed as `dbo#Agent`(as a geopolitical entity), which is a confusion between the disjoint types Place and Agent. One of the one of the patterns detected in (ii), however, is what the authors call *metonymy*, which is a conflict arising from disjoint but related interpretations of the same concept. In particular, they make

the example of dbo#family, which is used to related instances of dbo#Species and its property specializing concepts. However, dbo#Species are aligned to the type Organism, because "species in DBpedia include species as well as individual exemplars of a species (for example, famous race horses)". Although this case seems to exemplify a type/instance confusion, the authors arrive at it by, once more, detecting disjoint types in the domain/range of properties, as opposed to explicitly identifying anti-patterns related to this problem. Moreover, they seem to have a somewhat lenient approach with respect to these problems: "[t]he metonymy anti-pattern is difficult to resolve, because it is due to ambiguities that seem widespread in human language. Metonymy seems related to human propensity for an economy of means... [we try] to accommodate this 'power of ambiguity"'. We take here a radically different approach in this respect by advocating that these problems can cause logical contradictions and conceptual confusion, and by proposing concrete means to detect and correct them.

Acknowledgments

This research is partly funded by Brazilian funding agencies CNPq (grants 313687/2020-0, 407235/2017-5) and CAPES (grant 23038.028816/2016-41). Claudenir M. Fonseca and Giancarlo Guizzardi are supported by the NeXON Project (Free University of Bozen-Bolzano).

References

1. Almeida, J.P.A., Fonseca, C.M., Carvalho, V.A.: A comprehensive formal theory for multi-level conceptual modeling. In: Conceptual Modeling. pp. 280–294. Springer (2017)
2. Brasileiro, F., Almeida, J.P.A., Carvalho, V.A., Guizzardi, G.: Applying a multi-level modeling theory to assess taxonomic hierarchies in wikidata. In: Proc. 25th International Conference Companion on World Wide Web. pp. 975–980. WWW '16 Companion (2016)
3. Carvalho, V.A., Almeida, J.P.A.: Toward a well-founded theory for multi-level conceptual modeling. Software & Systems Modeling **17**(1), 205–231 (2018)
4. Fonseca, C.M., Almeida, J.P.A., Guizzardi, G., Carvalho, V.A.: Multi-level conceptual modeling: Theory, language and application. Data & Knowledge Engineering **134**, 101894 (Jul 2021). https://doi.org/10.1016/j.datak.2021.101894
5. Foxvog, D.: Instances of instances modeled via higher-order classes. In: Workshop on Foundational Aspects of Ontologies, 28th German Conf Artificial Intelligence. pp. 46–54 (2005)
6. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A.: Sweetening WORDNET with DOLCE. AI magazine **24**(3), 13–13 (2003). https://doi.org/10.1609/aimag.v24i3.1715
7. Gangemi, A., Guarino, N., Oltramari, A.: Conceptual analysis of lexical taxonomies: The case of WordNet top-level. In: Proc. FOIS 2001. pp. 285–296 (2001). https://doi.org/10.1145/505168.505195
8. Guizzardi, G., Almeida, J.P.A., Guarino, N., Carvalho, V.A.: Towards an Ontological Analysis of Powertypes. In: Proc. Joint Ontology Workshops 2015 Episode 1: The Argentine Winter of Ontology. CEUR Workshop Proceedings, vol. 1517. CEUR-WS.org (2015)
9. Paulheim, H., Gangemi, A.: Serving DBpedia with DOLCE–more than just adding a cherry on top. In: International Semantic Web Conference. pp. 180–196. Springer (2015)
10. Ravin, Y., Leacock, C.: Polysemy: Theoretical and computational approaches. OUP (2000)
11. Wetzel, L.: Types and tokens: on abstract objects. MIT Press, Cambridge, Mass (2009)
12. Wikidata: Help:items — Wikidata. https://web.archive.org/web/20210127110938/https://www.wikidata.org/wiki/Help:Items (2021), [Online: 2-May-2021]