Conceptually Disentangled Classificatory Ontologies

Mayukh Bagchi[1], Subhashis Das[2]

[1]*DISI, University of Trento, Via Sommarive, 9, 38123 Povo, Trento TN, Italy.*
[2]*CeIC, ADAPT, School of Computing, Dublin City University (DCU), Dublin 9, Ireland*

Abstract

In mainstream knowledge organization, classificatory ontologies are widely employed for classifying, annotating and searching for data specific to a particular domain. The key bottleneck, however, remains the fact that these ontologies fail to encode heterogeneity in conceptual representations and are *representationally static*, i.e. they assume a *one-size-fits-all* ontological hierarchy for a particular domain. We argue that the above bottleneck fundamentally ignores the phenomenon of *Conceptual Entanglement*, i.e. the many-to-many entanglement between the source and the target conceptual representation existing independently within each of the following five levels: *Perception*, *Labelling*, *Semantic Alignment*, *Hierarchical Modelling* and *Intensional Definition.* To that end, we also introduce, at a high level, the notion of *Conceptual Disentanglement* which can be seen as a multi-level conceptual modelling strategy to enforce one-to-one correspondences disentangling the many-to-many entanglement within each of the above level, tuned to the purposive viewpoint of the chosen target reality.

Keywords

Conceptual Entanglement, Conceptual Disentanglement, Classificatory Ontologies, Semantics.

1. Introduction

Consider the motivating case of the *Infinity Coast*[1], one of the tallest buildings in Brazil. It can be conceptually described and represented in different ways, for instance, as a *conference venue* in a database for conference hosts, a *gourmet space* in a food destination database or as a *party hub* in an event management database. Notice that the conceptual representations for each of the above cases, while referring to the same entity, are nevertheless *semantically heterogeneous* (see [1, 2, 3] for more such examples). Such heterogeneity, whether in the aforementioned example, or, in general, for any real-world entity, occur due to such representations being, at their core, *cognitive constructs* [4], grounded in the very way in which (human) conceptualizations are causally generated from (human) experientiality [5, 6, 7].

In this work, we concentrate on classificatory ontologies [8, 9] which, while failing to encode instantiations of heteroegenity of the above kind, are still employed in mainstream knowledge organization for data classification, annotation and search purposes. The key bottleneck stems from the fact that such ontologies, as compared to descriptive ontologies [8], are grounded in

[1]https://en.wikipedia.org/wiki/Infinity_Coast

a *representationally static* formalism, i.e., they assume a *one-size-fits-all* ontological hierarchy for data of a particular domain (being grounded in knowledge classification schemes [10]), and thereby lack operational precision specific to knowledge-based AI systems [11].

Such a formalism, in effect, ignores the phenomenon of *Conceptual Entanglement*, *viz.*, a layered, many-to-many *entanglement* from the perceptual generation of concepts to their classificatory ontological formalization. We outline five ordered, functionally linked levels into which *Conceptual Entanglement* distributes. The first level of entanglement (*Perception*) is generated due to the different ways of perceiving different real-world entities [12, 13]. The second level of entanglement (*Labelling*) arises due to the different ways of linguistically labelling different perceived concepts. The third level of entanglement (*Semantic Alignment*) pertains to the different top-level ontological distinctions [14, 15] into which different labelled concepts can be semantically constrained to. The fourth level of entanglement (*Hierarchical Modelling*) instantiates as the different hierarchical ontological models into which different labelled concepts can be organized. The final level of entanglement (*Intensional Definition*) occurs due to the different ways in which different concepts in the hierarchy can be defined via attributes.

Our solution approach has two core assumptions. Firstly, we assume that there is *"only one 'real world' but many different descriptions of this world depending on the aims, methodology and terminology of the observer"* [16]. Secondly, we maintain that heterogeneity is *"a feature which must be maintained and exploited and [is] not a defect that must be absorbed in some general schema"* [17]. Based on the above assumptions, we propose *Conceptual Disentanglement* as a conceptual modelling strategy grounded in guiding *best practices*, following which the many-to-many entanglements in each level (and subsequently, in entirety) of formalizing a classificatory ontology can be *disentangled* to one-to-one semantic correspondences, while still accommodating the specific heterogeneity (purposive viewpoint) of the chosen target reality.

The remainder of the paper is organized as follows: Section (2) details the layered phenomenon of conceptual entanglement. Section (3) elucidates the proposed conceptual disentanglement strategy for classificatory ontologies. Section (4) concludes the paper with a brief summative discussion of the work by comparing it to the relevant state-of-the-art.

2. Conceptual Entanglement

A conceptual representation is defined as *"an abstract, simplified view of the world that we wish to represent"* [18] and is fundamentally *mental* in nature [19]. We adhere to an extended notion of conceptual representation which models concepts via a five-level stratification: *Perception*, *Labelling*, *Semantic Alignment*, *Hierarchical Modelling* and *Intensional Definition*. Founded in the aforementioned stratification, we define *Conceptual Entanglement* to be the many-to-many entanglement between the source and the target conceptual representation that is *ubiquitous* with respect to the above levels (individually as well as cumulatively across levels). We now elucidate the many-to-many conceptual entanglement as it instantiates in each level.

Perception: Concepts, universally regarded as the *"building blocks of thoughts"* [20], are aggregated and abstracted via the process of perceiving a target reality. Note that both the real-world referent as well as their relevant properties are *concepts* in our view. To that end, the fact

that the same *referent* or *property* can be perceived variously by different *agents* depending on the different viewpoints leads to, we argue, the many-to-many entanglement at the perceptual level [21]. For instance, the *Infinity Coast* can be perceived as a *conference venue* or a *party hub*, each of which can also be a valid perception for the *Epic Tower*[2].

Labelling: Given perception, the focus of the second level is on linguistically *labelling* the perceived concepts for human as well as machine interaction. Nevertheless, the activity of labelling is *non-trivial* due to the deep interaction between language and thought. We mention two highlights. Firstly, languages are *"itemized inventories"* of a target reality [22]. Secondly, every language generates a *similar but not same* labelling of a perceived concept [23] inducing the many-to-many entanglement. For instance, the same perception of the *Infinity Coast* as a 'conference venue* can be heterogeneously labelled as a *locus* or as an *emplacement* in English. Moreover, the same perception can have further divergent labellings in multilingual scenarios.

Semantic Alignment: The third level aligns the labelled concepts to top-level ontological distinctions such as whether it is an *independent* or a *dependent* concept, or, for instance, a *process* or an *event* [14, 15]. This is crucial given the well-established fact that the same (perceived and labelled) concept can be modelled in terms of different top-level ontological distinctions, and vice versa. For example, the same *Infinity Coast*, which can be modelled as a *building* (an independent concept), can also be modelled as a *party hub* (a dependent concept) given that it participates in the event of a party, and vice versa.

Hierarchical Modelling: The fourth level concentrates on modelling the labelled, semantically constrained concepts in a *taxonomical hierarchy*. We ground our hierarchy modelling design in the *four-step* ontologically well-founded classification theory by Ranganathan [24, 25]. The first step concerns deciding the many differentiating *characteristics* for classification at different depths in the hierarchical tree. The second step involves the *succession of characteristics* which determine how different concepts are organized and applied at different successive level in the hierarchy. The third step and the fourth step concentrates on the consistency of the organization of multiple concepts horizontally across multiple depths (each depth termed as an *array*) and vertically across multiple paths (each path termed *chain*) in the taxonomic tree respectively. For example, the *Infinity Coast* as a *building* can be equally classified based on the characteristic *colour* or *purpose*. Similarly, given the first classifying characteristic as *colour*, the next classificatory characteristic can be, for instance, the *square footage*. All of these characteristics are, in turn, applicable for hiearchically modelling many other concepts in the same/similar target reality.

Intensional Definition: The fifth and the final level of entanglement occurs from modelling the concepts at an intensional level, wherein, each individual concept in the taxonomic hierarchy is defined via their appropriate *attributes*, thereby rendering the hierarchical model as a formal *classificatory ontology* [8, 9]. The many-to-many entanglement at the intensional level is generated when, each concept, in the different classificatory ontologies modelled out of the same target reality, can be differently defined via a distinct set of attributes. Let us take an example. The notion of *Infinity Coast* as a *conference venue* can be characterized differently via the following two sets of attributes: *{number of rooms, number of seminar halls}* or *{year of establishment, star rating}*. Further, these same attributes can be employed to define other

---

[2]https://fgempreendimentos.com.br/empreendimentos/epic-tower

concepts in the same/similar domain.

To sum up, we mention an important observation. The many-to-many conceptual entanglement across the different levels of conceptual representation results in a non-trivial *incorrespondence* amongst its mental model (which is language-agnostic), classificatory ontological model (which, almost always, is expressed in a formal ontology language) and its underlying logical axiomatization (expressed, mostly, as a decidable fragment of first order logic).

3. Conceptual Disentanglement

We propose *Conceptual Disentanglement* as a conceptual modelling strategy to tackle the five-fold characterization of conceptual entanglement that instantiates in modelling classificatory ontologies. It refers to a set of guiding *normative principles* which, if considered as best practice for each of the five levels, can enforce *one-to-one* correspondences with respect to the many-to-many entanglement at each level while, concurrently, factoring in the required heterogeneity of the target reality that needs to be modelled. In effect, conceptual disentanglement can provide the conceptual modelling foundations based on which, later on, a methodology for dynamically harmonizing diverse conceptual representations into a single classificatory ontology can be developed. We now elucidate conceptual disentanglement strategy specific to each level.

Perception: Firstly, we concentrate on the norms which tackle the conceptual entanglement instantiated at the *Perception* level. We recommend the following best practices:

- At the outset, the *target reality* should be specified with precision in terms of their *spatio-temporal* coverage. In case the target reality is comprised of several smaller component target realities, it should be modelled as a disjoint union of the component realities (i.e. as a disjoint union of component spatio-temporal coverages).
- Given the precise specification of the target reality, the second sub-activity should determine not only the concepts which should be modelled within the chosen target reality but also their purpose-driven *viewpoints*.

The aforementioned best practices allows precise selection the *intended ontological commitment* of the target reality and thus avoid instances of overcommitment and undercommitment which frequently plague domain classificatory ontologies. This, in effect, reduces the many-to-many entanglement at the conception layer to a one-to-one correspondence. For example, fixing the geospatial coordinates and temporal duration of the *Infinity Coast* modelled from two viewpoints, that of a *cinema theatre* and a *party hub*.

Labelling: Secondly, we outline the guiding norms for conceptual disentanglement with respect to labelling concepts:

- Fixation of the base natural language(s), e.g., English, Portugese etc., and corresponding *controlled vocabulary* terms, i.e., a widely *inter-labeller agreed* terminology that can be exploited to *unambiguously* label the perceived concepts. International terminological standards for various domains can be utilized for this purpose. Such a choice forces a one-to-one correspondence out of the multiplicity of possible labellings in the selected language(s) on one hand, and neutralizes the effect of linguistically-grounded labelling conflicts such as *endonym* and *exonym* [26] on the other hand.

- Optionally, the next step, especially key in multilingual data classification and search scenarios [27], should be to add global, unambiguous identifiers (such as from Wikidata[3]) to disambiguate each such labelled concepts.

Semantic Alignment*:* For this level, an ontological analysis [28] should be performed with respect to each of the labelled concepts (including both *referents* and their relevant *attributes*) from the previous level. The key aim of the ontological analysis should be to determine the exact ontological nature of each labelled concepts via their semantic conformance to a specific top-level ontological distinction. For example, in a specific case, the analysis might result in aligning the concept of *Infinity Coast* to the top-level distinction of a *dependent concept*.

There are two crucial advantages behind disentangling the conceptual entanglement at the semantic alignment level. Firstly, the ontological analysis on labelled concepts disobfuscates the *'arbitrarity'* in the intended meaning of such concepts [29, 30] by ascertaining their exact ontological nature, and, in turn, streamlines a one-to-one correspondence between the labelled concept and the top-level distinction out of the many-to-many entanglement exisiting previously. Secondly, and more relevant from the application perspective, the semantic conformance to top-level ontological distinctions also facilitate linking the concepts in a classificatory ontology to the Linked Open Data Cloud[4], thus, making it (and the data it classifies and annotates) interoperable with a highly interconnected network of semantically classified data and knowledge.

Hierarchical Modelling*:* Given the conformance of the labelled concepts to the top-level ontological distinctions, we now focus on the conceptual disentanglement best practices while modelling the taxonomical hierarchy. As from Section 2, we also ground our solution in Ranganathan's classification theory [24, 25]. In particular, we exploit the mutually coordinating normative principles proposed by Ranganathan [24] (termed as *canons*), for each of the four steps of building a taxonomy (see Section 2), to disentangle the many-to-many entanglement immanent in each step:

- In the first step, we reduce the many-to-many entanglement in the selection of the differentiating characteristic to one-to-one correspondence by exploiting the canons of *relevance* (stating that such a characteristic should be relevant to the purpose at hand) and *ascertainability* (stating that such a characteristic should be perceptually ascertainable). For example, in the case of building a classificatory ontology for buildings for a local government, we fix *legal nature of building* as the first classification characteristic.
- The many-to-many entanglement for the second step of choosing the succession of characteristics is tackled by employing the canon of *relevant succession* which enforces that the selection of successive differentiating characteristics across the depths of a taxonomy should be founded solely on purpose. For instance, the second characteristic for the classificatory ontology on buildings could be *year of establishment* given the purpose is to aggregate timeseries data on real estate by a local government body.
- The many-to-many entanglement for the third step of organizing an array is tackled by the canon of *exhaustiveness* which ensures that all the concepts at a specific depth in the taxonomic tree is exhaustively classified at the next depth and thereby ensures the exclusivity of chosen purpose-driven differentiating characteristic(s).

---

[3]https://www.wikidata.org/wiki/Wikidata:Main_Page
[4]https://lod-cloud.net/

- Lastly, the many-to-many entanglement for the fourth step of organizing a chain is tackled by the canon of *modulation* which ensures that there are no missing *conceptual* links in any path of a taxonomy. For example, this canon ensures that all the paths in the classification of a domain of buildings are populated by concepts at all depths and facilitates ruling out missing links which indicates a many-to-many crossover in the succession of characteristics.

Intensional Definition: Given the disentanglement of the taxonomic hierarchy, the final many-to-many entanglement is fixated by precisely determining the attributes that ought to be encoded by the classificatory ontology for each concept in its hierarchy. For example, as from before, we fix the attributes of the concept of *Conference Venue* as the *{number of rooms, number of seminar halls}* given our purpose of accommodating co-located talks and tutorial speeches.

To sum up, note that the conceptual disentanglement across the different levels of conceptual representation enforces a novel *correspondence* amongst its mental model (which is language-agnostic), classificatory ontology (expressed in a formal ontology language) and its underlying logical axiomatization (expressed as a decidable fragment of first order logic). Also notice that the decision to *reuse* concepts *as is* from existing ontologies or knowledge classification schemes, at any level of conceptual representation as characterized above, is a decision completely dependent on the project team effectuating conceptual disentanglement.

4. Conclusive Discussion

Given the explication of the phenomenon of *Conceptual Entanglement* in classificatory ontologies and a possible solution strategy in the form of *Conceptual Disentanglement*, the next key question becomes the reuse and/or development of a methodology for engineering conceptually disentangled classificatory ontologies. The existing landscape of general-purpose ontology development methodologies (see [31] for a survey on early generation methodologies; also see [32, 33, 34, 4]), while being exceptionally rich, are not completely suitable to be exploited in our case for two particular reasons. Firstly, none of the methodologies *recognize*, in an entirety, the five-layered phenomenon of conceptual entanglement (given their difference in focus). Secondly, none of the above methodologies are tailor-made for classificatory ontologies, the difference of which with respect to descriptive ontologies have been established [35]. Moreover, the same reasons also hold for ontology development methodologies developed in the context of engineering ontologies in different domains, e.g., see [36] for healthcare, [37] for life sciences, [38] for industries, [39] for smart cities, [40] for education etc. This is our immediate future work.

In summary, the short paper introduced the novel phenomenon of *Conceptual Entanglement* in classificatory ontologies within the broad context of knowledge organization. It also proposed, at a high level, the conceptual modelling strategy of *Conceptual Disentanglement* as a solution to the above phenomenon.

Acknowledgement

References

[1] S. Das, F. Giunchiglia, Geoetypes: Harmonizing diversity in geospatial data (short paper), in: OTM Confederated International Conferences" On the Move to Meaningful Internet Systems", Springer, 2016, pp. 643–653.

[2] S. Das, Domain Modeling Theory and Practice, Ph.D. thesis, University of Trento, 2018.

[3] F. Giunchiglia, M. Bagchi, Representation heterogeneity, in: 1st International Workshop on Formal Models of Knowledge Diversity (FMKD), Joint Ontology WOrkshops (JOWO), Jönköping University, Jönköping, Sweden, 2022.

[4] M. Bagchi, A large scale, knowledge intensive domain development methodology., Knowledge Organization 48 (2021).

[5] M. Bagchi, A Knowledge Architecture using Knowledge Graphs, Master's thesis, Indian Statistical Institute, Bangalore, 2019.

[6] F. Giunchiglia, M. Bagchi, Object recognition as classification via visual properties, in: 17th International ISKO Conference and Advances in Knowledge Organization, Aalborg, Denmark, 2022.

[7] M. Bagchi, A diversity-aware domain development methodology, arXiv preprint arXiv:2208.13064 Accepted @ PhD Symposium, 41st International Conference on Conceptual Modeling (ER Conference). (2022).

[8] F. Giunchiglia, M. Marchese, I. Zaihrayeu, Encoding classifications into lightweight ontologies, in: Journal on data semantics VIII, Springer, 2007, pp. 57–81.

[9] A. R. D. Prasad, D. P. Madalli, Classificatory ontologies, The Hague, 29-30 October (2009) 223.

[10] P. Rafferty, The representation of knowledge in library classification schemes, KO Knowledge Organization 28 (2001) 180–191.

[11] M. Bagchi, Towards knowledge organization ecosystem (koe), Cataloging & Classification Quarterly 59 (2021) 740–756.

[12] F. Giunchiglia, M. Bagchi, Millikan + ranganathan – from perception to classification, in: 5th Cognition And Ontologies (CAOS) Workshop, Co-located with the 12th International Conference on Formal Ontology in Information Systems (FOIS) 2021, Bolzano, Italy, 2021.

[13] F. Giunchiglia, M. Bagchi, X. Diao, Visual ground truth construction as faceted classification, arXiv preprint arXiv:2202.08512 (2022).

[14] R. Arp, B. Smith, A. D. Spear, Building ontologies with basic formal ontology, 2015.

[15] S. Borgo, R. Ferrario, A. Gangemi, N. Guarino, C. Masolo, D. Porello, E. M. Sanfilippo, L. Vieu, Dolce: A descriptive ontology for linguistic and cognitive engineering, Applied ontology (2022) 1–25.

[16] INSPIRE, D2.8.ii.2 data specification on land cover – technical guidelines, https://inspire.ec.europa.eu/id/document/tg/lc (2013).

[17] F. Giunchiglia, Managing diversity in knowledge, in: ECAI 2006: 17th European Conference on Artificial Intelligence, volume 141, IOS Press, 2006, p. 4.

[18] M. R. Genesereth, N. J. Nilsson, Logical foundations of artificial intelligence, Morgan Kaufmann, Massachusetts, 2012.

[19] N. Guarino, D. Oberle, S. Staab, What is an ontology?, in: Handbook on ontologies, Springer, Berlin, Heidelberg, 2009, pp. 1–17.

[20] E. Margolis, S. Laurence, Concepts, in: E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy, Spring 2021 ed., Metaphysics Research Lab, Stanford University, Stanford, 2021.

[21] M. Bagchi, D. Madalli, Domain visualization using knowledge cartography in the big data era: A knowledge graph based alternative, in: International Conference on Future of Libraries: Jointly organized by Indian Institute of Management (IIM) Bangalore and Indian Statistical Institute (ISI) Kolkata, Bangalore; https://library.iimb.ac.in/conference2019/contributorspresentations, 2019.

[22] R. W. Brown, E. H. Lenneberg, A study in language and cognition., The Journal of Abnormal and Social Psychology 49 (1954) 454.

[23] L. Boroditsky, How language shapes thought, Scientific American 304 (2011) 62–65.

[24] S. R. Ranganathan, Prolegomena to Library Classification, Asia Publishing House, New York, 1967.

[25] S. R. Ranganathan, Philosophy of library classification, Sarada Ranganathan Endowment for Library Science, Bangalore, India, 1989.

[26] D. Perko, P. Jordan, B. Komac, Exonyms and other geographical names, Acta geographica Slovenica 57 (2017) 99–107.

[27] G. Bella, L. Elliott, S. Das, S. Pavis, E. Turra, D. Robertson, F. Giunchiglia, Cross-border medical research using multi-layered and distributed knowledge, in: 10th International Conference on Prestigious Applications of Intelligent Systems@ ECAI 2020, IOS Press, 2020, pp. 2956–2963.

[28] N. Guarino, C. Welty, Evaluating ontological decisions with ontoclean, Communications of the ACM 45 (2002) 61–65.

[29] N. Guarino, The ontological level: Revisiting 30 years of knowledge representation, in: Conceptual modeling: Foundations and applications, Springer, 2009, pp. 52–67.

[30] N. Guarino, The ontological level, Philosophy and the cognitive sciences (1994).

[31] M. Fernández-López, Overview of methodologies for building ontologies, in: IJCAI99 Ontology Workshop, volume 430, Citeseer, 1999.

[32] M. Fernández-López, A. Gómez-Pérez, N. Juristo, Methontology: from ontological art towards ontological engineering (1997).

[33] N. F. Noy, D. L. McGuinness, et al., Ontology development 101: A guide to creating your first ontology, 2001.

[34] M. C. Suárez-Figueroa, A. Gómez-Pérez, M. Fernández-López, The neon methodology for ontology engineering, in: Ontology engineering in a networked world, Springer, 2012, pp. 9–34.

[35] F. Giunchiglia, B. Dutta, V. Maltese, From knowledge organization to knowledge representation, KNOWLEDGE ORGANIZATION 41 (2014) 44–56.

[36] S. Das, S. Roy, Faceted ontological model for brain tumour study., Knowledge Organization

43 (2016).

[37] B. Smith, W. Ceusters, Ontological realism: A methodology for coordinated evolution of scientific ontologies, Applied ontology 5 (2010) 139–188.

[38] M. Poveda-Villalón, A. Fernández-Izquierdo, M. Fernández-López, R. García-Castro, Lot: An industrial oriented ontology engineering framework, Engineering Applications of Artificial Intelligence 111 (2022) 104755.

[39] P. Espinoza-Arias, M. Poveda-Villalón, R. García-Castro, O. Corcho, Ontological representation of smart city data: From devices to cities, Applied Sciences 9 (2018) 32.

[40] S. Das, D. Naskar, S. Roy, Reorganizing educational institutional domain using faceted ontological principles, Knowledge Organization 49 (2022) 6–21.