The Inescapable Relativity of Explicitly Represented Knowledge: An FCA Perspective

David Flater(✉)

National Institute of Standards and Technology, Gaithersburg, USA
david.flater@nist.gov

Abstract. Knowledge models are supposed to capture knowledge of lasting value in a reusable form. However, reuse of these models is hampered by arbitrary and application-specific constraints; any constraints that conflict with a new application must be altered or removed before the models can be reused. This article explores seven facets of conceptual relativity that would impact the use of Formal Concept Analysis formalisms to represent knowledge, demonstrating that the capture of application-specific constraints is inextricable from the modelling process.

1 Introduction

Knowledge models are intended to capture knowledge of lasting value in a reusable form. They can model the concepts relevant to a software project, a domain of discourse, or any world view of whatever scope. Their range of representations includes specialized ontology languages such as Web Ontology Language (OWL) [9], general-purpose modelling languages such as Unified Modeling Language (UML) [13], and the formalisms of Formal Concept Analysis (FCA) [5].

Reuse of knowledge models is hampered by arbitrary and application-specific constraints. Any constraints that conflict with a new application must be altered or removed before the models can be reused. For this reason, the modeller seeks to avoid mingling arbitrary and application-specific constraints with those believed to be universal. In a typical application of FCA, the algorithmically identified formal concepts would promulgate application-specific constraints with no warning to the practitioner. However, the problem actually is endemic to the process, and no language or technique can avoid it entirely. It is therefore important to be aware of the dangers.

This article explores seven facets of relativity in knowledge models, demonstrating that the capture of application-specific constraints is inextricable from the modelling process. The facets of conceptual relativity to be discussed are shown in Fig. 1.

The term *intent* is already well-established in the knowledge modelling domain. The terms *essence*, *identity*, and *unity* were previously introduced to

M. Kryszkiewicz et al. (Eds.): ISMIS 2017, LNAI 10352, pp. 580–586, 2017.
DOI: 10.1007/978-3-319-60438-1_57

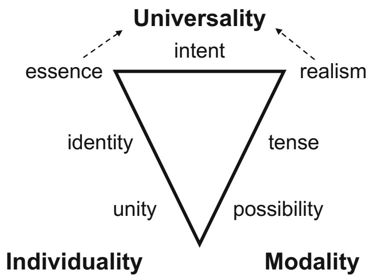

Fig. 1. Facets of conceptual relativity

the knowledge modelling community by Guarino and Welty [6,7]. The remaining terms have been introduced in parallel fashion to complete the framework needed to discuss conceptual relativity. The dilemmas confounding knowledge modellers today were first identified and explored by philosophers, so the corresponding terms from philosophical references have been applied.

As the diagram suggests, the facets are not independent; neither do they fall neatly into categories. However, generally speaking, essence, identity, and unity have to do with individuality—the factoring of the domain of discourse into separate things. Possibility, tense, and realism have to do with modality—the factoring of the domain of discourse into different ways of existing. Essence, realism, and intent have to do with universality—the determination of what is held constant. Essence is an individual perspective on universality (what universality means for individuals); realism is a modal perspective on universality (what universality means for existence).

The following sections examine the seven facets of conceptual relativity as they would impact the use of FCA formalisms to represent knowledge. The formal contexts and concept lattices of examples that have been redacted due to the page limit can be found in an unabridged technical report [3].

2 Essence

> A property of an entity is essential to that entity if it must hold for it. This is a stronger notion than one of permanence, that is, a property of an entity is not essential if it just happens to be true of it, accidentally [6].

In the context of this discussion, to refer to properties as being essential to an *entity* means, more accurately, that those properties are deemed logically necessary in some intensional classification or identification of that entity. Essence identifies conditions that are necessary; it does not address sufficiency.

The appropriateness of any given classification is relative to the application. For example, consider the classification of chemical elements. A research institute would need to identify different isotopes of helium in nuclear physics experiments, but in many industrial applications, it goes without saying that the helium cylinder contains mostly ^{4}He. Consequently, FCA of data from these different applications would lead to different sets of attributes for a concept called 'helium.' Manual reconciliation would conclude that a neutron count of 2 is essential to the industrial concept of helium, even if it was never stated explicitly in the data, while it is non-essential to the nuclear physics concept.

3 Identity

When something undergoes a change, whether or not it is considered the *same* thing afterwards depends on how that thing is identified.

Using nominal scales to transform many-valued attributes, the formal context shown in Table 1 models the decay of a ^{5}He atom at time t_1 to ^{4}He at time t_2. As yet, no commitment has been made regarding whether the ^{4}He is the same atom as the ^{5}He. This formal context merely models observations of phenomena (1 and 2).

The decision to identify the two observations with the same atom (one that mutated) can be modelled using the extension of FCA called Temporal Concept Analysis [14], which adds a time relation between the two observations (called "actual objects"). If one chooses instead to view the ^{4}He as a different atom (the product of the previous atom's decay), the time relation is deleted, but the formal contexts are *structurally identical*. This suggests, accurately, that the decision to identify the two observations with the same atom or with different atoms is somewhat arbitrary. It is a subjective interpretation of objective phenomena. In contrast, some modelling environments treat identity as a transcendental, non-qualitative property ('haecceity') and cannot accurately represent the relationships among alternate identities for the same object.

That the appropriateness of any given selection of identity criteria is relative to the application is easily shown using the classic 'ship of Theseus' example. A ship that has been repaired and restored is the same ship as far as navigation is concerned, but it is not the same from the forensic perspective, e.g., in an investigation of whether the ship was in compliance with regulations when an incident occurred. In the latter application, the ships "before the repair" and "after the repair" are treated as different objects.

Table 1. Formal context for ^{5}He decay

	t_1	t_2	2 protons	2 neutrons	3 neutrons
1	×		×		×
2		×	×	×	

4 Unity

Unity relates to the philosophical notion of boundaries: one can define a thing by selecting spatial and temporal boundaries, deciding what is part of it and what is not. When formalized, the process is analogous to the example in the previous section: one subjectively picks objects out of a soup of observations.

It is often if not always the case that the spatial and/or temporal boundaries of a thing as people conceive of it are *vague*. Any precise model of such boundaries, including one using FCA, necessarily adds arbitrary and application-specific constraints. However, one can use conceptual scaling to minimize the impact. For example, again considering the process of radioactive decay, it is unlikely that the precise instant at which an atom decayed (or the instants at which the process of decay began and ended) would be known. However, it would be known with some certainty that it had not yet decayed at time t_1, and that it had already decayed at time t_2. With appropriate scaling, one need only consider those time granules for which precise knowledge is available.

5 Possibility

Possible things is a way of saying "things that might actually exist, but that we do not *know* exist;" alternately, "things that could potentially exist someday, but that do not exist now."

The uncertainty about possible things is epistemic in nature. Any hypothesized class that does not intend a logical contradiction could possibly have instances. This uncertainty is not directly modelled by FCA extensions that apply possibility theory and deal with missing/unknown values (e.g., Ref. [1]) because the objects themselves are hypothetical and thus entirely missing from the data.

One way to model possible things in FCA is to represent existential assertions *about* hypothetical objects as objects, and then use attribute scaling to capture all of the modality of the truth values of those statements. Figure 2 shows the concept lattice for an epistemic modal logic that enables an accurate description of the state of knowledge or belief about an assertion. The names of the attributes have been prefixed by "know" to emphasize that a statement can be true without one knowing that it is true, but not vice-versa.

A context that does not support *unknown* clearly leaves the modeller with little choice other than to make invalid substitutions. But too many *unknown*s results in a vacuous model—all things are possible; nothing can ever be ruled out. The treatment of possible things thus becomes a compromise between the desire for a generally valid model and the desire for a model that is constrained enough to enable the application for which it was built.

6 Tense

Since it is the goal of modellers to capture knowledge of lasting value, the question of how to model the past and future as distinct from the present is often ignored. The resulting model is timeless in the sense of having no concept of time whatsoever. Things simply are as they are, unchanging; or, if things do change, the result is a *different* model. There is no formal connection between

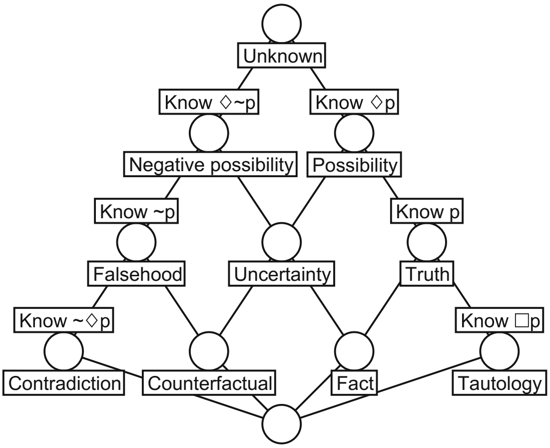

Fig. 2. Concept lattice for epistemic modal logic

the old and the new. This suffices if, in the application of interest, no contradictions result from instantiating all things past, present, and future as if they were contemporaries. However, it does not suffice for any application that needs to deal with change.

Those modellers who do model time choose to structure it in different ways. Tense logic structures time in terms of past, present, and future (a.k.a. the A-series or the tenser approach) [4]. UML sequence diagrams structure time in terms of earlier and later (a.k.a. the B-series or the detenser approach). Process Specification Language (PSL) [12] structures time in terms of reified time points (a.k.a. the four-dimensional approach). In FCA, these different structures of time can be integrated through appropriate scaling of attributes that indicate the times at which an object existed.

7 Realism

Classifying things is a process of abstraction, but some reject the claim that classes are abstractions of a higher level than the things classified.

Some modelling architectures segregate levels of abstraction (*fixed architecture*); others do not (*flat architecture*). The desirability of strict separation is a topic of debate. There are intuitively attractive notions that cannot be rendered faithfully using a fixed architecture. For example, consider the class that is called *class*. Intuitively, *class* is an instance of itself. But proponents of fixed architecture argue that it is confused thinking to identify any instance of a class with the class itself, and that doing so produces a model that has no sensible interpretation by man or machine [8,10], or at best an unconventional interpretation that does not integrate readily with conventional logic [11]. In any event, it certainly invalidates the set-theoretic interpretation of classification.

The decision to use a fixed architecture or not is a technical one, influenced by the relative expedience of expressing the concepts needed to serve particular applications. However, the barriers to translation between fixed and flat architectures are significant. In FCA, the concept whose intent consists of all attributes of an object g and whose extent consists of all objects having those attributes is called the *object concept* of the object g and is denoted by γg [5, Definition 22]. Formally, g and γg are not even comparable; one is an object, the other is a concept. FCA formalisms thus are best suited for a fixed architecture representation of knowledge, but this introduces constraints that some see as arbitrary.

8 Intent

In a static universe, extensional definitions are sufficient. If two concepts have the same extent, then they are interchangeable within the scope of the static universe and there is no value in distinguishing them. The value of intent is in making statements regarding possible and future individuals. Whether or not it is necessary to make statements about possible and future individuals or to distinguish concepts having the same extents is clearly application-specific, as is the selection of essential properties for intensional definitions.

FCA reduces the intensional/extensional dichotomy to a mathematical extreme, defining formal concepts in such a way that intensional definitions (in terms of necessary and sufficient attributes) and extensional definitions (in terms of objects) entail one other via a formal mapping. The intent that is derived from available data is forced to change when the data do. In knowledge representation generally, intensional definitions are invalidated when some possible individual that breaks the assumptions of the modeller becomes actual or becomes known. This problem is not intractable; it can be avoided by sacrificing the ability to map between intent and extent [2].

9 Conclusion

This article explored seven facets of relativity that would impact the use of FCA formalisms to represent knowledge, demonstrating that the capture of application-specific constraints is inextricable from the modelling process.

The semantic differences between models built for different applications can also be modelled and analyzed using FCA. By analyzing those differences, one can formally determine whether the applications are sufficiently compatible at the conceptual level to enable integration. That next step is explored in the unabridged technical report [3].

Acknowledgments. Thanks to all whose reviews and suggestions have improved this article, including Edward Barkmeyer, Peter Becker, Joachim Hereth Correia, Steven Fenves, Michael Grüninger, and the reviewers for the special session on knowledge discovery with FCA and related formalisms (FCA4KD++).

References

1. Ait-Yakoub, Z., Djouadi, Y., Dubois, D., Prade, H.: From a possibility theory view of Formal Concept Analysis to the possibilistic handling of incomplete and uncertain contexts. In: 5th International Workshop "What can FCA do for Artificial Intelligence?" August 2016. http://oatao.univ-toulouse.fr/17213/
2. Flater, D.: A logical model of conceptual integrity in data integration. J. Res. Natl. Inst. Stand. Technol. **108**(5), 395–402 (2003). doi:10.6028/jres.108.034
3. Flater, D.: Relativity of explicit conceptual models. NIST IR 7148, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, July 2004. doi:10.6028/NIST.IR.7148
4. Galton, A.: Temporal logic. In: Zalta, E.N. (ed.) Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, winter 2003 edn., December 2003. https://plato.stanford.edu/archives/win2003/entries/logic-temporal/
5. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Heidelberg (1999)
6. Guarino, N., Welty, C.: Evaluating ontological decisions with OntoClean. Commun. ACM **45**(2), 61–65 (2002)
7. Guarino, N., Welty, C.: Identity and subsumption. In: Green, R., Bean, C.A., Myaeng, S.H. (eds.) The Semantics of Relationships: An Interdisciplinary Perspective. Information Science and Knowledge Management, pp. 111–126. Springer, Dordrecht (2002). doi:10.1007/978-94-017-0073-3_7
8. Korzybski, A.: Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics, 5th edn. Institute of General Semantics, Englewood (1994)
9. OWL 2 Web Ontology Language document overview, 2nd edn. W3C recommendation, December 2012. https://www.w3.org/TR/owl2-overview/
10. Pan, J.Z., Horrocks, I.: Metamodeling architecture of web ontology languages. In: Proceedings of the Semantic Web Working Symposium, pp. 131–149, July 2001
11. Pan, J.Z., Horrocks, I.: RDFS(FA) and RDF MT: two semantics for RDFS. In: Fensel, D., Sycara, K., Mylopoulos, J. (eds.) ISWC 2003. LNCS, vol. 2870, pp. 30–46. Springer, Heidelberg (2003). doi:10.1007/978-3-540-39718-2_3
12. PSL Core Ontology (2004). http://www.mel.nist.gov/psl/psl-ontology/pslcore_page.html
13. Unified Modeling Language specification, version 2.5. OMG document formal/2015-03-01, Object Management Group, June 2015. http://www.omg.org/spec/UML/2.5/
14. Wolff, K.E., Yameogo, W.: Time dimension, objects, and life tracks. A conceptual analysis. In: Ganter, B., de Moor, A., Lex, W. (eds.) ICCS-ConceptStruct 2003. LNCS, vol. 2746, pp. 188–200. Springer, Heidelberg (2003). doi:10.1007/978-3-540-45091-7_13