

Default Physical Measurements in SUMO

Francesca Quattri

The Hong Kong Polytechnic University
Hong Kong

francesca.quattri@connect.polyu.hk

Adam Pease*

adam.pease@articulatesoftware.com

John P. M^cCrae

Universität Bielefeld
Germany

jmccrae@cit-ec.uni-bielefeld.de

Abstract

The following paper presents a further extension of the Suggested Upper Merged Ontology (SUMO), i.e. the development of default physical measurements for most of its classes (Artifacts, Devices, Objects) and respective children. The extension represents an arbitrary, computable and reproducible approximation of defaults for upper and middle-level concepts. The paper illustrates advantages of such extension, challenges encountered during the compilation, related work and future research.

1 Introduction

Over the last fourteen years SUMO (Pease, 2011; Niles and Pease, 2001) has been developed into a large, general-domain ontology, which currently¹ includes 20,000 terms and 80,000 axioms stated in higher-order logic (Pease and Schulz, 2014). SUMO provides an open source environment for the development of logical theories called SIGMA (Pease, 2011; Pease, 2003b). This enables the manipulation of different formal languages (including TPTP and OWL), (Adam Pease and Sams, 2003; Pease, 2003a). Among them, the logical formal language SUO-KIF has been selected for the development of knowledge-based (or KB) terms, through which SUMO can be searched. Another possible search of terms in SUMO is via the Princeton WordNet ®, to which the ontology has been fully mapped (Pease and Niles, 2003; Pease and Li, 2003; Pease and Murray, 2003).

In the first part of this paper, after introducing SUMO in generic terms, we explain the motivation behind the undergone extension of 300+ physical default measurements (the term ‘default’ is hereby used as synonym for ‘approximation’ or ‘estimation’). The second part deals with the advantages and issues encountered during the compilation of the defaults, and presents some practical examples of defaults and higher-order annotation. Related research and future work follow.

2 Default physical measurements in SUMO

The original intent behind the development of default physical measurements in SUMO is to provide factual peer-reviewed information about physical measurements of ontological classes. Almost all approximations of the default values have been established with reference to current ISO standards or norms set by governmental regulations. Only in the case that standard values are not provided or could not be retrieved, the compiler of the defaults has relied on personal judgment. In both cases, all defaults have been manually double-checked for validity by the compiler and the SUMO developer.

SUMO seems to be one of the first general-knowledge ontologies to provide extensive information on physical default measurements. Other data bases like DBpedia have (according to the authors’ knowledge) just recently started to provide a similar kind of information.² The physical defaults represent a big repository of approximated values based on physical properties, such as length, volume, size, width and

Same affiliation of the first author.

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Page numbers and proceedings footer are added by the organisers. Licence details: <http://creativecommons.org/licenses/by/4.0/>

¹As for the year 2014.

²<http://dbpedia.org/property/reference>

height. The approximation, as the term itself says, is partly arbitrary, computable, and comprehensively conducted. The measurements are formalized in minimum and maximum default values. The wording ‘maximum’ and ‘minimum’ should not be treated as the highest and lowest values attached to the respective *Artifacts*, but as some high or low values these *Entities* can own.

2.1 Advantages

We believe that the compilation of the defaults accomplishes three major advantages in the current format of the SUMO ontology:

1. Ontological formalization.
2. Objective adjustable values of physical properties.
3. Computable reproducible estimations of physical values.

Point (1) mirrors SUMO attempt as extensive ontology of general knowledge. Natural language fails in providing specificity for every single word and predicate, partly due to polysemy, synonymy as well as objective limitations of extensive precise formal description. We often refer to a term in vague sense and meaning, such as in the case of ‘car’ or ‘truck’. For further specification of the same, we tend to create new lemmas, derivatives and compounds. SUMO underscores a lemma in its definitional and ontological extent and the defaults consider the lemma as prototypical. As for its definitional extent, SUMO provides the definition of the lemma as reported in the Princeton WordNet®. As ontological and prototypical entity, the lemma is described in first-order and/or higher-order logic and thus transformed into a SUO-KIF KB term. In the case of *Truck*, the term is enlisted under *TransportationDevice* in SUMO. The following description in first-order logic (containing the quantifier ‘exists’) specifically states: “If a *TransportationDevice* is an *instance* of a *Truck*, then there exists a *kind of Object* such that a *kind of Object* is a *subclass* of *Object* and *kind of Object* is a *Cargo type* of *TransportationDevice*.”

```
(=>
  (instance ?T Truck)
  (exists (?L)
    (and
      (subclass ?L Object)
      (cargoType ?T ?L))))
```

Figure 1.: Example of a first-order axiom in SUO-KIF

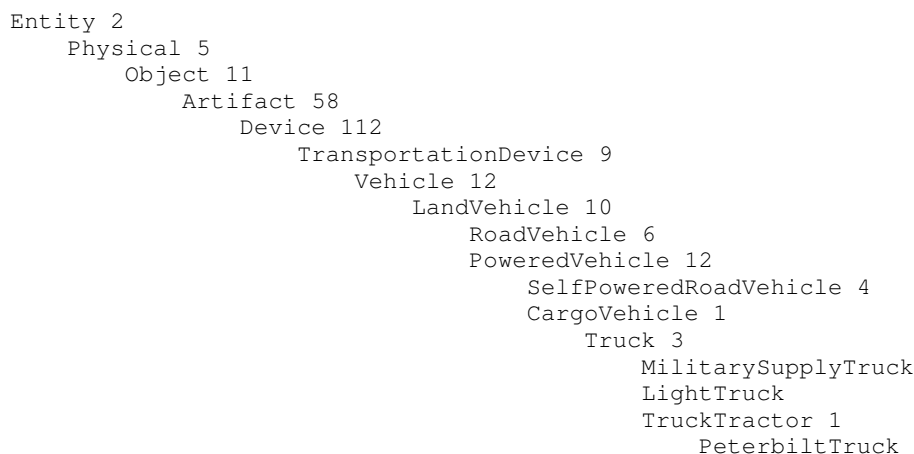


Figure 2.: Graphic documentation of the relation *subclass* for the term *Truck* with relative enumerated *direct-children*

Fig. 2 represents a graphic documentation of the same term as taxonomically listed (with the selected levels “above” and “below” *Truck* set to the value 10). The graph can be further extended to more levels, thus enabling a comprehensive look of all the branches that depart from the upper concept *Entity*. As for fig. 2., SUMO provides a specific taxonomy of the different kinds of *Truck*.

The default measurements in fig. 3 have been partly set by looking at standard measures for the same Artifact³:

```
;; Truck
(defaultMinimumLength Truck (MeasureFn 39 Foot))
(defaultMaximumLength Truck (MeasureFn Fn 49 Foot))
(defaultMinimumHeight Truck (MeasureFn 13 Foot))
(defaultMaximumHeight Truck (MeasureFn 15 Foot))
(defaultMinimumWidth Truck (MeasureFn 8.4 Foot))
(defaultMaximumWidth Truck (MeasureFn 9 Foot))

;; Vehicle
(defaultMinimumLength Vehicle (MeasureFn 13.5 Foot))
(defaultMaximumLength Vehicle (MeasureFn 14 Foot))
(defaultMinimumHeight Vehicle (MeasureFn 4.6 Foot))
(defaultMaximumHeight Vehicle (MeasureFn 4.8 Foot))
(defaultMinimumWeight Vehicle (MeasureFn 1 TonMass))
(defaultMaximumWeight Vehicle (MeasureFn 1.7 TonMass))
```

Figure 3.: Extensions of physical defaults for Truck and Vehicle

In fig. 4 the physical default values for CreditCard have been established according to the international standard ISO/IEC 7810:2003.

```
;; CreditCard
(defaultMinimumLength CreditCard (MeasureFn 3.4 Inch))
(defaultMaximumLength CreditCard (MeasureFn 3.4 Inch))
(defaultMinimumHeight CreditCard (MeasureFn 2.1 Inch))
(defaultMaximumHeight CreditCard (MeasureFn 2.1 Inch))
```

Figure 4.: Extensions of physical defaults for CreditCard

The (2) advantage in having physical default measurements is the objectivity of the properties they are calculated upon. The defaults are set on objectively comparable properties, such as height, volume, weight, length and width. These are all features of size and mass that can be counted and approximated, with different units of measures.

Finally, the (3) advantage that we reckon exists in having the defaults is their computability. Despite being relative and partially arbitrary measures⁴, the defaults are adjustable and reproducible, which makes them adaptable to representation models, peer-review and further estimations. We believe that this way of calculating defaults of physical Objects is certainly more reliable than other attempted methods (e. g. (Bennett, 2001):117-118).⁵

2.2 Issues encountered during the research

Some challenges were encountered during the compilation of default measurements.

- The defaults cover classes of upper concepts in SUMO, and part of their children, but not the predicates that can possibly collocate with them. For example, concepts like Aircraft or Helicopter are covered in SUMO, but not expressions like ‘light aircraft’ or ‘civilian helicopter’. Sometimes, SUMO already provides a logical description of these adjectives as incorporated in the concept itself, as in the case of MilitaryAircraft, SelfPoweredRoadVehicle, or PrintedBook (fig. 4), meaning that rather than specifying the predicate, a new term is created. SUMO users should bare in mind that the Artifacts in SUMO always aim at representing a

³As in the case of Truck the defaults have been established by looking at the standard sizes as set by the U.S. Department of Transportation and Federal Highway Administration, <http://www.ops.fhwa.dot.gov/freight/sw/index.htm>

⁴As previously discussed in the paper, the defaults have been assigned on a subjective basis in case standard defaults could not be retrieved/are not available. Also, the defaults sometimes apply to one country’s regulations, and are therefore not internationally valid. Finally, the defaults have been given with selected units of measures (e. g. inches instead of centimeters, or pounds instead of kilograms. This specified, one should bear in mind the intention of the default extensions, namely to provide an approximation of prototypical, not universal Artifacts.

⁵Bennett, in his study on physical objects and geographic concepts, tries to delimit the boundaries of vague entities by providing answers to size-related questions (e. g. “How large an area must a forest occupy? Are there any constraints in its shape? Must it be maximal or could it share a border with another region of forest?”). In SUMO, we believe that the defaults, through which some of these questions can be answered, are more reliable, since anchored to standard values.

prototypical form of the same Object, i. e. a kind that is possibly shared in the collective thinking. The representation for Book as showed below aims therefore at representing the possibly most commonly form of Book known, namely a printed and not an electronic version of the same.

```
(=>
  (instance ?BOOK PrintedBook)
  (exists (?SHEET1 ?SHEET2)
    (and
      (component ?SHEET1 ?BOOK)
      (component ?SHEET2 ?BOOK)
      (instance ?SHEET1 PrintedSheet)
      (instance ?SHEET2 PrintedSheet)
      (not
        (equal ?SHEET1 ?SHEET2))))))
(=>
  (and
    (instance ?ARTICLE1 Article)
    (instance ?BOOK Book)
    (subsumesContentInstance ?BOOK ?
      ARTICLE1))
    (exists (?ARTICLE2)
      (and
        (instance ?ARTICLE2 Article)
        (not
          (equal ?ARTICLE2 ?ARTICLE1))
          (subsumesContentInstance ?
            BOOK ?ARTICLE2))))))
```

Figure 5.: Comparison between the logical annotation for Book in SUMO with the collocational unit `printed + Book`

It needs to be specified that the concept of `Attribute` in SUMO is differently interpreted from the concept of predicate or adjective in natural language. Attributes in the Upper Merged Ontology are instances of upper classes, but there also exists classes of Attributes. The `Attribute` class can contain subclasses (e. g. `Female`, `Male`, `BiologicalAttribute`), but these have not been assigned default physical values. The motivation is basically that we cannot numerically define *abstracta*, such as gender, color, or emotions and feelings. In the case of abstract concepts, such as `StockMarket` or `InterestRate`, we have tried to figure out these, where possible, as physical objects (e. g. the place where financial transactions take place, or the sheet where rates are printed on).

Other sort of literally definable attributes (including comparative forms) are included in SUMO in the form of relations, which express, *inter alia*, equations and inequalities (`greaterThan`, `smallerThan`, `larger`, `earlier`, `interiorPart`, `temporalPart`, (Pease, 2011):113). Finally, what is defined in SUMO as `PhysicalAttribute` should not be confused with the physical default values added to the ontology. Instances of this class include `Compliance`, `Conductivity`, `Flammable`, `Inductance`, `MutualInductance`, `Resistivity`, `Stiffness`.

Despite the lack of a comprehensive cover of linguistically definable collocational compounds in SUMO (as above mentioned), we estimate that it is not impossible to approximate values for them, given the existence of defaults for the concept that carries the predicate. For instance, it can be derived that `BigHouse` (not enlisted in SUMO) is something that can be 1.9 times bigger than a `Studio`, or 0.1 times smaller than a `Mansion`, once the standard values for `House`, `Studio` and `Mansion` are given.

Given a partial ordering of gradable adjectives⁶ that apply to a particular noun, we could create axioms (thus inducing a productive process) which would then partition the physical space with respect to that particular adjective. The fact that we have axioms would eventually release us from defining defaults for each class. In other words, the most frequently an adjective collocates with a class or a subclass, the higher is the chance to develop an axiom(s) that enables us to calculate the defaults for these same classes automatically.

- SUMO provides ontological information regarding concepts in their a-contextual and unidiomatic form. SUMO terms are not polysemous, therefore there is no notion of reusing a term to mean something else. This also means that specific cases of use for a term in specific ontologies, or as applied to metaphorical/idiomatic expressions, are not taken into account (e.g. ‘turning tables’; ‘cleared table’). Instead, we specialize terms via subclassing and adding axioms on the subclasses term when a new term is needed for a specific domain.

⁶As interpreted by: (de Melo and Bansal, 2013; Schulam and Fellbaum, 2010; de Melo, 2014a; de Melo, 2014b).

- The defaults are based on arbitrary subjective approximations of prototypes. The provided information has been carefully peer-reviewed and the defaults can be used, re-used, or changed according to the user's needs. The intent is in fact to provide a basic estimation of the physical values for that concept. Furthermore, we have used specific units of measurements to carry on the approximations (e. g. inches versus centimeters, tons and pounds versus kilos). We acknowledge that this might hinder or slow down the reusability process.

3 Practical applications of defaults in linguistic disambiguation

Since the development of the first several hundreds physical default measurements, their applicability and usefulness has been tested in two research studies.⁷ The defaults have proven helpful in linguistic analysis, particularly in the disambiguation of vague terms, such as vague predicates and concepts, as well as more complex linguistic forms, such as similes and metaphors. The advantage of having physical defaults based on standards and norms has given further validity to the disambiguation process.

3.1 Default measurements and adjectives (*lemonOILS* and *SUMO*)

The use of first order logic seems to break in the case of adjectives. In a recent research, we therefore make an in-depth analysis of different kinds of attributes and how they can be represented in different ontology-lexicon interfaces (*lemonOILS* and *SUMO*), and discuss the implications of the modelling with application to ontology-based question answering.

3.2 Default measurements, metaphors and similes

In another current study (see previous footnote), we use default physical measurements to disambiguate similes from metaphors. Starting from the claim that the taught difference between metaphors and similes in terms of which has or does not have 'like' or 'as' in its form is not a linguistically and cognitively satisfactory statement, we design a computable model to test the validity of novel metaphors and similes and use the physical default measurements for our purpose.

4 Future work

The extension of physical default measurements in *SUMO* is not intended to be the last of its kind. In our future work, we plan for instance a better specification of dimensionality. During the compilation of the physical defaults, we have in fact sometimes encountered the challenge of defining first the geometrical property proper of the concept. For instance, taking a *Leaf*, do we usually refer to its length, or to its height? Google can help to a certain extent in cases like this. A better disambiguation of contextually dependent measurements (length versus height, or width versus length) is therefore needed. A further improvement includes the compilation of mostly all subclasses and their children in higher-order logic as KB terms, as well as the assignment to them of physical defaults. To enable an automatic productive process in the generation of automatic axioms (as mentioned in 2.2), both with respect to collocational forms and with regards to the similar physical defaults that may exist between parent and child, we still need to evaluate whether there should exist a mechanism for conflict resolution or overwrite. If we take for instance the example of *Snake*, we consider at the moment that this instance of *Reptile* most probably can inherit some of the properties of the parent, and viceversa. As showed in 2.1 (fig. 3) above though, this derivation does not seem so obvious or even applicable, since there might be prototypical properties that might appear for one concept, but not for the other, or given the too high discrepancy of measurements.

Finally, once this comprehensive framework of properties and intuitive specification of defaults has been created, we could conduct psycholinguistic empirical experiments to determine what are the defaults and prototypes and examples that different classes of human beings hold to be true. This could give us indication on how and if prototypicality overlays with dimensionality.

⁷Submitted accepted papers for the CogALex Workshop, COLING 2014, Dublin, Ireland and the CCLCC Workshop at ESSLI 2014, Tuebingen, Germany.

5 Conclusion

In this paper we present a current extension of the general-domain ontology SUMO, i. e. the compilation of default physical measurements for 300+ classes and subclasses. The aim of this extension is to provide a peer-reviewed reliable, reusable and reproducible estimation of physical values for the ontology. The defaults have already proven to be helpful in the disambiguation of vague predicates and concepts, as well as similes and metaphors. As open-source application, constantly updated and improved, it is planned to apply further changes to the SUMO ontology, which include an even more comprehensive development of physical defaults, as well as the inclusion of other defaults for other properties. Despite their approximation, the defaults represent a computational ground for representation models and further calculations.

References

- W.R. Murray Adam Pease and Michael Sams. 2003. Applying formal methods and representations in a natural language tutor to teach tactical reasoning. In *Proceedings of the 11th International Conference on Artificial Intelligence in Education (AIED) Conference*, pages 349–356. IOS Publications.
- Brandon Bennett. 2001. Application of supervaluation semantics to vaguely defined concepts. In Daniel R. Montello, editor, *Proceedings of the 5th International Conference on Spatial Information Theory (COSIT'01)*, number 2205 in LNCS, pages 108–123, Morro Bay. Springer.
- Gerard de Melo and Mohit Bansal. 2013. Good, great, excellent: Global inference of semantic intensities. *Transactions of the Association for Computational Linguistics*, 1:279–290.
- Gerard de Melo. 2014a. From linked data to tightly integrated data. LREC 2014 Workshop on Linked Data in Linguistics (LDL-2014). Invited speaker.
- Gerard de Melo. 2014b. Link prediction in semantic knowledge graphs. The Hong Kong Polytechnic University, March. Invited speaker.
- Ian Niles and Adam Pease. 2001. Towards a standard upper ontology. In Christopher A. Welty and Barry Smith, editors, *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS 2001)*.
- Adam Pease and John Li. 2003. Agent-mediated knowledge engineering collaboration. In *Proceedings of the AAAI 2003 Spring Symposium on Agent-Mediated Knowledge Management*.
- Adam Pease and W.R. Murray. 2003. An english to logic translator for ontology-based knowledge representation languages. In *Proceedings of the 2003 IEEE International Conference on Natural Language Processing and Knowledge Engineering*, pages 777–783.
- Adam Pease and Ian Niles. 2003. Linking lexicons and ontologies: Mapping wordnet to the suggested upper merged ontology. In *Proceedings of the IEEE International Conference on Information and Knowledge Engineering*, pages 412–416.
- Adam Pease and Stephan Schulz. 2014. Knowledge engineering for large ontologies with sigmakee 3.0. Submitted accepted version for journal paper.
- Adam Pease. 2003a. Mapping linguistic elements to logical expressions. In *Workshop on Ontological Knowledge and Linguistic Coding at the 25th Annual Meeting of the German Linguistics Society (Deutsche Gesellschaft für Sprachwissenschaft)*.
- Adam Pease. 2003b. The sigma ontology development environment. In *Working Notes of the IJCAI-2003 Workshop on Ontology and Distributed Systems*, volume 71.
- Adam Pease. 2011. *Ontology: A Practical Guide*. Articulate Software Press, Angwin, CA.
- Peter F. Schulam and Christiane Fellbaum. 2010. Automatically determining the semantic gradation of german adjectives. In *Proceedings of KONVENS*.