Making Common Sense of the Leuven Concept Database

Guendalina Righetti[1], Oliver Kutz[1], and Daniele Porello[2]

[1] Free University of Bozen-Bolzano, Bolzano, Italy
[2] University of Genoa, Genoa, Italy

The word "common-sense" groups different aspects of human knowledge which permeate our experience of the world and allow us to move therein. Common-sense knowledge includes our ability to distinguish between single objects and classes of objects, to distinguish between animate and inanimate things, but also more mundane knowledge: the fact that fish live only in water, or the fact that my dad is necessarily older than me. Common-sense knowledge is acquired by humans through experience and throughout life almost imperceptibly, and in an almost completely effortless way. Despite the long tradition of research (9; 8) investigating how to bring this kind of knowledge from human to machine, it is still a wide-open research question. At the same time, any progress in this field directly benefits a number of AI applications. Many envisaged practical applications, in particular, require complex inferences, which, in turn, require large common-sense knowledge bases. In practice, this need has often resulted in the use of structured lexical databases (3), semantic networks (16), or linked data (1), as a link between natural language and higher level semantic representations. Nevertheless, these repositories often show some level of ambiguity, which demonstrates the lack of a common agreement on the meaning of the lexical entries. In order to overcome this difficulty, a number of works have recently provided these databases with deeper semantic support (15; 6; 14; 5; 11). The key ideas behind those approaches is to make these repositories "ontology-like", as far as possible. Crucially, these works often use a top-down approach which propagates certain top level distinctions of a foundational ontology onto the more general entries in the database at hand, exploiting its given internal schema structure, general relations, etc.

We explore here a different direction, based on a detailed case study of the Leuven Concept Database (2). Starting from a statistical representation of concepts grounded in psychological data, we analyse the difficulties encountered in a formalisation process based on typical logical languages. The Leuven Database contains information, gathered by a group of psychologist at the University of Leuven, over features exhibited by 15 concepts, and provides evidence on human conceptualisation. The conceptualisations that emerge from the Leuven Database do not necessarily reflect a *good*, *normative* definition of the concepts involved: they aim at being good descriptions of what people have in mind when they think of those concepts, and of the meaning they associate with them. Therefore, the database is permeated by "common-sense information", and exhibits some of the basic ambiguities related to the use of natural language. For instance, some of the features of the concept *Fish* ("has gills", "lays eggs", "lives in the sea") relate to a somehow biological perspective on the concept. Other features describe instead the concept *Fish* in its relation to humans being—and maybe w.r.t. some subject's personal experience: some of them ("swims in aquarium", "is sometimes kept as a pet") focus on the 'pet dimension' of *Fish*, while others ("contains omega3", "is tasty") relate to the 'food dimension'. Also, features pertaining to different dimensions may be considered conflicting—at least at some level: does the fish live in the sea or in the aquarium? Is it a pet or is it tasty? Similar considerations apply to all the concepts in the database: a *Sport* is a hobby, is relaxing and is fun, but can also be a *Profession*, which in turn is defined as a source of stress and frustration (but also an activity which is advantageous for the society and the economy). *Clothes* protect against the cold, but can be a status symbol, protect from the rain but express your personality; a *Tool* is an aid, but you can injure someone with it; *Vehicles* are polluting, but they are environmentally friendly. This gives us an idea of the context sensitivity of everyday concepts (17), but reveals also their polysemy—the fact that those category labels are used as umbrella words for slightly different meanings (7). These conceptualisations, therefore, constitute an excellent point of observation on the challenges to be faced to make this information machine interpretable. We propose here a study which addresses exactly these difficulties.

In order to make the content of the Leuven Database machine comprehensible, a process of formalisation and translation into an appropriate logical language is needed. We exploit here standard Description Logic languages (such as OWL), as well as *weighted* Description Logic, which allows a more cognitively grounded modelling (10; 12). Being inspired by standard Prototype Theory (13), weighted DL allows to define (common-sense) concepts via weighted features, that need to add up to a given threshold, mimicking the behaviour of perceptrons (4). Computational, logic-based languages obviously impose some limitation in term of expressivity. The semantics of the subsumption relation, for instance, implies that all the elements of the sub-class are also elements of the super-class. Clearly, this is a quite strong requirement when we are dealing with natural language formalisations and everyday concepts. E.g., some features are described by people by means of expressions which emphasise their partial applicability to the class under consideration (e.g. sometimes, can be, etc). When this is not expressly stated, it is often implicit in the use of everyday language (e.g. birds "can fly", but this does not imply that a penguin is not a bird). To guide the translation, the features collected in the Leuven database are firstly grouped into different meta-categories, according to their grammar, and in terms of Aristotle's *square of opposition*, which enables to distinguish between the four categorical statements (*All* S are P, *Some* S are P, *No* S are P, *Some* S are *not* P). Another constraint in the translation is given through an alignment with a background foundational ontology, which is used as a skeleton to inform some of the formalisation choices in the process of translation. Besides the ambiguity of the language, a lot of "preliminary" knowledge is left implicit by subjects in the data analysed here. We argue, and present some examples, that setting a foundational ontology as background knowledge enables us to impose order and coherence (when possible) onto the information, helping also to disambiguate some of the hidden meaning within the data (e.g. the fact that birds "can fly" refers to an *ability* rather than a *modality*).

Bibliography

[1] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.G.: Dbpedia: A nucleus for a web of open data. In: The Semantic Web, 6th International Semantic Web Conference. vol. 4825. Springer (2007)

[2] De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M.J., Voorspoels, W., Storms, G.: Exemplar by feature applicability matrices and other dutch normative data for semantic concepts. Behavior Research Methods **40**(4), 1030–1048 (2008)

[3] Fellbaum, C.: Wordnet and wordnets. In: Brown, K. (ed.) Encyclopedia of Language and Linguistics. pp. 665–670. Elsevier, Oxford (2005)

[4] Galliani, P., Righetti, G., Kutz, O., Porello, D., Troquard, N.: Perceptron connectives in knowledge representation. In: Knowledge Engineering and Knowledge Management - EKAW 2020. vol. 12387. Springer (2020)

[5] Gangemi, A., Guarino, N., Masolo, C., Oltramari, A.: Sweetening wordnet with dolce. AI Magazine **24**(3), 13 (Sep 2003)

[6] Gangemi, A., Nuzzolese, A.G., Presutti, V., Draicchio, F., Musetti, A., Ciancarini, P.: Automatic Typing of DBpedia Entities. In: The Semantic Web – ISWC 2012. pp. 65–81. Springer Berlin Heidelberg (2012)

[7] Hofstadter, D.R.: Analogy as the core of cognition. The analogical mind: Perspectives from cognitive science pp. 499–538 (2001)

[8] Lenat, D.B.: Cyc: A large-scale investment in knowledge infrastructure. Commun. ACM **38**(11), 33–38 (Nov 1995)

[9] McCarthy, J.: Programs with common sense. In: Proceedings of the Teddington Conference on the Mechanization of Thought Processes. pp. 75–91. Her Majesty's Stationary Office, London (1959)

[10] Porello, D., Kutz, O., Righetti, G., Troquard, N., Galliani, P., Masolo, C.: A toothful of concepts: Towards a theory of weighted concept combination. In: Proc. of DL Workshop. vol. 2373. CEUR-WS.org (2019)

[11] Razniewski, S., Das, P.: Structured Knowledge: Have We Made Progress? An Extrinsic Study of KB Coverage over 19 Years. In: Proc. of the 29th ACM International Conference on Information & Knowledge Management. p. 3317–3320. CIKM '20, Association for Computing Machinery, New York, NY, USA (2020)

[12] Righetti, G., Porello, D., Kutz, O., Troquard, N., Masolo, C.: Pink panthers and toothless tigers: three problems in classification. In: Proc. 7th International Workshop on AI and Cognition. vol. 2483. CEUR-WS.org (2019)

[13] Rosch, E.H.: On the internal structure of perceptual and semantic categories. In: Moore, T.E. (ed.) Cognitive Development and Acquisition of Language, pp. 111–144. Academic Press, San Diego (1973)

[14] Schmidt, D., Pease, A., Trojahn, C., Vieira, R.: Aligning conference ontologies with SUMO: A report on manual alignment via wordnet. In: Proc. of the Joint Ontology Workshops 2019. vol. 2518. CEUR-WS.org (2019)

[15] Silva, V., Freitas, A., Handschuh, S.: Word tagging with foundational ontology classes: Extending the wordnet-dolce mapping to verbs. pp. 593–605 (11 2016)

[16] Speer, R., Chin, J., Havasi, C.: Conceptnet 5.5: An open multilingual graph of general knowledge. In: Proc. of the 31st AAAI Conference on AI. pp. 4444–4451. AAAI Press (2017)

[17] Yeh, W., Barsalou, L.W.: The situated nature of concepts. The American Journal of Psychology **119**(3), 349–384 (2006)