# Ontological analysis of the wikipedia category system

*Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia*

## Abstract

Copyright © 2018 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved We analyse violations of the transitivity principle of the Wikipedia category system, i.e. the situations where articles from a subcategory doesn't logically belong to its parent category. The causes of the violation have been analysed on the base of ontological modelling methodologies such as OntoClean. We propose a new approach to automatically eliminating the violations. This approach is based on analysis of the relation of ontological dependence between categories. As a theoretical foundation of such analysis we propose a new deflationistic interpretation of the essential account of ontological dependence. The proof of concept has been evaluated on the category C:Mathematics. We are going to apply the proposed approach to derive a new large-scale domains hierarchy from the Wikipedia category system, and use it to provide BabelNet and DBpedia with fine-grained domain annotations.

## Keywords

Categorization, OntoClean, Ontological dependence, Ontology validation, Wikipedia

## References

[1] ANSI/NISO Z39.19-2005 (R2010). Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies.

[2] Auer, S. et al. (2007). DBpedia: A Nucleus for a Web of Open Data. In Aberer K. et al. (Eds.), Proceedings of the 6th International Semantic Web Conference and the 2nd Asian Semantic Web Conference (ISWC 2007 + ASWC 2007). The Semantic Web. Lecture Notes in Computer Science, vol. 4825, pp. 722–735. Berlin, Heidelberg: Springer. doi:10.1007/978-3-540-76298-0_52.

[3] Bizer, C. et al. (2009). DBpedia: A Crystallization Point for the Web of Data. Journal of Web Semantics, 7 (3), 154–165. Elsevier. doi:10.1016/j.websem.2009.07.002.

[4] Correia, F. (2008). Ontological Dependence. Philosophy Compass, 3 (5), 1013–1032. Blackwell. doi:10.1111/j.1747-9991.2008.00170.x.

[5] de Melo, G., Weikum, G. (2010). MENTA: Inducing Multilingual Taxonomies from Wikipedia. In Proceedings of the 19th ACM international conference on Information and knowledge management (CIKM '10), pp. 1099–1108. New York: ACM. doi:10.1145/1871437.1871577.

[6] Elizarov, A.M. et al. (2014). Mathematical Knowledge Representation: Semantic Models and Formalisms. Lobachevskii Journal of Mathematics, 35(4), 348–354. Pleiades Publishing. doi:10.1134/S1995080214040143

[7] Elizarov, A. et al. (2016). Mathematical Knowledge Management: Ontological Models and Digital Technology. In Kalinichenko L. et al. (Eds.), Selected Papers of the XVIII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2016). CEUR Workshop Proceedings, vol. 1752, pp. 44–50.

[8] Elizarov, A. et al. (2017). Digital Ecosystem OntoMath: Mathematical Knowledge Analytics and Management. In Kalinichenko L. et al. (Eds.), Proceedings of the 28th International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2016). Communications in Computer and Information Science, vol. 706, pp. 33–46. Springer, Cham. doi:10.1007/978-3-319-57135-5_3.

[9] Fine, K. (1994). Essence and Modality. Philosophical Perspectives, 8, 1–16. Ridgeview Publishing Company. doi:10.2307/2214160.

[10] Fine, K. (1995). Ontological Dependence. Proceedings of the Aristotelian Society, 95, 269-290. Wiley.

[11] Flati, T. et al. (2014). Two Is Bigger (and Better) Than One: the Wikipedia Bitaxonomy Project. In Toutanova K., Wu H. (Eds.), Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), vol. 1, pp. 945–955. ACL. doi:10.3115/v1/P14-1089.

[12] Flati, T. et al. (2016). MultiWiBi: The multilingual Wikipedia bitaxonomy project. Artificial Intelligence, 241, 66–102. doi:10.1016/j.artint.2016.08.004.

[13] Galieva, A. et al. (2017). Toward Domain-Specific Russian-Tatar Thesaurus Construction. Proceedings of the International Conference IMS-2017, pp. 120–124. New York: ACM. doi:10.1145/3143699.3143716.

[14] Gangemi, A., Guarino, N., Oltramari, A. (2001). Conceptual Analysis of Lexical Taxonomies: The Case of WordNet Top-Level. In Welty C., Smith B. (Eds.), Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS 2001), pp. 285–296. New York: ACM. doi:10.1145/505168.505195.

[15] Gangemi, A. et al. (2002). Sweetening Ontologies with DOLCE. In Gómez-Pérez A., Benjamins V.R. (Eds.), Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2002). Lecture Notes in Computer Science, vol. 2473, pp. 166–181. Springer, Berlin, Heidelberg. Doi: 10.1007/3-540-45810-7_18.

[16] Gangemi, A. et al. (2012). Automatic Typing of DBpedia Entities. In Cudré-Mauroux P. et al. (Eds.), Proceedings of the 11th International Semantic Web Conference (ISWC 2012). Lecture Notes in Computer Science, vol. 7649, pp. 65–81. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-35176-1_5.

[17] Gella, S., Strapparava, C., Nastase, V. (2014). Mapping WordNet Domains, WordNet Topics and Wikipedia Categories to Generate Multilingual Domain Specific Resources. In Calzolari N. et al. (Eds.), Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014), pp. 1117–1121. ELRA.

[18] Guarino, N., Welty, C. (2000). A Formal Ontology of Properties. In Dieng R., Corby O. (Eds.), Proceedings of the 12th International Conference on Knowledge Engineering and Knowledge Management Methods, Models, and Tools (EKAW 2000). Lecture Notes in Computer Science, vol. 1937, pp. 97–112. Springer, Berlin, Heidelberg. doi:10.1007/3-540-39967-4_8.

[19] Guarino, N., Welty, C. (2009). An Overview of OntoClean. In Staab S., Studer R. (Eds.), Handbook on Ontologies (2nd ed., pp. 201–220). Springer, Berlin, Heidelberg. doi:10.1007/978-3-540-92673-3_9.

[20] Guizzardi, G. (2005). Ontological foundations for structural conceptual models. Enschede: CTIT.

[21] Hawke, P. (2017). Theories of Aboutness. Australasian Journal of Philosophy. Taylor & Francis. doi:10.1080/00048402.2017.1388826.

[22] Hjørland, B. (2016). Subject (of documents). In Hjørland B. (Ed.), Encyclopedia of Knowledge Organization. ISKO. Retrieved from: http://www.isko.org/cyclo/subject.

[23] Hjørland, B. (1992). The Concept of 'Subject' in Information Science. Journal of Documentation, 48 (2), 172–200. MCB UP Ltd.

[24] Hoffart, J. et al. (2013). YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. Artificial Intelligence, 194, 28–61. Elsevier. doi:10.1016/j.artint.2012.06.001.

[25] Hovy, E., Navigli, R., Ponzetto, S.P. (2013). Collaboratively built semi-structured content and Artificial Intelligence: The story so far. Artificial Intelligence, 194, 2–27. Elsevier. doi:10.1016/ j.artint.2012.10.002.

[26] ISO 25964-1:2011. Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval.

[27] Kirillovich, A. et al. (2017). RuThes Cloud: Towards a Multilevel Linguistic Linked Open Data Resource for Russian. In Różewski P., Lange C. (Eds.), Proceedings of the 8th International Conference on Knowledge Engineering and Semantic Web (KESW 2017). Communications in Computer and Information Science, vol. 786, pp. 38–52. Springer, Cham. doi:10.1007/978-3-319-69548-8_4.

[28] Koslicki, K. (2012). Varieties of Ontological Dependence. In Correia F., Schnieder B. (Eds.), Metaphysical Grounding: Understanding the Structure of Reality (pp. 186–213). Cambridge: Cambridge University Press.

[29] Koslicki, K. (2013). Ontological Dependence: An Opinionated Survey. In Hoeltje M. et al. (Eds.), Varieties of Dependence (pp. 31–64). Munich: Philosophia Verlag.

[30] Lehmann, J. et al. (2015). DBpedia: A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. Semantic Web Journal, 6 (2), 167–195. IOS Press. doi:10.3233/SW-140134.

[31] Loukachevitch, N. (2011). Thesauri in Information Retrieval Tasks. Moscow: Moscow University Press.

[32] Loukachevitch, N., Dobrov, B. (2004a). Development of Ontologies with Minimal Set of Conceptual Relations. In Lino, M.T. et al. (Eds.), Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04), pp. 1889-1892. ELRA.

[33] Loukachevitch, N., Dobrov, B. (2004b). Ontological Types of Associative Relations in Information-Retrieval Thesauri and Automatic Query Expansion. In Oltramari A. et al. (Eds.), Proceedings of the LREC Workshop on Ontologies and Lexical Resources in Distributed Environments (Ontolex 2004), pp. 24–29.

[34] Loukachevitch, N., Dobrov, B. (2014). RuThes Linguistic Ontology vs. Russian Wordnets. In Orav H., Fellbaum C., Vossen P. (Eds.), Proceedings of the 7th Conference on Global WordNet (GWC 2014), pp. 154–162. University of Tartu Press.

[35] Loukachevitch, N., Dobrov, B., Chetviorkin, I. (2014). RuThes-Lite, a Publicly Available Version of Thesauru of Russian Language RuThes. In Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue", pp. 340–349. Moscow: RGGU.

[36] Mahdisoltani, F., Biega, J., Suchanek, F.M. (2015). YAGO3: A Knowledge Base from Multilingual Wikipedias. In Proceedings of the 7th Biennial Conference on Innovative Data Systems Research (CIDR 2015).

[37] Mares, E.D. (2004). Relevant Logic: A Philosophical Interpretation. New York: Cambridge University Press.

[38] Medelyan, O. et al. (2009). Mining meaning from Wikipedia. International Journal of Human-Computer Studies, 67 (9), 716–754. Academic Press. doi:10.1016/j.ijhcs.2009.05.004.

[39] Moro, A., Navigli, R. (2012). WiSeNet: Building a Wikipedia-based Semantic Network with Ontologized Relations. In Proceedings of the 21st ACM international conference on Information and knowledge management (CIKM 2012), pp. 1672–1676. New York: ACM. doi:10.1145/2396761.2398495.

[40] Nastase, V. et al. (2010). WikiNet: A Very Large Scale Multi-Lingual Concept Network. In Calzolari N. et al. (Eds.), Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010), pp. 1015–1022.

[41] Nastase, V., Strube, M. (2008). Decoding Wikipedia Categories for Knowledge Acquisition. In Cohn A. (Ed.), Proceedings of the 23rd National Conference on Artificial Intelligence (AAAI 2008), vol. 2, pp. 1219–1224. AAAI Press

[42] Nastase, V., Strube, M. (2013). Transforming Wikipedia into a large scale multilingual concept network. Artificial Intelligence, 194, 62–85. Elsevier. doi:10.1016/j.artint.2012.06.008.

[43] Navigli, R., Ponzetto, S.P. (2010). BabelNet: Building a Very Large Multilingual Semantic Network. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010), pp. 216–225. ACL.

[44] Navigli, R., Ponzetto, S.P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence, 193, 217–250. Elsevier. doi:10.1016/j.artint.2012.07.001.

[45] Nevzorova, O. et al. (2014). OntoMath Ontology: a Linked Data Hub for Mathematics. In Klinov P., Mouromstev D. (Eds.), Proceedings of the 5th International Conference on Knowledge Engineering and Semantic Web (KESW 2014). Communications in Computer and Information Science, vol. 468, pp. 105–119. Springer, Heidelberg. doi:10.1007/978-3-319-11716-4_9.

[46] Niemann, E., Gurevych, I. (2011). The people's web meets linguistic knowledge: automatic sense alignment of Wikipedia and WordNet. In Proceedings of the 9th International Conference on Computational Semantics (IWCS 2011), pp. 205–214. Stroudsburg: ACL.

[47] Nuzzolese, A.G. et al. (2013). Towards the Natural Ontology of Wikipedia. In Blomqvist E., Groza T. (Eds.), Proceedings of the ISWC 2013 Posters & Demonstrations Track. CEUR Workshop Proceedings, vol. 1035, pp. 273–276.

[48] Ponzetto, S.P., Navigli, R. (2009). Large-Scale Taxonomy Mapping for Restructuring and Integrating Wikipedia. In Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009), pp. 2083–2088. San Francisco: Morgan Kaufmann Publishers.

[49] Ponzetto, S.P., Strube, M. (2007). Deriving a Large Scale Taxonomy from Wikipedia. In Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI 2007), pp. 1440–1445. AAAI Press.

[50] Ponzetto, S.P., Strube, M. (2011). Taxonomy induction based on a collaboratively built knowledge repository. Artificial Intelligence, 175 (9–10), 1737–1756. Elsevier. doi:10.1016/j.artint.2011.01.003.

[51] Ruiz-Casado, M., Alfonseca, E., Castells, P. (2005). Automatic Assignment of Wikipedia Encyclopedic Entries to WordNet Synsets. In Szczepaniak P.S. et al. (Eds.), Proceedings of the 3rd International Conference on Advances in Web Intelligence (AWIC 2005). Lecture Notes in Computer Science, vol. 3528, pp. 380–386. Springer, Berlin, Heidelberg. doi:10.1007/11495772_59.

[52] Simons, P. (1987). Parts: A Study in Ontology. Oxford: Clarendon Press. doi:10.1093/acprof:oso/9780199241460.001.0001

[53] Suchanek, F. M., Kasneci, G., Weikum, G. (2007). Yago: a Core of Semantic Knowledge. In Proceedings of the 16th international conference on World Wide Web (WWW 2007), pp. 697–706. New York, ACM. doi:10.1145/1242572.1242667.

[54] Tahko, T.E., Lowe, E.J. (2016). Ontological Dependence. In Zalta E.N. (Ed.), The Stanford Encyclopedia of Philosophy. Retrieved from: http://plato.stanford.edu/entries/dependence-ontological/.

[55] Thomasson, A.L. (1999). Fiction and Metaphysics. Cambridge: Cambridge University Press.

[56] Thomasson, A.L. (2008). Existence Questions. Philosophical Studies, 141(1), 63–78. Springer Netherlands. doi:10.1007/s11098-008-9263-8.

[57] Thomasson, A.L. (2009). Answerable and Unanswerable Questions. In Chalmers D.J. et al. (Eds.), Metametaphysics: New Essays on the Foundations of Ontology. New York: Oxford University Press.

[58] Titze, G. et al. (2014). DBpedia Domains: augmenting DBpedia with domain information. In Calzolari N. et al. (Eds.), Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014), pp. 1438–1442. ELRA.

[59] Toral, A., Muñoz, R., Monachini, M. (2008). Named Entity WordNet. In Proceedings of the 6th Conference on Language Resources and Evaluation (LREC 2008), pp. 741–747.

[60] Wu, F., Weld, D.S. (2008). Automatically Refining the Wikipedia Infobox Ontology. In Proceedings of the 17th international conference on World Wide Web (WWW 2008), pp. 635–644. New York: ACM. doi:10.1145/1367497.1367583.

[61] Yablo, S. (2014). Aboutness. Princeton: Princeton University Press.

[62] Zirn, C., Nastase, V., Strube, M. (2008). Distinguishing between Instances and Classes in the Wikipedia Taxonomy. In Zirn C. et al. (Eds.), Proceedings of the 5th European Semantic Web Conference (ESWC 2008). Lecture Notes in Computer Science, vol. 5021, pp. 376–387. Springer, Berlin, Heidelberg. doi:10.1007/978---540-68234-9_29.