Ontological Analysis of the Wikipedia Category System

Alexander Kirillovich[1] and Olga Nevzorova[1,2]
[1]*Kazan Federal University, Kazan, Russia*
[2]*Research Institute of Applied Semiotics of the Tatarstan Academy of Sciences, Kazan, Russia*

Keywords: Wikipedia, Categorization, Ontology Validation, OntoClean, Ontological Dependence.

Abstract: 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.

1 INTRODUCTION

Wikipedia is one of the largest knowledge bases on the Web. Wikipedia data are used in such tasks as word sense disambiguation, text categorization, calculation of semantic similarity, and machine translation. To automatically process information from Wikipedia, means of data structuring are needed.

The category system is the main meen for structuring information in Wikipedia. The categories specify the thematic classification of articles.

There are two types of categories:

- Set categories, for example `C:Cities` category that contains articles about various cities (*New York*, *Moscow*, *Kazan*, *Seville*, etc).
- Topic categories, for example `C:City` category that contains articles on city-related topics (*Urban planning*, *Urbanization*, *History of cities*, *Urban culture*, etc).

Each category can include subcategories and can be included in a parent category. Thus the category system is organized as an directed acyclic graph.

Categories can be grouped using meta-categories, for example `C:Writers` → `C:Writers_by_nationality` → `C:Russian_writers`.

Thus it is important that the category system demonstrate transitivity and nested subcategories be relevant to the parent category. However, building a chain of nested categories does not always satisfy this requirement. This fact causes problems while searching for articles. Therefore it is necessary to analyze the causes of violations of transitivity which leads to a discrepancy between the category and the article that is located in one of the nested subcategories.

We understand relevance as rating a category-subcategory pair in terms of preserving the volume of common and distinctive features. We assume that for relevant categories, the volume of common attributes exceeds the volume of differences, and in this sense one can speak of the transitivity property of categories.

However, the requirement of transitivity is often violated. For example, the category `C:Arithmetic` contains irrelevant subcategories such as `C:Colombian_people_executed_by_firing_squad` (see example 1).

Example 1: `C:Arithmetic` → `C:Ratios` → `C:Rates` → `C:Temporal_rates` → `C:Acceleration` → `C:Force` → `C:Motion_(physics)` → `C:Flight` → `C:Ballistics` → ... → `C:Projectile_weapons` → `C:Firearms` → `C:People_associated_with_firearms` → `C:Shooting_victims` → ... → `C:Colombian_people_executed_by_firing_squad`.

Kirillovich, A. and Nevzorova, O.
Ontological Analysis of the Wikipedia Category System.
DOI: 10.5220/0006961803580366
In *Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2018) - Volume 2: KEOD*, pages 358-366
ISBN: 978-989-758-330-8

The objective of this paper is to analyze the problem of transitivity violations in the Wikipedia category system and to propose an approach to solving it.

The paper is organized as follows. In Section 2, we briefly enumerate other projects of structuring Wikipedia data and note its limitations. In Section 3, we analyze Wikipedia category system on base of ontological modeling methodologies and reveal causes of transitivity violations. In Section 4, we propose an approach that corrects the category system by removing intransitive chains. Section 5 describes the direction of future work.

2 RELATED WORKS

There are many projects for extracting structured data from Wikipedia (Medelyan et al., 2009; Hovy et al., 2013), such as **DBpedia** (Auer et al., 2007; Bizer et al., 2009; Lehmann et al., 2015), **YAGO** (Suchanek et al., 2007; Hoffart et al., 2013; Mahdisoltani et al., 2015), **WikiTaxonomy** (Ponzetto and Strube, 2007; Ponzetto and Strube, 2011; Zirn et al., 2008), **WikiNet** (Nastase and Strube, 2008; Nastase et al., 2010; Nastase and Strube, 2013), **ORA** (Gangemi et al., 2012; Nuzzolese et al., 2013), **WiBi** (Flati et al., 2014; Flati et al., 2016), **MENTA** (de Melo and Weikum, 2010), **BabelNet** (Navigli and Ponzetto, 2010; Navigli and Ponzetto, 2012), **WiSiNet** (Moro and Navigli, 2012), **KOG** (Wu and Weld, 2008), as well as projects for **alignment of Wikipedia and WordNet** (Ruiz-Casado et al., 2005; Toral et al., 2008; Niemann and Gurevych, 2011; Ponzetto and Navigli, 2009; Gella et al., 2014; Titze et al., 2014).

However, these projects cannot completely substitute the Wikipedia category system in the case of subject retrieval, i.e. retrieval of articles on the particular subject (for example, all the articles related to *Ancient Greece*).

3 ANALYSIS OF TRANSITIVITY VIOLATIONS

We have analyzed the category system using the methodology of ontological modeling and have identified the causes of transitivity violations.

The category system can be considered as a thesaurus (ANSI, 2005; ISO 2011; Loukachevitch, 2011). In this case the categories will correspond to thesaurus concepts, and the relationships between the category and its subcategory will be viewed as standard ontological relationships.

Here are examples of such relations:

- Generic relation:
  - C:Cities_in_Europe → C:Capitals_in_Europe;
  - C:Software → ... → C:Operating_systems;
  - C:Mathematical_axioms → C:Axioms_of_set_theory;
  - C:Machines → C:Engines;
  - C:Wars → ... → C:Wars_involving_the_Soviet_Union;
  - C:Fiction_books → ... → C:Dystopian_novels.
- Instance relation:
  - C:Capitals_in_Europe → C:Moscow;
  - C:Intergovernmental_organizations → C:United_Nations;
  - C:Universities_and_colleges_in_Connecticut → C:Yale_University;
  - C:Operating_systems → C:Unix;
  - C:Fields_of_mathematics → C:Algebra;
  - C:Axioms_of_set_theory → C:Axiom_of_choice;
  - C:Abstract_strategy_games → C:Chess;
  - C:Engines → C:Internal_combustion_engine;
  - C:Wars_involving_the_Soviet_Union → C:World_War_II;
  - C:Dystopian_novels → C:Nineteen_Eighty-Four;
  - C:Organs → C:Brain;
  - C:Space_stations → C:International_Space_Station.
- Part-whole relation:
  - C:Moscow → C:Cities_and_towns_under_jurisdiction_of_Moscow → C:Zelenograd;
  - C:Yale_University → C:Yale_University_Library;
  - C:United_Nations → C:International_Atomic_Energy_Agency;
  - C:World_War_II → … → C:Attack_on_Pearl_Harbor;
  - C:Central_nervous_system → C:Brain;
  - C:Unix → C:Network_socket;

- C:Internal_combustion_engine → C:Pistons.

- Associative relation:
  - Science → object of study:
    - C:Botany → C:Plants.
  - Agent → counteragent:
    - C:Plants → C:Herbicides;
    - C:Violence → C:Nonviolence;
    - C:Communism → C:Anti-communism.
  - Value → measure instrument:
    - C:Temperature → C:Thermometers.
  - Activity → agent of activity:
    - C:Hunting → C:Hunting_dogs;
    - C:Military → C:Military_personnel.
  - Raw material → product:
    - C:Grape → C:Raisins;
    - C:Petroleum → C:Petroleum_products → Gasoline;
    - C:Textiles → C:Textile_arts → C:Weaving.
  - Other associative relations:
    - C:Death → C:Death_customs → C:Funerals;
    - C:Automobiles → C:Auto_racing;
    - C:Books → C:Book_arts → C:Bookbinding.

The meta-categories correspond to the Node labels.

The problem of ascribing relevant subcategories to a given category corresponds to the standard query expansion task.

Representing the category system in the form of a thesaurus allowed us to apply the methodologies for constructing information retrieval thesauri as well as the methodologies for verifying the ontology correctness, such as OntoClean (Guarino and Welty, 2009; Guarino and Welty, 2000; Guizzardi, 2005; Gangemi, A. et al., 2001; Gangemi, A. et al., 2002).

As a result, we determined that many cases of transitivity violations are caused by violations of rules for constructing a hierarchy of ontology concepts. The main reasons are:

- Incomplete inclusion of one category in another category:
  - *Analog film stock* is included in the *Digital technology* category: C:Digital_technology → C:Digital_media → C:Video → C:Film_and_video_technology → C:Film_stock. The reason is that the *Video* category is not fully included in the *Digital media* category, because there is also the analog video.
  - The English-language novel *Lolita* is included in the *Russian novels* category: C:Russian_novels → …→ C:Novels_by_Vladimir_Nabokov → C:Lolita. The same way, the *Novels by Vladimir Nabokov* category is not fully included in the *Russian novels* category.
  - *Japanese language* is included in *Korea languages*: C:Korea_languages → C:Buyeo_languages → C:Japonic_languages → C:Japanese_language.
- Errors in the application of fuzzy concepts:
  - *Electric chairs* are included in the *Consumer goods* category: C:Consumer_goods → C:Furniture → C:Chairs → C:Electric_chairs.
- Errors due to the application of ambiguous categories:
  - *Record charts* is included in the *Diagrams* category: C:Diagrams → C:Charts → C:Record_charts. In one case *Charts* is interpreted as diagrams and in other as music charts.
  - *Ship construction* is included in the *Real* estate category: C:Real_estate → C:Construction → C:Ship_construction.
- Using a concept in different senses:
  - Digital library *Lib.ru* is included in the *Buildings*: C:Buildings_and_structures → C:Buildings_and_structures_by_type → C:Libraries → C:Digital_libraries → Lib.ru. In one case *Libraries* is interpreted as buildings type, and in another as the social institute.
  - *Nihilism* is included in the *Biology*: C:Biology → C:Life → C:Philosophy_of_life → C:Nihilism. In one case *Life* is interpreted as a biological process, and in another as a social process.
  - *Snow* category is included in the *Liquids*: C:Liquids → C:Water → C:Forms_of_water → C:Snow. In one case *Water* is interpreted as a chemical substance

as such, and in another as a chemical substance in liquid state.

- Incompatible identity criteria:
  - *The Kaaba* (a Muslim sacred building) is included in the *Mathematical objects*: C:Mathematical_objects → C:Geometric_shapes → C:Elementary_shapes → C:Cubes → C:Cubic_buildings → C:Kaaba. The error is located in the chain: C:Cubes → C:Cubic_buildings. Cubic buildings, strictly speaking, are not cubes, since cubes and cubic buildings have different criteria for identity. A cube is an abstract, timeless and unchanging object. If the size of a cube changes, then it will be another cube. A cubic building is a concrete object that exists in time and space and retains identity under small modifications.
  - *Bermuda Triangle* is included in the *Geometric shapes*: C:Mathematical_objects → C:Geometric_shapes → C:Elementary_shapes → C:Triangles → Bermuda_Triangle.
- Confusion between concepts and signs:
  - *House of Habsburg* is included in *Words and phrases* category: C:Words_and_phrases → ... → C:Surnames_of_Swiss_origin → C:Swiss_families → C:Swiss_noble_families → C:House_of_Habsburg.
  - *Toxin* is included in the *Language* category: C:Language → C:Terminology → C:Biology_terminology → Toxin. The reason for the error is that the toxin is not a term. It is a word 'toxin' that is a term.
- OntoClean constraints vionations:
  - *Analgesic* is included in the *Illegal drugs* category: C:Illegal_drugs → C:Morphine → C:Analgesic. The reason for the error is that the *Illegal drugs* is not a type. It is a role, and it should not contain type categories.
  - *Optical bombsights* category is included in the *Office equipment* category: C:Office_equipment → C:Computers → … → C:Analog_computers → …→ C:Optical_bombsights.
  - Poison berry *Sambucus* is included in the *Foods* category: C:Foods → C:Fruit → C:Berries → C:Sambucus.

In the following cases, transitivity is violated not because of an error, but because of the very principle of organization of the Wikipedia categories system:

- Nontransitivity of the class-instance relation:
  - The *Santa María* ship is included in the *Ship types* category: C:Ship_types → ... → C:Exploration_ships → Santa_María_(ship).
  - The '*Ode'* poem is included in the *Literary genres* category: C:Literary_genres → C:Poetry → ... → Ode_(poem). This case is complicated by the real lexical homonymy (the ode as a literary genre).
- Nontransitivity of associative relation:
  - *Zelenograd*, a city under Moscow jurisdiction is included in the *Capitals in Europe* category: C:Capitals_in_Europe → C:Moscow → … → C:Zelenograd.
  - *Blondi the dog* is included in the *Nazi leaders* category: C:Nazi_leaders → C:Adolf_Hitler → Blondi.
  - Fictional *Galactic Empire* is included in the *Northern American countries* category: C:Northern_American_countries → C:United_States → C:American_people → … → C:George_Lucas → C:Star_Wars → … → Galactic_Empire_(Star_Wars).
  - *Languages of Djibouti* is included in the *Statistics* category: C:Statistics → C:Statistical_data_sets → C:Demographics_by_country → C:Demographics_of_Djibouti → C:Languages_of_Djibouti.
  - *Biological weapons* is included in the *Labour law* category: C:Labour_law → C:Labour_relations → C:Occupational_safety_and_health → C:Toxicology → C:Biological_weapons.
  - *Colombian people executed by firing squad* is included in the *Arithmetic* category: C:Arithmetic → C:Ratios → C:Rates → C:Temporal_rates → C:Acceleration → C:Force → C:Motion_(physics) → C:Flight → C:Ballistics → … → C:Projectile_weapons → C:Firearms → C:People_associated_with_firearms → C:Shooting_victims → ... →

`C:Colombian_people_executed_by_firing_squad`.

Thus, the transitivity violation in the Wikipedia category system is caused by two groups of reasons. The first group includes the reasons related to the violation of rules for constructing a hierarchy of concepts in an ontology. These violations can be eliminated by the authors of Wikipedia. The second group includes the reasons related to the very principle of organizing the system of Wikipedia categories, the main one of which is the nontransitivity of the associative relationship.

4 APPROACH TO ELIMINATION OF TRANSITIVITY VIOLATIONS

In this section, we propose a method for elimination of intransitive category-subcategory chains.

4.1 The Basic Idea

As we have demonstrated, one of the main causes of transitivity violations is the associative relation, which is not transitive.

Up-to-date methods of extracting transitive hierarchical structure from the category system (such as YAGO and WikiTaxonomy) detect associative relationships between categories and eliminate them all. A disadvantage of these methods is elimination of potentially relevant relationships, that don't violate transitivity. As a result, there is a need for a method, that eliminates associative relationships violating transitivity (e.g. *Statistics* → *Demography*), but keeps relationships not violating transitivity (e.g. *Education* → *Teacher*).

A proposed method is based on approach to establishing associative relationships in RuThes thesaurus (Loukachevitch, 2011; Loukachevitch and Dobrov, 2014; Loukachevitch et al., 2014; Loukachevitch and Dobrov, 2004a; Loukachevitch and Dobrov, 2004b). According to this approach associative relationship between two concepts doesn't violate transitivity if ontological dependence relation holds between these concepts.

In RuThes the notion of ontological dependence (Tahko and Lowe, 2016; Correia, 2008; Koslicki, 2012, Koslicki, 2013) is formalized according to modal-existential account: an object A is ontologically depends on an object B iff necessary if A exists then B exists (Simons, 1987, chap. 8, pp. 290–323; Thomasson, 1999, chap. 2, pp. 24–34).

The modal-existential account has several advantages, including simplicity and mathematical rigorous. A disadvantage is that its application requires a human participation. In this regard, it is not appropriate for the given problem.

Additionally, modal-existential account has been criticized on pure ontological grounds. Kit Fine has demonstrated, that this account is very rough approximation to the notion of ontological dependence and has counterexamples. As an alternative, Fine proposed an essential account. According to this account, *A* depends on *B* iff *B* is a constituent of the essence of *A*. Essence of object is defined as a collection of propositions that are true in virtue of the identity of this object. These propositions, in turn, constitute real definition of the object (Fine, 1994; Fine, 1995).

The notion *to be true in virtue of the identity of* calls for clarification. According to Fine's own interpretation, based on meta-ontological realism, this notion designates unanalyzed relation between an object and a proposition. We propose another interpretation, based on deflationist approach of Amie L. Thomasson. According to Thomasson, existence and identity criteria for an object *A* are application/co-application conditions for the term "A" (Thomasson, 2008; Thomasson, 2009). According to our interpretation, a proposition is true in virtue of identity of the object *A* if it is a part of application/co-application conditions for the term "A". And, accordingly, an object *A* is ontologically depended on an object *B* iff application/co-application conditions for the term "A" contain application/co-application conditions for term "B". Our interpretation demonstrates that the essential account is independent from meta-ontological assumptions.

So, employing the essential account we obtain the following criteria of ontological dependence: *A* ontologically depends on *B* iff *B* is ineliminably involved in the definition of *A*. This criterion suits better for automatic application.

As an approximation of a definition of an object, represented by any given Wikipedia category, we take the lead section of the main article of this category as well as of other language versions of this article. Involvement of an object in the definition of another object is roughly approximated by existence of a hyperlink between the two definitions.

4.2 Proof of the Concept

The proof of concept of the proposed method has been realized as follows:

- Check whether the relationship between the cate-

gory and its subcategory is associative. Assume that the relationship is associative if one of its participants is a topic category. The type of the category is detected by the method from WikiTaxonomy project.

- If the relationship is associative, then check whether the ontological dependence relation exists between the corresponding concepts on the base of the criterion described above. If the ontological dependence holds, then keep the relationship between the categories, and eliminate it otherwise.
- If the relationship is not associative, but taxonomic, then refer to YAGO that contains refined taxonomic relationships. If the corresponding relationship is contained in YAGO, then keep it, and eliminate it otherwise.

4.3 Evaluation

The proof of concept has been evaluated with category `C:Mathematics`. The choice of this category was motivated by the practical task of interlinking OntoMath^{Pro} ontology (Nevzorova et al., 2014; Elizarov et al., 2014; Elizarov et al., 2016; Elizarov et al., 2017) with DBpedia.

For the task of evaluation we recursively formed a list of subcategories of `C:Mathematics` category. Then we applied our method to eliminate supposedly irrelevant subcategories. The lists of the eliminated and the kept categories were analyzed manually. The assessor's task was to check whether the kept categories were relevant, and the eliminated categories were irrelevant. The assessment result is represented in Table 1.

Table 1: The result of the preliminary assessment of the proposed method with C:Mathematics.

Total	4281
True positives	2136
True negatives	650
False positives	1010
False negatives	485
Recall	0,814956
Precision	0,678957
F1 score	0,740766

5 CONCLUSION

We analyzed the causes for transitivity violations in the Wikipedia category system and proposed an approach to their elimination.

The main contributions of this paper are:

(1) an ontological analysis of the Wikipedia category system;

(2) a new approach to eliminating the violations of transitivity in this category system, based on analysis of the relation of ontological dependence between categories; and

(3) a new deflationistic interpretation of the essential account of ontological dependence, obtained by application of the meta-ontological framework of Amie Thomasson to the ontological dependence account of Kit Fine.

However, our work is on early stage and will be continued in the following directions:

1. Ontological Analysys of the Wikipedia Category System. We are going to conduct an ontological analysis in a more systematic way. In particular, it is supposed to find a numerical distribution of the category-subcategory relationships as well as of the causes for transitivity violations.

2. Aboutness Analysis of the Wikipedia Category System. We are going to complement our ontological analysis of the Wikipedia category system with an analysis based on the notion of aboutness. Aboutness is the relation between a meaningful item (such as a document or a proposition) and its subject matter. Subject matter is a central concept in Library and information science (Hjørland, 2016; Hjørland, 1992) and has been formalized in formal semantics and mathematical logic (Hawke, 2017). In our analysis, we would rely on the formalization of Yablo (2014), according to which subject matter is conceived as a partition of logical space and the part-whole relation between two subjects is defined as refinement relation between two partitions. With this formalisation in hands, we consider a category-subcategory chain as transitivity-preserving if subject matter, associated with the subcategory is a part of the subject matter, associated with the category.

Additionally, we are going to investigate correspondence between the notions of aboutness and ontological dependence. Our working hypothesis to be proved or disproved is that *A* is ontologically depended on *B* iff the subject matter associated with *A* is a part of the subject matter associated with *B*.

3. Formalization of the Deflationistic Interpretation of the Essential Account of Ontological Dependence. In this paper we proposeed a new deflationistic interpretation of the essential account of ontological dependence, where this

relation has been explained in terms of application/co-application conditions. However, the notion of application/co-application conditions itself requires clarification and formalization. We a going to formalize this notion on base of the situation semantics for relevant logic (Mares, 2004).

4. Refining of the proposed Approach. In particular, it is supposed to analyze other articles in the target category, as well as to use the context of the link within the definition.

5. Fine-grained Domain Annotation of DBpedia and BabelNet. We are going to apply the developed approach to derive a new large-scale domains hierarchy from the Wikipedia category system. In contrast to relatively small domains hierarchies of WordNet, BabelNet and DBpedia, the extracted hierarchy will contain a large number of fine-grained domains, such as *Ancient Greece*, *Moscow*, *Modal logic*, *Object-oriented programming* or *Star Wars*.

This hierarchy will be integrated into RuThes Cloud, a multilevel multilingual resource of the Linguistic Linked Open Data (Kirillovich et al., 2017; Galieva et al., 2017), and can be naturally applied to carry out fine-grained domain annotation of the resources, automatically extracted from Wikipedia, first of all, DBpedia and BabelNet.

ACKNOWLEDGEMENTS

This work was funded by the subsidy allocated to Kazan Federal University for the state assignment in the sphere of scientific activities, grant agreement no. 1.2368.2017, and by the subsidy of the Russian Government to support the Program of Competitive Growth of Kazan Federal University.

REFERENCES

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

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.

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.

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

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.

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

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.

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.

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

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

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.

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

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.

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.

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.

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.

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.

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.

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.

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

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

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

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

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.

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.

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

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.

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.

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

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.

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

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.

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.

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.

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.

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)*.

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

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.

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.

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.

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

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.

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.

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.

Nevzorova, O. et al. (2014). OntoMath^{PRO} 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.

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.

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.

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.

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.

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.

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.

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

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.

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/.

Thomasson, A.L. (1999). *Fiction and Metaphysics*. Cambridge: Cambridge University Press.

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

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.

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.

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.

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.

Yablo, S. (2014). *Aboutness*. Princeton: Princeton University Press.

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:10.1007/978-3-540-68234-9_29.