Barometr Regionalny
Tom 14 nr 3

Towards the Ontology of Places in Poland: an Example of the Mazowieckie Voivodship

Krzysztof Pancerz, Piotr Grochowalski
University of Rzeszów, Poland

Aneta Derkacz
University of Management and Administration in Zamość, Poland

Abstract
The main goal of our research is to build the ontology of places in Poland covering variety of aspects of places, mainly administrative and socio-economic. In the paper, we show a part of the ontology on the example of the Mazowieckie Voivodship. The ontology is being implemented using the OWL 2 Web Ontology Language.

Keywords: knowledge base, ontology, semantic relations, OWL, places in Poland

Introduction

The main goal of our research is to build the ontology of places in Poland. Modern IT tools, currently being used in various areas of administration and economy, require semantic and well-structured knowledge bases covering different aspects of information that is stored. Until now, information about places in Poland is gathered in resources of Central Statistical Office of Poland or other digital resources available on the Internet. First of all, this information is scattered and disconnected. Moreover, information is stored in the tabular form (for example, in relational databases) and divided according to its character as well as defined topics.

Knowledge engineering is a modern field within artificial intelligence that develops knowledge-based systems. Ontologies, as formal representations of knowledge, have recently gained popularity in knowledge engineering. In general, ontologies describe concepts in domains of interest as well as relationships that hold between those concepts (see Section 1). Two reasons seem to be the main source of this popularity. Firstly, there exist well-defined standards of languages for the ontology representation. Secondly, ontologies cover various semantic aspects of information.

The ontology proposed by us is being implemented in accordance with the OWL 2 Web Ontology Language specification.[1] The OWL 2 Web Ontology Language, informally OWL 2, is an ontology language for the semantic web with formally defined meaning. OWL 2 ontologies provide classes, properties, individuals, and data values. They are stored as semantic web documents.

We assume that, as a target, the ontology will represent rich information about different aspects of places in Poland, among others, administrative and socio-economic (see Section 2). Because of the large scale of the project, in the first instance, places that are capitals of voivodships or county seats are taken into consideration. Subsequently, the ontology will cover all places having city status (granted to them by way of legal acts) and all commune seats. Finally, all places in individual communes will be described in the ontology. In the paper, we show a part of our ontology on the example of the Mazowieckie Voivodship (see Section 2).

1. See: OWL 2 Web Ontology Language: Structural Specification and Functional-Style Syntax (Second Edition) by Boris Motic et al. W3C, 2012, [@:] https://www.w3.org/2012/pdf/REC-owl2-syntax-20121211.pdf.

One can find several ontologies covering selected aspects of the discussed problem (e.g., GeoNames) a geographical database containing information about over eleven million place names.[2] However, those ontologies do not include any specific information (for example, covering various socio-economic aspects) about places and regions in Poland. In general, they are focused on place names (toponyms).

1 Ontologies and Semantic Relations

During the last two decades, an increasing attention has been focused on ontologies. Ontological engineering refers to the set of activities that concern the ontology development process, the ontology life cycle, and the methodologies, tools and languages for building ontologies (Gómez-Pérez, Fernández-López, and Corcho 2004). Ontologies are currently used in knowledge engineering and data mining to capture knowledge about some domain of interest. In this paper, our domain of interest covers administrative, economic, and social aspects of places in Poland. There are many definitions and interpretations of the term "ontology" in the literature (Guarino and Giaretta 1995). We use mainly those proposed by Neches et al. (1991) and Kohler et al. (2006). Formally, the ontology can be represented by means of graph structures, called ontological graphs (Pancerz 2012). For a given ontology, an ontological graph $OG = (C, E, R, \rho)$ consists of

C—a nonempty, finite set of nodes representing concepts in the ontology,
E—a finite set of edges representing relations between concepts from C,
R—a family of semantic descriptions of types of relations (represented by edges) between concepts, and
ρ—a function assigning a semantic description of the relation to each edge.

Semantic relations are very important components in ontology modeling as they describe the relationships that can be established between concepts. A comprehensive review of the literature concerning semantic relations is given in (Nastase et al. 2013). As the authors noticed, almost every new attempt to analyze semantic relations leads to a new list of relations. WordNet (Fellbaum 1998), a kind of a linguistic ontology that is a large lexical database for English created at Princeton University and based on psycholinguistic theories, represents around a dozen semantic relations between concepts, including: synonymy, antonymy, hyponymy, hyperonymy, meronymy, and holonymy. Synonymy concerns concepts that have the same meaning or nearly the same meaning. Antonymy concerns concepts that have the opposite meaning. Hyponymy/hyperonymy determines narrower/broader meaning of concepts. Meronymy and holonymy define part/whole relationships between concepts.

In case of our ontology of places, we take into consideration three basic semantic relations between concepts:

- SUBCLASS-OF (hyponymy), also known as IS-A. If c SUBCLASS-OF c', it means that c is a kind of c' (c is a more specialized concept than c'), for example, city is a kind of place
- PART-OF (meronymy). If c PART-OF c', it means that c is a part of c', for example, commune is a part of county
- INSTANCE-OF. If i INSTANCE-OF c, it means that i is an instance (example) of c, for example, Radom is an instance of city

One can see that we can also consider inverse relations, SUPERCLASS-OF (hyperonymy), HAS-A (holonymy), HAS-INSTANCE for SUBCLASS-OF, PART-OF, and INSTANCE-OF, respectively.

Basic semantic relations are used in our ontology of places to describe relationships covering the administrative aspects, for example, categories of administrative districts, categories of places, categories of roads, etc. Moreover, we can distinguish many specific semantic relations describing relationships covering economic and social aspects of places (see Section 2). One of the key decisions to take in the ontology development process is to select the language in which the ontology will be implemented. Our ontology of places is built in accordance with the OWL 2 Web Ontology Language (shortly OWL 2). OWL 2 is the most recent development in standard languages defined by the World Wide Web Consortium (W3C). An OWL ontology consists of three components:

2. See: GeoNames, [@:] http://www.geonames.org/.

- classes
- individuals
- properties

Classes are representations of concepts from the domain of interest. Classes are interpreted as sets that contain individuals. Individuals (also known as instances) represent objects in the domain of interest. Individuals can be referred to as being instances of classes. Properties (also known as roles or attributes) are binary relations (relations between two elements) on individuals. Properties link two individuals together. There are two main types of properties in OWL 2:

- object properties linking an individual to an individual
- data properties linking an individual to a data value

In general, building ontologies is a complex and time consuming process. It is even more complicated if an ontology has to be implemented directly in an ontology language. There is a number of ontology development tools, see (Gómez-Pérez, Fernández-López, and Corcho 2004). In our project, we use Protégé (Musen 2015) — a free, open source, platform-independent environment for creating and editing ontologies and knowledge bases.[3]

2 The OWL Ontology of Places in Poland

In this section, we show selected parts of our ontology of places in Poland (defined classes and class hierarchies, identified individuals, identified properties linking individuals). The last part of this section consists of fragments of the ontology implemented by means of the OWL 2 Web Ontology Language. Ontology development is necessarily an iterative process.[4] One of the main steps is defining the classes and the class hierarchy. Classes represent concepts in the domain of interest. Table 1 includes distinguished classes representing categories of administrative districts in our ontology of places. Visualization of the hierarchy of classes corresponding to categories of administrative districts, using the Simple Ontology Visualization API (SOVA), is shown in figure 1. SOVA is a Protégé plug-in for full ontology visualization.[5] It is worth noting that the direction of arrows in SOVA visualization corresponds to the SUBCLASS-OF (hyponymy) relation, for example *Gmina miejska* (urban commune) is a subclass of *Gmina* (commune). One can see that there are some equivalent classes, *Miasto_na_prawach_powiatu* is equivalent to *Powiat_grodzki*.

Table 2 includes distinguished classes representing categories of places. Visualization of the hierarchy of classes corresponding to categories of places, using SOVA, is shown in figure 2. Table 3 includes distinguished classes representing categories of ways. Categorization was made according to the Polish law. In case of English terms, the WordNet hyponymy/hyperonymy relations were also taken into consideration (Fellbaum 1998). Visualization of the hierarchy of classes corresponding to ways, using SOVA, is shown in figure 3.

The next step is creating individuals (instances) of classes in the hierarchy and defining class assertions. Visualization of the individuals of counties in the Mazowieckie Voivodship, using SOVA, is shown in figure 4. Edges in SOVA visualization correspond to the INSTANCE-OF relation, for example, *Powiat_siedlecki* is an instance of *Powiat* (county). Figure 5 shows the individuals of motorways and expressways in the Mazowieckie Voivodship (namely, A2, S2, S7, S8, S17, S79).

In case of the individuals corresponding to categories of administrative districts, an important thing is to identify the property corresponding to the PART-OF (meronymy) relation. Figure 6 shows the PART-OF relation for the Legionowski County (Powiat Legionowski). Arrows (with dotted lines) corresponding to the PART-OF relation are added manually. One can see, for example, that:

- *Powiat_legionowski* is a part of *Województwo_Mazowieckie,*
- *Gmina_Wieliszew* is a part of *Powiat_Legionowski,* etc.

3. See also: A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools. Edition 1.3 by Matthew Horridge et al. The University of Manchester, March 24, 2011, [@:] https://mariaiulianadascalu.files.wordpress.com/2014/02/owl-cs-manchester-ac-uk_-eowltutorialp4_v1_3.pdf.

4. See: Ontology Development 101: A Guide to Creating Your First Ontology by Natalya F. Noy and Deborah L. McGuinness. Stanford University, 2001, [@:] http://protege.stanford.edu/publications/ontology_development/ontology101.pdf.

5. See: http://protegewiki.stanford.edu/wiki/SOVA.

Tab. 1. Classes representing categories of administrative districts

English Term	Polish Term
Country	Kraj
Voivodship	Województwo
County	Powiat
City with County Status	Miasto_na_prawach_powiatu / Powiat_grodzki
Commune	Gmina
Urban Commune	Gmina_miejska
Urban-Rural Commune	Gmina_miejsko-wiejska
Rural Commune	Gmina_wiejska

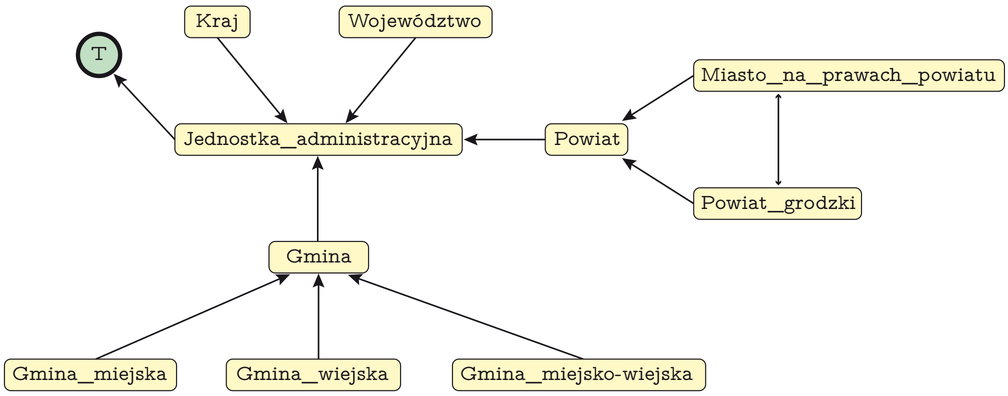

Fig. 1. The hierarchy of classes corresponding to categories of administrative districts

Tab. 2. Classes representing categories of places

English Term	Polish Term
Place	Miejscowość
City/Town	Miasto
Village	Wieś

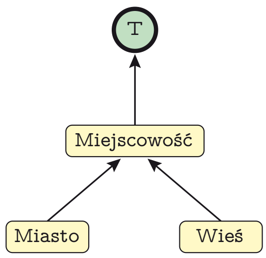

Fig. 2. The hierarchy of classes corresponding to categories of places

Tab. 3. Classes representing categories of ways

English Term	Polish Term
Way	Droga
Road / Route	Droga (publiczna)
Motorway / Freeway	Autostrada
Expressway	Droga_ekspresowa
National Road	Droga_krajowa
Voivodship Road	Droga_wojewódzka
County Road	Droga_powiatowa
Commune Road	Droga_gminna

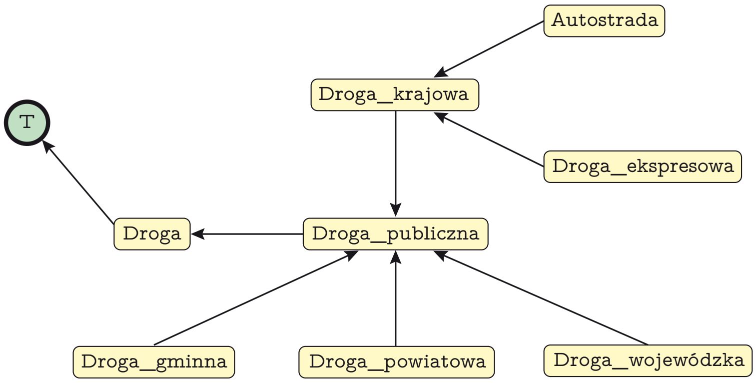

Fig. 3. The hierarchy of classes corresponding to categories of ways

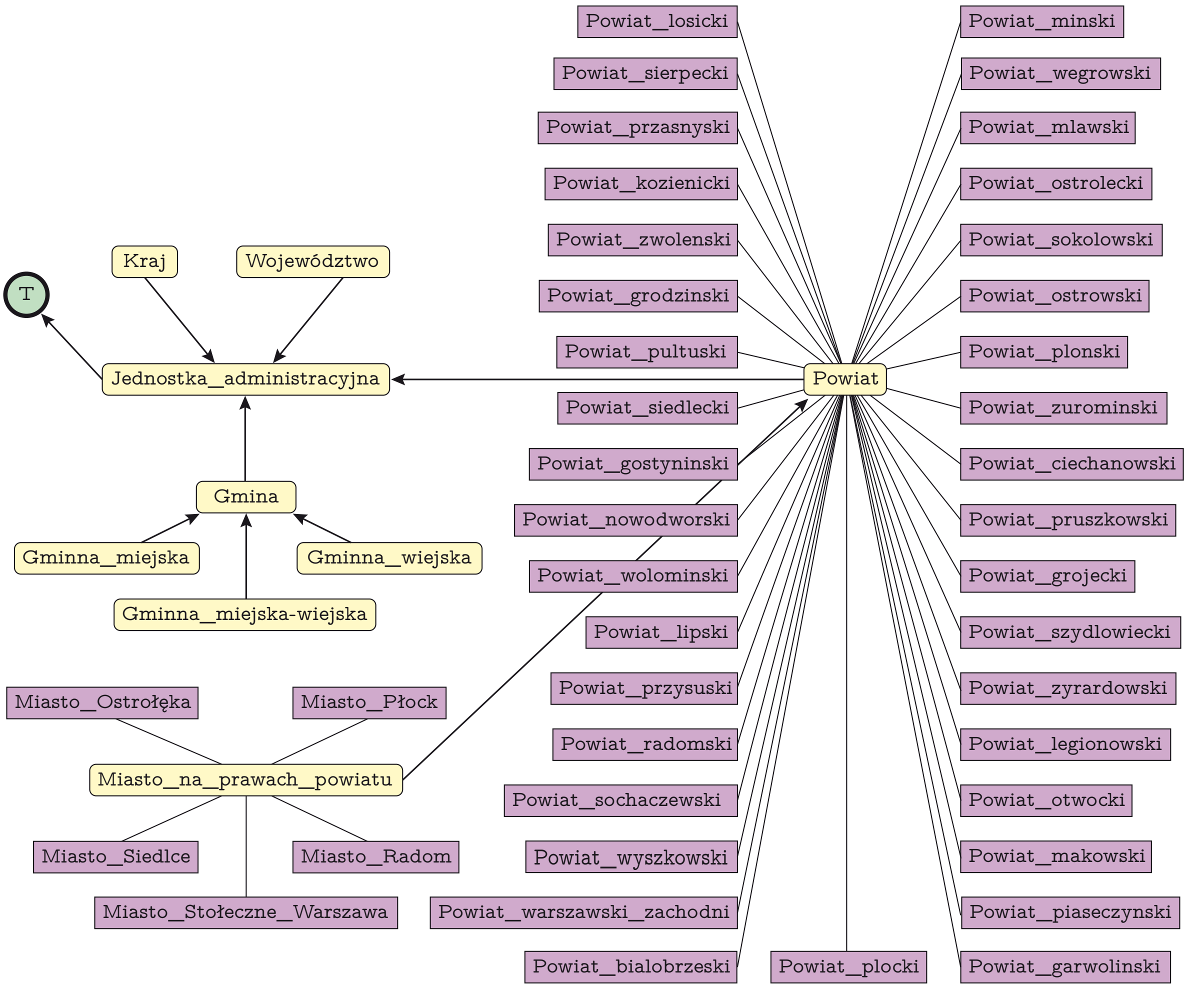

Fig. 4. The individuals of classes corresponding to counties

As it was mentioned in Section 1, we can distinguish many specific semantic relations describing relationships covering economic and social aspects of places. For example, we can assign places with motorways and expressways by the *posiada_dostęp_do* (hasAccessTo) relation (it means that there exists an entrance to the motorway or expressway in a given place).

Another important relations from the economic and social points of view are for example:

- hasAirport
- hasRailwayStation
- hasUniversity
- hasHospital
- hasCourt

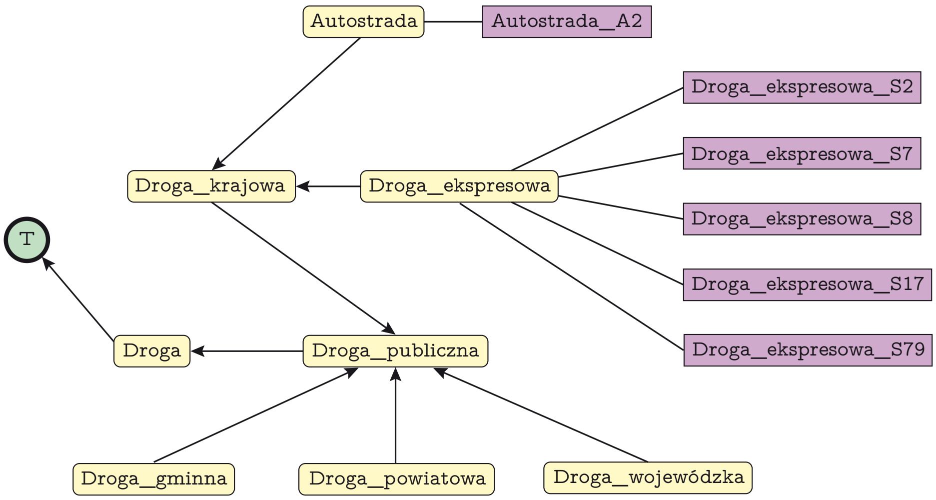

Fig. 5. The individuals of classes corresponding to motorways and expressways

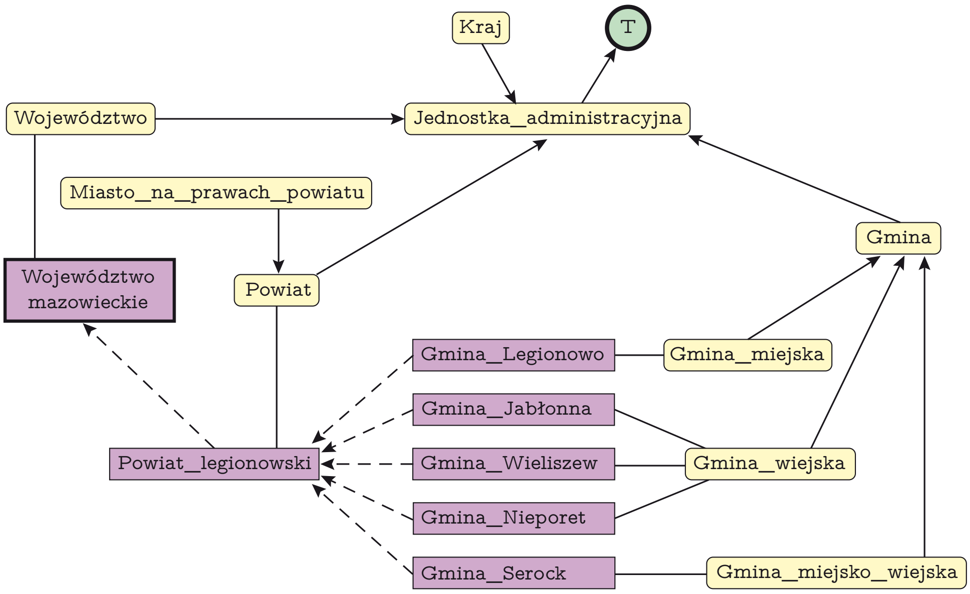

Fig. 6. The PART-OF relation for the Legionowski County

- hasPostOffice
- hasShoppingCentre/hasMall
- hasCinema
- hasTheatre
- hasConcertHall
- etc.

We take into consideration also data properties describing places, for example:

- hasPopulation
- hasArea
- hasCarPlatePrefix (where applicable)
- etc.

As it was mentioned earlier, our ontology of places in Poland is implemented by means of the OWL 2 Web Ontology Language. Now, we present some fragments of the OWL ontology using functional syntax:

- Declaration of classes:
```
Declaration(Class(:Jednostka_administracyjna))
Declaration(Class(:Kraj))
```

```
Declaration(Class(:Województwo))
Declaration(Class(:Powiat))
Declaration(Class(:Miasto_na_prawach_powiatu))
Declaration(Class(:Gmina))
Declaration(Class(:Gmina_miejska))
Declaration(Class(:Gmina_miejsko-wiejska))
Declaration(Class(:Gmina_wiejska))
```

- Declaration of the hierarchy of classes:
```
SubClassOf(:Gmina_miejska:Gmina)
SubClassOf(:Gmina_miejsko-wiejska:Gmina)
SubClassOf(:Gmina_wiejska:Gmina)
```
- Declaration of individuals:
```
Declaration(NamedIndividual(:Województwo_Mazowieckie))
Declaration(NamedIndividual(:Powiat_legionowski))
Declaration(NamedIndividual(:Powiat_lipski))
Declaration(NamedIndividual(:Miasto_Radom))
```
- Class assertion:
```
ClassAssertion(:Województwo:Województwo_Mazowieckie)
ClassAssertion(:Powiat:Powiat_legionowski)
ClassAssertion(:Powiat:Powait_lipski)
ClassAssertion(:Miasto_na_prawach_powiatu:Miasto_Radom)
```
- Declaration of object properties:
```
Declartion(ObjectProperty(:posiada_dostęp_do))
ObjectPropertyDomain(:posiada_dostęp_do:Miejscowość)
ObjectPropertyRange(:posiada_dostęp_do:Droga_ekspresowa)
ObjectPropertyRange(:posiada_dostęp_do:Autostrada)
```
- Declaration of data properties:
```
Declaration(DataProperty(:prefiks_numeru_rejestracyjnego))
DataPropertyDomain(:prefiks_numeru_rejestracyjnego:Powiat)
DataPropertyRange(:prefiks_numeru_rejestracyjnego xsd:string)
```
- Object property assertion:
```
ObjectPropertyAssertion(:posiada_dostęp_do:Konotopa:Droga_ekspresowa_S2)
ObjectPropertyAssertion(:posiada_dostęp_do:Legionowo:Droga_ekspresowa_S2)
ObjectPropertyAssertion(:posiada_dostęp_do:Majdan:Droga_ekspresowa_S2)
ObjectPropertyAssertion(:posiada_dostęp_do:Opacz_Mała:Droga_ekspresowa_S2)
ObjectPropertyAssertion(:posiada_dostęp_do:Warszawa:Droga_ekspresowa_S2)
```
- Data property assertion:
```
DataPropertyAssertion(:prefiks_numeru_rejestracyjnego:Powiat_Legionowski „WL")
```

Conclusions

In the paper, we have shown the part of the ontology of places in Poland built by us. Building the OWL ontology is the first stage of our research. In the future, we plan to create a graph database to store the ontology. The graph database is a database that uses graph structures for semantic queries with nodes, edges and properties to represent and store data. Next, we will create an intelligent search engine for our ontology. The engine will be based on computational intelligence methods, especially rough set methods (Pawlak 1991). In the past, we used some rough set methods for browsing ontology data (Suraj, Grochowalski, and Pancerz 2013). The ontology built by us can be used in various socio-economic research as the knowledge base. Moreover, it may constitute a basis for search engines and other computer tools used in the real-estate market. We are planning to make the created ontology publicly available.

Acknowledgments

This work was conducted using the Protégé resource, which is supported by grant GM10331601 from the National Institute of General Medical Sciences of the United States National Institutes of Health.

References

Fellbaum, C. ed. 1998. *WordNet. An Electronic Lexical Database, Language, speech, and communication*. Cambridge, Mass: MIT Press.

Gómez-Pérez, A., M. Fernández-López, and O. Corcho. 2004. *Ontological Engineering. With Examples from the Areas of Knowledge Management, e-Commerce and the Semantic Web*. London: Springer-Verlag.

Guarino, N., and P. Giaretta. 1995. "Ontologies and Knowledge Bases Towards a Terminological Clarification." In *Towards Very Large Knowledge Bases*, edited by N.J.I. Mars, 25–32. Amsterdam: IOS Press.

Kohler, J., S. Philippi, M. Specht, and A. Ruegg. 2006. "Ontology Based Text Indexing and Querying for the Semantic Web." *Knowledge-Based Systems* no. 19 (8):744–754. doi: 10.1016/j.knosys.2006.04.015.

Musen, M.A. 2015. "The Protégé Project: A Look Back and a Look Forward." *AI Matters* no. 1 (4):4–12. doi: 10.1145/2757001.2757003.

Nastase, V., P. Nakov, D. Ó Séaghdha, and S. Szpakowicz. 2013. *Semantic Relations Between Nominals. Edited by G. Hirst, Synthesis Lectures on Human Language Technologies*. Toronto: Morgan & Claypool.

Neches, R., R. Fikes, T. Finin, T. Gruber, R. Patil, T. Senator, and W.R. Swartout. 1991. "Enabling Technology for Knowledge Sharing." *AI Magazine* no. 12 (3):36–56.

Pancerz, K. 2012. "Toward Information Systems over Ontological Graphs." In *Rough Sets and Current Trends in Computing*, edited by J. Yao, Y. Yang, R. Słowiński, S. Greco, H. Li, S. Mitra and L. Polkowski, 243–248. Berlin; Heidelberg: Springer.

Pawlak, Z. 1991. *Rough Sets. Theoretical Aspects of Reasoning about Data, Theory and Decision Library Series D, System Theory, Knowledge Engineering, and Problem Solving*. Dordrecht-Boston: Kluwer Academic Publishers.

Suraj, Z., P. Grochowalski, and K. Pancerz. 2013. "Knowledge Representation and Automated Methods of Searching for Information in Bibliographical Data Bases: A Rough Set Approach." In *Rough Sets and Intelligent Systems — to the Memory of Professor Zdzislaw Pawlak*. Vol. 2, edited by A. Skowron and Z. Suraj, 515–538. New York: Springer.

Wykaz identyfikatorów i nazw jednostek podziału terytorialnego kraju. Podział terytorialny z dnia 1.01.2015 r. 2015. Zeszyty Metodyczne i Klasyfikacje. Warszawa: Główny Urząd Statystyczny.