

Geomatic Concepts in Agriculture Thesauri

O. Čerba, K. Jedlička

University of West Bohemia, Czech Republic

Anotace

Zemědělské tezaury (například AGROVOC nebo NAL Agriculture Thesaurus) jsou velmi velké a robustní systémy formalizovaných znalostí. Jsou zaměřeny především na informace z oblasti zemědělství, nicméně i ony používají fragmenty informací a znalostí z dalších oblastí. A tak zahrnují i základní pojmy z oboru geomatiky. I přesto, že geomatické pojmy nejsou hlavní složkou výše uvedených tezurů, hrají velmi důležitou roli v procesu detailního popisu zemědělských a dalších pojmů obsažených v tezaurech (zejména v procesu jejich měření, pozorování nebo mapování).

Tato studie posuzuje geomatické koncepty uvedené v AGROVOC a NAL tezaurech z hlediska geomatiky (ale s ohledem na metodiku vývoje a údržby tezurů). Zaměřuje se na hodnocení podskupiny pojmů souvisejících s geomatikou a příbuzných vědeckých disciplín, jako je kartografie, fotogrammetrie, GIS a dálkový průzkum Země. Autoři studovali definice pojmů, jejich hierarchie, vztahy a i odkazy na další informační zdroje. Výsledkem je krátký seznam doporučení, jak zlepšit a obohatit výše uvedené tezaury z hlediska konceptů geomatické domény, což může vést ke zlepšení kvality tezurů a jejich informační hodnoty.

Klíčová slova

Geomatika, thesaurus, sémantika, zemědělství, koncept.

Abstract

The agriculture thesauri (e.g. AGROVOC or NAL Agriculture Thesaurus) represent very large and robust systems of formalized knowledge. They are primarily focused on information related to agriculture. But they also use fragments of geomatic information and knowledge in a form of concepts and their terms. These concepts include general terms of all parts of geomatics as well as data instances (such as particular methods). Even though these concepts are not the main component of above-mentioned thesauri, the concepts from geomatic domain play very important role in a process of detail description of agricultural and other concepts (including processes of their measurement, observation or mapping) contained in thesauri.

This paper assess geomatic concepts in AGROVOC and NAL Agriculture Thesaurus from the view of geomatics (but with a respect to methodologies of thesauri development and maintenance). It means evaluation of the subset of concepts related to geomatics and close scientific disciplines such as cartography, photogrammetry, GIS science or remote sensing. Authors look into definitions of concepts, their hierarchy, relations and links to other information resources. As the result there is a short list of recommendations how to improve and enrich the above-mentioned thesauri from the view of concepts from geomatic domain. It can enhance the quality of thesauri and their information value.

The paper introduces the fundamental terminology (terms thesaurus, geomatics and concept) and related researches. Then a description of mapping of concepts in particular tools follows. The results of mapping are summarized in the part focused on the most frequent imperfections. The last section (with the exception of the final conclusion) presents the set of recommendations concerning usage of concepts from geomatic domain in agricultural thesauri.

Key words

Geomatics, thesaurus, semantics, agriculture, concept.

Introduction

Thesauri belong to very powerful semantic tools. They provide lot of information classified to groups based on broader and narrower term relations. The agriculture thesauri such as AGROVOC, CABI or NAL (National Agricultural Library's) Agriculture Thesaurus represent very large and robust systems of formalized knowledge from agriculture domain. They are primarily focused on information related to agriculture. But they also use fragment of geomatic information and knowledge in a form of geographic concepts. These concepts includes general terms connected not only to geomatics (e.g. "border", "Earth", "accuracy") as well as specific geomatic concepts (such as "remote sensing", "spatial data" or "photogrammetry"). Even though these concepts are not the main component of above-mentioned thesauri, the concepts from geomatic domain play a very important role in a process of detail description of agricultural and other concepts (including processes of their measurement, observation or mapping) contained in thesauri.

This paper assess geomatic concepts (concepts related to geomatics and close scientific disciplines such as cartography, photogrammetry, GIS science or remote sensing) in AGROVOC and NAL Agriculture Thesaurus semantically. Authors look into definitions of concepts, their hierarchy, relations and links to other information resources. As the result of research there is a list of recommendations how to improve and enrich above-mentioned thesauri from the view of concepts from geomatic domain. It can enhance the quality of thesauri and their information value.

The paper introduces the essential terminology (terms: thesaurus, geomatics, semantics and concept) and related researches focused on comparison of concepts and their evaluation. Than a description of methodology based on semantic factoring, similarity quantification and formal conceptual analyses follows. The results of evaluation are summarized in the part focused on the most frequent imperfections. The last section (with the exception of the final conclusion) presents the set of recommendations concerning using concepts from geomatic domain in agricultural thesauri.

Materials and methods

This paper deals with several terms that could not be frequently used in the agriculture domain.

Therefore authors consider to publish their definition and explanation necessary.

Geomatics – scientific and technical interdisciplinary branch focused on collecting, distributing, storing, analysing, processing and presenting of geographical data or geographical information – is the young discipline of science dealing with spatial data and spatial technologies. Just the orientation to "spatial" is very important to the relation between geomatics and agriculture. Geomatics provides a large portfolio of data, technologies (including Global Navigation Satellite System), analyses and map outputs, which are essential in many sectors of agriculture such as precision farming, satellite farming, site specific crop management, sustainable agriculture or landscape and rural development.

Thesaurus is a lists of terms or concepts (notice: Terms and concepts represent different entities in the ontological and semantics theory, but for purposes of this article they will be used as equivalent words, because each term will be represented by concept), which is based on similarity of meaning of particular terms. It usually contains a hierarchical structure of terms or concepts (using the principle of broader and narrower concepts) and relations to synonyms and sometimes antonyms. Thesauri also provide an explicit explanation or definitions of terms similarly to dictionaries. More detail information on thesauri is available for example in the article *Approaches To Thesaurus Production* (Michiels, Noel, 1982) or *Thesaurus Maintenance, Alignment and Publication as Linked Data: The AGROVOC Use Case* (Caracciolo et al., 2011).

In the article the two essential agriculture thesauri are compared. AGROVOC – multilingual agricultural thesaurus is managed by Food and Agriculture Organization (FAO) of the United Nations (UN). It consists of over 32 000 concepts available in 21 languages. AGROVOC is based on the SKOS (Simple Knowledge Organization System) standard (Bechhofer, Miles, 2009). AGROVOC is very open. It means not only providing of Linked Data, but above all a huge number of publication (e.g. Soergel et al., 2006, Sini et al., 2008 or Caracciolo et al., 2013) describing development and research of the thesaurus.

The National Agricultural Library's Agricultural Thesaurus – online vocabulary tools of agricultural terms is produced by National Agricultural Library, United States Department of Agriculture, and the Inter-American Institute for Cooperation

on Agriculture as well as other Latin American agricultural institutions. It contains about 98 000 terms in English and Spanish. The thesaurus uses SKOS standard. This thesaurus is also available as Linked Data. More information on NAL is published in document Alonso (2007).

Semantics as a research of meaning and its studying belongs among key tasks of contemporary sciences connected with information technologies, including geomatics. Particular studies are focused on a development of semantics models (Huang et al., 2010, Zhang, Xu, 2011, Wang et al., 2011, Zhang et al., 2014), collecting of semantic data and information (Hazman et al., 2009, Pazienza et al., 2012), linking of semantic data resources (Severino, 2007, Lauser, 2008, Lopez-Pellicer et al., 2010), semantic similarity (Bae et al., 2014, Batet et al., 2014, Jiang et al., 2014) or development of semantic data resources in various domains (An, Zhao, 2007, Ping, Yong, 2009, Li et al., 2013, Gimenez et al., 2013). Majority of above mentioned document, which are connected to geographic domain, constitutes the main theoretical background of this article as well.

The methodology of comparison of selected geomatic concepts in agriculture thesauri is composed of two basic steps – selection of suitable concepts and their processing. The selection of concepts is based on personal experience of authors as well as on domain dictionaries such as online Terminological dictionary of surveying and cadastre or International GIS Dictionary (McDonnell, Kemp, 1995). The final set contains 34 concepts: “accuracy”, “border”, “cartography”, “coordinate”, “dimension“, „Earth“, „feature“, „generalization“, „geodesy“, „geography“, „geoinformatics“, „geomatics“, „GIS“, „GNSS“, „GPS“, „imagery“, „landscape“, „map“, „map projection“, „mapping“, „metadata“, „morphology“, „photogrammetry“, „projection“, „raster data“, „remote sensing“, „scale“, „spatial data“, „surface“, „surveying“, „thematic map“, „topography“, „vector“ and „vector data“. The set includes common terms (e.g. “border”, “feature”), technologies (e.g. “GPS”, “GNSS”), sciences (e.g. “cartography”, “geodesy”) and specific geomatic concepts (e.g. “spatial data”, “map projection”).

Selected terms were processed by both descriptive and quantitative methods to test their occurrence in thesauri, similarity and position in the hierarchy of thesauri. To the comparison there are used descriptive methods of graphical scheme of hierarchies (based on broader and narrower

concepts, which are the key components of SKOS standard) and other relations such as related concepts (similarly to Severino, 2007). Concepts were also processed by quantitative methods focused on computation of similarity using Formal Concepts Analysis (Wille, 1992), which is described with use of mathematical structures in Kavouras, Kokla (2007). The tested concepts are decomposed into primitives, which are in this case based on definitions (and other explicit descriptions) published in both thesauri. Based on the occurrence of primitive concepts the similarity according Tversky (1997).

Results nad discussion

In total 34 concepts connected to geomatics mentioned in previous parts of this article were compared. The both thesauri contain 21 of the set of tested concepts (61,76%, Table 1). It is result better than average, but it is necessary to mention two additional facts:

1. There are only 14 concepts (41.18%, Table 1) which are found in both thesauri simultaneously – “cartography”, “dimension”, “geodesy”, “geography”, “geomatics”, “gis”, “gps”, “imagery”, “landscape”, “mapping”, “photogrammetry”, “remote sensing”, “scale” and “topography”.
2. None of thesauri contain such a fundamental geomatic concepts such as “map”, “coordinate” or “border”. Moreover these concepts are quite general and overlap to other thematic domains, including agriculture.

A scant occurrence of concepts is not sufficient from the view of semantics. Users have to have more detail explicit information available to realize context and meaning of concepts. Thesauri allow to add descriptions or definitions to each concepts. But as it is evident from following table (Table 1), taking advantage of explicit descriptive information is not sufficient.

The following results are based on comparison of 2 concepts (“cartography” and “photogrammetry”) defined in both thesauri. These concepts are processed by methods described in the part Methodology.

Authors deal with two pairs of definitions:

“cartography”

- AGROVOC: The art and science of the production of maps. This includes

| Thesaurus | Number of concept | Number of defined or described concepts |
|---------------|-------------------|--|
| AGROVOC | 21 | 9 (42.86% from concepts contained in AGROVOC; 26.47% from all tested concepts) |
| NAL | 21 | 11 (52.38% from concepts contained in AGROVOC; 14.29% from all tested concepts) |
| AGROVOC & NAL | 14 | 2 (42.86% from concepts contained in AGROVOC; 5.88% from all tested concepts) |

Source: own processing

Table 1: Overview of concepts and their description in thesauri.

the construction of projections, design, compilation, drafting and reproduction.

- NAL: The art, science and technology of mapmaking.

“photogrammetry”

- AGROVOC: The science of obtaining reliable measurements from photographs.
- NAL: The science of deducing precise measurements from photographs.

All definitions are decomposed to single terms:

“cartography”

- AGROVOC (10): art, science, production, map, construction, projection, design, compilation, drafting, reproduction.
- NAL (4): art, science, technology, mapmaking.

AGROVOC & NAL (2): art, science.

“photogrammetry”

- AGROVOC (5): science, obtaining, reliable, measurement, photography.
- NAL (5): science, deducing, precise, measurement, photography.
- AGROVOC & NAL (3): science, measurement, photography.

Decomposed terms of particular definition are processed by statistic computation of similarity, which is expressed as a percentage (0% means completely different concepts, 100% same concepts). As it is evident from definitions and their decomposition the similarity of concept “photogrammetry” is higher (42.86%) than the similarity of concept “cartography” (16.67%). Taking into consideration specificity of tested concepts both values are insufficient.

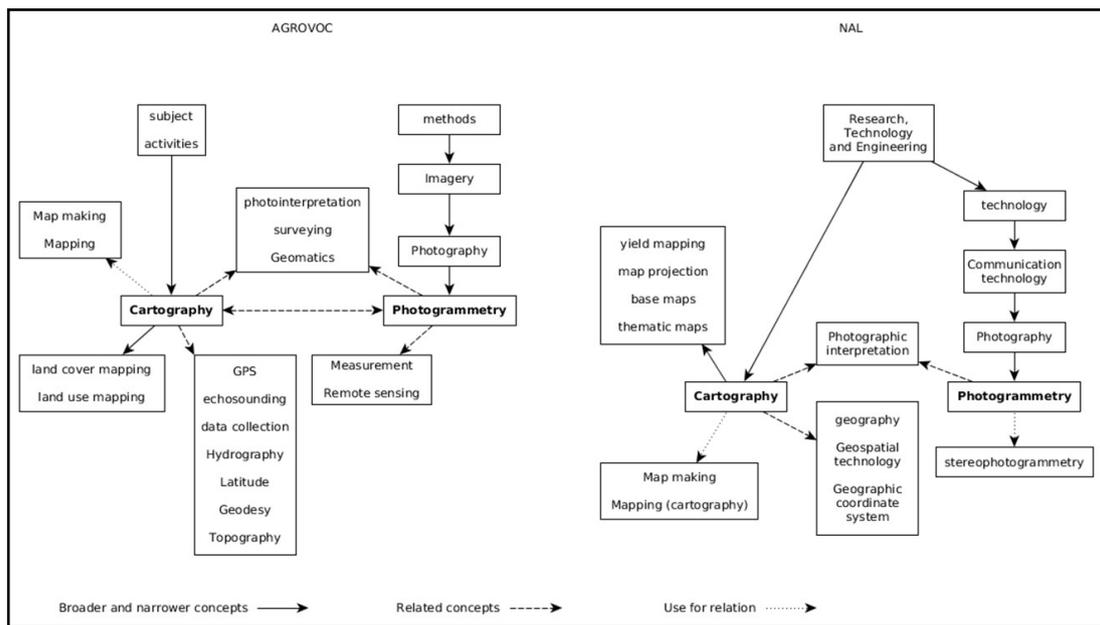
Previous results point to a speculation that both

pairs of concepts are composed from more or less independent concepts (even though concepts express the same entity). This guess is supported by the Formal Concept Analyses, which is connected with semantic factoring of definitions. Its results show that concepts have several common terms, but they are not interconnected by any relation of subordination or superiority.

Testing of various relations (broader and narrower concepts, related concepts and relation Used for) have to find other types of connections of evaluated concepts. Following scheme (Figure 1) presents a comparison of hierarchies and relations of both concepts in each thesaurus.

From the Figure 1 it is evident, that similarly to definitions the classification of concepts and relations are quite different in the AGROVOC and NAL Agriculture Thesaurus. There are following heterogeneities, which are characteristic not only for tested concepts (this opinion is based on the detail research of the AGROVOC that is not published now).

- Even though both concepts represents branches of science or technical activities, their classification into broader concepts is totally different, especially in case of AGROVOC. But also independent position of the concept “cartography” in NAL is not correct. This difference is important in case of concept “Photography”. In both thesauri it is the nearest broader concept to “photogrammetry”, but AGROVOC classifies this concept to methods and NAL to technologies.
- The development of narrower concepts seems to be non-systematic. It is evident from the sets of narrower concepts to the “cartography” in both thesauri. NAL deals with small subset of key cartographic



Source: own processing

Figure 1: Relations of concepts in the NAL thesaurus.

concepts and AGROVOC use relations to two (relatively marginal) types of mappings.

- The last group of heterogeneities comes from the related concepts. Even though cartography and photogrammetry are very similar scientific disciplines, according to the tested thesauri they have only one joint related concept – “photointerpretation” or “Photographic interpretation” (thesauri deals with two different concept, but in this case they can be considered to be equivalent). Moreover only AGROVOC contains the relation between “cartography” and “photogrammetry”. Relation to other related concepts are similarly to narrower concepts very non-systematic and in many cases not correct (e.g. very weak relation between “cartography” and “echosounding”).

Conclusion

This article describes analyses of concepts connected to geomatics and their implementation in two the most important agricultural thesauri (AGROVOC and NAL Agricultural Thesaurus). The results could be summarized into following points:

1. Occurrence of geomatic concepts in agricultural thesauri is slightly above-

average, but there is a lack of common concepts.

2. Explicit specification of geomatic concepts in agricultural thesauri is very poor.
3. The similarity between definitions is very poor. Even though both concepts are very specific and well (but not uniformly) defined.
4. There are two pairs of interconnected definitions without any relations of subordination or superiority.
5. Both tools contain some errors, but NAL looks more consistent.
6. Inconsistency in hierarchies even though both thesauri are focused on the same topic and use same rules, principles and standards. For example thesauri contain quite specific concepts such as “thematic map” or “GPS”, but not concepts “map” or “GNSS” which should their broader concepts logically.

Results of the limited research described in this article show that

- ...it is necessary to stop a spontaneous and subjective development and modification of semantic tools. They have to be created in very close cooperation of domain experts with use of standardized (recommended) methods.
- ...users as well as developers have to become

aware that universal concepts do not exist, because they are context-dependent. These contexts proceed from many aspects such as language convention, culture, domain, education etc. Users need to feel context or they need to be able to find it in a semantic tool. Tools have to provide context by explicit semantic description and relations.

The future steps of this research will be focused on evaluation of other types of geographic concept and semantic tools. It should lead into a creation of recommendations for users and administrator of semantic tools and a development of a more efficient and correct way of storing of geographic concepts with use of more detailed relations and multiple representation.

The high quality of implementation of geographic concept is very important from a view of general communication or sharing information. It is

crucial in many processes such as various types of agriculture connected to geoinformation technologies, implementation of INSPIRE directive in Europe (details see in Řezník, 2013) and other standardization activities or educational activities that need to handle well described information (details see in Janečka et al., 2011).

Acknowledgements

This contribution was supported by:

- the European Regional Development Fund (ERDF), project „NTIS - New Technologies for Information Society“, European Centre of Excellence, CZ.1.05/1.1.00/02.0090
- the project EXLIZ – CZ.1.07/2.3.00/30.0013, which is co-financed by the European Social Fund and the state budget of the Czech Republic.

Corresponding author:

Ing. et Mgr. Otakar Čerba, Ph.D.

University of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic

Phone: +420604475579, E-mail: ota.cerba@gmail.com

References

- [1] Alonso, Y. The National Agricultural Library – Providing Tools For Education And Development In Agriculture. Best Practices in Government Information: A Global Perspective. 2007, 205. ISBN 978-3-598-11769-5.
- [2] An, Y., Zhao, B. Geo ontology design and comparison in geographic information integration. In Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on (Vol. 4, pp. 608-612). IEEE. ISBN 978-0-7695-2874-8.
- [3] Bae, M., Kang, S., Oh, S. Semantic similarity method for keyword query system on RDF. Neurocomputing. 2014, 146, p. 264-275. ISSN 0925-2312.
- [4] Batet, M., Harispe, S., Ranwez, S., Sánchez, D., Ranwez, V. An information theoretic approach to improve semantic similarity assessments across multiple ontologies. Information Sciences. 2014, 283, p. 197-210. ISSN 0020-0255.
- [5] Bechhofer, S., Miles, A. SKOS simple knowledge organization system reference. W3C recommendation, W3C, 2009.
- [6] Caracciolo, C., Morshed, A., Stellato, A., Johannsen, G., Jaques, Y., Keizer, J. Thesaurus Maintenance, Alignment and Publication as Linked Data: The AGROOVOC Use Case. In Metadata and Semantic Research. 2011 (p. 489-499). Springer Berlin Heidelberg. ISBN 978-3-642-24731-6.
- [7] Gimenez, P. J., Tanaka, A. K., Baião, F. A. A geo-ontology to support the semantic integration of geoinformation from the National Spatial Data Infrastructure. In GeoInfo. 2013, p. 103-114. ISSN: 2179-4820.
- [8] Hazman, M., El-Beltagy, S. R., Rafea, A. Ontology learning from domain specific web documents. International Journal of Metadata, Semantics and Ontologies. 2009, Vol. 4, No. 1, p. 24-33. ISSN 1744-263X.

- [9] Huang, Y., Deng, G., Wu, X., Zhao, Z. Research on Representation of Geographic Feature Based on Geo-Ontology. In *Intelligent Systems and Applications (ISA), 2010, 2nd International Workshop on* (p. 1-5). IEEE. ISBN 978-1-4244-5874-5.
- [10] Janečka, K., Berzins, R., Charvát, K., Dzerve, A. On How to Build SDI Using Social Networking Principles in the Scope of Spatial Planning and Vocational Education. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6783 LNCS (PART 2). 2011. p. 78-92. ISBN 978-3-642-21887-3.
- [11] Jiang, Y., Wang, X., Zheng, H. T. A semantic similarity measure based on information distance for ontology alignment. *Information Sciences*. 2014, Vol. 278, p. 76-87. ISSN: 0020-0255.
- [12] Kavouras, M., Kokla, M. *Theories of geographic concepts: ontological approaches to semantic integration*. CRC Press. 2007. ISBN 9780849330896.
- [13] Lauser, B., Johannsen, G., Caracciolo, C., Keizer, J., van Hage, W. R., Mayr, P. *Comparing human and automatic thesaurus mapping approaches in the agricultural domain*. 2008.
- [14] Li, J., Liang, Y., Wan, J. Geo-Ontology-Based object-oriented spatiotemporal data modeling. In *Pervasive Computing and the Networked World*. Springer Berlin Heidelberg. 2013, p. 302-317. ISBN 978-3-642-37015-1.
- [15] Lopez-Pellicer, F. J., Silva, M. J., Chaves, M. Linkable geographic ontologies. In *Proceedings of the 6th Workshop on Geographic Information Retrieval*, ACM. 2010, p. 1, ISBN: 978-1-60558-826-1.
- [16] McDonnell, R., Kemp, K. K. *International GIS dictionary*. John Wiley & Sons, 1995. ISBN 9780470236079.
- [17] Michiels, A., Noel, J. Approaches to thesaurus production. In *Proceedings of the 9th conference on Computational linguistics*. 1982, Vol.1, p. 227-232.
- [18] Pazienza, M. T., Stellato, A., Tudorache, A. G., Turbati, A., Vagnoni, F. An Architecture for Data and Knowledge Acquisition for the Semantic Web: The AGROVOC Use Case. In *On the Move to Meaningful Internet Systems: OTM 2012 Workshops*, Springer Berlin Heidelberg. 2012, p. 426-433. ISBN 978-3-642-33618-8.
- [19] Ping, D., Yong, L. Building place name ontology to assist in geographic information retrieval. In *Computer Science-Technology and Applications, 2009. IFCSTA'09. International Forum on IEEE*. 2009, Vol. 1, p. 306-309. ISBN 978-1-4244-5423-5.
- [20] Řezník, T. Geographic information in the age of the INSPIRE directive: Discovery, download and use for geographical research. *Geografická informace v době směrnice INSPIRE: Nalezení, získání a využití dat pro geografický výzkum, Geografie-Sborník CGS*. 2013, Vol. 118, 1, p. 77-93.
- [21] Severino, F. The term development in the thesauri of international organisations. *The European Journal of Development Research*. 2007, 19, No. 2, p. 327-351. ISSN 1743-9728.
- [22] Sini, M., Lauser, B., Salokhe, G., Keizer, J., Katz, S. The AGROVOC Concept Server: rationale, goals and usage. *Library Review*. 2008, 57, No. 3, p. 200-212. ISSN 0024-2535.
- [23] Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., Katz, S. Reengineering thesauri for new applications: the AGROVOC example. *Journal of digital information*. 2006, 4, No. 4. 26 p. ISSN 1368-7506.
- [24] Tversky, A. Features of similarity. *Psychological Review*. 1977, 84, No. 4, p. 327-352. ISSN 0033-295X.
- [25] Wang, Z., Li, M., Li, F. Geo-ontology model based on description logic. In *Geoinformatics, 2011 19th International Conference on* (p. 1-5). IEEE. ISBN 978-1-61284-849-5.
- [26] Wille, R. Concept lattices and conceptual knowledge systems. *Computers & mathematics with applications*. 1992, Vol. 23, No. 6, p. 493-515. ISSN 0898-1221.

- [27] Zhang, C., Cheng, J., Liu, J., Pang, J., Liang, C., Huang, Q., Tian, Q. Object categorization in sub-semantic space. *Neurocomputing*, 2014. ISSN 0925-2312.
- [28] Zhang, X., Xu, J. Construction of geo-ontology knowledge base about spatial relations. In *Spatial Data Mining and Geographical Knowledge Services (ICSDM)*. 2011 IEEE International Conference on IEEE. p. 234-237. ISBN 978-1-4244-8352-5.