Scientific
Research

An Empirical Investigation of Common Sense of Land Use from a Statistical Approach

Yuki Hanashima

Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Japan
Email: hanashima@geoenv.tsukuba.ac.jp

ABSTRACT

Recently, ontological study has been one of the key concerns of geographic information science, a number of studies have been conducted in both of philosophical and knowledge engineering approach. Some studies pointed out the importance of human cognition and social context for development of ontologies. This paper presents empirical investigation of common sense of land use categories for development of suitable ontologies for each cultural or speech communities. Distinctions and characteristics in perceiving land use categories were described by a psychological method that was submitted to Japanese graduate and undergraduate students. In addition the results were analyzed using correspondence analysis, a statistical technique for categorical data. This analysis serves to clarify the dominant determining factors for land use categories.

Keywords: Empirical Investigation; Land Use Category; Common Sense; Ontology

1. Introduction

Semantic issues have always been a key concern in geographic information science (GIScience) because semantic interoperability plays a crucial role in the sharing and integration of geographic information [1]. Although the Open Geospatial Consortium (OGC) and ISO/TC211 provide certain standards supporting the deployment of geospatial web services regarding semantic interoperability, these standards address the interoperability issue at the syntactic level. They are therefore limited in terms of semantics and do not provide a consistent model for the semantic integration geospatial services [1,2]. Ontology has been identified as an explicit specification of a conceptualization contributing to the establishment of semantic interoperability. Until recently, research related to ontologies in GIScience has been broadly divided into the philosophical approach and the knowledge-engineering approach. The philosophical approach has addressed top-level ontologies for the geographic domain; the knowledge-engineering approach has addressed ontologies as application-specific and purpose-driven engineering artifacts [1].

Within the philosophical approach, a number of theories about geospatial ontologies have been discussed (e.g., formalization of ontology [3]). While the conventional approaches to study ontologies had been based on the objectivist point of view premised on a real world independent of human cognition and social context, the sub-

jectivist point of view has been focused on ontology research in geographic information [4,5]. Mapping human cognition or categories to ontologies is the most rational basis for data integration or sharing [6,7]. In fact, it is necessary to develop suitable ontologies for each cultural or speech communities. Common sense is critically reflected by background that people of each community are thinking, speaking and perceiving in every day. Mark and Turk investigated common sense of landscape categories in the language of the Yindijibarndi people [8]. The series of studies about ontologies presented by these authors constitutes one of the few research efforts to investigate common sense within a perspective of human cognition and cross-linguistics for the development of ontologies [8,9]. In contrast, Mark et al. and Smith and Mark employed questionnaire method to obtain empirical evidence of the influence of human cognition on geographical categories [10,11].

In order to develop valid geospatial ontologies, they must be investigated in each language domain because sharing spatial data and attributes requires language stability across cultures and geographies—assumptions that are seldom true [12,13]. In addition, the targets of previous studies about geospatial ontologies are natural objects such as landscape [8]; insufficient study has been conducted for artificial objects. The objective of this study is to empirically investigate common sense of artificial land use categories, such as public facilities in an urban area, in Japanese community. The questionnaire

method was used to investigate this common sense. The results of the questionnaire were then analyzed, using a statistical technique, to clarify the distinctions and characteristics of this common sense.

2. Empirical Investigations of Common Sense

2.1. The Questionnaire Method as an Empirical Investigation of Conceptualization

Questionnaire method was used as psychologically empirical investigation of some kinds of common sense or human cognitions. The questionnaire conducted in this study was submitted to 60 graduate and undergraduate students majoring in earth sciences at the University of Tsukuba, and contained 38 facility classes to classify into four classes: "Public facility", "Commercial facility", "Residence", and "Others" (**Table 1**). This questionnaire was applied in Japanese-language circumstance.

2.2. Natural and Artificial Land Use Categories

One of the best-known land use classification systems is the Land Cover Classification System (LCCS), developed

Table 1. Question table used in the questionnaire method.

Sample answer			
City hall	P	Japanese style hotel	
Shopping center		Horse race track	
Concert hall		Amusement park	
Apartment complex		Athletic field	
Art museum		Public school	
Library		Driving school	
Police box		Clinic	
Welfare house for aged		Aquarium	
Dormitory for diet members		Botanical garden	
Temple, shrine		Private tutoring school	
Kindergarten		Vocational college	
Wholesale market		Shinkin bank*	
Public bath		Broadcast station	
Movie theather		Railroad station	
Bus terminal		Stock terminal	
Medical center		Student dormitory	
Fire station		Electric plant	
Bank		Park	
Child care center**		Airport	
Cemetery			

P: Public facility, C: Commercial facility, R: Residence, O: Others. *Japan's Shinkin banks, commonly known as credit associations, are relatively small financial institutions that are privately held by members living near a bank's headquarters [14]; **In Japan, child care center is generally under private management and is different from a kindergarten that is licensed by the Ministry of Health, Labor and Welfare or the Ministry of Education, Culture, Sports, Science and Technology. Their facility functions are, however, almost same.

by the Food and Agricultural Organization of the United Nations [15]. The LCCS has detailed a classification based on natural land use categories, such as natural vegetation or agricultural land. Although classifications about artificial land use are important factor to describe human activity, would be more complex and arbitrary than natural land use categories. In this study, the "Public facility", "Commercial facility", "Residence", and "Others" were used as artificial land use categories, because these categories often appear in developed urban areas. The common sense underlying these categories would be complex and arbitrary because they would be deeply related to culture and history. However, because GIS applications or services in urban areas would produce many benefits, it is valuable to investigate the common sense of these artificial land use categories for development of its ontologies.

3. Relationships between Land Use Categories and Facilities

3.1. Correspondence Analysis

To investigate the common sense about the land use categories used in the questionnaire, correspondence analysis was applied. Correspondence analysis is a multivariate statistical technique for use with categorical data rather than quantitative data. Correspondence analysis has become increasingly popular in ecological, marketing and psychological research [16]. The basic idea of correspondence analysis is to reduce the dimensionality of a data matrix and visualize it in a subspace of low dimensionality, commonly two- or three-dimensional [17]. Both columns and rows can be visualized on the same plot [18]. The main components underlying correspondence analysis are mass, profile, and chi-square distance. Assume that the cross-tabulated data under examination are described formally by matrix $F = \left[f_{ij} \right]$ with size $I \times J$. The correspondence matrix P is denoted as a matrix in which all elements of F are divided by the grand total n, which is $P = 1/n F$. Next, row and column summaries of the *correspondence matrix* are defined

$$p_{i+} = \sum_{j=1}^{J} p_{ij}, \, p_{+j} = \sum_{i=1}^{I} p_{ij}$$

(called *mass* in correspondence analysis). The respective row and column *profiles* of P are defined as $q_{ij}^r = p_{ij}/p_{i+}$ (row profile), $q_{ij}^c = p_{ij}/p_{+j}$ (column profile).

Here, *the chi-square distance* between row i and row k, $d_{(i.k)}$ is denoted as

$$d_{(i,k)} = \sqrt{\sum_{j=1}^{J} \left\{ q_{ij}^r - q_{kj}^r \right\}^2 \Big/ p_{+j}} \qquad (1)$$

The chi-square distance for column elements can be calculated using column profiles in a way analogous to

that used to calculate the chi-square distance for row elements.

In this study, the "ca" function of the ca package in the R language [17] was used for correspondence analysis. This function employs the singular-value decomposition (SVD) as a solution for the correspondence analysis. The ca package outputs the scores for the row and column elements and the cumulative contribution ratio for each axis based on the chi-square distances. As shown by Equation (1), the chi-square distances can be calculated for rows and columns separately. Accordingly, the scores are based on the scalar products of the row vectors and column vectors, which depend on the lengths of the vectors and the angles between them rather than the absolute distance between the vectors [19].

3.2. Results of Correspondence Analysis

The result obtained from the questionnaire was used as input data for this analysis (**Table 2**). This result can be visualized in a two dimensional scatter plot (**Figure 1**). In this figure, the top and right axes represent the land use categories (column elements in **Table 2**). The bottom and left axes represent the score of the facility classes (row elements in **Table 2**). The contribution ratios corresponding to each axis are 53.91% for the first (vertical) axis and 34.02% for the second (horizontal) axis. The cumulative contribution ratio is 87.93% in two-dimensional visualization. Although the first axis is generally plotted horizontally, the second axes were rescaled for improved visual quality in this study.

In **Figure 1**, the land use categories "Public facility", "Others", and "Commercial facility" are arranged in a straight line along the horizontal axis. "Residence" is located on the opposite side from the three categories and above them on the vertical axis. The contribution ratio of the vertical axis is larger than that of the horizontal axis. "Residence" is thus clearly distinguished from the other three categories. In other words, this can be considered that "Residence" is far from other categories in regard to semantic similarity. The facility classes located around "Residence" are "Apartment complex", "Student dormitory", and "Dormitory for Diet members", all of which have function of housing. In addition, "Welfare house for aged" is closer to these classes than are the other facility classes. In Japan, welfare houses for the aged is in a variety of types, such as nursing care facilities or lifetime care service facilities, but the housing function is common to all. In fact, the primary function of "Welfare house for aged" is not simply housing, whereas "Apartment complex", "Student dormitory", and "Dormitory for Diet members" serve as primarily housing. Therefore it can be considered that "Welfare house for aged" is located far from "Residence" in the plot than these other facility classes in virtue of priority of housing function.

The "Others" category was used if a facility class could not be classified into the other three categories. "Others" is located in the middle between "Public facility" and "Commercial facility". Therefore, it appears to classify a facility class into "Others" in case of not decide whether to classify the facility as a "Public facility" or a "Commercial facility". "Grave site", "Botanical garden", and "Child care center" are located around "Others."

Table 2. Results from the questionnaire.

|	P	C	R	O
Shopping center	0	60	0	0
Wholesale market	1	54	0	5
Japanese style hotel	1	47	1	11
Apartment complex	2	0	57	1
Driving school	5	30	0	25
Private tutoring school	5	34	0	21
Stock exchange	6	40	0	14
Student dormitory	6	3	47	4
Bank	6	37	0	17
Shinkin bank	7	38	0	15
Movie theater	7	49	0	4
Amusement park	8	46	0	6
Dormitory for Diet members	11	1	44	4
Broadcast station	11	28	0	21
Public bath	12	41	0	7
Aquarium	12	43	0	5
Cemetery	12	4	0	44
Vocational colleges	14	24	0	22
Child care center	14	23	0	23
Horse race track	16	36	0	8
Welfare house for aged	17	12	12	19
Temple, Shrine	21	0	2	37
Botanical garden	25	29	0	6
Clinic	30	16	0	14
Electric plant	30	5	0	25
Kindergarten	38	6	0	16
Medical center	41	9	0	10
Police box	43	1	0	16
Airport	43	9	0	8
Concert hall	48	9	0	3
Bus terminal	48	5	0	7
Public school	49	0	0	11
Art museum	51	6	0	3
Fire station	51	0	0	9
Railroad station	52	3	0	5
Athletic field	55	3	0	2
Park	57	0	2	1
Library	59	1	0	0
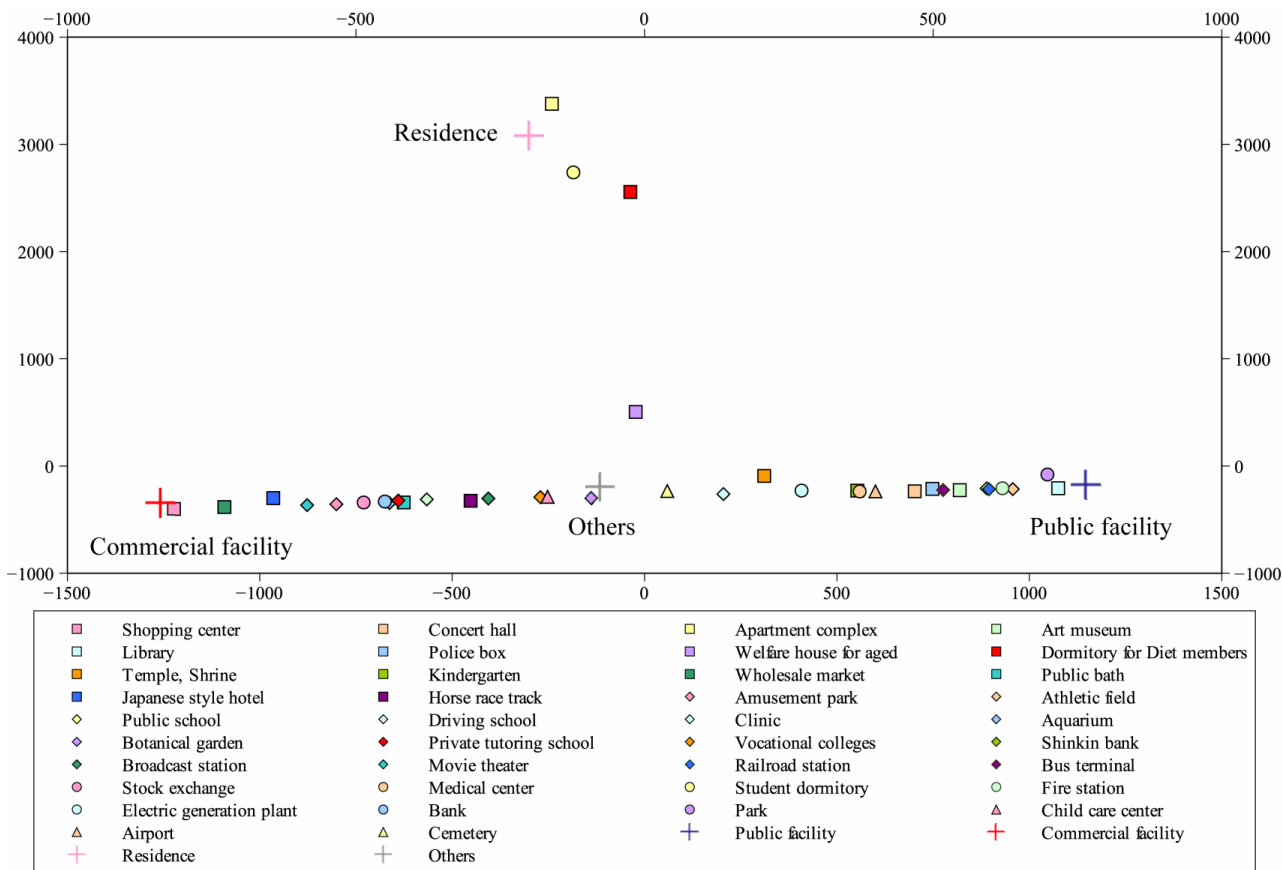
Figure 1. Scatterplot of the results from the correspondence analysis.

The facility classes located around "Public facility" include "Library," "Park," and "Athletic field". These facilities are available to all citizens free or low cost. In contrast, the facility classes located around "Commercial facility" include "Shopping center," "Wholesale market," and "Japanese style hotel". These facilities have economic activity as their primary function.

4. Determining Factors for the Land Use Categories

4.1. Categorization of the Facility Classes

To investigate the relationships between the land use categories and the facility classes, the facility classes were categorized in terms of establishment agent and establishment purpose (**Table 3**). The term "Establishment agent" was defined whether to use taxes to establish the facility. "Public" indicates taxes were used, and "private" indicates taxes were not used. The term "Establishment purpose" is used to categorize the primary function of these facilities. This categorization was developed based on a land use database named Digital Map 5000 (land use) published by the Geospatial Information Authority of Japan (GSI) to divide the facilities into more detailed categories.

4.2. Comparative Analysis Based on "Establishment Agent" and "Establishment Purpose"

Figure 2(a) shows the relationships between the land use categories and the facility classes in terms of "Establishment agent." This figure shows that "public" and "public/private" facilities concentrate around "Public facility" and "Residence" and "private" facilities are located around "Commercial facility" and "Others". In contrast, the relationships in terms of "Establishment purpose" show that there are several categories that are concentrated (**Figure 2(b)**). For example, the "Residence" characteristics category is concentrated around the "Residence" land use category, and "transportation" is concentrated around "Public facility".

The dominant determining factor for land use categories can be inferred from the relationships in terms of "Establishment agent" and "Establishment purpose." The three facility classes "Apartment complex", "Student dormitory", and "Dormitory for Diet members" are concentrated around the "Residence" land use category. Although their "Establishment agent" is not in common, their "Establishment purpose" is in common. This means that "Establishment purpose", not "Establishment agent",

is the dominant determining factor for the land use category "Residence". The housing function is a more strongly determining belonging to "Residence" than the use of taxes or no taxes to establish the facilities. Similarly, "Wholesale market" is close to "Commercial facility," although its "Establishment agent" is "public". Therefore, the facilities to be built for economic activity would be easily classified into "Commercial facility" irrespective of taxes use for its establishment.

5. Conclusions

This study presented an empirical investigation of common sense of land use categories using the questionnaire method and a statistical technique. Although the land use categories used in this study are limited, the several characteristics and distinctions of these land use categories were clarified. In addition, the dominant determining factors for the land use categories were confirmed. Although some
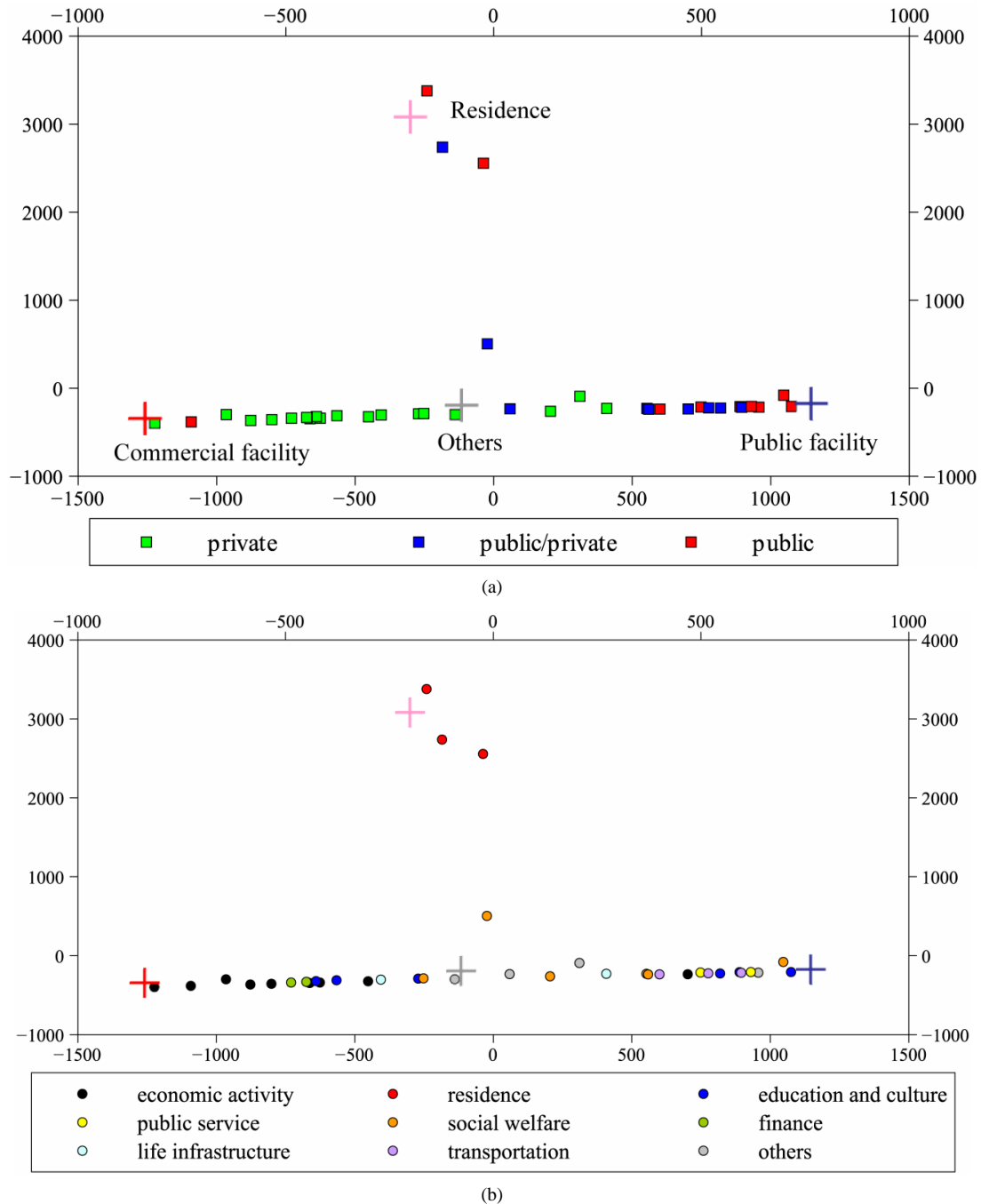

(a)

(b)

Figure 2. Scatterplots of the results of the correspondence analysis expressed in terms of "Establishment agent" (a) and "Establishment purpose" (b).

Table 3. Characteristics of the facility classes for determining factors of land use.

	Establishment agent	Establishment purpose
Shopping center	Private	Economic activity
Wholesale market	Public	Economic activity
Japanese style hotel	Private	Economic activity
Apartment complex	Public	Residence
Driving school	Private	Education and culture
Private tutoring school	Private	Education and culture
Stock exchange	Private	Finance
Student dormitory	Public/Private	Residence
Bank	Private	Finance
Shinkin bank	Private	Finance
Movie theater	Private	Economic activity
Amusement park	Private	Economic activity
Dormitory for Diet members	Public	Residence
Broadcast station	Private	Life infrastructure
Public bath	Private	Economic activity
Aquarium	Private	Economic activity
Cemetery	Public/Private	Others
Vocational colleges	Private	Education and culture
Child care center	Private	Education and culture
Horse race track	Private	Economic activity
Welfare house for aged	Public/Private	Social welfare
Temple, Shrine	Private	Others
Botanical garden	Private	Others
Clinic	Private	Social welfare
Electric plant	Private	Life infrastructure
Kindergarten	Public/Private	Social welfare
Medical center	Public/Private	Social welfare
Police box	Public	Public service
Airport	Public	Transportation
Concert hall	Public/Private	Economic activity
Bus terminal	Public/Private	Transportation
Public school	Public	Education and culture
Art museum	Public	Education and culture
Fire station	Public	Public service
Railroad station	Public/Private	Transportation
Athletic field	Public	Others
Park	Public	Others
Library	Public	Education and culture

land use data were published by national survey institution, the definitions of land use category are not standardized. Therefore the data integration between these data is much difficult. This result can contribute to develop land use ontologies in Japanese community, and to standardize the definitions of land use categories.

Egenhofer and Mark proposed a naive geography con-

cerned with formal modeling of common sense about the geographic world and the design of GIS applications and services for people without special geographic sense or training [20]. Contributions to naïve geography would be made by various types of research achievements, such as the perception and cognition of space and studies of the relationship between natural language and perceptual representation [21]. Although this study is a case study of limited land use categories, advanced studies in related common sense areas would serve to contribute applications and services based on naïve geography.

REFERENCES

[1] P. D. Donato, "Geospatial Semantics: A Critical Review," *Proceedings of* 10*th International Conference on Computational Science and Its Applications* (*ICCSA*), Fukuoka, 23-26 March 2010, pp. 528-544. doi:10.1007/978-3-642-12156-2_40

[2] J. Brodeur, Y. Bedard, G. Edwards and B. Moulin, "Revisiting the Concept of Geospatial Data Interoperability within the Scope of Human Communication Processes," *Transactions in GIS*, Vol. 7, No. 2, 2003, pp. 243-265. doi:10.1111/1467-9671.00143

[3] N. Guarino, "Formal Ontology, Conceptual Analysis and Knowledge Representation," *International Journal of Human and Computer Studies*, Vol. 43, No. 5-6, 1995, pp. 625-640. doi:10.1006/ijhc.1995.1066

[4] P. Agarwal, "Ontological Consideration in GIScience," *International Journal of Geographical Information Science*, Vol. 19, No. 5, 2005, pp. 501-536. doi:10.1080/13658810500032321

[5] N. Schuurman, "Formalization Matters: Critical GIS and Ontology Research," *Annals of the Association of American Geographers*, Vol. 96, No. 4, 2006, pp. 726-739. doi:10.1111/j.1467-8306.2006.00513.x

[6] J. Brodeur, Y. Bédard, G. Edwards and B. Moulin, "Revisiting the Concept of Geospatial Data Interoperability within the Scope of Human Communication Processes," *Transactions in GIS*, Vol. 7, No. 2, 2003, pp. 243-265. doi:10.1111/1467-9671.00143

[7] F. T. Fonseca, M. J. Egenhofer, P. Agouris and G. Câmara, "Using Ontologies for Integrated Geographic Information System," *Transactions in GIS*, Vol. 6, No. 3, 2002, pp. 231-257. doi:10.1111/1467-9671.00109

[8] D. M. Mark and A. G. Turk, "Landscape Categories in Yindjibarndi: Ontology, Environment, and Language," *Proceedings of* 03' *Conference on Spatial Information Theory*, Ittingen, 24-28 September 2003, pp. 28-45. doi:10.1007/978-3-540-39923-0_3

[9] D. M. Mark, A. G. Turk and D. Stea, "Progress on Yindjibarndi Ethnophysiography," *Proceedings of* 07' *Conference on Spatial Information Theory*, Melbourne, 2007, pp. 1-19. doi:10.1007/978-3-540-74788-8_1

[10] D. M. Mark, B. Smith and B. Tversky, "Ontology and Geographic Objects: An Empirical Study of Cognitive Categorization," *Proceedings of* 99' *Conference on Spatial Information Theory*, Hamburg, 25-29 August 1999,

pp. 283-298. doi:10.1007/3-540-48384-5_19

[11] B. Smith and D. M. Mark, "Geographical Categories: An Ontological Investigation," *International Journal of Geographical Information Science*, Vol. 15, No. 7, 2001, pp. 591-612. doi:10.1080/13658810110061199

[12] A. Frank, "Multi-Cultural Aspect of Spatial Knowledge," *Proceedings of* 3*rd International Conference on Geospatial Semantics*, Mexico City, 3-4 December 2009, pp. 1-8. doi:10.1007/978-3-642-10436-7_1

[13] N. Schuurman, "Formalization Matters: Critical GIS and Ontology Research," *Annals of the Association of American Geographers*, Vol. 96, No. 4, 2006, pp. 726-739. doi:10.1111/j.1467-8306.2006.00513.x

[14] M. M. Spiegel and N. Yamori, "Dterminations of Voluntary Bank Disclosure: Evidence from Japanese Shinkin," *CESifo Working Paper*, No. 1135, 2004, p. 39. http://www.cesifo-group.de/portal/page/portal/ifoHome/b-publ/b3publwp/_wp_by_number?p_number=1135

[15] FAO, "Land Cover Classification System (LCCS): Classification Concept and User Manual," FAO, Rome, 2000.

[16] L. Doey and J. Kurta, "Correspondence Analysis Applied to Psychological Research," *Tutorials in Quantitative Methods for Psychology*, Vol. 7, No. 1, 2011, pp. 5-14.

[17] O. Nenadi and M. Greenacre, "Correspondence Analysis in R, with Two- and Three-Dimensional Graphics: The ca Package," *Journal of Statistical Software*, Vol. 20, No. 3, 2007. http://www.jstatsoft.org/v20/i03

[18] P. M. Yelland, "An Introduction to Correspondence Analysis," *The Mathematica Journal*, Vol. 12, 2010. http://www.mathematica-journal.com/2010/09/an-introduction-to-correspondence-analysis/

[19] M. Greenacre, "Correspondence Analysis in Practice," Academic Press, London, 1993.

[20] M. J. Egenhofer and D. M. Mark, "Naïve Geography," *Proceedings of* 95' *Conference on Spatial Information Theory*, Semmering, 21-23 September 1995, pp. 1-15. doi:10.1007/3-540-60392-1_1

[21] B. Jiang and X. Yao, "Location-Based Services and GIS in Perspective," *Computers*, *Environment and Urban Systems*, Vol. 30, No. 6, 2006, pp. 712-725. doi:10.1016/j.compenvurbsys.2006.02.003