Understanding Mental Categorization on a VGI System to Improve Data Management

João Vitor Meza Bravo*, Claudia Robbi Sluter*, Fernando Luiz de Paula Santil, Luciene Stamato Delazari*, Monica Cristina de Castro***

* Federal University of Paraná, Brazil; contact: jvmbravo@gmail.com
** State University of Maringá, Brazil

Abstract. This paper describes the initial efforts for providing knowledge support to VGI system managers, whose interest lies upon decision making about selecting or deleting geocategories. We have picked the Wikimapia® as an example, in accordance with the Goodchild's description of VGI systems, in order to generate a model for decision making based on analysis of mental categorization theory and structures of reasoning. Besides, considering that Wikimapia®'s content is organized in geo-categories, these data can supply a rich analysis of mental geocategorization. We have selected four touristic places which received a great amount of visitors per year to test our hypothesis: the more a place is visited the more fitted for use is the information about it. The results pointed out that there are informative and distinctive categories describing these selected places and the categories are more or less representatives depending on amount of visitors. Distinctive categories implies higher level of knowing about described place and lower abstraction while Informative categories implies lower level of knowing about place and high level of abstraction (ROSCH, 1975). Regarding these concepts, the main goal was to identify which categories are the most representatives. This can improve the VGI systems usability whereas data manager could select those categories achieving higher and lower levels of users' abstraction. Similarly, we focused on explain how the Wikimapia® users employed representative categories to describe features accordingly with its amount of visitors.

Keywords: VGI, Mental Categorization, data management, partonomy, taxonomy.

1. Background

The goal of this research was to provide knowledge support for VGI system managers at the moment they will have to make a decision about selecting or deleting geoinformation posted by users. VGI - Volunteered Geographic Information is defined by Goodchild (2007) as "citizens as sensors". In this new environment, anyone can post his/her own spatial information on the web. That means VGI users are potential mapmakers and their role as map users has changed: nowadays they can also produce geographic information.

In VGI systems the user or "produser" (BUDHATHOKI et al, 2008) has an important role, which is describing the world as seen by his/her own eyes. Before web 2.0 (O'REILLY, 2007) (COMODE & KRISHNAMURTHY, 2008), this was a task done only by cartographers. At that time, the purpose of the studies related to the spatial cognition and its categorization was to understand what kind of mental strategies individuals used while reading maps (LLOYD *et al*, 1996) (MACEACHREN, 1995). Nowadays people interact with the newer systems and, as mentioned before, they are able to disseminate spatial information themselves.

One of the most common VGI systems is Wikimapia® (GOODCHILD, 2007). When people post geoinformation on it, they are invited to create both the feature and the category, which describes the first element. Creating categories is an important subject as it can negatively affect the results of searching for features and, therefore, the user's trust on the website if not successful in his use of the system.

Nowadays, the categories moderators (CM) can decide which categories remain in the systems and which ones will be deleted. The deleted categories are considered non-representative for a specific kind of feature and this decision is based on the CM's personal criteria. The lack of formal criteria on decisions making about category status can introduce a level of uncertainty to the system.

Taking into account the VGI characteristics of Wikimapia® and the importance of geoinformation reliability, we propose to investigate how the Wikimapia®'s data about features and categories could be used by its CM in defining criteria to select and delete categories. Our hypothesis is that, the more a place is visited the more accurate and, therefore, fitted for use, the information is. This fact could be used to know the mental categorization of the users related to their abstraction levels of knowledge (ROSCH 1973) and also their reasoning structures, explained by the partonomy and taxonomy concepts (Tversky and Hemenway, 1984), which are related to the spatial characteristics of the posted location. The abstraction level of knowledge of spatial information could be taken as a criterion for the CM to make a decision about selecting or deleting categories.

The theory proposed by Rosch (1973), in which she defines two levels of categorization: informative and distinctive, made this kind of decision possible as the CM could classify the categories based on the concepts of abstraction levels, as well as the frequency that a place is visited.

2. VGI systems and the Theories of Mental Categorization and Classification of Spatial Information

2.1. Volunteered Geographic Information systems

Goodchild (2007) coined the term "VGI" that is similar, in meaning, to "neogeography" (TURNER, 2007) and "collaborative mapping" (BALRAM & DRAGIĆEVIĆ, 2006). The idea of collaborative mapping has emerged on the Web 2.0 context (GOODCHILD, 2007). On the Web 2.0 era, "any participant can be a content creator" and "numerous technological aids have been created to maximize the potential for content creation" (COMODE & KRISHNAMURTHY, 2008).

VGI systems, as a kind of Web 2.0 service, allow the users to create and to handle their own spatial data. In Wikimapia®, the data posted by the users need to be described and linked to a specific kind of content. The users can also create data classes under which, some data with similar characteristics, will be assembled. Features with similar characteristics define a Wikimapia® category. The categories created by the users have a strong association with their mental processes of understanding the world's objects. There is a kind of user who is allowed to eliminate features, categories and texts. They can also punish the spammers and people who damage the information or the system (WIKIMAPIA®, 2012). For example, someone who is "3 Regular User" on Wikimapia® is allowed to accomplish more tasks than the "1" or "2 Regular User", as this user has a bigger amount of time of work in the process of creating or fixing features and categories.

The way the users describe the categories can difficult to search for some features. Therefore, keeping the system as easier as possible on this information search is an important matter. Creating categories is not a simple task. This action impacts the system usability and it relies only on the user´s skills and cognition.

On the Wikimapia® system, for example, the CMs are chosen by a community of high level users. This approach for managing the system aims providing reliability on the shared information. Despite the importance of the categorization of spatial information in the VGI's context, there is no research results on a similar issue in the specialized literature.

2.2. Mental Categorization Theories and Classification of Spatial Information

MacEachren (1995) pointed that "maps depict categories rather than individuals" and also "to make maps that work, we must depict categories using methods that match the structures of human mental categorization". Rosch (1973) argued that categories are natural and their creation depends on the human knowledge and interaction with objects. The basic-level categories are often the starting point when people classify objects and also the world in feasible structures of mental categorization. Spatial structures are independent from each other and the differences among their categories are related to the abstraction's level of each individual.

Rosch (1973) proposed three levels of information structures: the subordinate, the basic and the superordinate. In order to understand the basic level of geographic categories, their structures and formation, Lloyd et al (1996) interviewed Geography students living in different places and studying at different universities. The authors aimed to discover the relationship between the structures described by Rosch (1973) and the structures on the mental categorization that resulted on understanding the space.

For this research, it is helpful to use Lloyd et al (1996) examples in order to explain the real world categories. They categorized the term "Place" at the superordinate level; the terms "country", "state", "city" and "neighborhood" at the basic level; and, finally, "United States", "Canada", South Carolina", Columbia, "Charleston", and other geographic places, at the subordinate one. From their examples, it is possible to realize what kind of structures represent each one of these levels. There is a hierarchy associated with each one of these structures which are based on the abstraction's level or, as said by Lloyd et al (1996), "the superordinate level is the most distinctive and the subordinate level is the most informative".

There are structures of reasoning associated with these levels of categorization and their abstraction's level relies on people's knowledge while handling and classifying data. In order to store classified data in their mental structures, individuals need to build a hierarchical context. This way of processing information is quite useful in the categorization of VGI shared data as it could improve the results of comparing and discarding redundant information tasks.

The knowledge about classifying categories created by Rosch (1973), and discussed by Lakoff (1987) and Lloyd et al (1996), was the key to associate the classes of mental processes with those presented by Tversky and Hemenway (1984): taxonomy and partonomy. Tversky and Hemenway (1984) stated that "knowledge organization by parts" is called Partonomy and "organization by kinds", Taxonomy. Partonomic regionalization of space relates the process of discovering and analyzing space parts with characteristics in common. The creation of parts or regions is similar to the process of defining differences in a subordinate level, as defined by Rosch (1996). Different from partonomy, taxonomic regionalization corresponds to regrouping similar things and this is the most informative level, named by Rosch (1973) as basic-level.

Studying these structures and classifications allowed us to identify the processes used by the VGI users by the time they created categories of places analyzed on Wikimapia®. The places are four touristic points and are described in the next item, which also presents the method adopted during this research.

3. Methods

The first step of the methodology was to gather and analyze information from Wikimapia®. At this stage we adopted the Suchan & Brewer (2000) qualitative method, based only on documental analysis. We selected a set of geographic features in order to describe the analysis' method of this research. These selected features shared a common characteristic: they are broadly known as touristic points and are highly visited on Wikimapia®. To facilitate the verification of our hypothesis, we ranked the features by their amount of visits per year and we also searched for the frequency that the information about those features had been edited. The next step was to search for the categories related to them, and which are the actual touristic points.

With the purpose of verifying the possibility of knowing about the level of users' abstraction while they generated a category, it is important to understand how the categories are described in the Wikimapia® website. Therefore, we also sought to understand their mental reasoning at the right moment they classified the spatial data they posted or edited on Wikimapia®. The data set for the analysis of our hypothesis had been completed with the data on the amount of features stored in each category.

Consequently, to understand users' mental reasoning, we used the concepts of taxonomy and partonomy that describe the structures of reasoning behind the process of spatial knowledge acquisition (Tversky & Hemenway

1984). The last step of our research method was to classify the Wikimapia® categories based on the theory of basic-level categories proposed by Rosch (1973). Thus, the geographic places were understood in accordance with a semantic reasoning and were classified by their meanings (Brown, 1958; Rosch, 1973; Lloyd et al, 1996).

4. Discussion

The table 1 shows the data used to verify our hypothesis: an estimated amount of visits per year to the selected places (geographic feature), the categories related to the features, and how many times each of those features were edited. The places we have selected for this research were: the Eiffel Tower, the Statue of Liberty, the Christ of Redeemer, the Botanical Garden (Curitiba, Brazil).

Place	Localization	Associated Categories	Estimation visits/year*	Information edition (amount)
Eiffel Tower	Paris (France)	tower, eiffel, antenna, interesting place	6.200.000	276
Statue of Liberty	New York – (USA)	statue, museum, monument, place with historical importance, donated / gifted object, interesting place, UNESCO World Heritage Site	3.800.000	88
Christ the Redeemer	Rio de Janeiro (Brazil)	statue, monument, panoramic view, interesting place	2.000.000	84
Botanical Garden	Curitiba (Brazil)	botanical garden	750.000	12

Table 1. Data used for the hypothesis' analysis.

*The estimation of visits per year was provided by the parks administration and were based on the last five years.

The result of the data classification based on their relationship with the structures of reasoning and the abstraction's levels is showed on table 2. On the column 'Categories (number of features)', we present the terms used as categories of Wikimapia® and the amount of features related to each category. Each of those categories was associated to an abstraction level, as defined by Rosch (1973). The abstraction's levels are related to different structures of reasoning, from partonomy to taxonomy. Taking 'Eiffel Tower' as an example, we discovered that the terms 'interesting place', 'antenna', 'tower', 'Eiffel' are related to it. The amount of features correlated to a category is an indicator of its representativeness, so the more a number of features the more representative the category is.

Structures of reasoning	Levels of Abstraction	Categories (number of features)
	EIFFEL TOWER	
Taxonomy	Superordinate	Interesting place (180692)
	Basic	Antenna (2569), tower (12835)
Partonomy	Subordinate	Eiffel (12)
	STATUE OF LIBERTY	
Taxonomy	Superordinate	Place with historical importance (59750), donated / gifted object (46), interesting place (180692)
	Basic	Statue (5575), museum (51082), monument (32654)
Partonomy	Subordinate	*Null*
	CHRIST THE REDEEMER	
Taxonomy	Superordinate	Panoramic view (3314), interesting place (180692)
	Basic	Statue (5575), monument (32654)
Partonomy	Subordinate	*Null*
	BOTANICAL GARDEN	
Taxonomy	Superordinate	*Null*
	Basic	Botanical garden (1051)
Partonomy	Subordinate	Botanical garden (1051)

Table 2. Features, Categories, Structures of reasoning and the Classification of categories by their levels of abstraction.

At table 2, the gray scale represents the amount of information processed at different structures of reasoning. Therefore, partonomy is represented by a lighter gray and taxonomy by a darker grey. This means we used a smaller amount of information when we developed partonomic reasoning than when the reasoning was taxonomic. These structures of reasoning are the way people organize the information about geographic features. A partonomic classification reveals characteristics of parts of the objects from the world. These characteristics are more distinctive than informative and, at partonomic level of reasoning, people are interested in describing these parts of the world with the same properties. For example, if the category 'Eiffel' was created from a partonomic reasoning of the geographic region as there are fewer features sharing this particular characteristic, then if some people want to link another feature to this very same category, they should be geographically near to the category 'Eiffel'. When the parts are near to each other, their cores have similar properties (Tversy and Hemenway, 1984).

Taxonomic reasoning is related to informative categories and, hence, allows the users to accumulate data of different geographic places into the same category. Taxonomic reasoning is based not only on the geographic distribution of a phenomenon but also on its abstract meaning. Consequently, the term “Interesting Place” does not reveal any particular characteristic of features that are related to it. These features are not necessarily near to each other, and their cores do not necessarily share the same properties (Tversy and Hemenway, 1984).

Based on the meaning of partonomy and taxonomy (Tversy and Hemenway 1984), and also on the levels of abstraction (Rosch 1973), we reasoned the classification based on partonomic reasoning with the subordinate level, and the taxonomic reasoning with the superordinate level (Table 2). For instance, in the categories of Eiffel Tower (Table 2), the term “interesting place” represents a more informative than a distinctive category as many places can be understood as an interesting one to be visited and this conclusion is supported by the number of features linked to it, e.g. 180692 by the time we wrote this paper. This same term, “interesting place”, is also linked to the categories named as Statue of Liberty and Christ the Redeemer, which are places where there are a great amount of visits per year. The terms “Place with historical importance", "donated/gifted object", "Panoramic view” were also classified under informative categories as they could be related to any place and the way a person links them to a specific place is context-dependent. For example, a “Seamark” and “the biggest building on a city” can be classified into the category “panoramic view” and, if they are historical marks of a specific region, they can be also classified as “a place with historical importance”. So, there are many places that could be regrouped on those two categories and this is the reason we understood them as informative categories.

The categories named as “Antenna”, “Tower”, “Statue”, “Monument”, did not provide us with any specific information about a place or a feature. We can visualize, in our minds, a shape or an image of a concept, or even an object of the world, independently on the context people handle them in Wikimapia®. These terms are related to every one of those places we used as our research data, except the Botanical Garden. This fact led us to a conclusion that not only people are interested in describing the features (distinctive category), but they are also concearned in creating an informative category that will allow others to find the same place.

For the Botanical Garden we could not see any informative category, but a distinctive one. According to its administration, the place is unknown for most of people who live in Curitiba and it is only known by a specific group of tourists and visitors. There are many botanical gardens in the world, what can be proved by the number of places linked to this category, e.g. 1051 (Table 2). One of those is located in this Brazilian city and it does not have a particular name, being called only as Botanical Garden. Then, taking into account not only the number of visits to this very place in Wikimapia® but also that this is not a well-known place even for the locals, we classified it at a basic-level in addition to the subordinate level.

This analysis of categories and its related abstraction's level on the information posted on Wikimapia® could be used by the CMs when they will need to choose one category, among some of them to be kept in the website. This decision could be based on a simple question: assuming the user is a tourist, which terms he/she would use on a search task? For example, a tourist could use the term "interesting place" if he/she wants to know about a generic place. Besides, he/she is not interested in a "place with a historical importance" or "donated / gifted object". Considering that the terms "interesting place", "place with a historical importance" and "donated / gifted object" are classified at the same level of abstraction, which is superordinate level, the CM could conclude that an "interesting place" is a more informative category than the others. Therefore, at the informative level, the CM could only keep the category "interesting place" as it is more representative of this abstraction level.

At the basic-level, the Wikimapia® user needs to visualize an image of a place and this visualization does not need to be accurate. So, the words "Antenna", "Monument", "Tower", "Statue", "Museum" are appropriated terms to describe the features at a basic-level. Botanical garden can also be classified at a basic-level because, as mentioned above, there are many botanical gardens in the world. This is a specific but also a generic term that describes a place. However, as "botanical garden" is the name of a specific place, we could state that this is a term used at a subordinate level as well. The term "Eiffel" is a more distinctive category. For example, if someone wants to know an "interesting place", that could be a "tower" and this tower could be the “Eiffel Tower”, that makes this tower a specific one. On Figure 1, we illustrated the abstraction levels about the Eiffel Tower and the mental images that could be associated with it at each level.

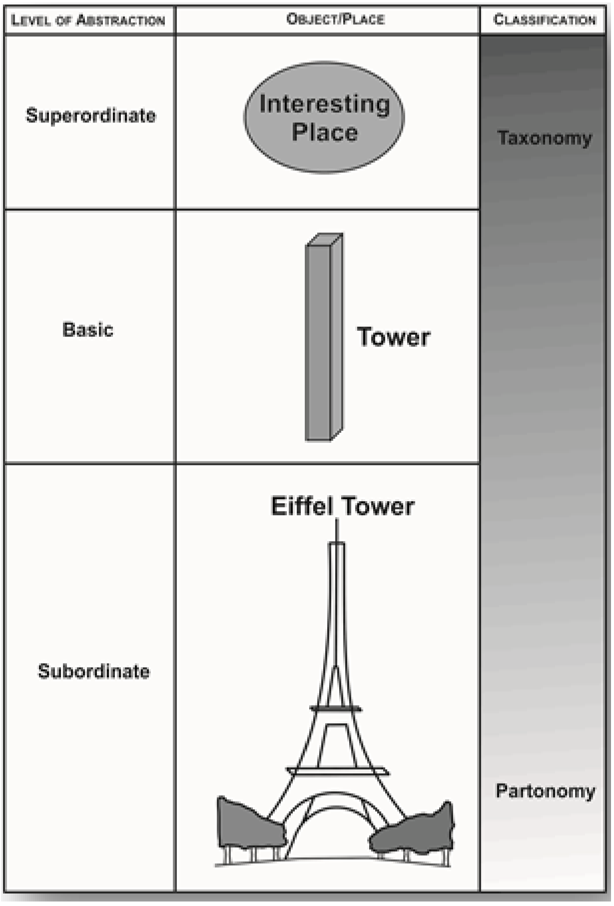

Figure 1. Scheme of Eiffel Tower categories at Wikimapia®.

Therefore, the quality of a created category will depend on its nature but also on the subjects skills. If the category is representative or refined, then only a special attention from CM could validate the user’s input. The flow-chart (Figure 2) shows this sketch.

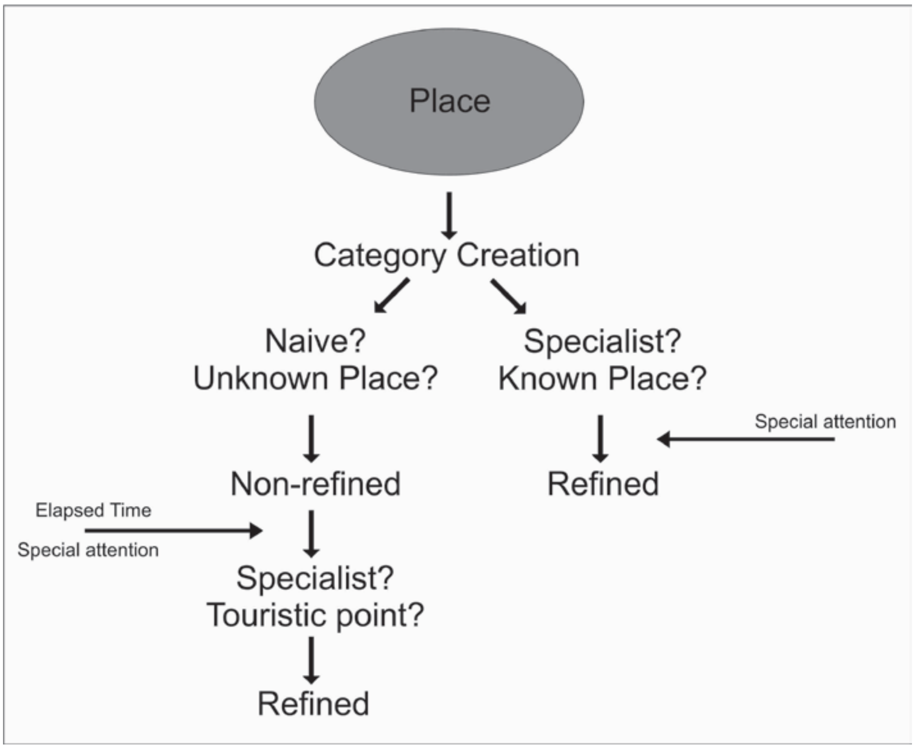

Figure 2. Flowchart of analysis.

5. Conclusion

The analysis of the results and the discussion of the hypothesis showed us that the efforts of a CM on deciding about the representative level of any kind of category could be minimized and the results could be improved. One possible way to achieve the improvement of the results is to use theories of mental categorization and classification of spatial information, as a knowledge support for decisions making.

Creating categories mean a lot to the system's usability as individuals are advised by CM to be careful when using this system tool. The possibility for users to create categories in Wikimapia® allowed us to study spatial categorization and cognition based on the data people generated by their own resources. That is to say that the analysis of some characteristics of the information posted on Wikimapia® made possible to us to understand how the users employ their mental categorization strategies while creating maps. This understanding could assist CM of Wikimapia® in making better choices while they have to decide which categories are to be related to the stored features.

References

Brown R (1958) How shall a thing be called? Psychological Review: 1 (65).

Budhathoki N R, Bruce B C, Nedovic-Budic Z (2008) Reconceptualizing the role of the user of spatial data infrastructure. GeoJournal: (72).

Cormode G, Krishnamurthy B (2008) Key differences between Web 1.0 and Web 2.0. First Monday: 6 (13).

Golledge R, Stimson R J (1987) Analytical Behavioural Geography. New York: Cioom Helm.

Goodchild M F (2007) Citizens as sensors: the world of Volunteered Geography. GeoJournal: (69).

Lakoff G (1987) Woman, fire, and dangerous things: what categories reveal about the mind. Chicago: University of Chicago Press.

Lloyd R, Patton D, Rex C (1996) Basic-Level Geographic Categories. The Professional Geographer: 2 (48), p. 181–194.

MacEachren A M (1995) How maps work: representation, visualization, and design. New York: The Guilford Press.

O'Reilly T (2007) What is Web 2.0: Design patterns and business models for the next generation of software. Communications & Strategies: 1 (65).

Rosch E (1973) Natural categories. Cognitive Psychology: (4), p. 328-350.

Rosch E (1975) Cognitive representations of semantic concepts. Journal of Experimental psychology: General: 3 (104), p. 192-233.

Rosch E, Mervis C, Gray W, Johnson D, Boyes-Braem P (1976) Basic objects in natural categories. Cognitive Psychology: 8, p. 382-439.

Shivanand B, Dragievi S (2006) Collaborative Geographic Information Systems. Idea Group Publishing, United States.

Turner A (2007) Neogeography - towards a definition. A weblog posting. Available at: http://highearthorbit.com/neogeography-towards-a-definition/, accessed in august, 2012.

Tversky B, Hemenway K (1984) Objects, Parts, and Categories. Journal of Experimental Psychology: General: 2 (113)

Wikimapia® (2012): http://wikimapia®.org . Accessed 20 october, 2012.