Enhancing citizen science participation in Geoweb projects through the instance-based data model

Roman Lukyanenko[1], Jeffrey Parsons[2], Yolanda Wiersma[3]

1 Faculty of Business Administration, Memorial University, rl1087@mun.ca
2 Faculty of Business Administration, Memorial University, jeffreyp@mun.ca
3 Department of Biology, Memorial University, ywiersma@mun.ca

Abstract

The increased popularity of user-supplied content on the Internet provides an opportunity to convert volunteered user information into a scientifically viable resource. Geoweb is a technology that combines geographic and user supplied data online. Yet, broader participation by members of the public in Geoweb is often constrained by varied levels of domain expertise of potential project participants. Further, if a Geoweb project is to have scientific utility, an additional constraint maybe placed upon participation – conformity with conceptual requirements of a given scientific domain. In this paper we propose an instance-based model of data collection and citizen science knowledge integration with scientific domain. To test the assumptions of the model, we propose an experiment to: (1) test of the applicability of basic-level categories to citizen science observations and (2) to determine if subjects are able to describe a phenomenon using a small number of observable features.

Background and Relevance

User-supplied content on the Internet is growing at a staggering rate (Anderson 2007). Projects like Facebook, Wikipedia, YouTube, and Twitter, have millions of users. Users are increasingly transitioning from a mere audience to creators and shapers of the online content. The willingness of people to share information has the potential to facilitate engagement of citizen scientists online (Sieber 2006). According to Goodchild, for example, humans are effective sensors of biological and ecological change (Goodchild 2007). New Internet tools further allow users to reference their observations to a specific geographic location. This is the premise of Geoweb projects like OpenStreetMap, Google Maps, Wikimapia. A number of Geoweb projects have an explicit scientific component, such as eBird (www.ebird.com), E-Flora BC (www.geog.ubc.ca/biodiversity/eflora/), and iSpot (http://ispot.org.uk/).

Yet, a traditional approach to citizen participation in scientific data collection tends to be constraining due to the requirement to understand the language, structure and concepts of the scientific domain. In cases where a phenomenon needs to be identified, such as in many Geoweb projects with a biological component, traditional approaches rely on the need for a positive identification of species. These requirements may be unrealistic for novice volunteers, and limit the scope of potential participation.

To combat this limitation, we propose an approach to data collection and storage that does not require users to identify and classify observed phenomena. Instead, users will have an option to record any observable attributes associated with the sighting. The attributes will also serve as a bridge between the scientific domain of the project and representation of reality by its non-expert participants, thus allowing crucial information to be communicated and collected. We propose a data collection interface based on Bunge's ontology (Bunge 1977) and cognitive theory of basic categories (Rosch, Mervis et al. 1976).

By shifting the focus from a predefined classification to the instance and its attributes, we do not need to model a domain *a priori* in terms of classes of interest (Parsons and Wand 2000). It is sufficient to ensure that the application has a comprehensive collection of classes, and each class contains a set of qualifying attributes (properties of a phenomenon that establish its membership with possible classes). However, despite the theoretical power of the instance-based model there is limited research that deals with its practical implementation. In particular the issue of effective attribute management remains unaddressed. With potentially large number of attributes attached to each instance, collecting, storing and presenting attributes can be a challenge. Rosch et al. (1976) noted the reality is made of an "infinite number of discriminately different stimuli" (p. 382). This research further proposes an adoption of a cognitive concept of basic categories as a filter to manage large numbers of attributes.

In order for an instance-based approach to be effective, two broad assumptions should hold true: (1) users identify unknown phenomena using basic categories, and (2) there is a short list of identifying attributes of a phenomenon that users are capable of observing. In order to test these assumptions, we propose an experiment to verify the applicability of the basic categories to the online context of Geoweb applications. We also wish to determine if users observe a small number of identifying features of a phenomenon.

Methods

The subjects of the experiment are first year business students from Memorial University. The choice is justified by the diversity of the backgrounds as the first year business course is an elective for many undergraduate programs. Bailenson et al. warned about a tendency of undergraduate students to use scientific biological taxonomy when classifying phenomena (Bailenson, Shum et al. 2002). However, we argue that students of the first year business course are likely to have limited exposure to biological nomenclature. In addition we will control for prior taxonomoic knowledge via a pre-experimental questionnaire that asks for details on background knowledge in biology. The stimulus set contained 30 images of plants and animals likely to be encountered by citizen scientists. The stimuli belong to a variety of biological genera. Some contain relatively obscure and unknown species, such as boreal felt lichen, while others are more charismatic – bald eagles, moose, coyotes. Not all stimuli, therefore, can be positively identified by all subjects, which is a realistic representation of various levels of citizen science expertise (Coleman, Georgiadou et al. 2009). In the experiment, the subjects were instructed to identify or describe the species in the image displayed. These descriptions can be in terms of categories or attributes. The talk will present the findings of the experiments that offer support for the instance-based approach to citizen science participation.

Conclusions and Implications

The result of the experiment has implications for developing participatory models of citizen science. By testing ways users identify known and unknown phenomena in the context of a Geoweb project, we have gained experimental support for the instance-based attribute model. People tend to recognize basic categories when faced with both familiar and unfamiliar objects. Basic categories become a useful entry point of a citizen science interface. The intuitiveness of basic categories streamlines the process and increases interface usability. The interface is then designed around a set of attributes that permits inferences about subcategory membership and guide users through identification process. This approach allows for a broader audience to engage in scientifically-oriented projects, as non-experts do not need to know exactly the phenomenon that was observed. This has far-reaching implications for the scientific community

as data generated by many “eyes on the ground” increases the likelihood of rare or unusual species being detected. Some potential uses of data collected this way might be unanticipated. For example, long term data can be useful to identify benchmark conditions in the event of a natural or anthropogenic disaster (e.g., the Gulf oil spill, climate change issues etc.) and can guide preventive and restoration strategies.

References

Anderson, P. (2007). "All that glisters is not gold' - Web 2.0 and the librarian." Journal of Librarianship and Information Science **39**(4): 195-198.

Bailenson, J. N., M. S. Shum, et al. (2002). "A bird's eye view: biological categorization and reasoning within and across cultures." Cognition **84**(1): 1-53.

Bunge, M. A. (1977). The furniture of the world. Dordrecht ; Boston, Reidel.

Coleman, D. J., Y. Georgiadou, et al. (2009). "Volunteered Geographic Information: The Nature and Motivation of Producers." International Journal of Spatial Data Infrastructures Research **4**: 332-358.

Goodchild, M. (2007). "Citizens as sensors: the world of volunteered geography." GeoJournal **69**(4): 211-221.

Parsons, J. and Y. Wand (2000). "Emancipating Instances from the Tyranny of Classes in Information Modeling." ACM Transactions on Database Systems **25**(2): 228.

Rosch, E., C. B. Mervis, et al. (1976). "Basic Objects in Natural Categories." Cognitive Psychology **8**(3): 382-439.

Sieber, R. (2006). "Public participation geographic information systems: A literature review and framework." Annals of the Association of American Geographers **96**(3): 491-507.