Developing Community Phenotype Ontologies: Understanding Users' Preferences

Limin Zhang, Zuleima Cota, Hong Cui
University of Arizona
Tucson, AZ, U.S.A
liminz, zcota,
hongcui@email.arizona.edu

Hsin-liang Chen
Missouri University of
Science & Technology
Rolla, MO, U.S.A.
chenhs@mst.edu

Bruce Ford
University of Manitoba
Winnipeg, Canada
bruce.Ford@umanitoba.ca

Joel Sach, Jocelyn Pender, James Macklin
Agriculture Agri-Food Canada
Ottawa, ON, Canada
Joel.sach, jocylin.pender,
james.macklin@aaf.ca

Anton Reznicek
University of Michigan
Ann Arbor, MI, U.S.A.
reznicek@umich.edu

Julian Starr
University of Ottawa
Ottawa, MB, Canada
julian.starr@uottawa.ca

ABSTRACT

This poster reports preliminary user-testing results on four different methods to add terms to a phenotype ontology. A total of 31 graduate students from UA iSchool and three senior botanists participated in two different experiments. Results suggest the Quick Form and WebProtege are preferred by biologists and WikiData and Wizard are not preferred for different reasons.

Author Keywords

Ontology construction; user studies; phenotype data management; biodiversity informatics

ASIS&T Thesaurus

Knowledge organization systems, bioinformatics, user generated content.

INTRODUCTION

Phenotypes are critical for describing species, studying function, and understanding organismal evolution, but only a very small amount of descriptive data are provided with clear semantics via ontological annotations. The lack of such computable data is due to the high cost of manual annotation, incomplete phenotype ontologies, and high inter-curator variations. We are investigating ways to enable biologists to directly contribute to ontology construction with the purpose of producing semantically clear computable phenotype data for large scale biological projects. Here we report preliminary findings on users' preferences for adding terms to their research community.

ontologies. Four methods were evaluated, Quick Form, Wizard, WikiData, and WebProtege. These methods cover the full range of the ontology construction landscape, from simple web-based forms, to a set of guided questions, an open knowledge base used by some biology ontologies, and a well-known editor designed for an ontology engineer. Graphic-based ontology editors were not selected for this user study because our biologist collaborators had ruled out these options in our first project meeting.

RELATED WORK

Several other projects share the same goal as ours – to enable biological authors to produce computable data. These include TaxonWorks (http://taxonworks.org/) and Morph*D*Base (https://www.morphdbase.de/), but neither conducted usability research. Prior work that compared the usability of different ontology editors (e.g., Norta et al., 2010; Khondoker, 2010; Alatrish, 2013;) either tested only ontology experts or used tasks unconnected to phenotypes.

RESEARCH QUESTION

What are biology users' relative preferences among Quick Form, Wizard, WikiData, and WebProtege, in terms of the usability, the support for recording the full semantics of a term, and biology users' confidence in their ability to use t method(s)?

RESEARCH DESIGN

The user interface of the four methods evaluated are at http://shark.sbs.arizona.edu/add2ontologymodular/public/leaf/hong/carex.

Thirty-six participants were recruited from a required graduate-level course in the UA iSchool and assigned arbitrarily into a Latin Square with four groups. Only completed the experiment. Participants first filled out a survey regarding their experience with controlled vocabulary editors and wikis. After watching a 3-6 minute video demo of each method, they rated them using a 5-point Likert-Scale in terms of their usability, their support for recording full semantics of a term, and the user's confidence in her ability of use the methods. Around three days later, they completed a hands-on task adding the term "leaf blade" to an ontology with over 2000 terms, followed by rating these methods again using the same scale. This task involved adding leaf blade's synonym, part of, and has part relationships with other existing terms or new terms.

In addition to the student experiment, three botanists at different career stages also participated in a group think-aloud experiment using the same "leaf blade" task to rate the four methods.

This usability study employs common biological knowledge that is familiar to the general public. More studies will be conducted to involve more advanced biology knowledge with biology students and other botanists as participants. Participants with diverse backgrounds help reveal different usability issues.

RESULTS

The sum of scores (ranks) of all 31 participants on the three aspects of the four methods are reported in Table 1. The "*" indicates a statistically significant Friedmann test at 99.9% confidence level on the ranks, meaning the ranks of the four tools were significantly different.

	Easy to use*	Helpfulness	Confidence*
Rankings after watching the video			
Quick Form	108	80	112
Wizard	98	77	109
WebProtege	57	81	74
WikiData	47	72	69
Rankings after hands-on task			
Quick Form	115	75	130
Wizard	90	88	110
WebProtege	65	86	93
WikiData	42	71	70

Table 1. Sum of student participants ranks.

The think-aloud session with botanists revealed similar preferences, but new interesting observations. Botanists agree that WikiData is the most challenging method. While agreeing that Quick Form is the easiest method, they recognize that it does not record computer consumable semantics of the terms and relationships. And for the latter task, all three botanists agree that a WebProtege-like tool has the potential. While students liked Wizard, botanists find the questions asked by Wizard too detailed.

User-method interaction logs recorded by the experiment platform and other questions in the surveys on the specific features of different methods will be analyzed in more depth in the near feature and discussed at the poster session during the ASIST meeting.

DISCUSSION

In terms of ease of use, the two rounds of ranking by student participants produced a similar result: Quick Form is the easiest, followed by Wizard and WebProtege, and WikiData was deemed the most difficult. The hands-on experience resulted in a slight adjustment in the scores of WebProtege (scored a bit higher) and Wizard (scored a bit lower), but did not change the overall ranking.

In terms of the methods' support/helpfulness for recording the full semantics of the terms, the scores are similar across different methods and the ranking is not statistically significant.

In terms of the user's confidence in applying the methods to add terms to ontologies, participants felt equally confident using Quick Form and Wizard, either before or after the hands-on task.

The three botanists' observation on the tools covered the ease of use, like the student participants, but went beyond. They acknowledged that WikiData's page-based organization schema does not provide the user with a full picture of all the terms in an ontology, and the process of adding a term could become rather involved (need to add several other terms and visit several other pages). While the users are guided by Wizard's questions, the detailed questions have the potential of inviting the user to over-think, thus rising the introduction of errors to the ontology. For many botanists, Quick Form is the preferred one, however, it cannot fulfill some botanists' demands to to convert the information collected by Quick Form to machine consumable semantics. They are also fond of

WebProtege for several reasons: (1) WebProtege's class hierarchy shows all terms in an ontology upfront and this familiar set of the terms makes botanists feel at home right away. (2) WebProtege makes adding terms that are needed in asserting a new relationship easy – typing the needed term and it is added right on the spot.

CONCLUSION

Usability studies involving iSchool students and botanists show that Quick Form and WebProtege are preferred methods for users with limited knowledge of ontology construction.

REFERENCES

Alatrish, E. S. (2013). Comparison some of ontology. *Journal of Management Information Systems*, *8*(2), 018-024.

Khondoker, M. R. et al. (2010). Comparing ontology development tools based on an online survey. *WCE 2010, London, UK*.

Norta, A., et al. (2010). Utility Evaluation of Tools for Collaborative Development and Maintenance of Ontologies. *14th IEEE IEDOCConference Workshops*, 207–214. https://doi.org/10.1109/EDOCW.2010.30