Graduate School of Library and Information Science
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Measuring variation in curators' GO annotations through a controlled multi-MOD study

W. John MacMullen
University of Illinois, 501 E. Daniel St, MC-493, Champaign IL 61820-6211 USA wjohn@uiuc.edu

Abstract

This work employed a multi-methods approach combining a prospective controlled experiment with broad contextual analysis (including observations, document content analysis, workflow analysis, and interviews) to investigate the nature and extent of variation in human-curated annotations of the scientific literature using the Gene Ontology (GO), a standardized cross-organism controlled vocabulary. Data obtained to date include 3,500 GO annotation instances generated by 23 biological curators. Substantial inter-annotator instance variation was observed, and curators' instances exhibited considerable semantic heterogeneity.

RESEARCH QUESTIONS

1. **How significantly do curators differ in annotation outcomes?** This question was investigated by conducting individual and consensus GO annotation experiments and comparing differences in curators' outcomes at different levels (annotation, instance, and attribute).

2. **Do differences in curators' educational-, training-, and research backgrounds influence their GO annotation performance?** This question was investigated by comparing the formal GO annotation data from the individual and consensus annotation experiments with demographic data collected from questionnaires, and in individual semi-structured interviews.

3. **Do differences in curators' personal annotation behaviors influence their GO annotation performance?** This question was investigated by comparing formal GO annotation data from individual and consensus annotation experiments with data on personal annotation behaviors (such as workflows and resources employed) that were obtained from individual interviews with curators, observations, and artifact analyses.

RESEARCH DESIGN

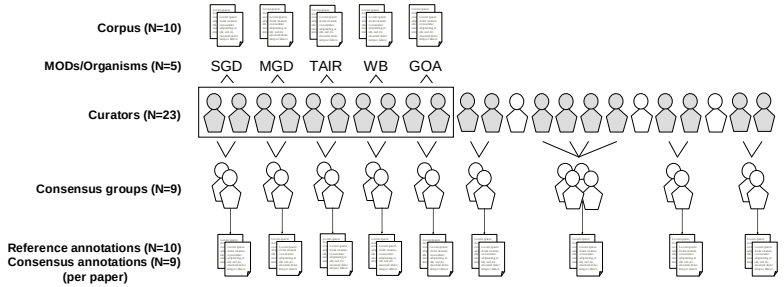
- A prospective controlled experiment was conducted, with 23 scientific curators from 11 model organism databases (MODs), who annotated a corpus of 10 articles at full redundancy, with curators paired by database to create consensus annotations. Consensus annotations created by organism experts were used as reference standards against which the remaining annotations will be compared.

- Contextual data about the curators' backgrounds, experience, and workflows were also collected, via document content analysis, workflow analysis, and individual curator interviews. Hand-annotated paper articles from six participants of the study were obtained for analysis.

Acknowledgments

- This work benefited from discussions with curators from multiple model organism databases at the 2005 and 2006 Gene Ontology Annotation Camps held at Stanford University. The Stanford Department of Genetics and the *Saccharomyces* Genome Database (SGD) provided travel support for this project.
- W.J.M. was supported in part by an Individual Biomedical Informatics Fellowship (F37 LM009194) from the National Library of Medicine, and an unrestricted research gift from Microsoft Research to the Annotation of Structured Data research group in the School of Information and Library Science at the University of North Carolina at Chapel Hill.

RESEARCH DESIGN



QUANTITATIVE DATA OBTAINED

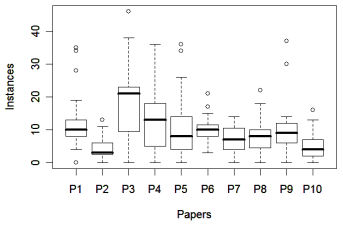
Individual annotations per paper	23
Consensus annotations per paper	9
Individual annotations	230
Consensus annotations	90
Total annotations	320
Individual annotation instances	2,289
Consensus annotation instances	1,258
Total annotation instances	3,547

CONTEXTUAL DATA OBTAINED

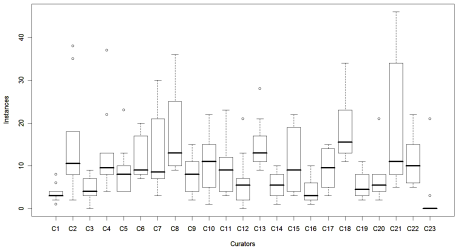
Audio:
- 6 individual curator interviews (20-40 minutes duration)

Documents:
- Six sets of 10 manually annotated articles, plus additional personal notes (~630 pages total)
- 23 personal data questionnaires about background and experience

Range of annotation instances per paper (all curators)



Range of annotation instances per paper (all curators)



P1 individual and consensus attribute counts compared to Reference annotation

		Consensus			Individual		
	Reference	Low	High	Mean	Low	High	Mean
Instances	4	4	38	16	4	35	13
Genes	4	2	19	6	3	19	5
Terms	1	1	10	6	1	15	6
Evidence codes	1	1	4	3	1	4	3
Aspects	1	1	3	2	1	3	2

Instance heterogeneity inhibits pairwise comparisons (partial P1 data)

	Gene product	GO aspect	GOID	GO term	Evidence code
Reference	PPM2	BP	31591	wybutosine biosynthesis	IMP
	TRM12	BP	31591	wybutosine biosynthesis	IMP
	TYW1	BP	31591	wybutosine biosynthesis	IMP
	TYW3	BP	31591	wybutosine biosynthesis	IMP
Consensus 2	NFS1	BP	31591	wybutosine biosynthesis	IMP
	PPM2	BP	31591	wybutosine biosynthesis	IDA
	PPM2	BP	31591	wybutosine biosynthesis	IMP
	TRM12	BP	31591	wybutosine biosynthesis	IDA
	TRM12	BP	31591	wybutosine biosynthesis	IMP
	TYW1	BP	31591	wybutosine biosynthesis	IDA
	TYW1	BP	31591	wybutosine biosynthesis	IMP
	TYW3	BP	31591	wybutosine biosynthesis	IDA
	TYW3	BP	31591	wybutosine biosynthesis	IMP
Consensus 8	NFS1	BP	31591	wybutosine biosynthesis	IDA
	PPM2	BP	31591	wybutosine biosynthesis	IDA
	PPM2	BP	50843	S-adenosylmethionine catabolism	IDA
	TRM12	BP	31591	wybutosine biosynthesis	IDA
	TRM12	BP	50843	S-adenosylmethionine catabolism	IDA
	TYW1	BP	31591	wybutosine biosynthesis	IDA
	TYW3	BP	31591	wybutosine biosynthesis	IDA
	TYW3	BP	50843	S-adenosylmethionine catabolism	IDA

Consensus group annotation data for paper P1 Reference annotation genes

Gene	Obs	Distinct groups	Distinct terms
PPM2	36	9	18
TYW3	24	9	10
TYW1	30	8	11
TRM12	25	8	13
Others	32	5	6

Pairwise comparison of P1 consensus annotations to Reference annotation

	True positives	False positives	False negatives
Reference annotation	4	-	-
Consensus 2	4	30	0
Consensus 3	0	12	4
Consensus 4	4	6	0
Consensus 5	4	13	0
Consensus 6	4	13	0
Consensus 7	1	11	3
Consensus 8	0	38	**4**
Consensus 9	4	3	0

Poster available at: http://macmullen.com/conferences/biocurator