











**Table 2.** Average annotation time per paper with different levels of semantic support

set and level of semantic support	available tags	$\bar{t}$ (min)
<b>SET 1</b> (no semantic support)	$\emptyset$	90
<b>SET 2</b> (partial semantic support)	enzyme, organism, pH, temperature	65
<b>SET 3</b> (full semantic support)	enzyme, organism, pH, temperature	56

## 5 Conclusions

We presented our ongoing development of a semantic infrastructure for enzyme data management. In the context of biofuel research, our system targets the automatic extraction of knowledge on fungal enzymes from genome research literature. Preliminary experiments show that semantic support allows for a significant decrease in manual curation time. However, future work is needed to evaluate the impact of such a system on the quality of the curated data.

**Acknowledgments.** Funding for this work was provided by Genome Canada and Génome Québec.

## References

- Alex, B., Grover, C., Haddow, B., Kabadjov, M., Klein, E., Wang, X.: Assisted curation: does text mining really help. In: Pacific Symposium on Biocomputing. vol. 13, pp. 556–567 (2008)
- Ananiadou, S., McNaught, J.: Text Mining for Biology And Biomedicine. Artech House, Inc., Norwood, MA, USA (2005)
- Bontcheva, K., Cunningham, H., Roberts, I., Tablan, V.: Web-based Collaborative Corpus Annotation: Requirements and a Framework Implementation. In: New Challenges for NLP Frameworks. pp. 20–27. ELRA, Valletta, Malta (May 22 2010)
- Caporaso, J.G., Deshpande, N., Fink, J.L., Bourne, P.E., Cohen, K.B., Hunter, L.: Intrinsic evaluation of text mining tools may not predict performance on realistic tasks. In: Pacific Symposium on Biocomputing. vol. 13, pp. 640–651. World Scientific Publishing (2008)
- Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In: Proc. 40th Anniversary Meeting of the ACL (2002)
- Federhen, S.: The Taxonomy Project. In: McEntyre, J., Ostell, J. (eds.) The NCBI Handbook, chap. 4. National Library of Medicine (US), National Center for Biotechnology Information (2003)
- International Union of Biochemistry and Molecular Biology: Enzyme Nomenclature 1992. Academic Press, San Diego, California (1992)
- Murphy, C., Powlowski, J., Wu, M., Butler, G., Tsang, A.: Curation of characterized glycoside hydrolases of fungal origin. Database 2011 (2011)
- Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Söhngen, C., Stelzer, M., Thiele, J., Schomburg, D.: BRENDA, the enzyme information system in 2011. Nucleic Acids Res. 39, (Database issue):D670–676 (2011)
- Shadbolt, N., Berners-Lee, T., Hall, W.: The semantic web revisited. IEEE Intelligent Systems 21(3), 96–101 (2006)
- The UniProt Consortium: The Universal Protein Resource (UniProt). Nucleic Acids Research 37(D), 169–174 (2009)
- Witte, R., Gitzinger, T.: Semantic Assistants – User-Centric Natural Language Processing Services for Desktop Clients. In: 3rd Asian Semantic Web Conference (ASWC 2008). LNCS, vol. 5367, pp. 360–374. Springer, Bangkok, Thailand (2009)
- Witte, R., Kappler, T., Baker, C.J.O.: Ontology Design for Biomedical Text Mining. In: Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences, chap. 13, pp. 281–313. Springer (2007)