

take into account this difference, which technically needs the integration of subcategorization information in the NLP tool.

Another question raised is if **possible** actions should be considered as actionable phrases or not, as in the example sentence below:

If you click on the button that shows a window on the bottom left corner, and then ...

Those phrases can be enriched with a “condition” type (i.e. denoting a conditional statement), such as in [10].

5.2 Observations about the linking method

In our setting we have used standard Information Retrieval (IR) techniques, searching over an index including the text of other documents found in the KB.

A first improvement would be to link to sections of a document rather than the full document. As we are already able to perform the corresponding segmentation, we also plan to perform section-based indexing to validate our hypothesis.

Furthermore, the titles of such sections are very informative (in a similar manner to the titles of the documents). We believe therefore that processing in a customized manner different parts of the documents (for instance assigning different weights for specific parts of the document, as we have done for the titles) would be beneficial.

The distinct separation between relevant and irrelevant results observed in the linking experiments indicates that techniques such as binary classification, or learning-to-rank, could give us a better result than the experimentally selected threshold used in this study. This of course would require more annotated data, which we could use as seeds for training a classifier. Although the annotation process is time-consuming, and very often not trivial, statistical machine learning is still worth investigating. The same is true for the extraction component, although special attention should be given to the considerations mentioned in 5.1.

6. CONCLUSION

We presented a method for enriching community-specific procedural knowledge entries that can be found on the Web. We have achieved that by linking text fragments describing how to perform a certain action. Our experiments show that such fragments can be efficiently extracted, and that they allow for higher linking performance than state-of-the-art methods. We are releasing the dataset we used, together with our annotation, hoping that this will help in fostering further research on this subject.

Based on an analysis of the evaluation results, we also proposed a number of possible improvements, such as taking into consideration modalities during extraction and performing section-specific indexing for linking.

In addition, because of the lack of space, we did not have the opportunity to speak of the representation model that can be used to store and structure the extracted information, however releasing such a model, as linked open data, is in our immediate plans.

7. REFERENCES

- [1] S. Ait-Mokhtar, J.-P. Chanod, and C. Roux. Robustness beyond shallowness: Incremental deep parsing. *Nat. Lang. Eng.*, 8(3):121–144, June 2002.
- [2] S. Bird, E. Klein, and E. Loper. *Natural Language Processing with Python*. O’Reilly Media, Inc., 1st edition, 2009.
- [3] C. Brun, V. Nikoulina, and N. Lajos. Linguistically-adapted structural query annotation for digital libraries in the social sciences. In *Proceedings of the 6th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, LaTeCH ’12, pages 55–64, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.
- [4] Y. Gil. Human tutorial instruction in the raw. *ACM Trans. Interact. Intell. Syst.*, 5(1):2:1–2:29, Mar. 2015.
- [5] B. Hachey, W. Radford, J. Nothman, M. Honnibal, and J. R. Curran. Evaluating entity linking with wikipedia. *Artificial Intelligence*, 194:130 – 150, 2013. Artificial Intelligence, Wikipedia and Semi-Structured Resources.
- [6] P. Pareti, E. Klein, and A. Barker. A Semantic Web of Know-how: Linked Data for Community-centric Tasks. In *Proceedings of the 23rd International Conference on World Wide Web Companion*, pages 1011–1016, 2014.
- [7] P. Pareti, B. Testu, R. Ichise, E. Klein, and A. Barker. Integrating know-how into the linked data cloud. In K. Janowicz, S. Schlobach, P. Lambrix, and E. Hyvönen, editors, *Knowledge Engineering and Knowledge Management*, volume 8876 of *Lecture Notes in Computer Science*, pages 385–396. Springer International Publishing, 2014.
- [8] C. Paris, K. V. Linden, and S. Lu. Automated knowledge acquisition for instructional text generation. In *Proceedings of the 20th Annual International Conference on Computer Documentation*, SIGDOC ’02, pages 142–151, New York, NY, USA, 2002. ACM.
- [9] F. Roulland, S. Castellani, N. Hairon, and P. Valobra. Method and system for linking textual concepts and physical concepts, June 14 2012. US Patent App. 12/967,210.
- [10] Z. Zhang, P. Webster, V. Uren, A. Varga, and F. Ciravegna. Automatically extracting procedural knowledge from instructional texts using natural language processing. In N. C. C. Chair), K. Choukri, T. Declerck, M. U. DoÄşan, B. Maegaard, J. Mariani, A. Moreno, J. Odiijk, and S. Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC’12)*, Istanbul, Turkey, may 2012. European Language Resources Association (ELRA).