Beyond the Book: globalization in literature from a digital perspective

Floor Buschenhenke, Karina van Dalen-Oskam
Huygens ING
karina.van.dalen@huygens.knaw.nl

Carlos Martinez-Ortiz
Netherlands eScience Centre
c.martinez@esciencecenter.nl

Marijn Koolen
Institute for Logic, Language
and Computation,
Universiteit van Amsterdam
marijn.koolen@uva.nl

Globalization is an important research topic in many fields, also in literary studies. In our project Beyond the Book, funded by the Netherlands eScience Center, we try to find a way to measure how "international" a novel is. The idea behind our project is that the topic and content of a novel from one country may have an appeal to readers in other countries if the novel, for instance, reflects a certain amount of cultural knowledge that is also available to readers in the other countries. It could also be the case that readers from other countries do not look for shared cultural elements, but prefer to read about exotic other cultures. This will depend on "fashions" in literature.

To establish the level of shared cultural knowledge or exoticness we use Wikipedia as a reference frame. Wikipedia is currently widely used in the analysis of cultural diversity. For example, Warncke-Wang *et al.* (2012) describe a way to compare different-language Wikipedias based on their size, amount of shared entries, and amount of unique entries per language. They also pay attention to the flow of translations between the Wikipedias. This results in an intriguing overview of content-related differences between Wikipedias. Eom & Shepelyansky (2013) provide rankings of persons based on their entries in different-language Wikipedias. Pfeil *et al.* (2006) analyse cultural differences based on measurable practices in collaborative authoring in Wikipedia. Their findings include some correlations with four cultural influences as proposed by Hofstede (1991). Comparable work on all five of Hofstede's dimensions of cultural differences (power distance, collectivism versus individualism, femininity versus masculinity, uncertainty avoidance, and the later added long-term versus short-term orientation) has been done by Hara *et al.* (2010).

Our own approach consists of a more detailed way to look at shared and unique knowledge. For this presentation we focus on the analysis of Wikipedia entries about proper names (named entities) that are found in novels. In particular, we are looking at the contributions to English Wikipedia entries from different countries as our source data. If entries dealing with topics show relatively more edits from users from certain countries compared to users from other countries, we interpret this as a sign of a cultural preference for the topic in those countries. The different cultural preferences as gathered from Wikipedia are then compared to a corpus of Dutch novels. In our presentation we will describe the pipeline of tools that we developed, show some first results, and present our ideas on next steps that could lead to more knowledge about the impact of globalization in literature.

References

Eom, Y., Shepelyansky, D. L. (2013) 'Highlighting entanglement of culture via ranking of multilingual Wikipedia articles', *PLoS ONE* 8(10): e74554. doi:10.1371/journal.pone.0074554

Hara, N., Shachaf, P., Hew, K. F. (2010) 'Cross cultural analysis of the Wikipedia community', *Journal of the American Society of Information Science and Technology* 61 (10), pp. 2097-2108.

Hofstede, G. H. (1991) *Cultures and organizations: Software of the mind*, London: McGraw-Hill.

Pfeil, U., Zaphiris, P., and Ang, C. S. (2006) 'Cultural differences in collaborative authoring of Wikipedia', *Journal of computer-mediated communication*, vol. 12, no. 1 [Online]. Available: http://jcmc.indiana.edu/vol12/issue1/pfeil.html

Warncke-Wang, M., Uduwage, A., Dong, Z., and Riedl, J. (2012) 'In search of the ur-wikipedia: Universality, similarity, and translation in the wikipedia inter-language link network', *WikiSYm '12: Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration*, ACM, 2012, pp. 20:1–20:10. [Online]. Available: http://doi.acm.org/10.1145/2462932.2462959