Graduiertenkolloquium Angewandte Informatik

Making Authorship and Social Interactions Transparent in Collaboratively Written and Revisioned Text – The Use Case of Wikipedia

Dipl.-Medienwissenschaftler Fabian Flöck
GESIS – Leibnitz-Institute for the Social Science, Mannheim

When readers judge the credibility of a newspaper report or a blog post, one major factor is **who** wrote the piece in question. Yet, when reading the products of digital collaborative writing systems, individual author attribution is usually not available.

Even more telling for assessing the quality of a text – of a newspaper or a Wikipedia article – might it be to understand **how** it was composed, i.e., the (collaborative) processes behind the content production. For Wikipedia, we argue that certain (and which) meso-level "social mechanisms" exist that are typical to emerge between editors writing a document together – such as claiming ownership of a collaborative document or non-cooperative behavior – which can impair the workflow for creating high-quality content. Still, they cannot be sufficiently analyzed, as appropriate methods for mining and explicitly representing editor interactions are lacking.

We hence present a method for revisioned, collaborative writing systems to **extract the fine-grained interactions of editors with each other and with the content over time and represent them in a way that best models the underlying reality of the socio-technical system**, based on the raw text revisions produced by the editors.

The first part of this approach is an algorithm to **mine authorship** of single text tokens from the revision history of a document. The solution we present is the first to be evaluated at over 95% correct attributions; and it decreases execution time by at least one order of magnitude compared to previous techniques.

Next, through a user pre-study, we establish that state-of-the-art detection of editor interactions is based on **modeling disagreement** (reverts) at a very coarse-grained level and isn't able to capture large portions of occurring disagreements.

Based on those insights we present an extension of the authorship algorithm to **mine interactions** of users with each other on top of the accurate authorship and change attributions to infer detailed "agreement" and "disagreement" relations.

Finally, leveraging the extracted data, we built two working prototypes of **web-based visualizations** for end users, which make editor-editor interactions, authorship and additional collaboration metrics transparent.

To conclude we outline how our solutions can be used in other collaborative writing systems and what further research can be conducted using the mined data.

Termin: Freitag, 10. Juni 2015, 14.00 Uhr

Ort: Englerstraße 11, 76131 Karlsruhe
Kollegiengebäude am Ehrenhof (Geb. 11.40), 2. OG, Raum 231
(Hinweise für Besucher: www.aifb.kit.edu/web/Kontakt

Veranstalter: Institut AIFB, Forschungsgruppe Wissensmanagement

Zu diesem Vortrag lädt das Institut für Angewandte Informatik und Formale Beschreibungsverfahren alle Interessierten herzlich ein.

Andreas Oberweis, Hartmut Schmeck, Detlef Seese, Rudi Studer (Org.), York Sure-Vetter