

Wikipulse – Automatically Generating an Online Newspaper from Wikipedia Articles

Tobias Fütterer, Peter A. Gloor, Tushar Malhotra, Harrison Mfula, Karsten Packmohr, Stefan Schultheiß

More and more, user-generated content is complementing conventional journalism. While we don't think that CNN or New York Times and its professional journalists will disappear anytime soon, formidable competition is emerging through humble Wikipedia editors. In earlier work (Becker 2012), we found that entertainment and sports news appeared on average about two hours earlier on Wikipedia than on CNN and Reuters online. In this project we build a news-reader that automatically identifies late-breaking news among the most recent Wikipedia articles and then displays it on a dedicated Web site.

Our task consists of three parts:

1. find relevant articles,
2. reformat the articles in news style format and
3. display them on a Web page.

The core of our system consists of an algorithm that automatically finds the most news-worthy articles on Wikipedia grouped by news related categories (e.g. Current Events, Sports, Politics etc.).

Besides obvious selection criteria such as the most recently edited and searched articles, in the first step, we also employ algorithms from earlier work (Fuehres 2012), where we discovered that building an article network based on “shared-editorship” links – two articles obtain a link if the same editor edits both of them – points out the most important recent articles.

In the second step we reformat the Wikipedia articles, which are written in factual history-style, into a more journalistic-style, using automatic abstract generation techniques.

The final step consists of displaying the articles in a reader-friendly online newspaper.

References

Becker, Melina, The actuality of online encyclopedias – an empirical analysis (published in German: Die Aktualität von Online-Enzyklopädien – Eine empirische Analyse am Beispiel Wikipedia), Diploma Thesis, University of Cologne, 2012

Fuehres, H. Gloor, P. Henninger, M. Kleeb, R. Nemoto, K. Galaxysearch: Discovering the Knowledge of Many by Using Wikipedia as a Meta-Search Index. Proceedings Collective Intelligence 2012, April 18-20, Cambridge, MA