

De la fouille de textes à la recommandation de lectures – Applications sur les plateformes d’Openedition.org

Patrice Bellot*

*Aix-Marseille Université – CNRS – Univ. Toulon – EHESS – Université d’Avignon
LSIS UMR 7296 / CLEO OpenEdition UMS 3287
patrice.bellot@univ-amu.fr

La recommandation de lectures peut être réalisée selon une analyse des comportements des internautes (qu’est-ce qui est lu et par qui) ou selon une analyse des contenus eux-mêmes. La première approche est la seule possible lorsque les contenus sont difficilement accessibles ou lorsque leurs méta-données sont trop peu nombreuses. Dans le cas contraire, l’accès au contenu textuel (qu’il soit écrit ou transcrit de l’oral) rend possible une extraction des mots clés, une catégorisation thématique, une analyse de sentiments...

Comme cela a été montré lors de la tâche *Social Book Search* des conférences CLEF, l’impossibilité de l’accès au texte intégral des contenus peut être compensée par une analyse des commentaires des internautes qui synthétisent les opinions et les contenus, insistant sur leurs aspects les plus notables.

Dans le cadre de nos travaux sur les contenus des plateformes d’édition en ligne OpenEdition.org, nous avons proposé différentes solutions pour mettre automatiquement en relation des contenus (création de liens à partir de citations automatiquement détectées) (Ollagnier et al., 2016), pour identifier des textes critiques sur ces contenus (classification en genre) (Benkoussas et al., 2014), pour leur appliquer une analyse des sentiments (polarité des critiques, identification des aspects abordés) (Hamdan et al., 2014) et pour exploiter des requêtes longues et verbeuses pour une recommandation guidée (Chaa et al., 2016; Benkoussas et al., 2015). Ces solutions, supervisées ou non, emploient des approches numériques du traitement automatique des langues et de la recherche d’information.

Au-delà des approches elles-mêmes, je présenterai les différents prototypes mis en place sur les collections en Sciences Humaines et Sociales d’OpenEdition.org ainsi que les nombreux verrous qui restent ouverts.

Références

- Benkoussas, C., P. Bellot, et A. Ollagnier (2015). The impact of linked documents and graph analysis on information retrieval methods for book recommendation. In *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, WI-IAT 2015, Singapore, December 6-9, 2015 - Volume I*, pp. 385–392.
- Benkoussas, C., H. Hamdan, P. Bellot, F. Béchet, et E. Faath (2014). A collection of scholarly book reviews from the platforms of electronic sources in humanities and social sciences ope-

- nedition.org. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014.*, pp. 4172–4177.
- Chaa, M., O. Nouali, et P. Bellot (2016). Verbose query reduction by learning to rank for social book search track. In *Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum, Évora, Portugal, 5-8 September, 2016.*, pp. 1072–1078.
- Hamdan, H., P. Bellot, et F. Béchet (2014). Supervised methods for aspect-based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval@COLING 2014, Dublin, Ireland, August 23-24, 2014.*, pp. 596–600.
- Ollagnier, A., S. Fournier, et P. Bellot (2016). Cascade de crfs et SVM pour la détection de références bibliographiques diffuses dans les articles scientifiques. In *CORIA 2016 - Conférence en Recherche d'Informations et Applications- 13th French Information Retrieval Conference. CIFED 2016 Colloque International Francophone sur l'Écrit et le Document, Toulouse, France, March 9-11, 2016, Toulouse, France, March 9-11, 2016.*, pp. 139–152.