

Visualisation de mesures agrégées pour l'estimation de la qualité des articles Wikipedia

Fanny Chevalier*, Stéphane Huot** et Jean-Daniel Fekete***

* Microsoft Research - INRIA Joint Center
Bât 490, Université Paris-Sud, F-91405 Orsay Cedex
fanny.chevalier@inria.fr

** Université Paris-Sud & CNRS - INRIA
LRI, Bât 490, F-91405 Orsay Cedex
stephane.huot@lri.fr

*** INRIA
Bât 490, Université Paris-Sud, F-91405 Orsay Cedex
jean-daniel.fekete@inria.fr

Résumé. Wikipedia, devenue l'une des bases de connaissances les plus populaires, pose le problème de la fiabilité de l'information qu'elle dissémine. Nous proposons *WikipediaViz*, un ensemble de visualisations basé sur un mécanisme de collecte et d'agrégation de données d'édition Wikipedia pour aider le lecteur à appréhender la maturité d'un article. Nous listons cinq métriques importantes, déterminées lors de sessions de conception participative avec des experts Wikipedia pour juger de la qualité, que nous présentons au lecteur sous forme de visualisations compactes et expressives, dépeignant le profil d'évolution d'un article. Nos études utilisateur ont montré que *WikipediaViz* réduisait significativement le temps requis pour évaluer la qualité en maintenant une bonne précision.

1 Introduction

Wikipedia devenue l'un des dix sites web les plus visités dans le monde (Lee et Bill (2007)) avec plus de 14 millions d'articles répartis dans 270 versions localisées. Le concept fondamental de Wikipedia, « l'encyclopédie libre que tout un chacun peut éditer »¹, a montré être un excellent moyen de rendre l'information plus exhaustive, plus précise et maintenue à jour. Paradoxalement, les articles, par leur caractère volatil, sont sujets à contenir de l'information non vérifiée et il devient ainsi crucial d'aider les lecteurs, en particulier néophytes, à évaluer la qualité et la fiabilité des articles.

Pour estimer la maturité et la qualité d'un article, les utilisateurs fréquents de Wikipedia ont l'habitude de collecter des indices dans les pages de discussion, les historiques, et l'apparence visuelle des articles. Les utilisateurs novices quant à eux n'ont pas ce réflexe, et sont même parfois ignorant du fait que la qualité peut être discutable (Bryant et al. (2005)). Un moyen de pallier à ce problème serait d'attribuer une note à chaque article pour informer de sa maturité ou sa qualité. Un tel système, non automatique, permettant de classer les articles selon leur qualité et leur importance a été récemment intégré à Wikipedia². Néanmoins, le processus de classification effectué « à la main » est long et fastidieux (seul 1% des articles est noté). Il n'y a ainsi pas encore de moyen direct d'évaluer la qualité des articles Wikipedia.

Dans cet article, nous présentons des indicateurs faciles à comprendre, en automatisant la collecte de données jugées importantes pour évaluer la fiabilité d'un article. La qualité des articles étant essentiellement liée au procédé d'édition et de relecture et ainsi à leur maturité —