

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 —

Estimation de la qualité des articles Wikipedia



FIG. 1 – Visualisations WikipediaViz dépeignant le profil d’un article

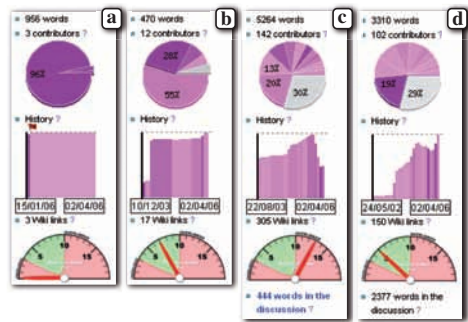


FIG. 2 – Exemples de profils types : (a) Ébauche, (b) Bon Début, (c) Bon article et (d) Article de qualité

elle ne relève pas d’une autorité — nous proposons de visualiser les informations recueillies dans le but d’améliorer la visibilité et la transparence d’un tel processus de rédaction collaborative. L’objectif est donc d’aider l’utilisateur à appréhender le fonctionnement de Wikipedia et de lui fournir d’une manière simple et intelligible les données relatives au processus d’édition à travers lequel sont passés les articles qu’ils consultent. En particulier, nous sommes intéressés par les *utilisateurs occasionnels*, qui ne maîtrisent pas les moyens de récolter ces indices, qui parfois ne sont pas conscients des problèmes de fiabilité, et qui ne sont familiers ni des statistiques, ni des techniques de visualisation.

Notre travail a d’abord consisté à identifier les indicateurs importants permettant de juger de la maturité d’un article Wikipedia et ainsi de sa qualité. Nous avons travaillé avec des administrateurs Wikipedia, des contributeurs fréquents et des sociologues pour établir des métriques objectives qu’ils considéraient comme fondamentales pour dresser le profil d’un article. Trouver et interpréter ces données n’est pas trivial et requiert d’être familier de l’environnement de Wikipedia. Nous avons donc conçu *WikipediaViz*, cinq indicateurs visuels qui représentent ces métriques dans la colonne gauche de l’interface standard de Wikipedia (Fig. 1).

2 État de l’art

Bien que de nombreux utilisateurs n’accordent pas une entière confiance à Wikipedia, une étude qualitative comparant Wikipedia et l’encyclopédie traditionnelle Britannica a conclu que les deux étaient à peu près similaires en termes de qualité (Giles (2005)). Outre ses résultats — qui sont difficiles à prouver formellement — cette étude montre la difficulté d’évaluer la qualité des articles Wikipedia.

Des études quantitatives dont l’objectif est d’évaluer la qualité de Wikipedia d’un point de vue plus objectif en utilisant des métriques basées sur les données d’édition de Wikipedia. En utilisant simplement le nombre de mots d’un article comme mesure de qualité (Blumenstock (2008)) ou le nombre d’éditions et le nombre de contributeurs uniques d’un article, le nombre

1. <http://www.wikipedia.org>
 2. <http://en.wikipedia.org/wiki/Wikipedia:1.0>

de contributeurs uniques traduit la « diversité » d'un article (les articles de qualité ont substantiellement plus de contributeurs impliqués (Wilkinson et Huberman (2007b)) et le nombre d'éditions sa « rigueur ». Dans (Adler et de Alfaro (2007)), ces deux mesures sont combinées pour définir la notion de *réputation* d'un auteur. D'autres mesures uniques, basées sur l'agrégation de plusieurs indicateurs ont été proposées pour prédire la qualité d'une contribution Druck et al. (2008) ou encore la véracité du contenu d'un article Dondio et al. (2006). Par ailleurs, le travail collaboratif dans les pages de discussion dédiées — où les changements sont souvent discutés avant d'être introduits dans l'article (Viegas et al. (2007)) — a montré jouer un rôle critique pour la qualité des articles (Kittur et Kraut (2008); Wattenberg et al. (2007)).

Toutes ces études montrent que les métriques qui mettent en évidence le processus d'édition peuvent être utilisées comme indicateurs pour estimer la qualité d'un article. Il a été prouvé que révéler ces données affecte la confiance qu'accorde le lecteur en la fiabilité des articles (Kittur et Kraut (2008); Wattenberg et al. (2007)). Cependant, les résultats de classification automatique d'article par qualité n'ont pas été satisfaisants (précision de 90% Stvilia et al. (2005)) et les erreurs de classification nombreuses introduisent une méfiance des utilisateurs. En complément des méthodes analytiques, des visualisations ont été proposées dans le but d'améliorer l'expérience des lecteurs de Wikipedia.

Les deux visualisations History Flow (Viégas et al. (2004)) et WikiDashboard (Suh et al. (2008); Pirolli et al. (2009)) dépeignent l'évolution d'un article dans le temps. History Flow montre la taille de l'article avec les changements de son contenu (ajouts et suppressions) et leurs auteurs. Bien qu'il offre des informations très précises concernant l'histoire d'un article et les contributions des utilisateurs, History Flow nécessite une large zone d'affichage ainsi qu'une certaine expérience pour être compris, comme l'ont indiqué nos experts Wikipedia. WikiDashboard fournit les graphes d'activité hebdomadaire d'un article ou d'un auteur, mais ne prétend pas montrer des indices sur la qualité. Il n'est pas trivial pour le lecteur d'interpréter ces graphes d'activité comme tels. Les Chromogrammes (Wattenberg et al. (2007)) sont des visualisations conçues pour comprendre l'activité des contributeurs prolifiques de Wikipedia. Mais ici encore, ces visualisations ne sont pas destinées aux lecteurs occasionnels mais plutôt pour une étude sociologique de Wikipedia.

Ces visualisations permettent une analyse très précise des articles et auteurs, mais elles nécessitent une large zone d'affichage pour accéder aux détails. Par ailleurs, History Flow et Chromogrammes ne sont pas compatibles avec la lecture des articles eux-mêmes mais plutôt destinés à une analyse sociale des méta-données.

En résumé, il y a eu des efforts pour définir et évaluer la qualité des articles Wikipedia mais aucune solution n'a encore été proposée pour l'exposer à l'utilisateur occasionnel. À l'inverse, des outils permettent la visualisation et l'analyse de certains aspects des données Wikipedia, mais ils sont destinés aux utilisateurs expérimentés. Il existe donc un réel besoin d'outils pour aider les lecteurs occasionnels à prendre conscience que le contenu de Wikipedia est volatil et ainsi les aider à estimer l'impact sur la qualité et la fiabilité des articles. Comme nos utilisateurs cibles ne sont experts ni en analyse statistique, ni en visualisation, notre objectif premier est de fournir rapidement des informations sur l'état et l'évolution des articles tout en maintenant un bon compromis entre simplicité et pertinence des mesures et des visualisations associées, sans pour autant sacrifier leur justesse dans ce contexte. Ainsi, la première étape de notre travail a consisté à identifier comment les experts gèrent les problèmes de qualité et ce dont ils auraient besoin pour les aider dans cette tâche.

3 Conception participative avec les experts Wikipedia

Nous avons organisé deux sessions de conception participative avec 4 administrateurs Wikipedia, 2 contributeurs majeurs et 2 sociologues. Le principal objectif était d'identifier les besoins des utilisateurs, en particulier lorsqu'ils naviguent comme simple lecteur et ainsi lister les indices sur lesquels nos experts se basent pour estimer la qualité d'un article.

La première discussion a révélé l'importance de la question de l'estimation de la qualité dans la communauté Wikipedia. Tous ont demandé des outils qui reflètent les problèmes de qualité lorsqu'ils consultent Wikipedia. Les sessions ont abouti à deux méthodes pour résoudre ce problème : les méthodes objectives, qui exposent des données factuelles, et les méthodes subjectives qui tentent d'agréger des scores pour fournir un niveau de qualité global. Les deux approches semblent utiles, mais nous avons choisi de nous concentrer sur l'approche objective car elle semble mieux répondre à l'esprit de Wikipedia (neutre, vérifiable et factuel).

Tous les participants se sont accordés à dire qu'ils ne consultent plus un article sans avoir au préalable vérifié un certain nombre d'informations via les pages d'historique et de discussion. Ils se fient aussi au style et au format de l'article pour estimer sa qualité, en se basant sur les recommandations de Wikipedia, regardant si l'article cite des références et est bien intégré dans l'encyclopédie (liens internes).

La plupart des critères exposés par les experts sont étroitement liés à la maturité des articles. La corrélation entre maturité et qualité dans Wikipedia a été montrée dans (Bruls et al. (2000)) et les auteurs mentionnent l'idée de calculer un niveau de maturité à partir des mesures. Cependant, la valeur obtenue montre une forte déviation standard parmi les articles et ne semble pas fiable sans une supervision. Il est délicat de calculer automatiquement une unique valeur reflétant la qualité d'un article.

3.1 Cinq mesures pour évaluer la maturité/qualité

Nous avons listé cinq mesures objectives qui sont utilisées par les experts pour se faire une opinion initiale d'un article et qui ont été prouvées corrélées à la qualité dans la littérature.

Nombre de mots. Intuitivement, on peut supposer que les articles de qualité sont longs (détaillés et complets). Il est moins certain que les articles longs soient bons. Blumberg a montré que la taille des articles était l'un des indicateurs les plus efficaces pour différencier les articles de qualité des autres (Blumberg (2008)), mais il mentionne plusieurs contre-exemples qui dissuadent d'utiliser cette mesure seule comme mesure de qualité. Il serait possible de considérer des techniques de traitement automatique des langues naturelles afin d'évaluer la diversité du vocabulaire, repérer les mots mal orthographiés ou ignorer les mots de liaisons. Cependant, cela nécessiterait un traitement pouvant s'avérer coûteux, pour fournir un surplus d'information au lecteur peu au fait de telles méthodes d'analyse automatique.

Nombre de contributeurs et taux de contribution. Le nombre de contributeurs distincts et le taux de leur contribution restant dans la version courante sont des indices importants car ce sont de bons indicateurs de l'investissement de la communauté, les articles de qualité dans Wikipedia ayant généralement un grand nombre de contributeurs distincts (Wilkinson et Huberman (2007a,b)).

Nombre et taille des éditions. Idéalement, une page atteint son niveau de qualité maximal quand elle a été éditée un certain nombre de fois pour être complétée et corrigée (Wilkinson et Huberman (2007b)). De plus, la taille des contributions montre s'il y a eu d'important changements ou simplement des corrections mineures. Lorsque l'on prend en compte l'aspect temporel (comme il est décrit dans Viégas et al. (2004); Suh et al. (2008)), on peut appréhender l'histoire de l'article : peu d'éditions sur une longue période traduit habituellement qu'il a atteint un bon niveau de complétude ou que le sujet suscite peu d'intérêt ; des pics d'activité peuvent traduire l'influence d'évènements d'actualité (élections, sport, ...).

Nombre de références et de liens internes. Révéler les sources d'informations est reconnu être un important facteur influençant la confiance (Stvilia et al. (2005)) car cela offre la possibilité de vérifier et compléter l'information. Ainsi, les utilisateurs familiers sont suspicieux lorsque des faits ne sont pas étayés par des références dans l'article. Ils observent aussi le nombre de liens vers d'autres articles Wikipedia (appelés *liens internes* ou *wikiliens*) que contient un article qui montre l'intégration de l'article dans l'encyclopédie. Le Manuel de Style de Wikipedia³ encourage l'utilisation de liens vers d'autres articles pour aider les utilisateurs à assouvir leur curiosité ou approfondir leurs recherches. Un article *sous-référencé* (faible nombre de liens) peut suggérer qu'il est récent et pas encore bien intégré à l'encyclopédie. À l'inverse, un article *sur-référencé* peut mener le lecteur vers des articles peu pertinents.

Taille et activité de la discussion. Les pages de discussion sont un espace d'échange d'idées sur le contenu et le style d'écriture d'un article. Elles sont aussi utilisées pour planifier l'évolution de la page. La coopération et la coordination ont un fort impact sur la qualité d'un article (Wilkinson et Huberman (2007b,a); Viegas et al. (2007)) et les utilisateurs fréquents ont toujours un regard sur la page de discussion lorsqu'elle existe. Mis à part le contenu, ils observent sa taille, qui les informe de controverse potentielles, et son activité récente pour identifier si des consensus sont en cours d'adoption.

4 Les visualisations WikipediaViz : conception itérative

Les mesures listées dans la section précédente ne peuvent pas être obtenues directement dans Wikipedia : certaines s'obtiennent via une navigation fastidieuse et approfondie à travers l'interface de Wikipedia, les autres nécessitent un traitement lourd sur l'ensemble des révisions.

Nous avons automatisé la collecte et le calcul des mesures et modifié l'interface de Wikipedia en ajoutant cinq visualisations pour dépeindre le profil d'un article (Fig. 1). Chaque visualisation correspond à un objet dynamique HTML ou une image, insérés dans le panel gauche, sous le logo de Wikipedia. Les visualisations sont conçues pour être compactes, simples et expressives. En termes d'interaction, seules des infobulles sont disponibles pour accéder aux détails, la possibilité de navigation pouvant détourner le lecteur de sa tâche initiale (lire l'article). Une documentation est accessible pour chacun des indicateurs en cliquant sur le symbole « ? » associé. La documentation contient plusieurs exemples de profils types illustrant comment interpréter les indicateurs de WikipediaViz.

3. <http://en.wikipedia.org/wiki/Wikipedia:ManualofStyle>

Nous avons itéré deux fois pour aboutir à une interface utilisable, la première version ayant révélé que l'une des visualisations était trop difficile à interpréter pour un utilisateur néophyte.

Nombre de mots. Le nombre de mots est affiché textuellement puisqu'il consiste en une simple valeur numérique et qu'il est impossible de positionner par rapport à un « standard ».

Contributions des auteurs. Cette visualisation combine dans la même vue le nombre d'auteurs, leur taux de contributions et le nombre d'éditions.

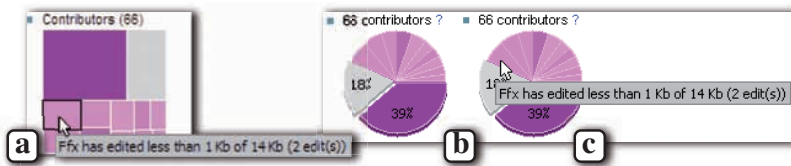


FIG. 3 – Visualisation interactive des contributions des auteurs contribution : Treemap (a) et graphique circulaire (b-c).

Première itération La répartition des contributions des auteurs d'un article était initialement visualisée à l'aide d'un Treemap carré (Bruls et al. (2000)) (Fig. 3a). Le carré complet représente la taille de l'article et chaque carré à l'intérieur la contribution d'un auteur. La surface d'un carré est proportionnelle à la quantité de texte qui reste dans la version affichée de l'article. Les carrés sont colorés en fonction du nombre d'éditions effectuées par l'auteur correspondant, avec une teinte d'autant plus foncée que l'utilisateur a édité souvent. Pour éviter l'encombrement de la visualisation, les petites contributions qui seraient traduites par des carrés de quelques pixels (moins de 1/100ème de la surface totale) sont agrégées et affichées comme un unique carré gris. Mais l'étude initiale a révélé que les utilisateurs occasionnels avaient des difficultés à comprendre la représentation en Treemap.

Deuxième itération. Nous avons simplifié la visualisation pour la rendre la plus simple et expressive possible. Nous avons opté pour un graphique circulaire (Fig. 3b), qui est une représentation populaire (« camembert »), facile à interpréter. La visualisation affiche les mêmes informations que le Treemap, avec la même approche de distribution spatiale et de code couleur.

Les visualisations supportent l'exploration de détails concernant les contributions en positionnant le curseur sur une portion du « camembert » : une infobulle indiquant le nom du contributeur, la taille agrégée de sa contribution et le nombre d'éditions apparaît alors (Fig. 3c).

L'obtention de ces données via l'interface standard de Wikipedia aurait nécessité le parcours intégral de l'historique et il est impensable de calculer le taux de contribution à la main dans la mesure où il faut retrouver l'auteur de chaque caractère de l'article. Le graphique circulaire permet d'identifier immédiatement les articles largement retravaillés (les articles de qualité tendent à être co-écrits par quelques contributeurs majeurs et édités par de nombreux autres) et à l'inverse des articles dans leur forme quasi-initiale (bien que ce ne soit pas une preuve de faible qualité).

Évolution. La Figure 4 montre la frise chronologique représentant l'évolution de la taille de l'article. L'intervalle de temps est affiché sous le graphique dont la barre la plus à droite correspond à la version courante ; la taille est représentée par la hauteur d'une barre. L'échelle de temps est calculée de façon à ce que la frise exploite l'espace de la visualisation, avec une limite basse de 3 pixels pour une période. Lorsque plusieurs changements ont lieu dans une même période de temps, la taille moyenne est représentée. À l'inverse, quand il n'y a pas de changements durant plusieurs périodes, la dernière taille est affichée jusqu'au changement suivant. Les barres sont colorées selon le même code que précédemment (une teinte claire correspond à peu d'éditions, foncée à de nombreuses modifications). On obtient plus de détails en positionnant le curseur sur l'une des barres : apparaît une infobulle indiquant les dates de début et de fin de la période et le nombre d'éditions qui ont été effectuées (Fig.4b).

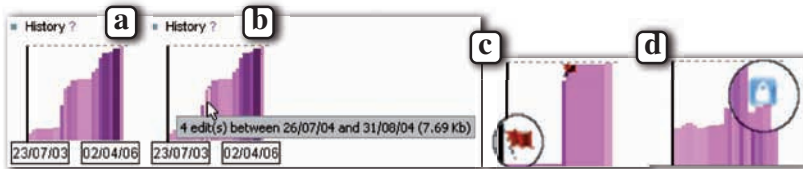


FIG. 4 – Visualisation interactive de l'évolution de la taille d'un article au cours du temps.

La page d'historique d'un article permet de retrouver ses versions antérieures et ainsi d'avoir un aperçu de son évolution. Mais il est très difficile d'identifier les événements extraordinaires alors que la frise attire l'attention sur les périodes de stress (e.g. beaucoup d'éditions en un court laps de temps) et les accidents (e.g. vandalisme, ou restructuration). La maturité peut se traduire sur la frise par la stabilisation de la taille à la fin de vie de l'article après quelques étapes importantes d'évolution identifiables par un changement brutal de taille et de couleur. La frise permet aussi de repérer les *guerres d'édition* : deux contributeurs changent à de nombreuses reprises la même portion d'un article à intervalles de temps très courts. Ces événements se traduisent par une succession de barres de taille stable et de teinte foncée.

La frise a été augmentée lors de la deuxième itération pour présenter au moyen de petites icônes (Fig. 4c-d) les événements exceptionnels dans la vie de l'article : les bannières et les protections. Dans Wikipedia, des bannières peuvent être placées dans un article pour attirer l'attention du lecteur sur certains aspects (e.g. l'article est une ébauche ou il ne contient pas suffisamment de références externes). Nous représentons l'ajout d'une bannière avec un drapeau rouge, positionné au dessus de la barre de temps correspondante. Un drapeau rayé indique une bannière ôtée (Fig. 4c). Un événement plus critique concerne la protection. Les articles sujets aux controverses ou au vandalisme peuvent être semi-protégés (seuls les utilisateurs enregistrés peuvent les modifier) ou protégés (une autorisation doit être accordée par un administrateur). Un événement de protection est représenté par un cadenas sur la frise (Fig. 4d), rayé lorsque la protection est levée. Ces icônes permettent d'identifier d'un coup d'œil les événements importants qui sont très difficiles de retrouver autrement.

Liens internes. Nous avons conçu une « jauge » qui montre la densité de liens internes par page (pourcentage de mots qui sont des wikiliens) (Fig. 5). Le nombre de wikiliens contenus

Estimation de la qualité des articles Wikipedia

dans la page est aussi affiché textuellement. Les couleurs de fond de la jauge expriment trois zones de densité : sous-référencement, bon référencement et sur-référencement.



FIG. 5 – Jauges de liens internes (sous-référencé, bien référencé, sur-référencé)

La zone rouge à droite (plus de 10%) indique un sur-référencement. La valeur 10% vient des recommandations de style de Wikipedia stipulant qu'un article est sur-référencé lorsqu'il excède cette valeur. La zone rouge à gauche indique un sous-référencement (moins de 3%). Nous avons déterminé ce seuil après analyse des articles de qualité de Wikipedia, supposés respecter les directives de style. L'étude a montré que plus de 80% des 345 articles de qualité ont une densité de wikiliens entre 3 et 10% (4% ayant une densité supérieure à 10%). Nous avons déduit qu'un seuil de 3% était représentatif des articles sous-référencés.

Il n'est pas évident d'identifier le sur-référencement ou sous-référencement via l'affichage standard de Wikipedia, en particulier lorsque l'article dépasse la fenêtre du navigateur alors que notre indicateur affiche l'information de manière concise en haut de page.

Taille de la discussion et indicateur d'activité. Les lecteurs peuvent facilement accéder aux pages de discussion d'un article dans Wikipedia. Cependant, les lecteurs novices les consultent rarement (Bryant et al. (2005)). Pour mettre en évidence l'existence d'une discussion, nous avons inséré un simple indicateur en bas de notre profil de visualisations : si elle existe, sa taille est indiquée sous forme textuelle. De plus, si de nouveaux commentaires ont été ajoutés au cours des deux dernières semaines, l'indicateur devient un lien vers la page de discussion. Ainsi, nous attirons l'attention du lecteur sur l'activité récente de la discussion et l'encourageons à la consulter.

Profils types des articles. Un échantillon extrait d'une de nos copies de travail de la base de données Wikipedia (Avril 2006) est montré sur la Figure 2. On remarque rapidement que le premier de ces profils est plus récent que les autres et qu'il a un faible nombre de contributeurs. C'est le profil typique d'une *Ébauche* : créé récemment, avec très peu de contributeurs parmi lesquels l'un a rédigé presque l'intégralité du contenu, une frise plate, peu de wikiliens et pas de page de discussion. Dans la plupart des cas, ces articles sont aussi très courts, celui-ci contient environ 1000 mots. Un lecteur occasionnel peut penser qu'un article d'une telle ampleur traduit un certain niveau de maturité alors que son profil indique le contraire. L'article b est en train d'atteindre un certain niveau de maturité, avec un contributeur prédominant, une taille qui a atteint un plateau et une bonne densité de liens internes. Un détail non négligeable est l'absence de page de discussion. On peut considérer que cet article est un *Bon Début*. Enfin, les profils c et d sont ceux d'articles mûrs avec un nombre important de contributeurs. Ils diffèrent par la densité de liens et l'activité de la discussion. Le profil d indique une bonne densité de liens. Il y a quelques entrées dans la page de discussion, mais pas d'activité récente. Cela peut indiquer

que l'article a atteint une phase de stabilisation, révélatrice d'une bonne qualité. À l'inverse, le profil c affiche une taille très variable, une forte densité de liens et une activité récente de la discussion. Même si cet article peut être considéré comme mature, le profil indique que l'article change encore et le lecteur devrait prendre en considération cette information.

5 Étude utilisateur et discussions

Nous avons mené trois études afin d'évaluer l'efficacité des visualisations WikipediaViz pour aider l'utilisateur à améliorer leur première impression de la qualité d'un article. Deux études pilotes ont été menées avant que le système de classification des articles selon leur qualité soit intégré à Wikipedia. En l'absence de vérité terrain, il était difficile et risqué de dresser des conclusions. Ces deux sessions ont pourtant révélé d'importants points. Nous n'avions par exemple pas anticipé que la présence des visualisations ne garantissait pas alarmer le lecteur de problèmes potentiels de qualité, un nombre non négligeable de sujets ayant admis ne pas avoir essayé de les utiliser pour remplir la tâche. Pour la troisième étude, nous avons utilisé la classification récemment introduite dans Wikipedia⁴ comme référence pour comparer le niveau de qualité assigné aux articles avec la qualité estimée par l'utilisateur.

Conditions expérimentales 24 sujets bénévoles (19 hommes, 5 femmes) âgés de 22-62 ans ont participé à l'expérimentation menée dans une bibliothèque et musée scientifiques (14 lecteurs occasionnels de Wikipedia, 8 utilisateurs fréquents, et 2 ne connaissant pas Wikipedia). La tâche était d'évaluer la qualité de 24 articles, 6 pour chaque niveau de qualité parmi Article de Qualité (AdQ), Bon Article (BA), Bon début (BD) et Ébauche (Eb) (**facteur Qualité**). Les articles ont été présentés avec différentes interfaces (**facteur Technique**) : l'interface Wikipedia (A) et les visualisations WikipediaViz uniquement (V), i.e. le contenu de l'article était masqué, seul son titre était visible. La Qualité et la Technique ont été contrebalancées en utilisant un carré Latin. Nous avons enregistré la réponse de l'utilisateur et mesuré le Temps de réponse.

Hypothèses. Notre hypothèse était que les visualisations pouvaient aider le lecteur à estimer la qualité d'un article, à la fois en Temps et en Précision. Cela se traduirait par une corrélation entre la Technique et à la fois le Temps et la Précision. On définit la Précision (P) comme étant la distance entre la qualité de l'article (de 1 pour AdQ à 4 pour Eb) et la réponse du sujet.

Résultats et discussion. L'analyse de la variance n'a révélé aucun effet significatif de la Qualité et la Technique sur P . La courbe de P par rapport à la Qualité et à la Technique (Fig. 6a) montre que P est très similaire pour les différentes Techniques et Qualités : l'estimation de la qualité est toujours bonne avec ou sans les visualisations. Même s'il ne valide pas notre hypothèse, ce résultat montre que le profil visuel est suffisamment expressif pour évaluer la qualité avec une précision similaire à la condition où le contenu de l'article est visible.

L'analyse de la variance a montré un effet significatif de la Technique et la Qualité sur le Temps ($F_{1,23} = 31.60$, $p < 0.0001$ et $F_{3,69} = 15.52$, $p < 0.0001$). Les comparaisons post-hoc de moyennes par paires (t-test de Student, $\alpha = 0.050$) a indiqué que le temps requis pour

4. <http://en.wikipedia.org/wiki/Wikipedia:1.0>

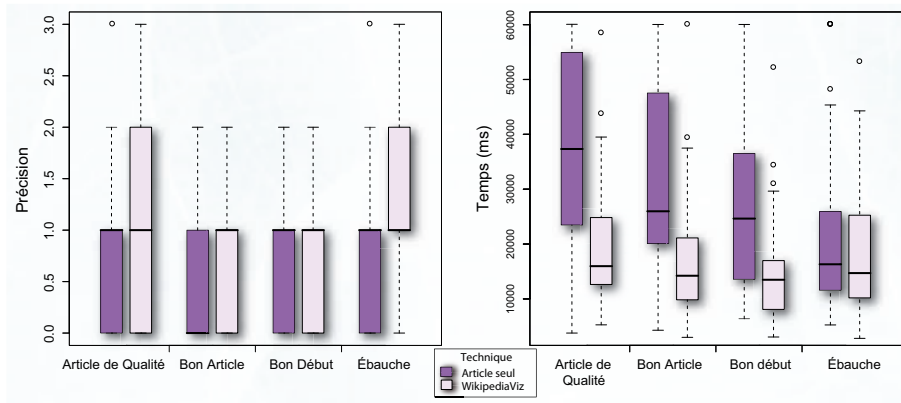


FIG. 6 – (a) Précision de la réponse et (b) Performance, Temps par Technique et Qualité.

accomplir la tâche était plus court lorsque le profil visuel seul était visible (V, moyenne=16.8s) que lorsque le contenu était visible (A, moyenne=29.7s). L'analyse post-hoc pour la Qualité (Tukey HSD, $\alpha = 0.050$) a montré que AdQ et BA nécessitaient moins de temps pour être estimés que BD et Eb (Fig. 6b). Les articles de faible qualité (petit contenu et visualisations caractéristiques) semblent donc plus faciles à reconnaître que les articles de grande qualité qui nécessitent une analyse plus approfondie.

Les résultats de cette étude ne satisfont pas exactement nos hypothèses. Cependant, ils démontrent que WikipediaViz réduit le temps requis pour estimer la qualité par un facteur de 2 sans pénaliser la précision.

Évaluation subjective. Nous avons interviewé les participants après l'expérimentation pour recueillir les impressions sur les visualisations. Tous ont trouvé que l'affichage d'informations révélant le processus d'édition d'un article était utile. Plusieurs sujets ont mentionné qu'ils ont beaucoup plus appris sur les caractéristiques d'un article grâce aux visualisations qu'à la simple vision de son contenu, mais il leur a fallu du temps pour comprendre ce que les visualisations exprimaient. Cela signifie que l'utilisation de WikipediaViz requiert un apprentissage, mais que les utilisateurs occasionnels sont capables d'apprendre comment les utiliser efficacement. Certains participants auraient toutefois souhaité une simple mesure agrégée telle que « bon » ou « mauvais » dans laquelle ils auraient une confiance aveugle sans avoir aucune idée de la manière de la calculer.

6 Détails sur l'Implémentation

WikipediaViz est implémenté en PHP et intégré comme un plugin dans le système Media-wiki. Il utilise la base de données standard Wikipedia avec des tables additionnelles que nous calculons pour visualiser rapidement la frise et les contributions des auteurs. Ces tables sont actuellement stockées dans notre copie locale de la base de données Wikipedia et ont nécessité

trois semaines de calculs. Si nos visualisations étaient menées à être intégrées à Wikipedia, le calcul incrémental des données agrégées prendrait un temps négligeable et peu d'espace de stockage. Cependant, si l'on considère la charge actuelle des serveurs Wikipedia, cela pourrait être un problème à considérer.

Pour générer les visualisations, nous utilisons essentiellement des balises HTML, à l'exception du graphique circulaire, qui est une image générée à la volée (et mise en cache en utilisant le même mécanisme que le système de cache des pages générée par la base de données Wikipedia).

Le code est disponible comme logiciel gratuit et la version expérimentale est également disponible pour récolter des retours de la communauté Wikipedia, initialement pour un nombre limité d'administrateurs, et à terme, pour le reste des internautes.

7 Conclusion

Nous avons décrit le processus de conception itératif de WikipediaViz : cinq visualisations destinées à améliorer l'interface standard de Wikipedia pour les lecteurs occasionnels. Ces visualisations ont été conçues avec des administrateurs Wikipedia, des contributeurs majeurs et des sociologues. Elles montrent les mesures importantes pour estimer la qualité des articles, mais difficiles à collecter via l'interface standard que ne maîtrisent pas les novices. Nous avons montré que les visualisations améliorent significativement le temps requis pour évaluer la qualité des articles sans dégrader la précision, qui reste bonne.

Nos résultats d'expérimentation ne sont pas définitifs et nécessitent une étude plus approfondie mais ils révèlent déjà des problèmes importants de confiance aveugle en Wikipedia que nous n'avions pas anticipés. En ce sens, nous pensons que WikipediaViz peut contribuer à une meilleure utilisation de Wikipedia et adresser quelques uns des problèmes fréquents tels que la prolifération de bannières et l'exploration approfondie pour comprendre le profil des articles.

Maintenant que WikipediaViz a montré être utile, nous avons commencé à implémenter une version « live » où les valeurs des mesures sont calculées à la volée. L'infrastructure, qui nécessite de maintenir une version de Wikipedia à jour est lourde et complexe et demandait des efforts de mise en place sans même être sûrs de l'efficacité des visualisations. Nous espérons que ce sera un premier pas vers l'inclusion de WikipediaViz dans le système standard de Wikipedia.

Enfin, des études longitudinales supplémentaires sont nécessaires pour comprendre l'acceptation et l'utilisation de WikipediaViz par les lecteurs occasionnels et plus généralement approfondir notre réflexion sur le fait que les visualisations simples peuvent être efficacement utilisées pour améliorer l'expérience de l'utilisateur lorsqu'il accède à un contenu dynamique, mais l'évaluation du bénéfice reste une tâche difficile.

Références

- Adler, T. et L. de Alfaro (2007). A content-driven reputation system for the wikipedia. In *WWW'07*, pp. 261–270.
- Blumenstock, J. (2008). Size matters : word count as a measure of quality on wikipedia. In *WWW'08*, pp. 1095–1096.

- Bruls, M., K. Huizing, et J. J. V. Wijk (2000). Squarified treemaps. In *In Proc. of the Joint Eurographics and IEEE TCVG Symp. on Visualization*, pp. 33–42.
- Bryant, S. L., A. Forte, et A. Bruckman (2005). Becoming wikipediaian : transformation of participation in a collaborative online encyclopedia. In *GROUP'05*, pp. 1–10.
- Dondio, P., S. Barrett, S. Weber, et J. Seigneur (2006). Extracting trust from domain analysis : A case study on the wikipedia project. *Autonomic and Trusted Computing*, 362–373.
- Druck, G., G. Miklau, et A. McCallum (2008). Learning to predict the quality of contributions to wikipedia. In *WIKIAI 08*, pp. 7–12.
- Giles, J. (2005). Internet encyclopaedias go head to head. *Nature* 438(7070), 900–901.
- Kittur, A. et R. E. Kraut (2008). Harnessing the wisdom of crowds in wikipedia : quality through coordination. In *CSCW '08*, pp. 37–46.
- Lee, R. et T. Bill (2007). Wikipedia users. Pew Internet and American Life Project.
- Pirolli, P., E. Wollny, et B. Suh (2009). So you know you're getting the best possible information : a tool that increases wikipedia credibility. In *CHI '09*, pp. 1505–1508.
- Stvilia, B., M. B. Twidale, L. C. Smith, et L. Gasser (2005). Assessing information quality of a community-based encyclopedia. In *ICIQ'05*, pp. 442–454.
- Suh, B., E. H. Chi, A. Kittur, et B. A. Pendleton (2008). Lifting the veil : improving accountability and social transparency in wikipedia with wikidashboard. In *CHI'08*, pp. 1037–1040.
- Viegas, F. B., M. Wattenberg, J. Kriss, et F. van Ham (2007). Talk before you type : Coordination in wikipedia. *Hawaii International Conference on System Sciences 0*, 78a.
- Viégas, F. B., M. Wattenberg, et D. Kushal (2004). Studying cooperation and conflict between authors with history flow visualizations. In *CHI '04*, pp. 575–582.
- Wattenberg, M., F. Viégas, et K. Hollenbach (2007). Visualizing activity on wikipedia with chromograms. In *INTERACT 2007*, pp. 272–287.
- Wilkinson, D. M. et B. A. Huberman (2007a). Assessing the value of cooperation in wikipedia. *Firstmonday*.
- Wilkinson, D. M. et B. A. Huberman (2007b). Cooperation and quality in wikipedia. In *WikiSym '07 : Proceedings of the 2007 international symposium on Wikis*, pp. 157–164.

Summary

Wikipedia has become one of the most used knowledge bases worldwide; the problem of the trustworthiness of the information it disseminates becomes central. With *WikipediaViz*, we introduce five thumbnail visualizations integrated to the Wikipedia layout that can keep casual Wikipedia readers aware of important meta-information about the articles they read. The design of *WikipediaViz* was inspired by two participatory design sessions with Wikipedia experts. They explained the clues they used to quickly assess the trustworthiness of articles and expressed the need for tools with a view of raising quality issues for them as readers. Our user study shows that *WikipediaViz* significantly reduces the time required to assess the quality of articles while maintaining a good accuracy.