

Les références bibliographiques dans Wikipédia

Gilles SAHUT

Doctorant, LERASS, Université Paul Sabatier, Toulouse 3.

Résumé : La crédibilité de Wikipédia repose en grande partie sur le référencement de ses contenus. Les articles de cette encyclopédie collaborative intègrent donc des références bibliographiques. Nous proposons ici une synthèse des travaux sur l'usage de ces références pour la recherche d'information et son évaluation. Nous récapitulons également les résultats des recherches portant sur le nombre et la nature des sources citées dans l'encyclopédie. Leur grande diversité suscite des interrogations quant à l'existence d'une hiérarchie documentaire sur Wikipédia. Nous proposons alors des pistes de recherche visant à mieux comprendre les spécificités de ce modèle éditorial. Comment les règles relatives au référencement des sources sur Wikipédia évoluent-elles ? De quelle manière la communauté s'approprie-t-elle ces règles et les applique-t-elle ?

Mots-clés : Wikipédia, référencement, sources d'information, jugement de crédibilité, recherche d'information.

Toute entreprise encyclopédique a pour objectif d'élaborer une synthèse des savoirs existants à des fins didactiques et suppose, par là même, la reformulation d'informations. Ce processus est à l'œuvre dans Wikipédia. Cependant, les conditions d'élaboration de cette encyclopédie collaborative sont spécifiques. Elle est souvent définie comme l'encyclopédie à laquelle « *tout le monde peut participer* », ce qui génère, depuis le milieu des années 2000, de nombreux débats sur sa crédibilité. Face à cela, l'encyclopédie énonce le principe de vérifiabilité comme l'une des règles inhérentes à son écriture et son édition : « *Une information ne peut être mentionnée que si les lecteurs peuvent vérifier qu'elle a déjà été publiée par une source ou référence de qualité* » [1]. En ce sens, on peut considérer que les références bibliographiques¹³ figurant en notes ou à la fin d'un article de

¹³ Nous considérons ici qu'une référence bibliographique est « *[l']ensemble des éléments de données nécessaires pour identifier un document ou une partie de document de tout type, sur tout support (livre, article, site web, etc.)* » Boulogne, Arlette (coord.). *Vocabulaire de la documentation*. ADBS Éditions, 2004

Wikipédia, constituent des indices – voire des preuves – d'une reformulation de sources publiées. Or le paysage informationnel actuel est caractérisé par la multiplication du nombre de documents accessibles et leur extrême hétérogénéité. Le développement du web - et plus encore du web social – a favorisé la production de sources autoritatives résultant d'un processus d'auto-publication et émanant d'un auteur qui construit « *lui-même les conditions de sa reconnaissance dans l'univers électronique* » [2] ; ces documents cohabitent avec d'autres, issus de modèles éditoriaux plus traditionnels. Ce constat suscite donc une interrogation sur la nature et les caractéristiques des sources citées dans Wikipédia, l'encyclopédie étant parfois accusée de reposer sur des sources douteuses¹⁴.

Afin de souligner l'intérêt de cette thématique, les usages des références bibliographiques faits par les lecteurs de cette encyclopédie seront tout d'abord analysés. Nous synthétiserons ensuite les recherches récentes portant sur le nombre et la nature des sources citées dans Wikipédia. Nous proposons enfin des pistes de recherche suivies actuellement dans le cadre d'un doctorat en sciences de l'information et de la communication.

Les usages des références disponibles sur Wikipédia

Les mesures d'audience sur le web indiquent que Wikipédia est actuellement l'un des dix sites les plus visités au monde¹⁵. Cette notoriété se retrouve chez les jeunes, population dont les usages numériques sont les plus étudiés. Toutes les enquêtes sont unanimes, Wikipédia est devenue une source de premier plan pour la grande majorité des élèves et des étudiants, que ce soit pour leurs recherches personnelles ou dans le cadre de leur travail scolaire ou universitaire [3] [4] [5]. Elle est majoritairement exploitée lors de la phase initiale de la recherche d'information, sa consultation paraissant nécessaire afin de cerner le sujet de la recherche et de comprendre les termes qui le composent [4] [5] [6]. Ainsi il est particulièrement fréquent que les jeunes activent les liens hypertextes présents dans la bibliographie ou les notes des articles de Wikipédia. Grâce à ces références, ils accèdent à d'autres documents en ligne jugés pertinents pour leur recherche [5] [7] [8] [9]. On retrouve donc chez les jeunes une stratégie particulière de recherche

¹⁴L'une des premières polémiques autour de la Wikipédia française est née à la fin de l'année 2006 d'un problème lié à une référence présente dans l'article consacré à l'affaire Dreyfus. Un livre violemment antidreyfusard figurait dans la bibliographie accompagné d'une recommandation « *Ouvrage fondamental à consulter en priorité* ». Plusieurs journalistes (Daniel Garcia puis Pierre Assouline notamment) s'en sont indignés sur leurs blogs, générant ainsi des débats sur la fiabilité de cette encyclopédie.

¹⁵ Selon Alexa, Wikipédia.org figure au 7^e rang des sites les plus visités au monde en juin 2011.

d'information, le « chainage » qui avait été repérée par Ellis chez les chercheurs en sciences sociales à la fin des années 1980 [10]. Ainsi Wikipédia n'est pas seulement une ressource encyclopédique, elle peut être considérée comme une passerelle vers d'autres documents. En un sens, elle fait office de répertoire de ressources sur un thème alors que la plupart des annuaires généralistes « historiques » du web ont disparu.

Les références bibliographiques ne sont pas utilisées dans le seul but de repérer de nouveaux documents. Elles sont également susceptibles d'être examinées afin de formuler un jugement de crédibilité¹⁶. Les études témoignent sur ce point de pratiques évaluatives qui diffèrent. Certaines indiquent que rares sont les jeunes qui les prennent en compte afin d'évaluer le degré de crédibilité des articles de Wikipédia [6] [12].

Celle de Lim et Simon souligne que le thème de la recherche d'information est susceptible d'influencer cette pratique : quand celui-ci est jugé sérieux (par exemple les questions portant sur la santé), une certaine attention serait accordée aux références afin de formuler un jugement de crédibilité. En revanche, pour des sujets relatifs aux loisirs, les références ne seraient pas prises en compte. Néanmoins, selon l'aveu même des auteurs, ce résultat mérite confirmation [9]. La prise en compte des références semble également dépendre du degré de confiance accordé à Wikipédia.

D'après une expérimentation menée en psychologie cognitive, le nombre de références présentes dans un article de Wikipédia a une incidence positive sur sa crédibilité quand les personnes accordent un haut degré de confiance à l'encyclopédie collaborative. Celles qui doutent de la fiabilité de l'encyclopédie collaborative sont davantage attentives à la qualité des références [13]. Sur un autre plan, la participation au projet Wikipédia et la connaissance de son mode de fonctionnement paraissent avoir un impact net sur les pratiques évaluatives. Ainsi les jeunes contributeurs à la version chinoise de Wikipédia accessible à Hong-Kong réussissent à formuler des jugements de crédibilité argumentés et informés même s'ils disposent de peu de connaissances sur le thème de l'article. Ils s'appuient alors sur les indices périphériques tels les références et les liens externes qui sont disponibles [14].

Les études recensées convergent donc sur un point : lors de la consultation des articles de l'encyclopédie, les usagers –notamment les jeunes– exploitent fréquemment les références citées quand elles pointent vers des documents en ligne. Il arrive aussi que celles-ci soient analysées

¹⁶ Mobilisé dans une pluralité de discipline, le concept de crédibilité fait l'objet de définitions et d'approches multiples. Au sein du contexte des études sur la recherche d'information, il est possible de considérer le jugement de crédibilité comme une composante du jugement de pertinence. Il comporte principalement deux dimensions : la perception de l'expertise de la source et celle de sa bienveillance (ou son intention de dire la vérité). De multiples variables entrent en compte dans sa formulation : les connaissances de l'individu, son expérience, la situation dans laquelle il se trouve [9] [11].

à des fins d'évaluation de l'article. Néanmoins, la fréquence de cette pratique et les facteurs la conditionnant ne sont pas encore connus avec précision. La constitution de listes bibliographiques placées en notes ou à la fin des articles de Wikipédia permet donc d'attribuer une nouvelle visibilité aux documents cités et les rend disponibles pour de nouveaux usages, ce qui suscite alors une interrogation sur le nombre, les caractéristiques et la qualité de ces sources.

Nombre et type de sources référencées sur Wikipédia

L'étude des références dans Wikipédia s'inscrit dans un débat scientifique plus général : celui de l'évaluation de la qualité informationnelle de l'encyclopédie. Depuis les articles de Liu [15] et de Giles [16], les recherches scientifiques sur ce sujet empruntent principalement deux voies. La première méthode est celle de l'analyse d'un corpus d'articles de Wikipédia par des experts dans un domaine donné. Ceux-ci ont alors recours aux critères d'évaluation de l'information habituellement usités dans le cadre des sciences de l'information (exactitude, lisibilité, complétude...) ¹⁷. Une approche comparative avec d'autres sources est parfois mise en œuvre. La seconde approche vise à identifier des variables (nombre de contributeurs à l'article, nombre d'éditions de l'article, profil des contributeurs ...) afin de proposer des modélisations statistiques permettant de prédire la qualité d'un article ¹⁸. L'analyse des références citées dans Wikipédia ¹⁹ s'inscrit plutôt dans la première approche, ce qui n'exclut pas le recours à un outillage statistique.

En 2007, l'étude exploratoire de Willinsky réalisée à partir d'un corpus de 100 articles choisis de manière aléatoire indique une moyenne de 1,68 références par article. Ce chiffre peu élevé pouvant en partie s'expliquer par le caractère aléatoire du choix des articles. Alors que les encyclopédies généralistes traditionnelles sont centrées sur des thèmes académiques, Wikipédia est ouverte à la culture populaire : l'échantillon de Willinsky comprend ainsi tout une gamme de sujets (personnages de bandes dessinées, émissions télévisées, albums de groupes de rock plus ou moins connus...) qui n'ont pas fait l'objet d'une consécration académique.

Il est possible que pour de tels sujets, les contributeurs se fondent uniquement sur leurs connaissances personnelles et ne jugent pas nécessaire d'indiquer leurs sources. Willinsky relève cependant que des entrées plus classiques comportent également un nombre réduit de

¹⁷ Par exemple [16] [17] [18]

¹⁸ Par exemple [15] [19] [20]

¹⁹ Précisons que les études citées dans cette communication porte sur la Wikipédia en langue anglaise à l'exception de celle d'Huvila qui porte sur la Wikipédia en langue suédoise 2010 [26].

références [21]. L'étude réalisée par le même auteur l'année suivante offre un point de vue radicalement différent. Elle met en évidence que les éditeurs de l'encyclopédie collaborative ont intégré un nombre élevé de références pointant vers des écrits scientifiques numériques en accès libre (5000 liens vers arXiv.org, 1000 vers Social Science Research Network) mais aussi vers des banques de données partiellement ou totalement payantes (2400 liens vers Science Direct, 4500 vers JSTOR...). Dans le domaine philosophique, beaucoup d'articles comportent des références à la *Stanford Encyclopedia of Philosophy*, ressource gratuite en ligne élaborée par des universitaires selon un modèle scientifique. Par ailleurs, les textes de cette encyclopédie sont également mentionnés dans les pages de discussion accompagnant des entrées philosophiques de Wikipédia, nourrissant d'intenses échanges argumentatifs entre contributeurs [22].

Pour ce qui est des sciences exactes, Nielsen démontre à partir de méthodes statistiques que les articles de ce domaine de connaissance sont majoritairement dotés de références bibliographiques de travaux de recherche ; ce qui n'exclut pas la présence de sources non scientifiques tels des quotidiens d'actualité. Il relève également que le nombre de ces références scientifiques a tendance à augmenter au fur et à mesure de la croissance de l'encyclopédie et que les revues les plus citées sont celles qui, à l'instar de *Nature* ou *Science*, bénéficient d'un facteur d'impact élevé dans le *Journal Citation Reports*. D'après ces éléments, on pourrait considérer Wikipédia comme une source de bonne qualité dans le domaine scientifique [23]. En 2010, Haigh a recensé et caractérisé les références de 50 articles sur des thèmes relatifs aux sciences médicales et à la santé. En moyenne, chaque entrée comportait 50 références. Plus de la moitié de celles-ci indiquaient des sources jouissant d'une bonne réputation parmi les spécialistes du domaine.

Selon l'auteur, il est fort probable que, dans certaines disciplines telles l'anatomie et ou la physiologie, les sources mentionnées dans Wikipédia soient de nature plus scientifiques que celles utilisées pour la réalisation de certains manuels universitaires [24]. Toutefois, ces conclusions ne peuvent être généralisées. A l'issue d'une analyse d'un corpus d'articles sur l'histoire des pays, Luyt et Tan concluent que Wikipédia ne devrait pas être utilisée pour la recherche de références.

Outre le fait que leur nombre soit jugé peu élevé (en moyenne, 10,16 références par articles), les auteurs soulignent un décalage avec les pratiques de la publication scientifique dominante en histoire : alors que dans cette discipline, l'écrit de recherche est encore essentiellement diffusé sur support imprimé, les sources citées dans Wikipédia sont majoritairement issues du web. Par ailleurs, ils notent l'importance des citations de médias d'actualité en ligne (11,9% du total des références), ainsi que celle des sources relevant de sites gouvernementaux des différents pays (7,3%), de l'Etat fédéral américain ou d'institutions qui lui

sont liées²⁰ (9,8%) ; ceci ne correspond évidemment pas aux canons de l’édition savante. Luyt et Tan relèvent enfin que l’écrasante majorité des références est en langue anglaise (91%) alors que les articles analysés traitaient de sujets internationaux. Dès lors, des interrogations surgissent quant à la crédibilité et la neutralité des sources utilisées et, par ricochet, quant à celles des articles concernés. Le risque est ici une absence de pluralisme et l’imposition d’un point de vue nord-américain sur les histoires nationales [25].

Huvila a abordé la question de la référence au sein de Wikipédia sous un autre angle. Il s’est intéressé non pas à l’encyclopédie en tant que produit éditorial mais à ses contributeurs²¹ et aux sources qu’ils utilisent afin de participer à l’écriture d’un article. Peu étonnant au regard des résultats précédemment évoqués, la majorité des informations sont issues du Web, les « Wikipédiens » faisant, comme l’ensemble des internautes, un fort usage des moteurs de recherche et en particulier de Google. Toutefois, l’éventail des sources exploitées - citées et/ou reformulées - s’avère large (livres parfois trouvés dans des bibliothèques, presse en ligne, revues scientifiques, articles de Wikipédia en d’autres langues, connaissances personnelles qui sont mobilisées sans faire l’objet de références formelles...).

Cette diversité documentaire peut se comprendre par les profils différenciés des participants. En effet, Huvila identifie cinq types de contributeurs :

- les « *investigateurs* » qui mènent des recherches approfondies sur des thèmes correspondant à leurs centres d’intérêt à partir de médias d’actualité, documents vulgarisés ou factuels ;
- les « *surfeurs* », soucieux d’économiser leur temps et leurs efforts qui fondent leurs contributions sur des sources facilement repérables sur le web ;
- les « *matérialistes prudents* » qui s’appuient prioritairement sur les connaissances issues de leurs propres expériences ;
- les « *scientifiques* », souvent des doctorants ou des jeunes chercheurs qui proposent des états de la question fondés sur des écrits de recherche ;
- les « *éditeurs* » qui ne recherchent pas d’informations pour leur contribution, se concentrant prioritairement sur les tâches administratives, les corrections grammaticales, la traduction d’articles de Wikipédia en langue étrangère [26].

²⁰ Le Département d’Etat, la CIA ou encore le site Country Studies réalisé par un département de la bibliothèque du Congrès et parrainé par le Ministère des armées des Etats-Unis.

²¹ Suédois en l’occurrence.

Pistes de recherche

La nature et la qualité des références trouvées dans Wikipédia sont donc diverses. Elles varient selon les thématiques et les domaines de connaissance abordés mais également dans le temps, Wikipédia se caractérisant par une instabilité temporelle [27]. A l'instar d'autres types de publications [28], l'encyclopédie constitue un espace de médiations hybrides qui offre un enchevêtrement de références documentaires aux statuts différents. Ce constat peut entretenir des doutes sur la crédibilité de l'encyclopédie.

Il suscite parallèlement des interrogations sur les évolutions passées et à venir de cette entreprise éditoriale. Une hiérarchie des sources se dessine-t-elle au sein de la communauté wikipédienne? Le modèle éditorial de Wikipédia évolue-t-il afin de se rapprocher des standards académiques en vigueur pour l'évaluation des sources utilisées et citées? Ou reste-t-il marqué par un principe égalitaire pouvant impliquer des appréciations indistinctes de la valeur documentaire?

Wikipédia peut être qualifiée d'espace documentaire participatif et évolutif dans lequel la communauté participante établit, selon des principes démocratiques, ses propres règles de fonctionnement et les recommandations à suivre lors de l'élaboration des articles [29]. Le dispositif technique sur lequel s'appuie l'encyclopédie comporte un système d'archivage qui garde trace des débats accompagnant les diverses prises de décision communautaires. Dès lors, il devient possible de mener une étude diachronique afin d'identifier les transformations des règles relatives à la pratique citationnelle, de repérer leurs moments d'émergence, de consolidation, voire de remise en question et de tenter de cerner les facteurs qui ont présidé à leur évolution.

De manière complémentaire, les modalités d'appropriation de ces règles par les contributeurs de l'encyclopédie méritent d'être étudiées ainsi que les éventuelles résistances à leur application, leurs détournements ou leurs contournements. Il s'agira alors d'examiner les fonctions assignées aux références bibliographiques par les participants à l'encyclopédie dans le cadre des discussions accompagnant les éditions successives des articles. Nous entendons là analyser la manière dont les différentes sources citées sont qualifiées, discutées, différenciées afin de discerner si une hiérarchie documentaire émerge au fil du temps et réussit à s'imposer au sein de la communauté «wikipédienne».

Bibliographie

- [1]. Wikipédia (2011). Wikipédia:Vérifiabilité.
<http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:V%C3%A9rifiabilit%C3%A9>.
Consulté le 10/06/2011.

- [2] BROUDOUX E., GRESILLAUD S., LE CROSNIER H. & LUX-POGODALLA V. , Construction de l’auteur autour de ses modes d’écriture et de publication. In Actes Conférence Hypertextes Hypermédiâs H2PTM’05 : Créer, jouer, échanger: expériences de réseaux. Hermès, 2005.
- [3] LUCKIN R., LOGAN K., CLARK W., GRABER R., OLIVER M., & MEE A., Learners’ use of Web 2.0 technologies in and out of school in Key Stages 3 and 4: reseach report. Becta, 2008.
- [4] HAMPTON-REEVES S., MASHITER C., WESTAWAY J., LUMSDEN, P., DAY H. & HEWERTSON H., Students’ Use of Research Content in Teaching and Learning. University of Central Lancashire, 2009.
- [5] EISENBERG M. & HEAD A., How today’s college students use Wikipedia for course-related research. First Monday, 2010, 15(3).
- [6] SUNDIN O. & Francke, H., In search of credibility: Pupils’ information practices in learning environments. Information Research, 2009, 14(4).
- [7] LUYT B. et al., Youth Perception and Usage of Wikipedia. Information Research, 2008, 13(4).
- [8] LIM S., How and Why Do College Students Use Wikipedia? Journal of the american society for information science and technology, 2009, 60(11).
- [9] LIM S. & SIMON C., Credibility judgment and verification behavior of college students concerning Wikipedia. First Monday, 2011, 16(4).
- [10] ELLIS D., A behavioural approach to information retrieval system design. Journal of documentation, 1993, 45(3).
- [11] RIEH S. Y. & DANIELSON D. R., Credibility: A multidisciplinary framework. Annual review of information science and technology, 2007, 41(1).
- [12] FORTE A. & BRUCKMAN A., Learning information literacy in the age of Wikipedia. In Proceedings of the International Conference of the Learning Sciences, 2008, Utrecht, Netherlands,.
- [13] LUCASSEN T., NOORDZIJ M. L., & SCHRAAGEN J. M., Reference Blindness: The influence of references on trust in Wikipedia. In Proceedings of the ACM WebSci’11, June 14-17 2011, Koblenz.
- [14] CHAN C., CHAN P. D. J. Y. & CHAN A. D., Young Wikipedians’ perceptions of Wikipedia. In Wikimania 2011 Haifa.
- [15] LIH A., Wikipedia as Participatory Journalism: Reliable Sources? Metrics for evaluating collaborative media as a news resource. In 5th International Symposium on Online Journalism 2004.
- [16] GILES J., Internet encyclopaedias go head to head. Nature, 2005, 438(7070).
- [17] CHESNEY T., An empirical examination of Wikipedia’s credibility. First Monday, 2006, 11(11).
- [18] ELVEBAKK B., Philosophy Democratized? First Monday, 2008, 13(2-4).
- [19] WILKINSON, D. M., & HUBERMAN, B. A. Assessing the value of cooperation in Wikipedia. First Monday, 2007, 12(4-2).
- [20] JACQUEMIN P., LAUF A., POUDAT C., HURAUULT-PLANTET M. & AURAY N., La fiabilité des informations sur le web: le cas Wikipédia. In CORIA 2008, 2008, 227-235.
- [21] WILLINSKY J. What open access research can do for Wikipédia. First Monday, 2007, 12(3).
- [22] WILLINSKY J., Socrates Back on the Street: Wikipedia’s Citing of the Stanford Encyclopedia of Philosophy. International Journal of Communication, 2008, 2.

Les références bibliographiques dans Wikipédia

- [23] NIELSEN, F. A. Scientific citations in Wikipedia. *First Monday*, 2007, 12(8-6).
- [24] HAIGH, C. A. Wikipedia as an evidence source for nursing and healthcare students. *Nurse Education Today*, 2011, 31(2).
- [25] LUYT B. & TAN D., Improving Wikipedia's credibility: References and citations in a sample of history articles. *Journal of the American Society for Information Science and Technology*, 2010, 61(4).
- [26] HUVILA I., Where does the information come from? Information Source Use Patterns of Wikipedia. *Information Research*, 2010, 15(3).
- [27] ENDRIZZI L., Wikipédia : de la co-rédaction au co-développement de la communauté. In Chartron, G. Broudoux, E. (dir.). *Actes de la conference DocSoc, 2006*. ADBS, 2006.
- [28] COUZINET V., Médiations hybrides: le documentaliste et le chercheur en sciences de l'information. ADBS, 2000.
- [29] ZACKLAD, M. Espace documentaire participatif et gouvernance. In Congress of the European Regional Science Association (47th Congress) and ASRDLF (Association de Science Régionale de Langue Française, 44th Congress) PARIS - August 29th - September 2nd, 2007.