

Wikipedia, une encyclopédie en communs

Alexandre Monnin

► **To cite this version:**

Alexandre Monnin. Wikipedia, une encyclopédie en communs. Mokrane Bouzeghoub ; Rémy Mosseri. Les Big Data à Découvert, CNRS Editions, pp.2, 2017, À découvert. <hal-01374484>

HAL Id: hal-01374484

<https://hal.inria.fr/hal-01374484>

Submitted on 17 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Wikipédia, une encyclopédie en communs

Alexandre Monnin

L'encyclopédie en ligne Wikipédia constitue l'encyclopédie mondiale de référence. À ce titre, ses racines plongent jusqu'au Moyen Âge (Isidore de Séville), en passant par l'époque des Lumières (Diderot et d'Alembert). L'espace ouvert par le numérique a modifié en profondeur la chaîne éditoriale qui préside à la création d'une encyclopédie comme Wikipédia. Prolongeant les possibilités techniques offertes par le wiki, inventé dès 1994 par Ward Cunningham, Wikipédia s'est dotée de règles visant à ouvrir l'ensemble de ses articles à la contribution tout en ménageant une place à l'expertise. Celle-ci s'est déplacée de l'auteur, figure solitaire responsable de sa production, vers les sources, qui, seules garantissent le bien-fondé de contributions désormais pseudonymes. L'importance prépondérante des sources dans Wikipédia (y compris comme idéal régulateur de pratiques) se mesure au paradoxe dont elle allège la tension : si chacun a désormais la possibilité de contribuer à l'édifice mondial du savoir c'est moins en tant qu'auteur, porteur d'une singularité, qu'en tant que scripteur, au sens le plus large du terme. Outre l'écriture de textes, cette figure se manifeste dans des activités telles que la compilation de citations, l'établissement de bibliographie ou encore le choix de références.

Neutralité et vérifiabilité

Wikipédia repose sur une tripartition des sources, dont découle l'ensemble de ses règles de contribution. Les sources primaires, tels les témoignages, y sont globalement dépourvues d'autorité. Les sources secondaires occupent, en revanche, une place centrale : c'est sur elles que sont censées reposer les contributions des Wikipédiens. La règle veut qu'elle soit même proscrite au motif qu'aucun travail original n'est acceptable sur Wikipédia. Si les sources secondaires doivent acquérir une crédibilité, celle-ci le sera obligatoirement ailleurs, sous peine d'autoréférentialité. Enfin, l'encyclopédie elle-même sur le modèle d'une « source tertiaire », conçue à la manière d'un rassemblement de sources secondaires (au passage, le texte des articles ne rentre dans aucune de ces trois cases). Le caractère « N P O V » (*neutral point of view*), règle la plus connue et sans doute la moins bien comprise de Wikipédia, établit moins la neutralité de toute forme de savoir que la nécessité de déléguer l'adoption d'un point de vue à des sources faisant autorité. Ces choix sont ensuite pondérés en fonction de la crédibilité des sources qui s'y rattachent. Ce qui revient à articuler une pluralité de points de vue lors de la rédaction de chaque article. La vérité une et unique est donc bannie de Wikipédia au profit de la vérifiabilité. Ce qui n'est pas sans évoquer une forme de « positivisme des sources », au motif que les faits sont moins l'expression de réalités intangibles que de savoirs scientifiques en constante évolution. En effet, les sources secondaires, à partir desquelles les « faits » sont construits, se conçoivent avant tout sur le modèle des articles scientifiques. Leur validité est temporaire et la factualité, sur Wikipédia, transitoire. Pour toutes ces raisons, l'open access constitue la condition sine qua non d'une authentique démocratisation de la contribution des usagers. À cet égard, les « bibliothèques pirates », dont la finalité est d'étendre l'open access et les communs du savoir, quitte à franchir les limites de la légalité, n'en répondent pas moins de manière optimale aux impératifs découlant des règles éditoriales de Wikipédia, en particulier la nécessité d'accéder au plus grand nombre possible de sources secondaires. Ainsi le site [aaaaarg.fail](#) a-t-il

développé des outils de citation permettant de référencer et d'accéder avec une extrême précision aux passages des ouvrages cités en générant des URI correspondantes.

De l'accumulation à l'interpénétration

Deux lectures de cette entreprise s'offrent à nous. La première, partagée par une partie de la communauté des contributeurs, y voit la poursuite du projet des Lumières, sur le mode positiviste d'une accumulation infinie du savoir. L'autre perspective, bien illustrée par Bouvard et Pécuchet, pointe le caractère illusoire de cette accumulation de sources, dont les multiples points de vue laissent les personnages démunis et en butte aux échecs successifs. Le roman s'achève sur une note sceptique, lorsque les deux anti-héros endossent à nouveau leurs habits de copistes pour se livrer au seul exercice encore envisageable au milieu des contradictions du savoir, à savoir la rédaction d'un dictionnaire... des idées reçues !

Une troisième lecture nous semble toutefois à considérer, à partir de l'architecture du Web. Dans cette perspective, rompant avec les conceptions documentaires héritées de l'hypertexte, les URI sur lesquelles repose le système de nommage du Web identifient ce que l'on appelle des « ressources ». Autrement dit, des objets, et ce quelle que soit leur nature. À ces objets sont associées des « représentations http » (les « pages »). Constamment calculées et recalculées, celles-ci évoluent sans cesse et leurs contenus, à l'image de leurs formats, ne connaissent aucunes limites. Ces représentations, en droit sinon de fait, ne sont jamais identiques les unes aux autres. Chacune en revanche se doit d'être fidèle à la ressource identifiée par l'URI y donnant accès. Les ressources sont relativement stables tant qu'on les isole de leurs représentations. Elles apparaissent alors comme le facteur en fonction duquel ces dernières sont générées de manière régulière, au double sens temporel et normatif du terme. En revanche, pour peu que l'on appréhende les ressources en lien avec leurs représentations et les sous-réseaux extrêmement divers qui s'y rapportent, les frontières de celles-ci se font plus floues et le Web, comme appareil de désignation, dresse alors le portrait de collections en devenir qui s'interpénètrent les unes les autres, rompant, et c'est tout leur intérêt pour qui peut parcourir ces nouveaux réseaux, avec les frontières relativement claires des objets du sens commun (mésoscopiques, c'est-à-dire de grain moyen).

Formalisation du sens commun

L'Intelligence Artificielle, sur son versant logique, visait précisément à formaliser ledit sens commun, au moins depuis les travaux de McCarthy et Hayes, afin de permettre aux machines de raisonner à partir du même univers de discours que les humains. Un projet comme CYC (pour enCYClopédie), né dans les années 1980, prolongeait cet effort en mettant au travail une poignée d'experts en ingénierie des connaissances afin de formaliser les axiomes du sens commun nécessaires à la bonne compréhension du savoir encyclopédique. Aujourd'hui encore, ces formalisations issues de l'IA que sont les ontologies informatiques privilégient une représentation de grain moyen (paradoxalement issue de la philosophie, comme en atteste la prévalence des modèles hérités de la substance aristotélicienne).

Il existe cependant d'autres manières d'aboutir à un accord sur des collectifs d'objets, comme DBpédia, jeu de données extraites à partir de Wikipédia, qui constitue à ce jour le projet-phare du Web de données (cf. V.4). DBpédia propose, depuis 2007, des connaissances structurées destinées à stabiliser une caractérisation des objets s'appuyant à la fois sur des prédicats du sens commun (issus

d'ontologies informatiques) et des sources secondaires (cf. II.5). Le tout afin de fournir un référentiel en constante évolution aux applications du Web. En somme, si le Web constitue un appareil de désignation à grande échelle, Wikipédia permet aux humains (sans oublier les robots, responsables d'environ 30 % des actes d'éditions sur Wikipédia en français) de négocier, à partir de sources établies, un tri répertoriant les entités composant un monde commun. À son tour, DBpédia étend cet effort aux machines au moyen des langages et ontologies du Web de données.

Référence bibliographique

- A.MONNIN – Du cycle de vie des données au cycle de vie des objets, in L. CALDERAN, B. HIDOINE et J. MILLET (dir.), Le document numérique à l'heure du web de données, ADBS éditions, 2012.