Wikipedia & the Structure of Knowledge

Wikipedia is an online free encyclopedia, editable by anyone, and its self-reported mission is to create a “comprehensive written compendium that contains information on all branches of knowledge.” Wikipedia has over 6 million articles on a dazzling network of subjects in 285 different languages. A searcher of knowledge can often turn to Wikipedia as a source for a quick fact, and in the wide world of the internet and all the information it contains, Wikipedia feels reliable and ever-present.

The Wikipedia project is also always expanding: More articles are constantly being created and updated, and these articles are integrated into the existing collection. This infinite promise of Wikipedia makes it in some ways similar to the Library of Babel, a concept from a short story of the same name written by Jorge Luis Borges in 1941. This library is infinitely vast and contains all books in all languages about every subject that has ever existed: it “enshrines all information.” However, while all knowledge can be found on the shelves in the Library of Babel, each specific book is a tiny point in the expanse of information and essentially impossible to find.

If this property became true of Wikipedia, it would be just as difficult to use: Someone looking for something would not be able to navigate the website easily to find it. Would Wikipedia, in striving for the goal to make itself a compendium on all knowledge, and to constantly update its metaphorical shelves with new knowledge as it is discovered, render itself useless? Thankfully, this is not likely. As Wikipedia is right now, even with an astounding 6 million articles, it is easy to use and usually trustworthy. In its current state of constant expansion, it is still as easy to navigate as ever, and this likely will not change. Thus, while some properties of Wikipedia – the structuring of networks and its ability to grow without bound – create certain parallels to the infinite library of Borges’ imagination, clearly there must be differences too. So, to what extent is Wikipedia, an information phenomenon of the 21st century, similar to the Library of Babel, this infinite, mythical library of all knowledge, and what makes Wikipedia a useful institution, while the Library is not?

Contents

1 The Library
2 Wikipedia vs the infinite library
   2.1 Wikipedia as an encyclopedia
   2.2 Wikipedia as the Library
      2.2.1 All the paths
      2.2.2 The compendium
   2.3 Wikipedia as a filter
      2.3.1 Notability
      2.3.2 Falsehoods
      2.3.3 Languages of Knowledge
3 Conclusion: Wikipedia’s usefulness
4 References
The Library [edit]

The infinite Library of Babel is arranged in hexagonal galleries all identical to each other, filled with bookshelves on which sit all books that could ever exist about all subjects, in the past, present, and future[3]: Twenty bookshelves line 4 out of 6 sides in each hexagon, 5 bookshelves per wall. Each book contains 410 pages, on each page is 40 lines, on each line approximately 80 letters. These letters consist of 25 “orthographic symbols”[3]: the comma, period, and space, plus 22 letters of an alphabet.

Some of these books detail reality, some of them contain only falsehoods[9]. Borges describes this library as a place that tempts people to search for truth – people who, upon hearing that the library contains all knowledge, embark on quests to find a book containing the information they seek – but the truth is always beyond reach. Because the library contains all knowledge, no one can find any of it. It is nothing but an indecipherable pile of information – as James Gleick calls it in his book The Information: A History, A Theory, A Flood, a “perfect case of information glut[2].” There does exist in this library a “perfect compendium,” a book that keeps a true record of every book on every shelf. To have it would be unthinkable power: To know where every book is kept, and thus the access to the knowledge inside them – all knowledge that ever existed and will exist. Numerous individuals have tried to find this book using various methods, but it eludes them all. And without this compendium, the Library of Babel, while a storage system of information, is one that contains all of it but doesn’t enable a seeker to find anything at all.

Wikipedia vs. the infinite library [edit]

Wikipedia seems to imitate the infinite library in many ways. It too is a way to store information, but much more successfully than the Library of Babel. The way it is built has similarities with the structure of the Library of Babel; its mission, however, deviates from the concept ways that makes it usable.

Wikipedia seeks to create a compendium that “contains information on all branches of knowledge,” and does not aim to truly encapsulate all knowledge. In doing this, it creates for itself a goal to organize information in a more reasonable way than the Library by eliminating all the distracting or untruthful information that exists in the entire universe of knowledge. In addition, Wikipedia organizes its knowledge in a way that is navigable – a person can usually find what they are looking for. This is an important aspect of what sets Wikipedia apart from the Library of Babel.

Wikipedia as an encyclopedia [edit]

Wikipedia is, at its most basic level, an encyclopedia, and throughout history, there have been many discussions about how encyclopedias should be organized. In the 18th century, for instance, alphabetization was the preferred method of organization. Then, in the 20th century, hierarchical information storage was favored. However, Vannevar Bush, who wrote an influential essay on information structuring, believed that this was not ideal either: He thought that the information hierarchy of classes and subclasses was too constraining, and instead suggested associative indexing, in which there would be “trails” of information created as a person traversed through the world of knowledge.

Wikipedia is structured much like this. While it does have more ‘traditional’ pages that are lists of subjects where articles are organized in alphabetical order or in a hierarchical manner in which users can access each broad category’s page to find its subcategories, Wikipedia also utilizes links to allow users to access related topics within an article. Specific keywords with enough significance to warrant a Wikipedia page of its own are in blue text instead of black so that readers can click on the link, which brings them to that page.

The linking tool in Wikipedia is useful for readers in many ways: If someone doesn’t have enough background knowledge on a certain topic, they can click on the link to that topic to strengthen their understanding before continuing their perusal of the first article. Another possibility is that if a reader is done with an article but is still interested in related topics, they can look for the keywords with blue text to continue their knowledge-gathering on these adjacent articles.

In this way, Wikipedia reflects Bush’s ideas: Its hyperlinking capabilities allow users to “shift their center” of “investigation” and create their own information trails based on what they are interested in or need to know (in fact, there has been a game created out of this: players all begin on the same Wikipedia page, are given another topic keyword that has a page of its own,
and they race to find the blue links scattered across articles that, when clicked on one after another, will create a path to eventually reach the given topic: The first person to reach the page of the given keyword wins the game. This linking gives Wikipedia a more natural ordering of information, where relevant information can be accessed from its “neighboring” subjects. A person, instead of searching through catalogues or alphabetized books to find the next most relevant topic, can access it by one click of a mouse. George Landlow, an internet analyst, builds on Bush’s idea: He claims that Wikipedia is a cultural revolution, and that the ideas of hierarchy and linearity should no longer be the center of information organization, but rather “multilinearity, nodes, links and networks.” The links in Wikipedia that create these networks of information have revolutionized information systems.

Wikipedia as the Library [edit]

All the paths [edit]

The linking of Wikipedia articles resembles the paths of the infinite library a lot. Gleick compares the pathways Wikipedia’s links create to labyrinths. For example, he points out that Wikipedia has so many articles that use the same name or keyword, that it began using “disambiguation pages.” Under particular pages whose titles have multiple meanings, Wikipedia inserts a sentence at the top of the article that states “For other uses, see [keyword] (disambiguation).” Clicking on this link brings the user to a page that is a compilation of articles that use this keyword so that the user can find the one they are interested in. With so much information compiled on Wikipedia and its generous use of hyperlinks within each article, the paths one can take on the site will diverge in infinitely many places and go on their own separate, and still infinite, new stream of links.

These never-ending paths are quite similar to those in the Library of Babel. People search the library hexagon by hexagon, looking for deep truths about humankind, time, and the universe, jump from one topic to another, one shelf to another, and never reach an end. These infinite pathways are a physical representation of the vast networks of knowledge Wikipedia holds. For the people in the library, though, their endeavors are fruitless and much more daunting than a ‘click-on-the-next-interesting-link’ user of Wikipedia.

The compendium [edit]

A key difference between Wikipedia and the Library of Babel that makes Wikipedia more usable is the accessibility of the “perfect compendium.”

The perfect compendium contains a catalogue of all other books in the library. Such a thing existing literally (a Wikipedia page that lists all 6 million articles and provides links to each, for example) wouldn’t necessarily be useful in the real world, as not many people might have the time or interest in scrolling through 6 million links to 6 million articles. Wikipedia does have catalogues, divided into categories, with each category linking to subcategories, but this probably isn’t the perfect compendium as Borges may have imagined it. In the modern age of technology, though, information is at the fingertips of anyone who comes searching even without this website physically existing, and there is a useful way to catalogue enormous amounts of information: Gleick states that “search—engines, in cyberspace—find needles in haystacks.” The search engine is the catalogue of catalogues.

Two search engines are relevant here: Wikipedia’s internal search engine, and a broader internet search engine. Both function in the same way: They allow a user to pinpoint the ‘location’ of the topic they seek in cyberspace, then give them the ‘directions’ to that place. The difference is that Wikipedia’s search engine returns only Wikipedia articles, while a larger search engine gives the searcher more resources (although Wikipedia’s article on that topic, if it exists, is often at the top or near it). A person need only search a keyword, and a relatively short, helpful, and comprehensible catalogue of relevant topics is handed over to that person. The search engine dilutes the information that reaches a person; it filters out the unnecessary clutter.

This perfect compendium (nearly perfect, as it relies on human-built algorithms and is thus still prone to error, and it only contains information and knowledge that humankind has discovered thus far, unlike the real perfect compendium, which also contains all information that will ever exist), the search engine, effectively allows a person to leap out of the entire Library (although this is counterintuitive because the Library is infinite) to look at it from ‘above.’ While inside Wikipedia articles, clicking on links one after another to find one particular, infinitesimal piece of information is comparable to searching
meaninglessly through endless shelves, but using the search engine, the catalogue of catalogues, is equivalent to teleporting out of the Library from any point, magically finding the coordinates of the book a user wants, then teleporting back in to that spot. Wikipedia, in this way, is infinitely more usable than the Library of Babel.

**Wikipedia as a filter** [edit]

If the only difference between Wikipedia and the Library of Babel were the existence of its (near) perfect compendium, Wikipedia would be the Library except every person had a copy of the perfect compendium (and, of course, the fact that Wikipedia doesn’t contain knowledge not yet discovered), but it still may not be a useful institution. Another of its important features is that it is not saturated with “un-notable” articles and falsehoods. A community of Wikipedians, so-called “deletionists,” try to preserve the sanctity of Wikipedia by watching out for articles that in their view should not exist, then propose to delete them. They prefer having high-quality articles rather than a greater quantity of them (as opposed to “inclusionists,” who prefer to keep a larger number of articles).

First, they make sure that poorly written articles with questionable information or sources – potential falsehoods – are not included. Second, they try to ensure that all topics that have Wikipedia articles have a certain caliber of notability – significance to the world. Subjects that are not as well known or not impactful to society, in deletionists’ eyes, should not have Wikipedia articles.

While their work is rather controversial (there are often fights between inclusionists and deletionists over what should or should not be kept), the way deletionists deviate from the idea of the Library of Babel is still rather important: It sets Wikipedia up as a ‘reasonable’ filter (that is the goal, at least) for ‘important and truthful’ knowledge, and tries its best to keep out what, by their definition, shouldn’t be included. This is also partly what makes Wikipedia a useful source of information rather than a chaotic mess.

**Notability** [edit]

Deletionists want all articles on Wikipedia to be “notable.” This means that the minute details of the world, although they are to be found in the Library of Babel, are nowhere to be seen on Wikipedia.

One may argue that the (near) perfect compendium means that everything can be included in Wikipedia regardless of notability – since a person can just search for it in the compendium, and information that is useless would just naturally remain un-accessed.

Gleick, however, seems to disagree and cites psychological studies that suggest “too much information” can inherently be detrimental. The “information-load paradigm” theorizes that individuals can only absorb and utilize a certain amount of information, and researchers claimed too much information would not only fail to be advantageous, but that it would actually be disadvantageous and lead to impaired judgement.

A similar idea also applies to Wikipedia: Although a person might not be overloaded with information, as they can choose the Wikipedia link-paths they venture down, the institution itself can experience information-load. The (near) perfect compendium becomes a lot less perfect if notability was not observed when constructing Wikipedia’s information systems. For example, Jimmy Wales, a Wikipedia administrator who leans “inclusionist,” suggests that if it weren’t for “maintenance and reliability” concerns, he would be fine with the development of a biography article on Wikipedia for every person on the planet. The maintenance issues can be serious: Oversaturation of useless information could lead to a deterioration of the very usefulness Wikipedia can claim over the Library of Babel. For example, a common first name is “Will,” and a common last name is “Smith,” so there are probably a great many people named “Will Smith” in the world. When a person tries to find an article on “Will Smith,” though, they are most likely searching for the famous actor. If, when they put “Will Smith” in the search engine, many useless articles about barely-known Will Smiths come up, it might be considered detrimental to that person. This is ‘impaired judgement’ of the encyclopedia: If given too much information about the same topic keyword, the results can be skewed. According to the way search engines work, the most popular or notable articles are near the top of the results, whereas articles that attract less traffic will be further down the list. A person (if they already knew what they were searching for), could simply pick out the Will Smith they wanted from the top of the stack, but it would be, as Wales says, a lot more
work to organize these articles, which may not be worth it, especially if a great number of articles about persons named “Will Smith” were barely accessed.

**Falsehoods** [edit]

Out of the reasons why deletionists favor erring on the side of caution when deciding what is worthy of a Wikipedia page, excluding falsehoods is one of the most significant, and most understandable.

Firstly, the inclusion of false information, as the Library of Babel does, is dangerous. Particularly in the last few years in the United States, the rise of misinformation has been a significant concern. The internet, and particularly social media, has been filled with falsehoods that lead to serious consequences, many of which resulted in the loss of human life. Wikipedia, in aiming to eradicate falsehoods from its articles, not only makes itself a more useful resource, but also a safer one.

Additionally, eliminating falsehoods from the library of Wikipedia serves a purpose in maintaining its credibility. In the Library of Babel, people who sought to keep only the books containing truth “condemned entire walls” and destroyed millions of books potentially containing precious knowledge. With Wikipedia’s explicit goal of being a truthful resource and the refusal to systematically shelve falsehoods next to the facts, it no longer reflects the entirety of the real world as a “mirror” like the Library or the internet do, but it saves itself from the attacks of disbelievers. If the deletionists were not a prominent voice in favor of maintaining the truthfulness of the institution, Wikipedia’s shelves could be destroyed as well.

**Languages of Knowledge** [edit]

An unfortunate difference between the Library and Wikipedia is its lack of inclusion of other languages and cultures. The Library contains all books in all languages, and Borges makes specific references to individuals who believed certain lines of a particular book to be written in Yiddish or Portuguese, which actually turned out to be “a Samoyed-Lithuanian dialect of Guarani with inflections from classical Arabic”.

Wikipedia is lacking in this area. Many non-English speaking countries don’t use Wikipedia as much as the English-speaking population of the world does, or perhaps haven’t even come online: They don’t yet have the internet infrastructure to access Wikipedia in the first place. So they are more likely to not receive proper representation.

Deletionists are a part of this problem: They are gatekeepers, blocking insignificant material that inconveniences the institution and its users out. However, they often also filter what many consider to be useful and important topics to certain groups. Heather Ford, writer of “The Missing Wikipedians,” describes a culturally significant meme character named Makmende from Kenya whose article was deleted from Wikipedia because it had “no reliable sources, and no claims of notability” according to administration. The Kenyan Wikipedians were frustrated, because Makmende was a meme understood by Kenyans, and wasn’t necessarily meant for Wikipedia’s central audience and administration that historically includes mostly Western people (English Wikipedia has been and is still is the largest out of any language).

Many may ask why Kenyan Wikipedians didn’t just abandon the idea that English Wikipedia should have an article on him, and instead just write an article in Swahili. Ford answers this: One important reason is that Swahili Wikipedia doesn’t have many active editors, and Wikipedians living in Kenya view the English version as more reciprocal. When they contribute, they are practically guaranteed to receive something in return (they will more likely be able to find a Wikipedia article when they look something up on English Wikipedia). Swahili Wikipedia doesn’t have as large of an active, dedicated, editor population. In these editors’ view, the creation of an article on Makmende in the English-speaking Wikipedia space was more important than creating one for the Swahili version. But he was still removed from that space for not being ‘significant enough’. And thus compared to the English version, some Wikipedias in other languages and about other cultures aren’t receiving as much representation as they would in a truly reflective mirror of the world. Makmende was ultimately accepted and his article is here to stay, but deletionists, although they serve an important purpose for the English version of Wikipedia, should continue to be more aware of the sensitivities surrounding cultural inclusion.

Thus, Wikipedia is not a perfect filter of the ‘real world.’ The human gatekeepers of the shelves of Wikipedia have arguments amongst themselves about what to include and exclude, and they may sometimes reach ‘wrong’ conclusions. If Wikipedia is to be ever-advancing towards its goal of becoming a compendium of all knowledge (of the world), then it must also work
towards the inclusivity of articles written in all languages as the Library of Babel does, and articles written about the landmarks significant to other cultures. It is currently dominated by articles written in English and about the Western world. There is much to look forward to, however: The fastest growing parts of this online encyclopedia are in the other languages of the world. The work is never done, but the last few years have launched Wikipedia in the correct direction.

**Conclusion: Wikipedia’s usefulness** [edit]

Wikipedia’s function in the real world is thus far more practical than the Library of Babel. The two primary factors that make this so: First, Wikipedia can utilize modern technology to create a (near) perfect compendium of its collection of knowledge. The search engine allows a person to flit across the entire institution of Wikipedia without having to spend eternity stumbling through the endless shelves. Second, Wikipedia requires the filtering of falsehoods and the test of notability. Articles strive for accuracy and significance. This is what makes Wikipedia manageable; the danger of coming across a book of falsehoods or insignificant clutter, something that could block the truth from reach, is far less prominent. Wikipedia is thus a useful, diluted, navigable version of the infamous infinite Library of Babel, and is a revolutionary phenomenon of information structuring.

**References** [edit]

Talk: Wikipedia & the Structure of Knowledge

From Fake Wikipedia, not a real encyclopedia (but still free)

This is not a real Wikipedia article, but rather an exploration of Wikipedia’s formatting for an essay written by Audrey Yang for the class “We are not in a post-fact world”: Wikipedia & the Construction of Knowledge.

### Conception of the idea [edit]

@ry13702: Note: “ry13702” is not my NYU net ID, nor is it my Wikipedia username. ry13702 (talk) 00:04PM May 3, 2021 (CST)

@ry13702: Wikipedia’s feature for linking articles has been one of my favorite features for quite a long time. Throughout this course, I found myself actually thinking about this feature a lot, and how it fits in with how existing knowledge is organized in the world and especially in Wikipedia, and how new knowledge is integrated into what we can dramatically choose to call metaphorical “libraries.” ry13702 (talk) 05:15PM May 3, 2021 (CST)

@ry13702: Most recently, in our last readings on the Library of Babel, the comparison between this fictional, mythical library and Wikipedia fascinated me. This essay is essentially an exploration on “linking” and how Wikipedia, which is a vast collection of knowledge that utilizes this feature, compares to the *truly* infinite but also more chaotic Library of Babel. ry13702 (talk) 05:17PM May 3, 2021 (CST)

@ry13702: Actually, I might change the topic to “how Wikipedia compares to the library of babel” broadly, with a focus on linking. I think there’s some content that can be interesting to talk about related to Wikipedia & the library of babel, but isn’t about linking. ry13702 (talk) 12:58PM May 4, 2021 (CST)

@ry13702: Perhaps what I want to explore is in what ways Wikipedia can be considered an accessible, diluted version of the Library of Babel, that is not as perfect, but makes up for it by being useful. Perhaps this is the answer to my question of “how are they similar/different?” That might be a worthwhile path to go down. ry13702 (talk) 13:30PM May 4, 2021 (CST)

@ry13702: I think I have settled on this ^ concept. Through the process of writing & rereading the class texts, I’ve come to the conclusion that Wikipedia is similar in function to the library of babel, but its purpose is quite different. This is probably what makes it useful rather than a complete mess. ry13702 (talk) 17:42PM May 6, 2021 (CST)

@ ry13702: Actually, to amend this previous statement, perhaps the way modern technology works is instrumental in Wikipedia’s functionality, too. Without search engines I think we’d still be quite lost.
Why the Wikipedia article formatting [edit]

@ry13702: I found myself really enjoying the last essay I wrote in this format, so I decided to use it again for my final essay. I think there are a lot of reasons for this:

  @ry13702: It looks pretty and is more interesting to read than a traditional essay, in my opinion. ry13702 (talk) 05:20PM May 3, 2021 (CST)
  @ry13702: It's organized into distinct sections with subheadings, so it is clear what each paragraph/group of paragraphs is meant to say. This helps me (as a person with a pretty short attention span) stay engaged instead of staring at a mass of text organized in long paragraphs, which seems more daunting. ry13702 (talk) 05:20PM May 3, 2021 (CST)
  @ry13702: The “linking” idea also applies to my mock Wikipedia page: I can connect the ideas that I find to be relevant to each other by putting internal hyperlinks within the document. Writing my essay this way is also constructive to keeping track of my own ideas because my thought process is set up kind of like this: where ideas jump in one after another in my brain and I rush to write them down before the next one pops in my head and I forget it. ry13702 (talk) 13:04PM May 5, 2021 (CST)
  @ry13702: I don't know how I'm going to do this in Google Docs, however, as copy-pasting ruins the formatting. Perhaps I should hyperlink them once I import into Google Docs? ry13702 (talk) 13:40PM May 3, 2021 (CST)

@ry13702: The talk page that the professor suggested seemed like a really fun idea to explore. Documenting ideas as they come to me will be beneficial as edit because I'll be able to look back and see my previous thought process to build upon them. ry13702 (talk) 05:23PM May 3, 2021 (CST)

@ry13702: Something I've realized in my time at university so far is that writing is a process of discovery. While in high school, I would mechanically develop a thesis and then find the necessary evidence to support it (which is exactly what one reading from a few weeks ago warned us not to do), but both Writing the Essay and this class have allowed me the room to discover deeper truths on a topic as I’m writing. On Wikipedia, the talk page is usually for a dialogue between writers and editors of a page to develop the article, but instead of collaborating with other people, as Wikipedians do, this essay is a collaboration with myself (past, present, and future selves) – so the talk page consists of the different ideas I have had at different times, amendments I’ve wanted to make on my own essay, and discoveries I’ve made and realizations I’ve come to about the topic that came through the process of writing. This process has been, to me, as significant and interesting as a conversation with another writer or editor would be. ry13702 (talk) 21:35PM May 19, 2021 (CST)

All the types of “linking” [edit]

@ry13702: I've also always liked the idea of knowledge & concepts connecting to each other. There are a couple things I wanted to “link:

1. As I stated previously, I like the linking format of Wikipedia
2. I like essay topics that combine several sources and find common ground between them (here, I chose to connect the O'Sullivan text from the very beginning of the course, Fister & Ford, who I also used in my last two essays, and Borges & Gleick, who are the last two texts). ry13702 (talk) 20:52PM May 9, 2021 (CST)

@ry13702: The linking in this article is in the same format as Wikipedia: For example, any instance of “Library of Babel” in blue hyperlinked text links to the Wikipedia page on the subject. ry13702 (talk) 14:27PM May 21, 2021 (CST)
I've decided to state my ideas throughout the “essay” as facts, for two reasons:

1. A typical essay (for example the ones that we wrote in Writing the Essay) does this
2. Wikipedia’s format doesn’t really allow for “opinions”
   a. So although the essay consists of my own opinions and the evidence I gathered from our various readings, they’ll be stated as fact. ry13702 (talk) 18:20PM May 6, 2021 (CST)
   b. @ry13702: Update: the professor & I decided to pose questions in the lead rather than propose a thesis statement as an essay might or use completely third person with no questions as Wikipedia might, to better express the points I am trying to make. ry13702 (talk) 18:20PM May 6, 2021 (CST)

I’m lacking in some sources for the compendium section, although I don’t think that resources might actually be necessary. I’m stating commonly known facts about search engines, then making an analogy to the compendium as Borges describes it, so I’m not writing anything that would need to be fact-checked, I think. ry13702 (talk) 19:12PM May 6, 2021 (CST)

This idea of Wikipedia versus the library of Babel really is a rather beautiful metaphor for this entire class: in the whole world of information, Wikipedia creates an imperfect but nonetheless wonderfully useful compilation out of the chaos. The library of Babel is all the information and misinformation that exists in the world, and people, whilst traversing it, can be led in all sorts of wrong directions. Some, as Borges says, go insane trying to find the truth, or condemn entire shelves of books based on falsehoods of one. In our reality, these people also believe these falsehoods sometimes, which leads to dangerous consequences (a nice parallel to my first essay of this class, I think!). Wikipedia, however, gives some semblance of order, and keeps us grounded with a very well planned-out system of organizing information, and reminds us that all is not lost. ry13702 (talk) 15:23PM May 9, 2021 (CST)