

WikiPaths: Re-Imagining Information Space Through Semantic Play

Alejandro Abreu

Interactive Telecommunications Program

Tisch School of Arts

New York University

721 Broadway, 4th Floor

New York, NY 10003

alexabreu@gmail.com

ABSTRACT

WikiPaths is a hyperlink scavenger hunt game played on Wikipedia.org in which players try to find the shortest path between two, seemingly unconnected, encyclopedic entries. WikiPaths is not a separate website, but rather a casual game superimposed on Wikipedia itself. The game is written in JavaScript and runs as a Greasemonkey Firefox Add-on, allowing users to start a new game at their leisure as they also browse the encyclopedia. WikiPaths fuses normal browsing behavior with casual gaming, and explores how re-contextualizing an existing site can yield vital information about the connectivity of information and how people navigate information space.

KEYWORDS

Wikipedia, Greasemonkey, JavaScript, casual games, semantic analysis, virtual location, contextualization

INTRODUCTION

It is estimated that several hundred billion websites already exist on the Internet. With more website destinations than seconds in a lifetime, the web has reached a level of saturation requiring tools which highlight and make use of the overlapping relationships between sites. Internet search (i.e. Google) has made it possible to find the perfect result among the billions of candidates.

However, Internet search has also stripped the user of a sense for the relationships between related websites not hyperlinked or jointly accessed. To combat this loss of context, new tools have emerged which transform the web browser from a passive tool to an active one, capable of revealing extra information about pages not originally designed for by the creator of the site.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission.

Greasemonkey is one such tool which grants programmers the ability to write scripts that can quickly and dynamically modify how a user interacts with any given website. Many scripts have been written which run on the Greasemonkey platform providing functionality from language translation to e-mail organization.

One of the more powerful tools built on Greasemonkey is ShiftSpace, a collaborative editing layer available on top of any website for annotating pages while one browses. The power behind ShiftSpace is that it empowers users to add an information layer on top of a website which may be as simple as a pithy comment or as pertinent as a pointer to similarly themed websites. ShiftSpace allows users to add context to a singular browsing experience by incorporating supporting media layers which the original author may not have been aware. The original base content is made stronger though the addition of relevant information which users can then also navigate through.

Similarly, WikiPaths re-contextualizes how users may interact with Wikipedia – challenging them to focus their virtual *dérive* (defined by Situationist theory as passive movement through space) to reach a pre-determined end article in the fewest number of links. Wikipedia users often find themselves searching for a particular term, and then, after a few minutes, realizing that their curiosity has led them to an entirely unrelated topic. WikiPaths turns this behavior into a structured casual game which can be played all while the user simultaneously browses.

It is estimated that 200 million consumers play casual games online¹. The overwhelming majority of these players are not considered traditional ‘gamers’, and turn to casual games for their simplicity and the ability to achieve results quickly. WikiPaths builds upon the success of this style of game, further emphasizing the ‘pick-up-and-go’ nature of casual gaming by allowing players to start a game while they browse Wikipedia.

Yet, what appears to the player as a game is also designed to yield meaningful information about how Wikipedia users navigate the site and make connections between disparate pieces of information. When a user plays WikiPaths she

leaves behind a trail of nodes that describe her attempt to bridge the game's starting point with the destination. For instance, if a user were presented with Wikipedia entries for the Teenage Mutant Ninja Turtles and the Ottoman Empire, a clear route would most likely not be initially apparent to this user. Only through a series of logical choices about how the two topics connect to one another would the user be able to reach his goal.

As users play WikiPaths, the data representing each user's path is stored remotely. Once stored, it can be analyzed and compared to other players' interpretations of the same path; providing insight into both the organization and upkeep of Wikipedia as well as an individual's thought processes while traversing Wikipedia without search.

BACKGROUND

In 1941, Jorge Luis Borges published a short story entitled *The Library of Babel* in which he describes

an unbelievably large library containing all possible books. Within the "total" and "endless" reaches of the Library,"[t]here [is] no personal problem, no world problem, whose eloquent solution [does] not exist—somewhere ..." but also "[f]or every rational line or forthright statement there are leagues of senseless cacophony, verbal nonsense, and incoherency." As Borges describes it, the Library is the greatest imaginable source of information: it contains "The Vindications—books of apologiae and prophecies that would vindicate for all time the actions of every person in the universe and that held wondrous arcana for men's futures."^[2]

Although Borges' library contains all of humanity's works and knowledge, the size of the library itself makes finding even a remotely intelligible book nearly impossible. As James Grimmelmann further elaborates in his essay on Borges' work, "[t]he problem is access, not creation." The innumerable librarians of Borges' library watch over their respective hexagons containing mostly nonsensical works. It is believed, that one librarian, the Book-Man, at one point deciphered the map connecting all hexagons within the library, but both he and the "crimson hexagon" containing powerful magical books are always out of reach.

Here we come to the crux of the matter. There is no difficulty in ensuring that the Library contains a (near) copy of any book. But there is a large gap between the Library's containing a book and our being able to make use of it.^[2]

If the Library of Babel is projected onto the Internet, similarities begin to surface. Billions of websites already exist with thousands being created everyday. Although Google currently serves as the go-to index which yields the answers for our queries, such a tool is as opaque as the magical books in the "crimson hexagon" and as untouchable as the Book-Man in Borges' story.

Through WikiPaths I hope to continue the discussion of a method for mapping and organizing the Internet. Rather than depending on a closed system to generate results to queries, the system should be re-imagined to make transparent the underlying path from search term to destination. Humans have an ability to quickly adapt their cognitive map of information to find the associations

between data points which computers cannot. Although the Semantic Web has not yet materialized, I believe the motivations behind its design must be reevaluated. Tim Berners-Lee described his vision for the Semantic Web in 1999 as follows:

I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The 'intelligent agents' people have touted for ages will finally materialize.^[3]

Berners-Lee's vision echoes Nicholas Negroponte's earlier call for digital butlers capable of anticipating a user's every demand.

The best metaphor I can conceive of for a human-computer interface is that of a well-trained English butler. The "agent" answers the phone, recognizes the callers, disturbs you when appropriate, and may even tell a white lie on your behalf. The same agent is well trained in timing, versed in finding the opportune moments, and respectful of idiosyncrasies.^[4]

Theoretically, such a world sounds ideal; but only if humans are able to maintain a personal understanding and connection to the data delivered to them by the intelligent agents. For without this understanding, we move closer to the dilemma Borges foresaw so many years ago. What will happen when there are more butlers than knowledgeable users? Is such a system sustainable?

JUSTIFICATION

Redefining the way information should be indexed, linked, and retrieved is no small task; and the project here described does not attempt to do so. Rather, I hope to further the discussion through an idealized casual game played in a small, information rich, subsection of the Internet, Wikipedia.

Wikipedia, the volunteer produced and edited encyclopedia is the most popular general reference on the Internet⁵. According to its own entry on Wikipedia it currently contains 2.8 million English entries, and is edited several thousand times a day. It is often accused of being an inaccurate source by its detractors, yet a scientific comparison by Nature Magazine found Wikipedia's entries only marginally less accurate than those offered by Encyclopedia Britannica. More importantly, it is constantly evolving, mainly due to the volunteers who take on the challenge of monitoring and editing its millions of entries to ensure a neutral point of view. In his March 2009 New York Times article, *Wikipedia: Exploring Fact City*, Noam Cohen compared Wikipedia to a growing city.

But these concerns seem misplaced — Wikipedia can no more be completed than can New York City, which O. Henry predicted would be "a great place if they ever finish it." In fact, with its millions of visitors and hundreds of thousands of volunteers, its ever-expanding total of articles and languages spoken, Wikipedia may be the closest thing to a metropolis yet seen online.^[6]

Strikingly, Wikipedia, is a thriving metropolis – the kind of city urban reformer and writer, Jane Jacobs, would be happy to explore. The success of Wikipedia stems from its transparent volunteer-coordinated editing process. The citizens of Wikipedia, self-deemed Wikipedians, and the millions of visitors who traffic its articles are the street watchers Jacobs understood to be so vital to the safety and success of any city.

If we are to maintain a city society that can diagnose and keep abreast of deeper social problems, the starting point must be, in any case, to strengthen whatever workable forces for maintaining safety and civilization that do exist -- in the cities we do have. ... The first thing to understand is that the public peace -- the sidewalk and street peace -- of cities is not kept primarily by the police, necessary as police are. It is kept primarily by an intricate, almost unconscious, network of voluntary controls and standards among the people themselves, and enforced by the people themselves. No amount of policing can enforce civilization where the normal, casual enforcement of it has broken down.^[7]

It is this simple observation that humans are the best curators of their own space that Berners-Lee and the Semantic Web drivers must incorporate into their vision. Depending on an external and proprietary search-based tool to police or filter the Internet appears to be in contradiction to Jacobs' premise. A semantic web could result in a much more streamlined and efficient categorization of information. However, semantic indexing, undertaken by machines, results in a process further removed from human intuition and control.

The citizens and users of Wikipedia have proven to do an excellent job of creating and maintaining a tool used worldwide. Yet, if we have learned anything about the web, it is that it's never done evolving, and as it evolves, it grows stronger.

Evolvable systems -- those that proceed not under the sole direction of one centralized design authority but by being adapted and extended in a thousand small ways in a thousand places at once -- have three main characteristics that are germane to their eventual victories over strong, centrally designed protocols.^[8]

As described previously, through the use of tools like Greasemonkey, new paths for evolution have emerged. Rather than building anew, developers now have the option of building directly on top of existing sites and sources of information. Through layering and re-contextualization, developers can provide new experiences for users and also highlight information previously ignored or unseen. In the same way the Situationists, Guy Debord and Asger Jorn, asked people to imagine cities without previously constructed borders and boundaries, developers can also ask users to extract new meaning from the already established structures of the saturated Internet. It is through new analysis of the real, the everyday that new ideas and possibilities emerge. However, this potential might be lost

or severely weakened if machines alone inherit the full responsibility of maintaining, querying, and processing our information space.

Now that the drift offered a new way of surveying urban space, a new means of representing space on paper would have to be found. The 1956 Guide psychogéographique de Paris and 1957 Naked City served as alternative maps of Paris, Debord and Jorn representing the surreal disorientation of their drifts around Paris by scattering the pieces of map and the arrows showing their routes. ... In their maps, by stark contrast, Debord and Jorn attempted to put the spectator at ease with a city of apparent disorder, exposing the strange logic that lay beneath its surface.^[9]

Asking a user to think about Wikipedia differently is most likely asking too much; unless, the question is presented as a game. Herman Hesse's last work, *The Glass Bead Game*, was started in 1931 and published in 1943. In this book, Hesse described a futuristic society run by intellectuals who focus exclusively on education and on playing a game of the same name as his novel. Although never fully explained by Hesse, players of the *Glass Bead Game* are challenged to connect human knowledge into a woven symphony of ideas. The most skilled player of this game is deemed the Magister Ludi for his abilities to symbolically express the connections in humanity's collective learning and experience.

WikiPaths borrows this concept from the *Glass Bead Game* and challenges players to connect the nodes within the user-generated version of human knowledge. By encouraging users to connect articles in Wikipedia, WikiPaths will generate a rich dataset depicting users' thought processes and logical choices about how two discrete articles should be connected, transforming the analysis of Wikipedia's structure and organization into a crowd-sourced activity. This dataset can then be opened up to Wikipedia and other developers to create transparent tools which better match how humans think.

THE GAME

WikiPaths is a graphical user interface superimposed on that of Wikipedia. Logged in users are given the option of playing one of three different game types:

1. **Time Trial**
Players race through Wikipedia as fast as they can to reach the destination article.
2. **Distance**
Players attempt to make their way to the destination article using as few links as possible.
3. **Biathlon**
This mode combines the Time Trial and Distance modes for a unique and difficult challenge.

These different game types determine how users will be scored as they develop a path between two discrete Wikipedia articles. After selecting a game type, the game begins and users are presented with a random start and end article. They then begin their journey to reach the end article by clicking only outgoing Wikipedia links from the random starting point.

This is why the information gain is high at the start, as players get away to the same hubs, but decreases in the middle of the game; then the gain increases again, as they home in using the same common sense. In other words, the initial getting-away and the final homing-in are much more predictable after seeing game data than before, and the idea is to use the information gain to guess where the homing-in phase, and thus the relevant part of a single game path, starts. ^[10]



Figure 1. WikiPaths Interface

WikiPaths is based on the Wiki Game, a folk game which originated in Wikipedia's forums. As described on Wikipedia, Players (one or more) of the Wiki Game start on the same randomly selected article and must navigate to another pre-selected article (home) only clicking links within the current article. The goal is to arrive at the home article in the fewest number of clicks (articles) or the least amount of time. The single-player Wiki Game, known as Wikirace, WikiLadders or WikiClick, involves reducing one's previous time or number of clicks. Computers are fully capable of finding the shortest path between any two articles, but they are unable to infer relevance along the way. Humans, on the other hand, constantly adapt their knowledge base to draw a logical path from point A to point B.

In addition to WikiPaths, two formalized interpretations of the forum game exist, but these versions require users to navigate away from Wikipedia to separate websites in order to play. However, this approach severely limits the number of users who will play and contribute knowledge.

One of these games, Wikispeedia, also captures game data with the goal of translating human common sense into a hybrid semantic system for indexing information. Through experimentation, the Wikispeedia team in Montreal, Canada has found that their analysis of game data out-performs the classification results obtained by Latent Semantic Analysis, the most commonly used system for drawing semantic conclusions about data not readily understood by computers (i.e. human language). Their conclusions affirm that humans are still much better than machines at inferring relationships between distinct concepts.

The Wikispeedia researchers explain that as users play their game, they first attempt to reach a 'hub' or article which will contain many outgoing links to relevant information. Their findings show that these initial zooming out and subsequent homing-in processes yield pertinent information about the relationship between the two Wikipedia articles in question.

THE APPROACH

Both WikiPaths and Wikispeedia present the user with similar games with goals of capturing similar data. However, WikiPaths is less driven to transform said data into a new external computational semantic analysis tool, and more focused on presenting the user with a real-time map for use while navigating information space. As stated earlier, the notion that computers should attempt to take responsibility for the common sense process that underlies organizing, maintaining, and searching for information is incompatible with my goals. Rather, I believe users should take ownership of this practice, and that the semantic analysis of data like that generated by WikiPaths should be used to create systems that integrate human intuition into information navigation.

Throughout user-testing of WikiPaths, I have found that during the initial zooming out phase, described previously, players gravitate to geographical hubs. Players naturally categorize information by physical place while they decipher an approach to reach a given destination article. The notion of place being tied to information is readily found in the structure of Wikipedia. Researcher, Steven Dolan of Trinity College in Dublin, found that if lists and dates are ignored, the center of Wikipedia (point with the

shortest distance to all other articles) turns out to be the United Kingdom. This means that the Wikipedians have structured their encyclopedia to be traversed most easily in terms of place. To my mind, this is no coincidence. Place makes sense to humans, and technological innovation in web and mobile media are converging with user demand for tools which express data in terms of location.

Strikingly, there is no equivalent for information space. The semantic relationships connecting one piece of information to another are hidden from users and incorporated into products like Google's AdSense and Amazon's product recommendations. Yet, the need to know how to physically get from New York to Los Angeles is just as relevant when surfing between their virtual representations. Rather than magically jumping from site to site, users should be made aware of their semantic surroundings, adding context to their informational searches. Such a tool could provide users with a visual map of where they are in relation to corollary information allowing them to make inferences about how to best reach their goals.

NEXT STEPS & CONCLUSION

Newly launched, WikiPaths is beginning to generate data describing users' paths through Wikipedia. Not enough data has been generated from the game to draw conclusions on how to best implement a real-time visual semantic map for users browsing Wikipedia. However, all data generated by the game will be made available to the public in order to generate a discourse on how such a visualization should be designed.

Crowd-sourcing relies heavily on the size of the crowd, and WikiPaths will need to generate a considerable user base to generate meaningful results. To aid this growth, game features will be added to WikiPaths. Some of this functionality will include:

1. User-generated path functionality
2. Leader boards
3. An achievement system
4. Two player versus mode

In my research I have discovered small groups of researchers interested in how implementations of casual games like the Wiki Game can lead to a better understanding of an underlying information structure. The WikiSpeedia team, in particular, has shown that the data from their game yields better results than Latent Semantic Analysis, the most commonly accepted and validated tool for computing semantic relationships. However, making full use of this potential necessitates a fluid game interface, like that of WikiPaths, integrated into the browsing

experience. Future steps will include reaching out to these small teams as well as open source partners like ShiftSpace in order to fully realize a game capable of delivering a truer semantic understanding of Wikipedia and the rest of our information space.

It is my belief that casual games like WikiPaths will emerge as evermore useful resources -- strengthening our ability to access information in an increasingly saturated Internet environment. However, fully embracing the notions of crowd-sourcing and casual gaming will require new layered interfaces capable of transforming the existing browsing experience into a structured environment for data-rich play.

As the Internet has continued to grow, internet search has proven itself necessary. Yet the search, in its current form, has weakened the relationship between people and the information they seek. WikiPaths offers a new approach, one which re-engages human intuition in the navigation of information space. Fully realized, WikiPaths has the potential to provide users with a human centric map of semantic relationships, thus aiding in the navigation of and the discovery within Wikipedia and the rest of the Web.

REFERENCES

- ¹ Casual Games Market Report 2007. *Casual Games Association*, October 2007.
- ² Grimmelmann, J., Lewis T. Information Policy for the Library of Babel. *Maryland Journal of Business and Technology Law*, November 2007.
- ³ Berners-Lee, T. Fischetti, M. Weaving the Web. *Harper*, San Francisco, 1999.
- ⁴ Negroponte N., Asher M. Being Digital. *Random House Inc.*, New York, 1995.
- ⁵ Tancer, B. Look Who's Using Wikipedia. *Time*, May 2007.
- ⁶ Cohen, N. Wikipedia: Exploring Fact City. *New York Times*, March 2009.
- ⁷ Jacobs, J. The Death and Life of Great American Cities. *Random House Inc.*, New York, 1961.
- ⁸ Shirky, C. And Nothing to Watch: Bad Protocols, Good Users: In Praise of Evolvable Systems. *netWorker*, June.1998.
- ⁹ Sadler, S. The Situationist City. *MIT Press*, Cambridge, Mass, 1999.
- ¹⁰ West R., Pineau J., and Precup D. Wikispeedia: An Online Game for Inferring Semantic Distances between Concepts. *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)*, Pasadena, California, 2009.