Beyond the Desktop: A new look at the Pad metaphor for Information Organization

By Isaac Fehr

Abstract

Digital User interface design is currently dominated by the windows metaphor. However, alternatives for this metaphor, as the core of large user interfaces have been proposed in the history of Human-computer interaction and thoroughly explored. One of these is the Pad metaphor, which has spawned many examples such as Pad++. While the the Pad metaphor, implemented as zoomable user interfaces, has shown some serious drawbacks as the basis for an operating system, and limited success outside of image-based environments, literature has pointed to an opportunity for innovation in other domains. In this study, we apply the the design and interactions of a ZUI to Wikipedia, a platform consisting mostly of lengthy, linear, hypertext-based documents. We utilize a human centered design approach, and create an alternative, ZUI-based interface for Wikipedia, and observe the use by real users using mixed methods. These methods include qualitative user research, as well as a novel paradigm used to measure a user's comprehension of the structure of a document. We validate some assumptions about the strengths of ZUIs in a new domain, and look forward to future research questions and methods.

Introduction and Background

Windows-based user interfaces have dominated the market of multipurpose, screen-based computers since the introduction of the first windowed system in the Stanford oN-Line System (NLS)[3]. From Desktop computers to smartphones, most popular operating systems are based upon at least the window and icon aspects of the WIMP (Window, Icon, Menu, Pointer) paradigm. In these systems, information can be represented in two forms: that of an icon, and the "full view" of the information which is interpreted by, and shown in a window. This is seen most easily in interfaces that implement the paper paradigm, and allow the user to view and edit "documents." In a system based upon the window metaphor that implements the paper paradigm, as most do, every document has essentially two levels which it can be viewed at: an abstracted icon, often in the form of a generic thumbnail matching the document type, and static view of the document contained within a window.

These standards for interface design set the stage for decades of user interface innovations and improvements that helped to realize Douglas Engelbart's augmentation of human intelligence [4]. Users could now externalize much of the memory and many of the cognitive tasks that they were burdened with before. Among the many reasons for this success is are first, the desktop metaphor and subsequent paper paradigm acted successfully as user interface metaphors that helped people from many backgrounds and levels of experience with technology to adopt computers into their workflow and personal lives. Secondly, windows (specifically WIMP)-based interfaces allow the user to navigate through the the large and complex hierarchical file system in an abstracted, yet systematic way that allows for quick loading. Once the user has located the file they wish to view, they may open it, and position it within the desktop, relative to other windowed documents however they find most useful, just like on physical desktops. These features are relatively successful in terms of data organization and function, even on relatively weak machines.

However, with the increasing complexity of operating systems, and the number and types of tasks that are expected of them, we have recognized some of the weaknesses of the metaphors on which our technologies are built. First of all, is that the user of a windows based system is burdened with the task of organizing the system at both the file system level and the window level so that they can effectively retrieve the information they need.

However, the high level of abstraction of a file system can make it difficult to navigate. Because the user must often navigate a complex hierarchy, having to make a decision about what folder to click based only on the name, they have little information scent or other strategies to aid recognition, and must instead rely on recall of their file structure. This structure quickly becomes difficult and cumbersome to navigate through unless the user creates perfectly descriptive names for all folders and files, such that there would only be one logical place for any file to be in the system. We know, from the difficulties that other fields that deal with taxonomies, classification systems, and the like face, that this is an unrealistic expectation to put on the average computer user. It assumed both that the user has classification and organization skills, and that they will put the time and effort into applying them to the system. Predictably, this is not the case, and users often either spend large amounts of time looking for files, or relying on other tactics such as an inefficient search of the entire computer.

Once a user has located the icon for a file or program they wish to use, they must open it in a window. If the user only needs the one window open, organization is not a problem, since they can enlarge the window to fill the screen, and cover all non-relevant information. If the user must view two or three documents/programs at once, they can sometimes resize windows or rely on a tiling functionality to fit multiple documents on the screen at once. However, here we start seeing limitations of the paper paradigm and the windows metaphor:

1. because most documents are inspired by the paper paradigm, their windowed view shows a representation of a document with a certain fixed width and height in a static view. If the window is given less space, the contained document must either be geometrically resized, which can render its contents too small for the

user to utilize, or portions of the document will be out of view, which will require much 2-dimensional scrolling to reference.

2. For programs to be scalable, they often must have their own way of organizing their features and/or multiple documents within them. The premier example of where this breaks down is web browser tabs. The tabs are small representations of multiple documents that often have smaller labels and less space for text than desktop icons, reducing their distinguishability, and making it more difficult for user navigation.

3. Because of the need for application-specific functions and organization, these within-window user interface elements take space away from the document being viewed in the window, reducing its ability to be usefully resized to view it alongside other windows.

Another problem arises if the user is working with more windows than can fit within the screen space.   At this point they overlap, which seems like an elegant way to hide non-relevant information until it is needed; however, the fact that windows seem to lack any depth, and can be hidden behind an effectively limitless number of other, depthless windows, or even offscreen, can cause windows to be lost, and lead to fumbling on the user's part. This problem often requires additional user interface elements like a taskbar, or the ability to minimize windows, which suffers from the same organizational issues as tabs do within windows.

These issues with windows-based user interfaces have remained fairly consistent in the decades since they were first introduced. The multiple abstracted layers required to organize, navigate, and work within these interfaces often leads to inefficient use and a frustrating, cumbersome experience.   In part, these issues with organization and navigation arise because these systems do not utilize one of the most powerful aspects of the desktop and the paper that inspired them: a sense of space.

In the case of a physical desktop, the way that space is organized is contingent on the form of the items it contains, and the constraints derived from the amount of space on the desk. Often, at least in Western cultures, the 8 ½ " x 11" sheet of paper becomes the fundamental unit of organization on the desktop. Any documents being worked on or referenced are positioned close to the user so that they may comfortably read and write on those documents. These documents often lay flat on the desk for easy use. But off to the sides and further back on the desk, papers, books, tools, are also present. But these are organized differently. Items that require work are often piled up outside of the working scope, but still within arms reach for easy access when the current task is completed. An important aspect of these piles is that their form carries meaning in addition to the meaning derived from their position on the desk. The height of the stack may indicate the amount of work to be done. Papers or packets may be organized in different ways such as alternating orientation, or using tabs that stick out of the stack to indicate different things about the stack.

All the previously mentioned tactics for organization (x/y position, z position (stacking), orientation, tagging, grouping, etc), are enabled by the forms of the documents themselves, and the exploitation of this form is required for effective organization and use of a desktop. So it is ironic that interfaces inspired by the Desktop metaphor, and the subsequent paper and window metaphors do not utilize these rich aspects of documents in their digital representations. These areas which are lacking in the windows-based systems have been noted for a long time, and have been attempted to be addressed by another metaphor:

The Pad Metaphor

The Pad metaphor is an alternative to the Desktop and Windows metaphors, proposed to have strengths in some of the areas where they are weak. The Pad metaphor is fairly simple: imagine a whiteboard for which all points are within arms reach of a user, and the user is capable of comfortably writing to and retrieving from any micron-width portion of the board. [7] This theoretical interface would allow the user to utilize information at near-infinite levels of hierarchy, defined by the position and size they write at. If a user wants to add more details to an item they previously wrote, they just have to write it smaller underneath the previous item, or wherever they decide to put it. Information retrieval would also be greatly augmented by the sense of place inherent in the structure of the interface, the fact that all items are represented as themselves, thus the interface utilizes form and place in a way that current implementations of the Desktop metaphor have failed to.

However, because a user with the perfect perception and motor control required to use this whiteboard has not been discovered, implementations of the Pad metaphor require mechanisms to bridge the gap between the metaphor and user's capabilities. Typically, these solutions have come in the form of interactions that allow users to navigate a viewport through the space of the pad, and view the information at whatever level is appropriate for the required navigation, information retrieval, and information manipulation. In other words, these are zoomable user interfaces.

Where as desktop-based interfaces are defined by clicking to navigate between two levels of views (icon and window) and scrolling within windows, the zoomable user interfaces are defined by having zooming as the primary method of navigation. In large, unconstrained interfaces, this often necessitates the ability to pan in two dimensions as well.

One of the early large-scale implementations of this type of zoomable user interfaces was Pad++, a graphical, zoomable environment which supported zoomable versions of applications such as a web browser, drawing application, presentation software, and more. It also featured novel user interface features such as lenses and portals which augmented the user's ability to navigate a space and view the same information differently. While this system showed improvements over other interfaces in terms of users' levels of engagement, it never saw commercial success like many other zoomable user interfaces [1].

Benjamin B. Bederson, in his paper, “The Promise of Zoomable User Interfaces” reviews the history of ZUIs, including where they have succeeded, and where many have fallen short. In addition to implementational design suggestions such as maintaining aspect ratio and supporting scannability, he also addresses what we’ve learned about the use cases for ZUIs. Due to the large amount of navigation required for a ZUI, and problems that arise from unconstrained navigation such as getting lost in desert fog, Bederson suggests picking a smaller domain than entire user interface which can specifically use the benefits of ZUI. He suggests this appropriate use-case is defined by factors such as requiring a big-picture view as opposed to finding answers, being fundamentally spatial and visual (image-based as opposed to text or sound-based), and being composed of items which have a logical, smaller, recognizable, “zoomed-out” representation. [1]

An example of a ZUI which fits these suggestions and has seen great success is Google Maps. Maps, being fundamentally spatial due to their analogy to real space, make sense to navigate spatially in terms of panning and zooming. The interface takes little time to learn due to it being a fairly literal implementation of an interface for a physical map. However, if it were to follow the metaphor of a map too literally, it would run into the same issues of having too much information packed onto a tiny space, making readability and scannability into big problems. To address this issue, Google Maps implements a specific type of zooming to help focus attention, give the user a sense of where they should look for specific information without showing the information itself (information scent) and increasing scannability.

Semantic Zooming

Semantic zooming is a type of interaction in which a user interface element shows different information, and different representations of that information at different levels of zoom [2]. First shown in Pad, a precursor to Pad++, semantic zooming was imagined in terms of revealing lower-level text elements as a user zooms in. For example, when zooming on a document, it shows only a title, then maybe a summary then the whole text. For a calendar, it would show months, then day numbers, then more details including the full date and events.

Semantic zooming was seen as a way to reduce information overload, and increase element’s “information scent” and thus scannability [5]. In a effective implementation, a zoomed-out view of an object would contain a high-level representation of its content, possibly in the form of images, keywords, titles, previous actions, etc that is just enough to assist the user in quickly deciding whether or not they should zoom more into that element. While the information changes, it is also very important to maintain the sense of place with consistent proportions and relative position to other elements to use the benefits of the user’s cognitive place faculties. Semantic zooming is best thought of as an augmentation to geometric zooming, as semantic zooming often will require some form of geometric zooming to communicate the underlying metaphor to the user.

As was previously mentioned, Google Maps is one of the few examples of a widely-successful implementation of semantic zooming. As the user zooms in from a view that contains multiple states, to a view that contains one city, more and more roads are revealed, and the terrain changes from a plain green to a more detailed representation involving landmarks, and eventually buildings. It does so in place, keeping a consistent sense of place within the map as a whole, and revealing new elements at a point that feels natural. Doing so any sooner would unnecessarily clutter the map with elements too small to be useful, and doing so later could leave the user confused, staring at a featureless stretch of map, essentially lost in desert fog, with little indication that more detail exists at a more zoomed in level.

It is important to note that Google Maps is very heavily image-based, and a domain in which semantic zooming is such a clear solution to the problems of paper maps to the point that it is rarely even thought of as semantic zooming. It also falls clearly in line with B.B. Bederson's recommendations for use cases of zoomable user interfaces: navigating spatially organized data in a context where place and spatial navigation are central to the tasks afforded by the interface.

In this paper however, we will make the case that the restriction of ZUIs to strictly spatial and graphical domains is unnecessarily constraining to the design space. Specifically, we see many of the same problems such as information overload and "getting lost" arise in applications that involve navigating lengthy hypertext documents. Building on the lessons learned from previous ZUIs, we suggest that with an appropriate application of semantic zooming, a ZUI can make a large information space more navigable, particularly in the case of Wikipedia.

Wikipedia

Wikipedia, a hypertext-based, crowdsourced, free, online encyclopedia, is currently the 5th most visited website in the United States according to SimilarWeb as of March 2018. It primarily consists of articles which each center on a certain topic. Those articles contain a summary of the topic, subsections which describe different aspects of the topic, and references to source material. The sections can contain text, images, tables, other content, and most importantly: hyperlinks.

Hypertext, initially envisioned as a feature Vannevar Bush's theoretical persona computer, the Memex [6], and refined by many others is implemented in Wikipedia, and in webpages in general, as decorated text which, when clicked or otherwise interacted with, take the user to a new page or section of a page. Compared to encyclopedias, in which researching a topic referenced in one place required flipping through many volumes' alphabetic organization structure or referencing an index, hypertext greatly reduces the cost of learning new information, encourages using valid sources and references, and facilitates making connections between different fields of study.

However, hypertext, especially in its current form and function, is heavily dependent on the paper paradigm and thus defined as points that link different, entire documents together, and may be unnecessarily constrained by this feature. Hyperlinks serve to link together different sources and support self-directed learning and exploration between pages and topics on the internet. However, there are several aspects of their modern, standard implementations in web browsers which do not seem to support these actions. Here, we list our hypotheses for usability and design issues with Wikipedia, which in many ways are inherited from the paper paradigm, hypertext, and even more broadly, the windows and desktop metaphors

Lengthy, linear, hypertext documents lead to information overload and do not support information foraging

Wikipedia articles vary greatly in length and number of sections. Depending on the topic, they can range from a few, top-level sections to dozens of sections which can be up to three levels deep in the information hierarchy. As such, the user must decide what portions of the page they wish to read, as they will not always want to read the entire article like they would a book, especially if their task involves only a specific part of the topic, or getting a specific fact. Due to the variability in tasks and pages, different tactics must be adopted for searching for information, termed information foraging [5]. We will go into different tactics that we observed in the results section. Due to the amount of text in an article, as well as readability issues such as a high number of words per line, limited page structure information, and scrolling as the primary method of page navigation, information foraging on Wikipedia can be quite cumbersome.

Hyperlinks have a weak information scent

Hyperlinks are an elegant solution to linking multiple documents without changing the structure of those documents because they often take the form of decorated, in-line text. That same strength is also a weakness however, because the only hint as to what lies behind the link is the text itself. This text is often only one word or a short phrase, as making hyperlinks out of large sections of text is ugly, imparies readability, and brings up the possibility of overlapping hyperlinks. But because the only information scent afforded by the hyperlink is one or two words, it is difficult to evaluate whether or not that link will contain anything useful to the information forager, even if they recognize the topic. This can lead to the user spending time and energy switching pages only to not find anything useful, or choosing not to follow a link which contains information useful to them. Some websites (including Wikipedia recently) implement features such as a page-preview on hover. But still the issue remains of having to swap pages. This matters because switching to a new document in a windowed environment has a relatively high task-switching cost because each article is self contained, with its own organization and framing.

Hyperlinks are context-independent

Because the paper paradigm was carried over to webpages and thus Wikipedia, each topic on Wikipedia is written in a way such that it can stand alone from all others, including those that link to it, or those which it links to. A page can often be read from top to bottom to gain a relatively complete overview of the topic (at least as much as Wikipedia's editors have written), but as mentioned earlier, and as will be supported by our results, linear reading is not the reading strategy of all users for all tasks for all pages, particularly when it comes to hyperlinks.

When a user is reading about one topic, and clicks on a link within the topic, it is likely that the user wishes to learn about the new topic in the context of the old one, and probably return to the original topic (will be supported in results). However, clicking the link simply dumps the user at the top of the new topic in most cases, instead of carrying the user's context with them. This puts the burden of searching the new page for context-relevant information on the user, and leads to inefficiency and problems with distraction, commonly termed the “wikihole”.

The burden of document organization and history reconstruction is on the user

The final issue that we found to be within the scope of this paper is that the user is responsible for organizing their views, history, and activity of all the pages they visit on wikipedia. Inherited directly from the implementation of the paper paradigm in modern web-browsers, every link opens up a new document, within the same tab in the same window, in a new tab in the same window, or a new window. The user must choose which of these strategies to employ, use it, then organize the pages as they continue their information foraging task, moving, closing, opening, and reorganizing windows as they wish. This brings with it all of the previously discussed issues native to windowing systems and aspects of windows such as tabs. Because the user is responsible for organizing multiple pages, and also remembering items about page structure and page content from multiple pages, it is likely that the user will forget or misremember information as they continue to browse wikipedia.

Research Questions

At this point, we have discussed the history of modern user interface metaphors, reviewed lessons learned about the design and use cases of ZUIs, and identified a possible new use case for ZUIs (particularly semantic zooming) in large information spaces made up of many hierarchically-written hypertext documents, which is at odds with the current literature. At this point our research questions include:

1. What is the application for ZUIs in the realm of hypertext documents?
2. Will users find a ZUI for Wikipedia more engaging or better supportive of their information foraging tasks?

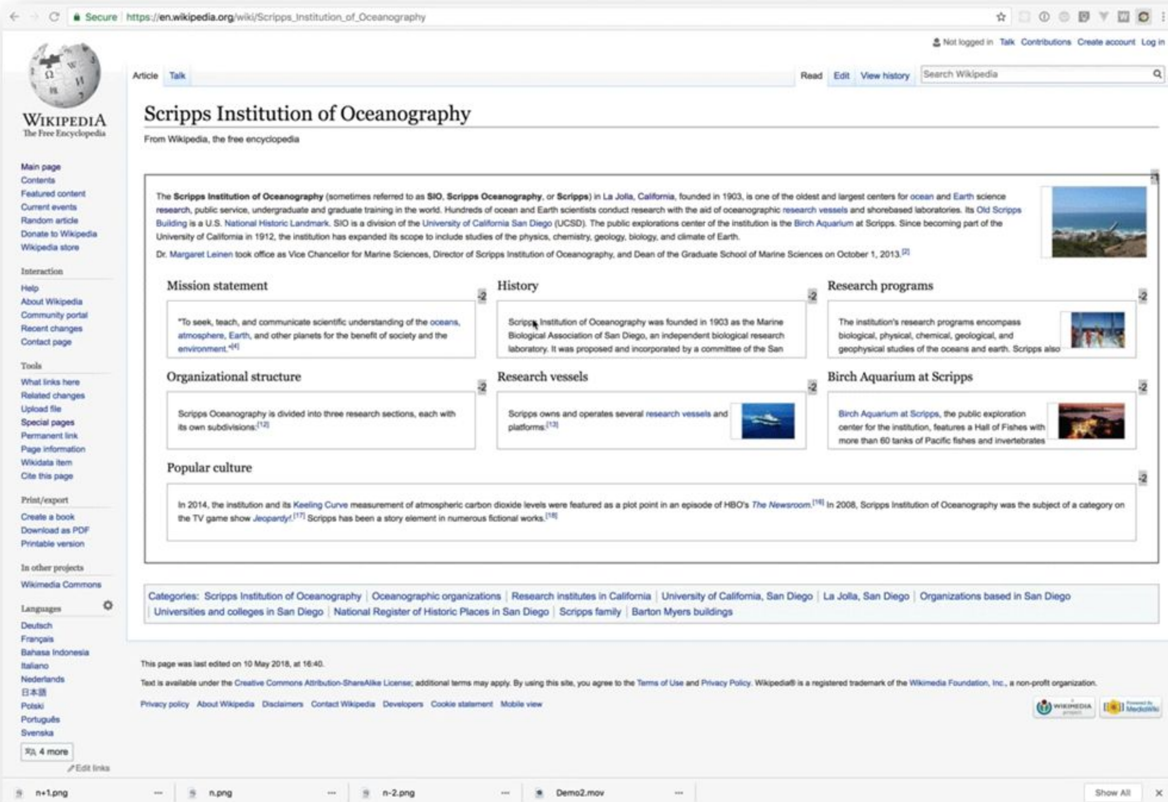
3. How do the established heuristics for ZUIs change in this context, particularly for semantic zooming?

Methods

Because of the abstract nature of user interface metaphors, and thus the difference between a zoomable user interface and a non-zoomable interface being rather hard to nail down, we decided that the best place to start would be to design a ZUI for a familiar platform that is stereotypical of the desktop, windows, and paper metaphors, and conduct a formative study to further develop research questions and hypotheses for the area.

As described previously, Wikipedia falls into this category, being a fairly literal implementation of a hypertext encyclopedia designed for and within the context of these metaphors, and remaining widely used and relatively unchanged since its popularization. [3] As such, we utilized a process inspired by human-centered design, involving an iterative development cycle for observing user behavior and needs, prototyping, and testing, to create an alternative research prototype interface to test our assumptions about use cases for ZUIs. That prototype was used in a mixed-methods, formative study resembling an A/B usability test between Wikipedia's standard user interface, and the prototype ZUI.

Prototype



The main goal of the prototype is to test a semantically zoomable user interface against a non-zoomable user interface for a platform and context that is familiar to the average user in order to observe naturalistic behaviors and how they change with a different interface. B. B. Bederson defines a ZUI as "systems that support that spatial organization and navigation among multiple documents or visual objects." [1] Because this is a very minimalistic definition of a ZUI, we will be interpreting it to also mean that zooming is the primary method of navigation, as opposed to unidirectional scrolling or using links to navigate between documents using tabs and windows. In order to test our assumptions in realistic use cases, we developed the ZUI in the form of a browser extension which changes the view of a loaded page from Wikipedia.

Approach

As such, our implementation of the ZUI differs minimally from Wikipedia's current user interface in regards to aspects which are not relevant specifically to a ZUI. This means keeping visual design (font-height, words per line, sidebars, etc) consistent, as to not confound our study by making Wikipedia more readable or otherwise user-friendly. The major difference in the prototype is that sections of the page are semantically zoomable, having three levels of view, from zoomed out to fully zoomed in, showing little information when zoomed out, and all information when zoomed in. When zoomed out, these sections are represented as "cards" which contain the section title, the beginning of the section text, and a thumbnail of the section's first image (if any) to align with the principle of semantic zooming that a zoomed-out view shows a little bit of information with the goal of communicating the section's contents to the user.

A major decision to make regarding the implementation of a ZUI was global zoom versus local zoom, and having a static viewport versus a mobile viewport. Global zoom is defined as a zoom event acting on the entire working "canvas", zooming all items equally, while a local zoom will only zoom the item being acted on, leaving the other items at the current level of zoom, as is somewhat similar to Fisheye zoom [9]. A global zoom better fits a more traditional idea of a zoomable user interface, but introduces performance problems due to every element being resized, and would cause constant reflow in a browser, changing the layout of the page, which is typically jaring to the user.

A static viewport in this implementation would mean that panning, particularly between articles or sections that are at the surface level of zoom is not allowed. Instead, the user would have to zoom out to a previous level, viewing the section that contains the previously viewed section, and it's sibling which the user is attempting to navigate to, and zoom into the sibling. Panning would only be allowed when a fully zoomed section's content does not fit on the page.

Because we are testing for the effect of a zoomable user interface, and not necessarily for a product up to market-ready usability standards, we chose to implement in the way that most directly tested our hypotheses about ZUIs, and introduced the least new usability and performance problems in a web environment (such as page reflow). As such, we chose to implement a local zoom with a static viewport, so that we could easily control the size of all elements, keep them in the same space relative to the screen as to not disorient the user, and to minimize actions on the DOM for performance purposes. This also had the added benefit of forcing the user to zoom to navigate the page instead of adopting strategies such as zooming globally to a full view of the document, then navigating it only by panning.

Design Rationale

Defining the act of semantically zooming on a document, or portion of a document which is composed primarily of hypertext is no trivial decision. In highly graphical interfaces, such as Google Maps, designing these levels of view require very skilled visual, interaction, and information designers. However, many of these decisions revolve around when to show and hide different labels, or what size and detail level to draw a road at. Through user testing and consulting literature concerning the number of items can be on screen without impairing task performance, there is a relatively well-defined path to adequately defining and designing these levels.

For text-based items, such as our domain of Wikipedia articles, what constitutes a "zoomed out" view of a section is somewhat unclear. It makes sense to think that a block of text, zoomed out, would maybe be one summary sentence, and then maybe a few keywords, and then the content would disappear, showing only a title. However, this design poses multiple challenges in terms of usability and technical feasibility.

For usability, replacing one text (keywords) with another (sentence) can be jarring to the user without a carefully choreographed transition that communicates to the user that the information is still the same, but merely at a more detailed view. Secondly, the implementation of this would require either a wizard of oz approach, an article structure in a way that has keywords for every sections, or advanced techniques such as natural language processing algorithms that can extract keywords or topic sentences. For the scope of this project, we are developing a solution that can stand on its own, and can support natural user navigation throughout wikipedia without the delays of a wizard of Oz approach. For the natural language processing approach, we believe that the concepts of zooming can be tested with a simpler implementation which, while not as supportive of information foraging tasks, will be easier to implement, more familiar to the average user, and not suffer from usability problems of replacing text and reflow.

This solution is simply to expand a section in terms of size. To fit the definition of semantic zooming in the simplest sense, less information will be shown in a zoomed out view, which is defined by how much text/image/etc content fits in a smaller interface element. As an element is zoomed, it shows more and more content until the section is in “full-view”, being big enough to display all of its content.

This solution, being technically feasible and designed to minimally test a ZUI on Wikipedia, may also lead to interesting user behaviors, such as supporting the ability to “peek” into sections, reading a little more content so they can decide whether or not they want to “dive in.” These findings will be discussed in the results section.

View, depth, and zoom

A ZUI implemented to navigate a hierarchical information space must have a consistent, mathematical framework to define the zooming transformation on an element. In geometric zooming, this takes the form of geometric scaling, in which the form of each element has a geometric transformation applied to it to enlarge or shrink it. This scaling factor can be defined as zoom. Zoomable elements of the ZUI must also have a base size, which defines their dimensions when no zoom has been applied. Let us refer to this as “depth.” For this implementation, any element nested within another is at one level deeper than its parent. For example, the summary at the top of the “French Revolution” Wikipedia page is at a depth of 0, the first level subsections such as “Causes” and “Constitutional Monarchy” will be at a depth of 1, second level subsections such as “Financial Crisis” are at a depth of 2, and so forth.

And finally, we will refer to the size calculated from the depth and the zoom, the form which the user sees the data at, as the “view.” View is calculated as zoom minus depth. For example, if an element is at a depth of 3, and has a zooming transformation applied of a value 2, the effective view will be -1, as in “one step zoomed out from the full data.”

All of these values are of an arbitrary range, which can be defined at that which makes sense in the context of the specific interface (0 to 1, 0 to 100%, etc). For the sake of simplicity, the prototype in this paper defines three different levels of view for each section of a wikipedia page: 0, -1, and -2. Any items at a view of 0 will be at full size, -1 slightly reduced, -2 greatly reduced, and less than 3 will be considered outside of the current scope, and not viewed. These views can be zoomed between between levels (-1 to 0) in which the representation of the element will change in a significant way. They can also be zoomed within levels (0.75 to .5), which will apply some sort of continuous zoom (showing slightly more information, without fundamentally changing the form or interactions of a section. This mix of between level and within level zooms allows flexibility in the representations of zoomed data, as well as a set of user interactions that effectively communicates the conceptual model of a zooming interface.

Name	**Description**
View	Zoom - Depth; the effective level of information at which the element is rendered for the user
Depth	How many layers “deep” an element is, also how much zoom must be applied to fully view an element’s data
Zoom	A transformation applied on an element’s depth that defines the view it is rendered at

Independents, Parents, and Children

Zooming an environment with a hierarchy or nested elements is an inherently recursive action in that the zoom of one element also affects all of the elements nested within it (children). This feature often goes unnoticed in geometrically zoomable interfaces, or those with global zoom, as the zoom simply needs to be applied to all elements equally. But in our case of a semantically zoomable interface which can zoom locally onto one section, the interactions between a parent and child during zooming become complex. Many questions arise such as:

1. What zooms, the object being acted on or its parent?
    a. What happens if the parent is too small to provide space for the child to zoom?

    b. Where are these boundaries drawn?

2. What happens to a zoomed elements siblings?
3. When do elements enter or leave the current scope (in local zoom with a fixed viewport)?
4. Can a parent's content be zoomed without zooming any of its children?

To address these questions, we developed three different labels that define how an object being zoomed interacts with other objects in the information hierarchy, and also aligns with the previously mentioned levels of view: Independent, Dependee, and Dependent

Name	**View Level**	**Interactions**
Independent (Empty Nester)	0 <= V < -1	Can be zoomed without affecting nested elements or elements it is nested in
Dependee (Parent)	-1 <= V < -2	Zooming effects this element, as well as all nested elements. This element's siblings are in view, but not affected by the zoom
Dependent (Child)	-2 <= V	All zoom events on this element are delegated to the parent, which zooms itself, this element, and this element's siblings

This system strictly defines the relationship and zooming interactions between all elements in view by their position in the information hierarchy relative to each other. The simple concepts are the Dependee (parent) and the Dependent (child). A child cannot be zoomed directly, as it is too far from the user, and would not have enough space to grow and show its full content. Instead, it delegates the zooming event to the parent, which can be zoomed independently of its siblings, taking space away from them, and thus gaining space for its own content and children to grow. For example, a user could attempt to zoom a element at the child level with a view of -2.5. The event would instead get passed to its parent, which is one layer higher than it, and thus has a view of -1.5 and is in the dependee/parent level. The parent gains a zoom transformation of +0.1, going from a view of -1.5 to -1.4 growing slightly in size and amount of information shown, and passing the zoom down to it's children, which changes from a zoom of -2.5 to -2.4.

But what happens when an element transitions between levels? Let us say that the previously described element, section 3: “Constitutional Monarchy”, is zoomed in, moving from a view of -1.01 to -0.91. This element has moved from the parent layer to the independent layer. Because the zoom was enacted while it was still a parent, the zoom is also passed down to all nested elements, moving them from a view of -2.01 to -1.91, shifting them from being children to parents. At this point, “Constitutional Monarchy”, the section that went from being a parent to an independent (empty-nester) effectively popped out all of its children once they became parents and are able to act on their own. The constitutional monarchy section is now at the 0 view layer (independent), and is in focus of the static viewport. Because the static viewport in a local zooming environment only focuses on one section at a time, the “Constitutional Monarchy” section is anchored to the top of the viewport, giving the most space to its content, and also showing its subsections at a lower level of view. All of “Constitutional Monarchy”’s sibling sections and parents are outside of the current scope, and thus not viewed except in signifiers such as breadcrumbs. In order to navigate to those sections, the user must zoom out. Instead the user could also zoom a subsection of the “Constitutional Monarchy” such as “3.1: National Constituent Assembly”, which, being at a parent level, will zoom to the independent level. This will push the “Constitutional Monarchy Section and all the siblings outside of the current scope.

Study

With our research questions and semantic zoomable user interface prototype for Wikipedia, we move forward into a formative, mixed methods study with the goal of validating assumptions about ZUIs and the problem space of navigating hypertext documents, piloting a new study paradigm, and informing future areas of research. Due to the formative nature of this study, and the focus on hypothesis generation, we are adopting a within subjects model similar to that of A/B usability tests. After an initial survey and screening of subjects, we had them complete the same series of tasks with both quantitative and qualitative measures in two sessions, one using Wikipedia (C1) and one using our prototype (C2).

Self-directed research task

Central to this study is user-directed research on Wikipedia, which is the first task in both sessions. In order to observe realistic use as accurately as possible, the user’s topic of research is self-chosen, and thus has intrinsic motivation to complete the task well. In the study’s pre-survey, the participants listed three topics that they would be interested in learning more about. Those three options were pre-screened by a researcher, and a topic was chosen for which a specific Wikipedia page exists, and that page has more than five sections and at least two levels of subsections (e.g. section 1 and 1.1). These requirements for page complexity were put in place so that we could study the effects of the structure of Wikipedia on user experience.

At the beginning of the study, the participant is informed that they will be spending 15 minutes researching a topic on Wikipedia chosen from those they listed in the pre-survey. They are then asked to explain why they are interested in the topic, and set goals for what they want to learn about the topic (taken from Wikipedia's questionnaire concerning how people use the website). [8] They are then instructed to spend 15 minutes researching the topic with those goals in mind. They are allowed to visit other pages on Wikipedia, but are asked to refrain from spending significant time on sources outside of Wikipedia. During this task the screen is being recorded, but the researcher leaves the room as to mitigate an observer effect.

Page Structure Task

After the duration of the task, the researcher re-enters the room, ends the recording, and gives the participant a short, quiz-like activity to complete. The user is instructed to draw out the structure of the page or pages they visited on Wikipedia as best as they could remember using the section titles as the primary unit of organization. This task was chosen for two reasons: 1) to get qualitative information about the user's conceptual model of Wikipedia pages and their structure, and 2) collect quantitative data on the user's recall of the page structure. This task was chosen over a classic content comprehension or content recall test so that it could 1) be scalably applied to any page they visit, and 2) test for hypotheses related to user's navigation and understanding of the structure of the page, instead of information recall.

User-led, recording mediated, retrospective talkaloud

The last portion of the study is what we call a user-led, recording mediated, retrospective talkaloud, which was designed with the goal of collecting qualitative data on user's understanding of their actions and the interfaces they use a la user testing without the observer effects that or distraction that come from doing a normal talkaloud. In this version of the talkaloud is a user-led interview in which they navigate through a screen recording of their self-directed research task in whatever way they wish, narrating their use and explaining the thought process behind their actions. For this study, the user is asked to explain what they were looking for in each section, if they found it or not, how they navigated the page, and why they read certain sections but not others. During the task, the researcher asks questions about navigation behavior that either was not explained by the participant, or needs further clarification.

Results

Eleven sessions were conducted for the standard Wikipedia user interface condition (n1 = 11) and five were conducted for the ZUI condition (n2 = 5). Each of these consisted of the previously described self-directed research task, structure task, and user-led, recording mediated, retrospective talkaloud. Here we will discuss the observed behavior in both conditions, and how it supported or did not support our hypotheses and assumptions about the current Wikipedia user interface, as well as what they imply about the promise of a zoomable user interface in the domain.

Qualitative

Information Overload

Apparent from the beginning are the issues with information overload in Wikipedia's standard user interface. When opening a page, and coming across a large subsection composed of only text, one participant remarked:

> "I usually lose my patience on wikipedia pages when they cluster information like this. I wish there were some highlights so I can better direct my eyes to the information that's important. And also its very disorganized."

While not all of the users commented on the amount of information causing problems of information overload, the behavior of each participant reflected it in one way or another, either in mistakes and inefficiencies they showed, or in strategies they employed to get around these shortcomings. Several users were noted by the researchers as not recognizing information on the screen which they were explicitly looking for. One participant experienced this while researching the causes of Alzheimer's Disease. While navigating through several pages related to this topic, they found themselves on the page for Tau Protein, searching for a section related to Alzheimer's Disease. They looked at the page's table of contents, didn't find what they were looking for, then started reading the page linearly, got frustrated, and used the browser's find text function. At this point, they discovered that there was a section listed in the table of contents titled "Hypothesis of Alzheimer's Disease." This issue, while hard to code for by researchers due to participants not defining an explicit goal for every part of their research, was hinted at pervasively throughout the data.

This issue is partly due to the visual design of Wikipedia pages. They have small text, minimal line spacing, and a large average number of words per line, which are typically harmful to readability. However, changing these attributes would greatly extend the length of the article, which is also likely to overwhelm the user and make skimming tactics and scrolling take more time. As such, a sensible direction for redesign could involve hiding information that is not currently needed, as the zoomed-out view of a ZUI is meant to do.

Often, users would find information which interests them, or which they were looking for upon visiting a section for the second or third time, seemingly not discovering it the first time. Additionally, issues with information overload can be inferred to be the reason for the variability and complexity of information foraging behaviors recorded.

Information Foraging

Here, we describe observed information foraging behaviors in open-ended research on Wikipedia, and how it relates to the user interface and underlying interface metaphors

High Variability in Navigation Strategies

What stood out immediately was the wide array of strategies employed to navigate Wikipedia, particularly in terms of moving from one page to another, or deciding what to read within one page. Variability in behavior is not always an indicator of faults in a user interface, as users have different needs and levels of experience. However, we also observed that people who used the default, or simplest methods of navigation also experienced frustrations with the system and tended to cover less ground in navigating the website.

Tabbing

The clearest distinction is between those who use the default behavior of hyperlinks in Wikipedia, opening a new topic within the same page, and those who open the page in another tab. Using the default behavior of a hyperlink, which happens by left clicking one, replaces the current page with the page being linked to, usually dropping the user at the top of the new page. Because links are encountered in the middle of a sentence, often in the middle of a paragraph in the middle of a section, clicking that link while reading will break the flow of the user's reading and possibly distract them from finishing a section and obtaining more relevant information past the first interesting link. Some users visited links in the same tab after finishing a section. However, this requires the user to remember to revisit the link and where where to find it while they are processing other information, which is likely to lead to errors. Additionally, using only one tab requires frequently using the back button, which has limited history traversal affordances because the forward button only leads to the page on which the back button was most recently pressed.

One participant stated that they were "annoyed" by having to constantly use the back button to learn about an artist, one of their albums, and each song on the album. In response, they changed their tactic to opening each page in a new tab, then visiting them one by one, and closing them when she was finished. Three of our participants explained there navigation strategy explicitly as opening new topics in new tabs as they are reading a section or page, and then visiting those pages one by one afterward. One participant described the linked pages as a "tree of information that I can jump out and into." So they used tabbing to go through each topic, evaluate if it was relevant to them, an stated, "it sort of helps me I guess track my sourcing as well as the route I've gone through already into that."

Throughout the study, it appeared that this tactic of opening new tabs while reading a section and visiting them afterward was the prefered method of those experienced in doing research on the web, and was adopted in response to frustration with the default behavior of hyperlinks. As noted above, we also found evidence that users navigate from a broad level to a specific one, treating their navigation like a tree, and seek a navigation strategy that lets them navigate that tree of activity, even minimally like new tabs do.

Due to technical limitations, we were not about able to test the envisioned implementation of opening new links in the spatial context of the same page with our ZUI prototype, so this feature of a ZUI will need to be explored in future research.

Skimming

Nearly all of our participants exhibited skimming behavior in one form or another, which we define as a tactic of scanning content in some way without reading all of it in order to test for the “information scent” of what they are searching for, and determine whether or not to read a section. Due to the relatively high time cost of reading a section fully, it is often more efficient to scan it first to decide whether it can be skipped. It is estimated that users only read an average of about 20% of the words on any given webpage[10], which points to skimming as a commonly used information foraging tactic.

From what we have seen, Wikipedia’s user interface does not well support the act of skimming, and thus information foraging. This problem overlaps with that of information overload, as several participants stated that it was difficult to skim lengthy blocks of text within one section with little organization. Most participants reported themselves as skimming sections at one point or another. Skimming also saw variability in tactics. Some users skimmed through the page, looking for text that “caught [their] eye” or looked “appealing”. Users often found subheadings and bullet points useful in understanding the main points, however, some found too many bullet points to be overwhelming as well. We will discuss how users decided to read a section later. Others skimmed the page using images to determine what they wanted to read. One user described this as being the intended way to use Wikipedia:

> “So I have a habit to look at these side examples because they're like, I think Wikipedia put these things here to give me a sense, like clear sense of what this whole bunch of words is talking about. So maybe I can just get a sense and jump through all these pages and I found this on this example.”

While these behaviors varied widely, they support the idea that information foraging and the scanning and skimming involved are fundamentally visual. Even skimming through text was based on “appeal” which users were unable to define clearly. Thus, an interface in which users are likely to skim should represent its content in a way that supports skimming.

In our prototype, which represents sections in minimized boxes that contain the first bit of text, the title, and an image (if one is in the section) seemed to better support fast “skimming” in that users would quickly scan each element before deciding what to read. If they were interested, they would expand the section slightly more and read enough to decide whether to “go into the section” or look somewhere else. This process was described as “fluid” by one of the participants. We believe that our implementation of a ZUI was successful in this regard because 1) it reduces the amount of information on a screen to that which provides the strongest information scent, and thus reduced information overload and 2) the zooming interaction gives the user more explicit control over where to look or “go” within a page than scrolling, and is more engaging overall.

It is worth noting that some users were sometimes surprised that they found something in one section after expanding it. This is due to the current implementation of the “zoomed out” view containing the first sentence or two, and thus relying on how well the article was written for its effectiveness. Other implementations need to be explored in the future.

Reading the First Paragraph

An observed reading strategy that falls in between reading and skimming was reading the first paragraph of each section. One user stated that they chose this tactic because they felt the first paragraph of a section often contained enough information to get a general understanding of the topic or decide whether or not to read it fully. If the section or the page was already short however, they would read the entire content.

This is likely because a common tactic in writing a lengthy section is to summarize the points in the first paragraph. The observation of this strategy supports our implementation of the “zoomed out” view of a section, but is heavily reliant on article writing style. Additionally, this tactic was often mixed in with skimming as well.

Non-Linear Foraging: From Broad to Specific

No matter what tactic participants used in C1 to navigate Wikipedia’s user interface, they mostly followed the pattern of starting with a broad search, navigating non-linearly through a page to find the topic of interest, and getting more specific until the user was finished, at which point they would return to the broad level of information. Specifically, this fits the tactics of skimming the whole page or skimming a lengthy section, then returning to a point of interest to dive deeper. This fits exactly the strategy of the participant who was researching meanings behind a particular artist’s songs. But instead of just searching for the songs, they started at the artist’s page, ignored all the content not relevant to their goal, found the album of interest, went to that page, then explored each song’s page one by one. One participant who showed this behavior described their tactic as “breadth then depth”, also saying, “I don’t go linearly through pages.”

As mentioned earlier, participants treated Wikipedia like a tree of information, and at each level they would look at all of the immediate options, make a decision about which path to go down, then continue down until they decided they were finished, at which point they would return to a page higher up in the tree, or even the summary of a current page, and look for links in that section to follow down to other topics.

Recognition-Directed Reading

Rather interestingly, and somewhat counterintuitive, we found that when we asked users to explain why they read one section and not another, they often stated that they remember seeing this topic in another place, hearing about it before, perceiving themselves as being informed on it, or otherwise recognizing the section. This recognition was often was sparked the interest to read a section. The exact reasons for reading a section included wanting to refresh their knowledge, check what they know, discover the authors' interpretations of the topic, or fill in gaps in their knowledge.

Users rarely states that they read a section which had a title or content which they did not recognize whatsoever, at least according to their narration of their activity. This seems counterproductive to acquiring new knowledge, but makes sense within the idea of information foraging, in which recognition is a powerful tool. The ZUI condition can theoretically better achieve this with a zoomed out view of an element containing elements helpful in recognition such as titles, thumbnail images, keywords, topic sentences, etc.

Use is Directed by Context

When a user follows a link from one Wikipedia page to another, they take their task, and their context with them. When clicking on "France" from the "French Revolution" page, a user does not start a new task, with a blank mind as the interface seems to assume by dumping the user at the top of the page, and giving them all of the Information to wade through, organized in a way that only makes sense from information architect's point of view. Instead, the users carry their knowledge, goals, and mindset along to the new page. And that mindset may not even be that of the French Revolution, but some other, more specific question that started the research.

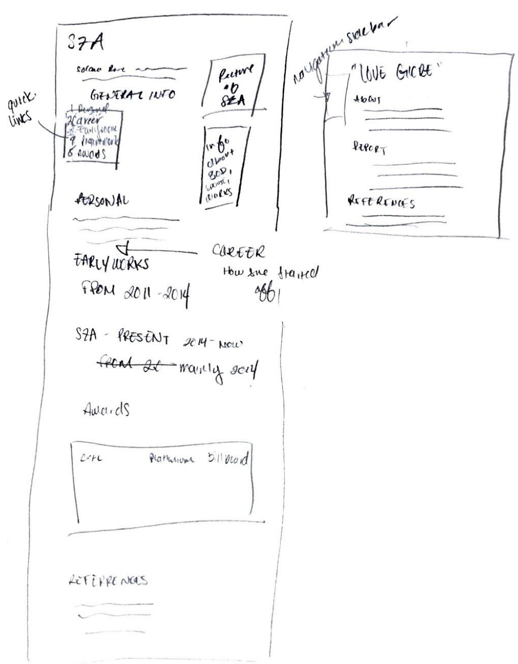
We saw this in the participant who was researching Alzheimer's Disease. They reported that the topic was personally important because the disease had affected their family, and so they were interested in causes, and possible treatments. While reading the section of the Alzheimer's Disease page related to causes, they found a reference to Amyloid Beta, and navigated to it's page, skipping the summary, and clicking the link in the table of contents related to Alzheimer's Disease. That section referenced the Tau protein, which they clicked the link to, and attempted to find a section on that page related to Alzheimer's as well.

We observed nearly every patient navigate through new pages, ignoring certain content and paying attention to others through a type of filter of the participant's browsing history and goals. One participant explored pages about a musician, their albums, and songs on their albums while only caring about the meaning and influence behind the work, disregarding all information about awards and other topics. Another participant was exploring films that were examples of the German Expressionism movement in the context of silent films.

A good user interface supports a user in their specific goals, and keeps their context of use in mind. This sort of design is difficult to create in an interface structured around linear documents, as documents are often self-contained, and switching between them requires switching into the context of the document, instead of the information's context changing to meet the user where they are at. A ZUI, by treating all the information in Wikipedia as an information space that can be navigated through one piece at a time, and bring new information from different articles into the user's current context can better support the user in these ways. We plan to explore this in future prototypes with features such as opening up a new topic within the page it is linked from, treating it as simply another subsection and opening it in the context of the current topic.

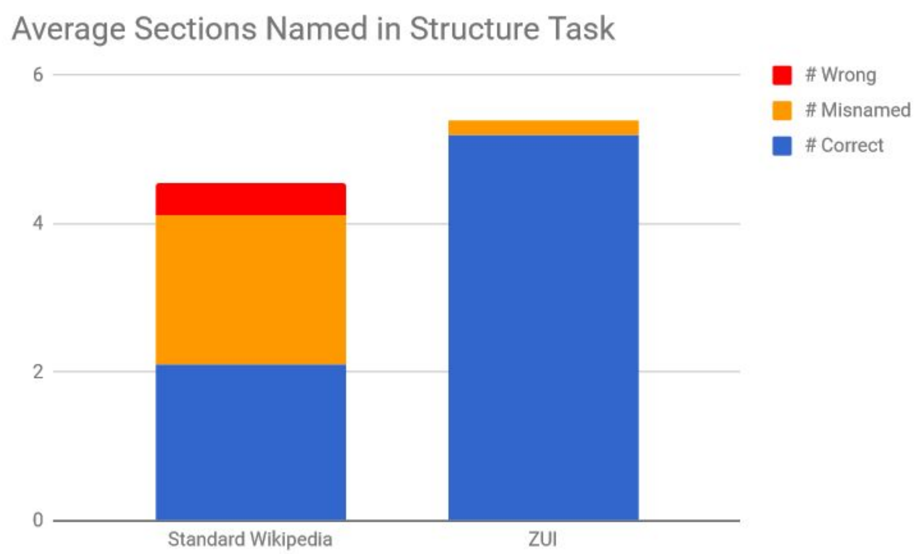
Quantitative

Structure Task

Standard Wikipedia	ZUI
	

As shown in the graph below, the number of sections a participant could name increased slightly from the standard Wikipedia condition (C1) to the ZUI condition (C2). However, there was a much higher incidence of misnamed or wrong section titles in the standard condition. Section titles were marked as correct if they had most of the title right (e.g. "Rankings" listed for the section “Rankings and Admissions”), Misnamed if they used a synonym (e.g. “On-Campus Housing” for “Residential Colleges”), and wrong if a section could not be found on a page, or it was listed in the wrong page.

These results will be controlled for order effects and be reproduced using more similar, strictly structured C1 and C2 tasks in future studies as opposed to the self-directed C1 research task of this study



The increase in section titles recalled, as well as those recalled correctly implies that the zoomable user interface is better supporting a user’s understanding of a page’s structure and the sections it contains. This is likely due to showing a “zoomed out” representation of the page that can often fit all 1st level subsections above the fold, showing them clearly with a prominent title, some text, and a thumbnail. In supporting recall and skimming, this seems like an obvious improvement over the standard table of contents

Discussion

This study does not bring many conclusions, but rather reaffirms the problems and assumptions underlying it. Wikipedia, and more broadly user interfaces that involve navigating a large information space made of hypertext documents, have a user experience which suffers from many of the symptoms of the user interface metaphors which are pervasive in modern technology. The concepts of the desktop, paper, and windows bring useful conceptual models, affordances, and constraints with them to interfaces they inspire. However, as capabilities of such technology, and the needs of their users surpass that which the desktop or piece of paper was ever capable of, it is time to question re-evaluate designs based on them, and expand our design space past them.

I am not suggesting that these metaphors are completely antiquated, and thus should be replaced by other models, such as the pad metaphor. Instead, I am merely using a redesign process in an appropriate domain as a way to generate knowledge about a particular topic in interface design, and use it as one example of the vast design space that lies outside of our cultural expectations of what an interface should be.

In regards to the implementation of a ZUI on a hypertext-rich information space, there is much more experimentation to be done to discover and further develop usability heuristics, and directions for future study

Future Work

The findings of this study are currently being used to iterate on the zoomable user interface prototype to resolve issues unrelated to specifically testing out the effect of a ZUI. We also look toward conducting small scale studies to test hypotheses about ZUI design: such as the applications of global versus local zooming, or the animation and transitions involved in zooming and how they influence the user's conceptual model of the interface.

Most interesting of these is the implementation of a "zoomed out" view of a text-based element. As mentioned earlier in the paper, future work will look at the effectiveness of showing representations derived from the text (summaries, key points, keywords, etc) instead of simply showing the as much of the text as will fit in the box, starting from the beginning. Designs such as these will be used together with the page structure task and other quantitative measures to obtain hard data on how a ZUI affects information foraging tasks, and develop guidelines for the implementation of these systems.

Assuming that we continue to find promise for ZUIs in the domain of navigating hypertext-rich documents, specifically in the case of Wikipedia, we also hope to re-explore features pioneered in older zoomable user interfaces, such as Pad++'s lenses and portals.

Conclusion

Windows, desktops, and paper have dominated our conceptual models of computers since they became a major consumer product, and thus have defined our interactions with them. While each of these user interface metaphors was a great innovation in terms of usability that brought millions of users to the world of computing, they have also been taken for granted as being *the* way computer user interfaces are. With this project, we hope to that our new look at an old alternative interface metaphor will may not only lead to an interesting and useful interface for Wikipedia, but will also expand our thinking about the design space for large-scale user interfaces.

Acknowledgements

I would like to sincerely thank Steve Zhao, Jim Hollan, Amy Fox, Philip Guo, and Marta Kutas for their endless support, advice, assistance, and inspiration that this project would not be possible without.

References

[1] Bederson, B. B. (2011). The promise of zoomable user interfaces. *Behaviour and Information Technology*, *30*(6), 853–866. https://doi.org/10.1080/0144929X.2011.586724

[2] Perlin, K., & Fox, D. (1993). Pad - An Alternative Approach to the Computer Interface. *In Proc. ACM SIGGRAPH*, 57–64. https://doi.org/10.1.1.33.1359

[3] Jeremy Reimer. (2005). A History of the GUI | Ars Technica. Retrieved June 14, 2018, from https://arstechnica.com/features/2005/05/gui/3/

[4] D. C. Engelbart. (1962). AUGMENTING HUMAN INTELLECT. Retrieved June 14, 2018, from http://www.1962paper.org/web.html

[5] Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*(4), 643–675. https://doi.org/10.1037/0033-295X.106.4.643

[6] Bush, V. (1945). As we may think. *The Atlantic Monthly*. https://doi.org/10.3998/3336451.0001.101

[7] Bederson, B. B., Stead, L., & Hollan, J. D. (1994). Pad++: Advances in Multiscale Interfaces. *Companion of the Conference on Human Factors in Computing Systems (CHI '94)*. https://doi.org/10.1145/259963.260379

[8] Singer, P., Lemmerich, F., West, R., Zia, L., Wulczyn, E., Strohmaier, M., & Leskovec, J. (2017). Why We Read Wikipedia. https://doi.org/10.1145/3038912.3052716

[9] Schaffer, D., Zuo, Z., Greenberg, S., Bartram, L., Dill, J., Dubs, S., & Roseman, M. (1996). Navigating hierarchically clustered networks through fisheye and full-zoom methods. *ACM Transactions on Computer-Human Interaction*. https://doi.org/10.1145/230562.230577

[10] Nielsen, Jakob. "How Little Do Users Read?" *Nielsen Norman Group*, Nielsen Norman Group, 6 May 2008, www.nngroup.com/articles/how-little-do-users-read/.