

Automated Webpage Redesign

Gabriella Brignardello
gabrig@stanford.edu

John Morgan
jmorg@stanford.edu

Milan Doshi
miland@stanford.edu

ABSTRACT

Currently, many websites that present us with data are inefficient to use, poorly formatted and organized, and just simply awful, hampering users' ability to interact with and understand the information. Because style, layout, and content vary greatly across websites, challenges to addressing this issue include generalizability and scalability. However, finding a solution to this problem will greatly improve user experiences and interactions online. In this paper, we present Webpage Handyman, a tool that allows users to re-visualize existing websites. Using auto layout techniques, which leverage style guidelines taken from tested design principles, as well as extracted semantic information, it makes the necessary, important design changes to the site. More specifically, the tool uses Node.js and Cheerio to parse webpage content, examines the depth of elements in the DOM and their positions relative to other elements, scales container widths relative to the page, finds existing or similar color schemes to style buttons using the CSS page, and makes the page navigation more distinguishable and easier to use. Given that there are many more transformations that can be made in order to improve poorly designed websites, it is important that Webpage Handyman is a tool that can be expanded upon. With a webpage parser and auto layout framework in place, there are an endless number of additional features that can be added to our tool, some of which we discuss in this paper.

INTRODUCTION

One of the most prevalent ways that individuals view data on a day-to-day basis is through the web. However, many sites that present data and information are extremely inefficient to use, poorly formatted and organized, and are just simply awful, which in turn hampers one's ability to understand the information these sites present.

To tackle this problem, we have developed an automated website redesign tool, Webpage Handyman. In order to automate the redesign process, this tool scrapes the website for its content, including CSS, HTML, and images and then creates a representation of the DOM. It then extracts the semantic information from a website, particularly tags and information relevant to them. Lastly, the tool uses auto layout techniques, which leverage style guidelines taken from fundamental design rules, as well as the the semantic information retrieved to make the necessary important design changes to the site.

With Webpage Handyman, users can re-visualize existing websites so that they can better interact with them, navigate on them, better understand the data, and have a better overall experience online.

RELATED WORK

The development process and functionality of our tool for re-visualizing existing websites will be building on three areas of prior work. These are: design principles for visual communication [1], automated layout techniques [2], principles for webpage usability [3], and design mining on the web [4].

Design Principles for Visual Communication

Agrawala, et al.'s paper [1] emphasizes the importance of following three steps when creating visualizations: (1) identifying the domain specific design principles; (2) putting these principles into use with automated visualization design systems; and (3) evaluating their effectiveness with metrics. Our project structure has been based on these three steps; we began our process by identifying a set of poorly designed sites and a set of well designed sites to establish what visualization aspects of webpages we want to focus on improving with our tool. We are also examining the most relevant related work (with this literature review). Then we will be creating our automated system that applies the design elements we have identified as most important. Lastly, we plan to conduct some user testing (as time is limited, it will not be as extensive as we would like, but we will still pursue an initial round) in addition to our own thorough testing as a means of evaluating the effectiveness of this tool.

Automated Layout Techniques

Because effective layout is one of the most important aspects of presenting information, it is necessary to explore what methods for automated layout currently exist. Lok and Feiner's paper [2] explores specifically constraint and learning-based methods and systems where the former uses a set of preset rules or constraints to lay out the content, while the latter is trained by either an expert or the user. Our work will be more in line with the constraint-based systems described as we will be developing a set of both abstract (or relational) and spatial constraints or rules, specifically related to overall layout, navigation, and hierarchical information, that will then be applied to a webpage in the redesign process through a user interface that will be used to both input a website and output the source code for the revised webpage.

Principles for Webpage Usability

In Neilson's book [3] about web usability, he explains that "people do not come to the Web for an 'experience,' they come for information," which ultimately leads him to assert two main principles on this matter: (1) web users want to find what they are looking for quickly, and (2) if they are

not looking for something specifically, they want to browse quickly and access information that they come across in a logical manner. For this reason, we are focusing our redesign on several web aspects that we have identified as consistently problematic in current “bad” websites: proper display of hierarchical information, overall style and spatial improvement, and clarifying the display of the navigation functionality. With these visual improvements, a user would be able to easily find what they are looking for and/or have a way of intuitively navigating the website and skimming for relevant information, increasing the webpage’s usability.

Design Mining the Web

Kumar, et al.’s paper [4] discusses knowledge discovery and web crawling techniques and how these can be effectively applied to webpages and their design data in order to improve automated design curation, as well as help users understand design demographics and support new data-driven design interactions. Specifically, it discusses several design mining principles (e.g. scalability, extensibility, completeness, and consistency) as well as order of operations (i.e. HTML to DOM to resources and derived features) that helped us develop and refine our semantic information extraction process.

While our tool will be based on the ideas from these prior works as well as additional online web design resources, we plan to take our tool one step further by applying these techniques simultaneously in order to take existing, poorly designed websites and transform the elements we choose into new versions that follow what we have recognized as highly regarded principles.

METHODS

Webpage Handyman allows users to re-visualize existing websites. As aforementioned, the tool uses auto layout techniques, which leverage style guidelines taken from tested design principles, as well as extracted semantic information to make the necessary, important design changes to the site. As style, layout, and content vary greatly across websites, challenges include generalizability and scalability. However, addressing this issue will improve user experiences and interactions online.

In summary, Webpage Handyman uses Node.js and Cheerio to parse webpage content, examines the depth of elements in the DOM and their positions relative to other elements, scales container widths relative to the page, finds existing or similar color schemes to style buttons using the CSS page, and makes the page navigation more distinguishable and easier to use. However, we will go into further detail about our approach in the following sections.

Identifying Relevant Design Principles

In order to identify the most relevant design principles that we could apply in our website re-visualization tool, we looked first at relevant works discussed above, like Nielsen’s book [3]. Then, we turned to online resources that

discussed fundamental and tested style guidelines and design principles [5, 6, 7, 8], as well as real-life examples of objectively “good” and “bad” websites.

The issues that we consistently saw on the websites that we deemed as “bad” include:

- Misused space and page width (i.e. drastically overused or underused)
- Poor layout of content (i.e. no whitespace, excessive text without breaks)
- Indistinguishable hierarchy of information (i.e. no visual, spatial, or stylistic separation between headers and body text or headers of different levels)
- Unclear navigation system (i.e. not visibly links, too small, misplaced on the page)
- Poor use of color palettes, fonts, and background imagery and coloring
- Overuse of graphics and imagery causing sensory overload

On the other hand, we classified websites as “good” for the following reasons:

- Appropriate and aesthetically pleasing use of space (i.e. whitespace and correct amount of page filled)
- Consistent color schemes and font usage
- User friendly and fluid design and layout of content
- Clearly identifiable navigation system (i.e. static bar across full width of top of the page page)
- Balanced use of imagery and text
- Visible and distinguishable headers and body text

From this, we synthesized our observations and review of prior works in order to create our own comprehensive set of style guidelines which is focused around four major website design principles:

1. Clear structure of hierarchical information
 - a. Distinct headers and breaks in content (i.e. distinguishable by font, color, and/or size)
 - b. “Call to action” items (e.g. buttons) should be made prominent but match page style
2. Easy to identify and understand navigation
 - a. Appropriate size and consistent styling across pages
 - b. Identifiable road map capabilities through link styling or full page width banner
3. Appropriate styling
 - a. Consistent typefaces and color schemes, and appropriate amount of variety
 - b. Maximizing use of space on the screen (while maintaining reasonable amounts of whitespace)

4. Proper use of space and intuitive layout of content
 - a. Use of Rule of Thirds, “F layout”, Golden Ratio (of “Divine Proportions”), and Gestalt Design laws
 - b. Overall clean, simple, and fluid design

While all of these address a visual component of websites, they were all chosen with the intention of making a website and its information easier to understand, interact with, and experience for users. Ultimately, usability and utility, supplemented by visual design, determine the success or failure of a website.

Initially, we set out to tackle all four of these points from our set of style guidelines; however, due to time constraints of the project, we executed two in their entirety — hierarchical information and navigation — and tackled some of styling. Layout improvements is the most complex of the four redesign principles that we identified as it requires implementation of natural language processing and artificial intelligence algorithms to correctly parse and organize the webpage content before reorganizing the page layout. In the subsequent sections, we will discuss specifically how we tackled each of the redesign principles from our set of style guidelines, including the algorithms and techniques used.

Webpage Parsing and Auto Layout Framework

We leveraged several existing libraries in order to scrape webpages for the information we would need to redesign them. Node.js, along with the Request module, allowed us to quickly point our server at a URL and access its HTML. Once we have made the request, we then use Cheerio, a server-side equivalent to JQuery, to parse the DOM and select elements based off of tags, attributes, and grab CSS.

Before editing the page, Node’s filestreams and Request allow us to download all related CSS and images that are necessary for the website, and then edit elements so that they point to the new versions of these elements. Because we dynamically scrape for all sources, websites can be viewed in browser from Webpage Handyman, or source code can be downloaded and re-hosted by the user after transformations are applied. With Cheerio, we can dynamically alter the HTML and CSS of a page and send it back to the client when the server finishes applying transformations.

This system accepts many URLs and is generalizable to several different web frameworks, making Webpage Handyman universal and lightweight.

Redesign Principle #1: Hierarchical Information

Visual hierarchy is one of the most important principles behind good web design as it is the order in which users will perceive the content and information of a webpage. Thus, it is an important for those who visit different webpages to be able to easily identify the hierarchy of the information and the importance of and relationships to each element on a page.

Often, information is set hierarchically in the DOM, but its styling and tags do not reflect this. One case of this is headers. If headers are styled similarly but are meant to convey different kind of divisions of the information, they make data more difficult to parse. In order to identify this case, we look at the depth of headers relative to the root of the DOM by traversing recursively for each element. We then examine common ancestors between elements to determine if headers are labeled correctly relative to their position and depth. If this has been done incorrectly, we adjust the labels in the DOM and the styling (i.e. font typeface and coloring) on the actual webpage to ensure that all of the elements properly reflect their characteristics (i.e. position and depth).

Furthermore, “call to action” items, like buttons, should be displayed prominently, yet appropriately with color schemes and styling that match the overall page theme. For example, we check the size configuration of each button on the webpage and if either the height or width of the buttons is too small, we make the appropriate size changes. Moreover, if a button is not properly styled to fit the overall theme of the website, we check the DOM for a prominent color used on the website and change the styling of the button to comply with the theme for the rest of the website.

With these transformations, the ranking of elements on a page as well as their associated relationships are more evident (i.e. headers vs. text, and distinguishable levels of headers), buttons are more prominent, and, overall, the visual hierarchy of information is clearer.

Redesign Principle #2: Navigation

A website’s navigation strongly affects how easily users can find and understand the information they are searching for. Being able to distinguish navigation from other information is important in the same way. Because the navigation system is essentially roadmap of the different information and pages contained within a website, it is particularly important to the success of a user’s experience.

In order to provide a clear and simple path to other pages and places on a website, it is crucial that the navigation is easily identified and accessed by users. In order to accomplish this, we looked at the current state of the navigation in the DOM and made changes that we believed would help the navigation stand out from the rest of the page and be easier to use. First, we adjusted the font, size, and color of the navigation text, as these adjustments made it easier to distinguish from other elements on the webpage. Moreover, we wanted to ensure navigation was easier to use by allowing users to see what navigation link they were currently hovering over. We checked the navigation link tags in the DOM and added new styling to each of these linked elements to handle when a user hovers over a specific navigation link.

With these transformations, the visibility and function of the navigation system improves, making it a better sitemap for users.

Redesign Principle #3: Styling

The style — which includes font typeface, coloring, background images, as well as usage of screen space — can greatly impact a user’s ability to understand and appreciate the information on a webpage. To improve this, we tackled several commonly found stylistic issues on webpages:

- Distracting and unnecessary background imagery and/or coloring
- Inconsistent use of font types and colors for both the headers and body text
- Unresponsive or preset div widths that result in underused screen space

Addressing these issues was relatively straightforward as we simply removed background imagery or coloring (and sometimes applied the coloring to the styling of other elements on the page, like buttons, in order to keep within the realm of the original creator’s theme). We found that in the kind of websites that needed redesigning, background images were almost never used effectively, and often times made viewing information on the page very difficult. Next, we standardized fonts and colors to Times New Roman and Black. Lastly, we made the page width responsive by setting it to scale with regards to the size of the view width in order to make use of the users screen space so that more information can be viewed simultaneously without having to scroll down excessively. Poorly designed websites often set arbitrary limits on how much room there is for content in order to show off things like flashy background images, but which can make dynamically generated content, like growing tables, harder to parse because they now require unnecessary scrolling and manipulation of the page.

Through applying the algorithms and techniques we described above that fall within the scope of the three redesign principles — hierarchal information, navigation, and styling — Webpage Handyman effectively transforms and re-visualizes individual webpages through a simple user interface that will be discussed in the next section.

RESULTS

Webpage Handyman User Interface

The interface that users will interact with was designed to be very user friendly and straightforward. In order to use Webpage Handyman, a user simply has to enter in the URL of the page that they would like to redesign into the text box and click “Redesign” (see Figure 1). The user can then choose to download the new source code (which includes the HTML, CSS, and all relevant and necessary resources bundled into a zip file) or preview the new webpage directly on our site. The preview appears in an iFrame below the buttons.

Additionally, we included “About” (see Figure 2) and “Contact” (see Figure 3) sections so that the user can gain a basic understanding of how the tool works and reach out to us with any questions or feedback.

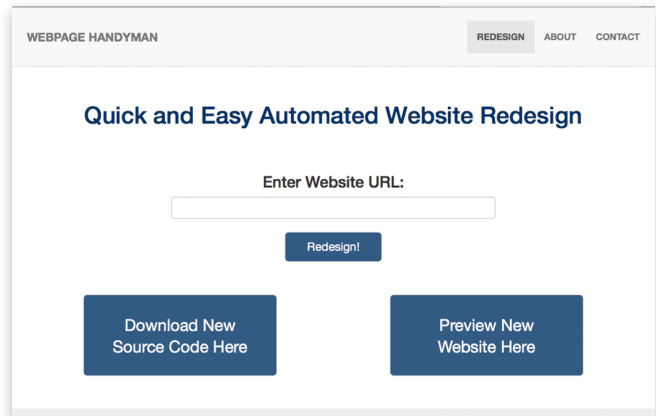


Figure 1. Webpage Handyman UI Redesign Section.

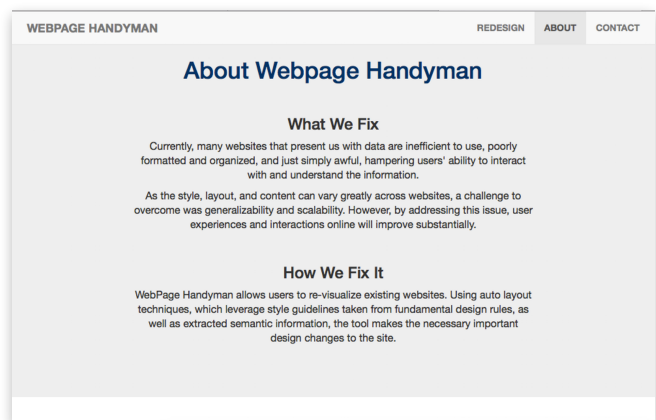


Figure 2. Webpage Handyman UI About Section.

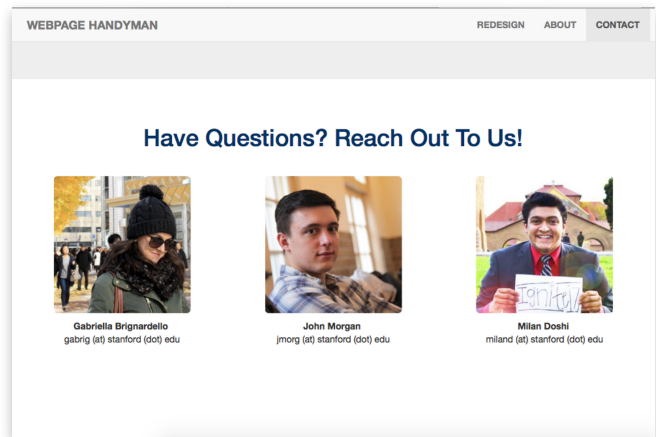


Figure 3. Webpage Handyman UI Contact Section.

Webpage Handyman Demo Transformations

In order to demo our tool, we identified several websites that could greatly benefit from application of our design principles. We also created our own poorly designed site as a way to show all of the implemented transformations in one place.

The original version of our full feature demo site (<http://bit.ly/full-feature-demo>) has issues with: background coloring, hierarchal information (i.e. header styling and

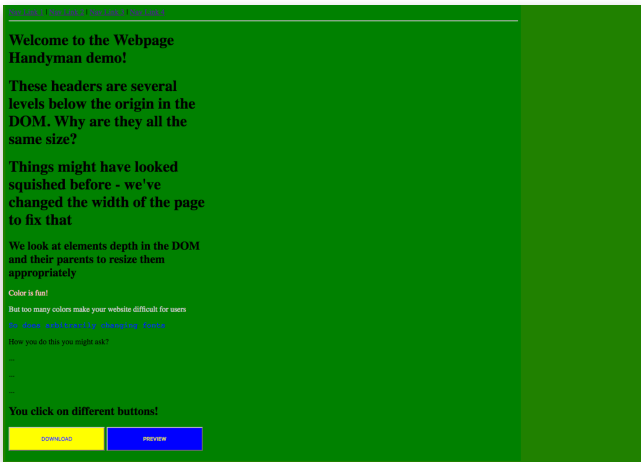


Figure 4. Original Full Feature Demo Site.

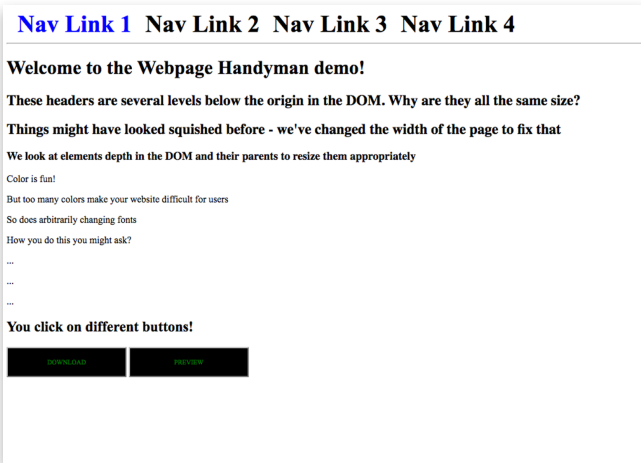


Figure 5. Redesigned Full Feature Demo Site.

differentiation), font types and colors, buttons, navigation, and inconsistent element widths (see Figure 4).

In our transformed version of our demo site, the headers have been restyled to show the depth and position of each element in the DOM, the background color has been eliminated but applied to the styling of the buttons (to maintain some of the thematic elements of the original site), the font type and color have been made consistent, the navigation has been made larger and more distinguishable in its purpose as a site roadmap, and the page width has been made responsive so that the content fills the screen (see Figure 5).

Two real-life websites that we transformed to show navigation and styling improvements were:

1. <http://games.stanford.edu/> (see Figure 6)
2. <http://www.dpgraph.com/> (see Figure 7)

In the re-visualized versions of both of these webpages, the navigation systems have been made more noticeable, the unnecessary backgrounds have been eliminated, the font type and color have been standardized throughout each of the pages, and the high level containers have been scaled so that the content fills the screen, making better use of the space (see Figures 8 and 9).

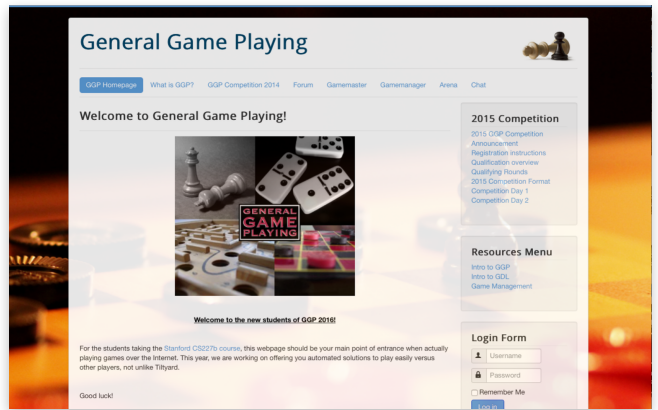


Figure 6. Original Stanford Games Demo Site.



Figure 7. Redesigned Stanford Games Demo Site.

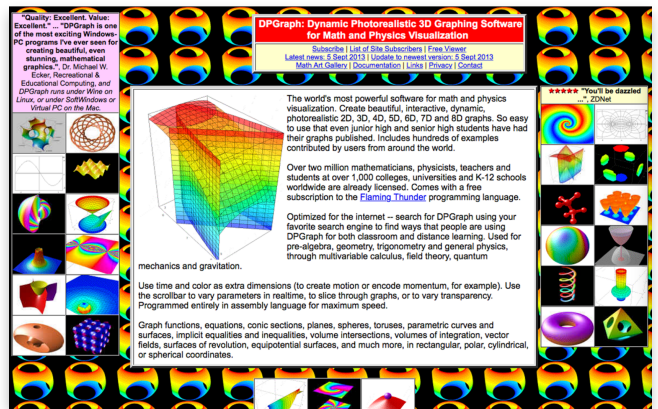


Figure 8. Original DP Graph Demo Site.

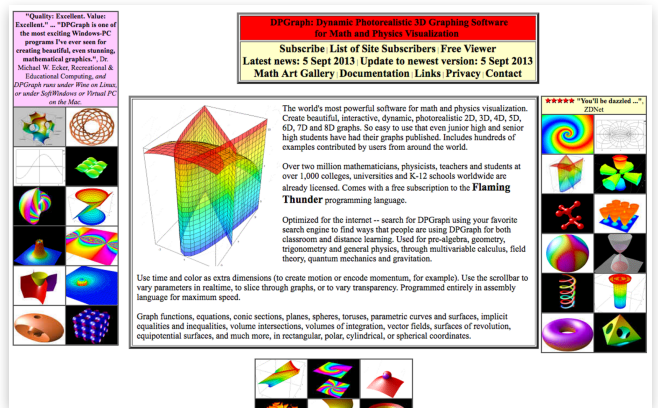


Figure 9. Redesigned DP Graph Demo Site.

DISCUSSION

While much of our explanations have been focused on user benefits, Webpage Handyman was designed with the intention of helping website creators as well. Through our tool, both website creators and users can learn about the visualization principles that we have selected as most important in web design and better understand how one should apply them. Furthermore, we heavily emphasize the importance of assisting users with quickly finding the information that they are seeking and/or making it easy for them to browse. For this reason, it is crucial that webpages maintain simplicity and clarity in their designs and layout.

In compiling many web design principles, we have also shown an important application of data visualization guidelines to the web. Our framework to parse data is also a useful extension of similar work, and can be used in the future to continue to mine information about websites.

FUTURE WORK

Although our tool was inspired from ideas and principles taken from prior works and online resources, the webpage parsing and auto layout framework that we have built is the most novel and impressive aspect of our tool. With this framework in place, there are a number of additional features that can be added to Webpage Handyman.

The most important, but most challenging of these is layout, which includes:

- Better identifying patterns
- Applying natural language processing to identify the importance of information, and then applying layout techniques based on a heuristic.
- Further improving utilization of space by incorporating strategic whitespace
- Properly partitioning content on the page
- Incorporating the “F Layout”, Rule of Thirds, the Golden Ratio (of “Divine Proportions”), and/or the Gestalt Design laws
- Automating content to conform to existing templates and systems if users desire, like Flexbox or Bootstrap

While there are many moving parts to implementing the layout redesign principle successfully, it can have dramatic effects on users’ experiences online. If a webpage can be reorganized and/or partitioned in an aesthetically pleasing and perceptually appropriate way, users will be able to more effectively view and appreciate the site’s information.

Another extension that we would be interested in pursuing is intelligently identifying and applying color schemes. This could be done by sampling colors on the page and comparing to existing, tested color schemes, and then applying transformations based on this scheme and existing patterns of colors. We also might want to better our heuristics for the elimination of color. For example, it might make sense to have a “Submit” button be green and a

“Cancel” button be red, which is something that our site currently would eliminate.

Furthermore, a huge improvement for users would be redesigning entire websites instead of just the page entered into the tool, which is currently all Webpage Handyman is capable of. Generalizing this would just take time to traverse internal links within a page and then send the user back a zip file containing the HTML, CSS, and other relevant and necessary resources for the entire website.

Lastly, since we offer so many different features, it would be ideal to allow users to toggle on and off particular redesign principles. This addition would be relatively simple to implement and would have important, positive effects on the user experience as it would give users the opportunity to customize the automated webpage redesign process to better meet their specific wants and needs.

We hope to continue to expand data visualization techniques outside of the traditional scope of the field in ways that improve the increasing amount of unstructured information users deal with on a day-to-day basis. While most work in data visualization focuses on information that conforms to scaffoldings which make the data easy to recognize and plot, web design poses an interesting challenge unlike what is seen in other visualizations because of its extreme variability.

REFERENCES

1. Agrawala, M., Li, W. and Berthouzoz, F. Design principles for visual communication. *Communications of the ACM* 54, 4 (2011), 60-69.
2. Lok, S. and Feiner, S. A survey of automated layout techniques for information presentations. *Proceedings of SmartGraphics* (2001).
3. Nielsen, J. *Designing web usability: The practice of simplicity*. New Riders Publishing (1999).
4. Kumar, R., Satyanarayan, A., Torres, C., Lim, M., Ahmad, S., Klemmer, S.R. and Talton, J.O. Webzeitgeist: Design mining the web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM (2013), 3083-3092.
5. Laja, P. 8 Effective Web Design Principles You Should Know. *ConversionXL*. <http://conversionxl.com/8-universal-web-design-principles-you-should-to-know/>.
6. Web Design Principles Of Successful Websites. *Cleverism*, 2015. <https://www.cleverism.com/web-design-principles-successful-websites/>.
7. Cao, J. The 5 Building Blocks of Visual Hierarchy in Web Design. *The Next Web*, 2015. <http://thenextweb.com/dd/2015/04/30/the-5-pillars-of-visual-hierarchy-in-web-design/>.
8. Friedman, V. 10 Principles Of Effective Web Design. *Smashing Magazine*, 2008. <https://www.smashingmagazine.com/2008/01/10-principles-of-effective-web-design/>.