2021

The Congressional Database: Designing a Web Application Using an HCI Approach

Liam R. Juskevice

The Congressional Database:

Designing a Web Application Using an HCI Approach

An Honors Paper for the Department of Computer Science

By Liam R. Juskevice

Bowdoin College, 2021

Abstract

The activities of the United States Senate are a topic of interest for researchers and concerned members of the public alike. Websites such as *GovTrack* and *Congress.gov* allow people to research specific bills among many other offerings. However, they have significant weaknesses regarding their ease of use and the way they organize and store data. The Congressional Database Project aims to provide an intuitive user experience navigating government data while storing the data in a consistent database. We approach this project from an HCI perspective in order to determine the best ways to improve the user experience. We have conducted a qualitative user study to test the effectiveness of our design and identify potential areas of improvement. This paper provides an in-depth overview of the design of the Congressional Database on the front end and back end. It then explains the methodology of our user study and discusses the implications of its findings.

Keywords: *U.S. Congress, congressional voting, congressional data, relational databases, web application development*

Acknowledgements

This thesis was only made possible by the guidance of my advisor Professor Mohammad T. Irfan and Professor Sarah M. Harmon. Dr. Irfan has guided me throughout my past two years spent working on this project. He has helped me countless times with this project and in many ways outside of it as well. Dr. Harmon's guidance and advice over this past semester has been invaluable. Many elements of this thesis, such as the design and conduction of our user study would not have been possible without her advice and guidance. Thank you, Dr. Harmon, your support has allowed me to take this thesis and project to the next level. Thank you both for helping me make this thesis into a reality.

Introduction

As technologies have advanced in the information era, so too has the way we engage with our government. One relatively recent development to that effect has been the movement for open government data (OGD). OGD refers to authoritative data collected and maintained by public entities that is unrestricted in its use and distribution (Wilson & Cong, 2021). The potential benefit of OGD initiatives include greater transparency and opportunities for public engagement. However, it is debatable whether these benefits have materialized in real world implementations of OGD. While it has become increasingly common for government bodies throughout the world to publicly release certain data, that act alone does not guarantee the aforementioned benefits. As Jannsen et al. (2012) point out, the value of OGD is created in its use. If an open data source is so difficult to use that only those with exceptional knowledge and time can feasibly make use of it, the benefits may become negligible.

With these issues in mind, the open government data on the United States Congress emerges as a flawed but usable example of such initiatives. In 2009 the Senate started publicly publishing data on votes cast in XML format (Hendler, 2009). This format is relatively easy to parse through for any interested party with some degree of programming knowledge; however, it is not something the average constituent could readily make use of. While public facing data surrounding the US Congress is readily available, the issue of presentability for general public and research use remains. Since the activities of the United States Congress impacts the country's entire population and even those abroad, open data attracts considerable interest. As a result, multiple websites and APIs have emerged to help the public, researchers, journalists, and more to work with and engage with the open data on Congress' activities.

These websites offer extensive information to help users navigate the activities of Congress. However, they tend to difficult to navigate and inefficient for analyzing information in bulk. For instance, *GovTrack.us* offers detailed insights on specific legislators when you select their names. If a user is not interested in a particular senator and instead wishes to know the makeup of a previous session of Congress, it would be incredibly difficult to figure out with *GovTrack*. If one wants to directly work with the open data on the United States Congress, they must make use of the API of a different party such as the ProPublica Congress API. These data sources certainly make it easier to analyze large amount of bills, roll calls, etc. but are often light on information compared to what a website like *GovTrack* has to offer. As a result, someone

interested in analyzing trends in congressional bills will likely need to gather data from multiple sources with a notable risk of contradictory information.

The Congressional Database

We have created the Congressional Database intending to address many of the issues previously identified with other sources of congressional data. It is a database that holds current and historical information on the United States Senate. This information includes roll call votes on bills, committees and their membership, bills, amendments, and more. At this time our data spans from the current session of Congress (117th) to the 93rd Congress (1973-1975). Within the years of Congress covered we aim to make the database as comprehensive as possible while limiting extraneous information.

This database was created in tandem with a website that will be hosted on CongressionalDatabase.com to present the database in an accessible and intuitive manner. Navigating the database's information on the website is made simple as the website is organized in a similar manner to the database itself. It grants issues, members, committees, and roll call votes, the primary tables the database is organized around, separate pages. The pages are designed with efficiency and minimalism in mind to make finding the necessary information easier for researchers and interested members of the public alike. Additionally, the website serves as a host to present our team's research findings to further engage the intellectual community.

We consider the design of the Congressional Database to fundamentally be an HCI problem. The main question is how can we provide the most satisfying and effective user experience possible for both the researchers and public we are targeting for this website? For our purposes we are defining "satisfaction" as the user getting what they desire from our web application, finding use of the interface to be pleasant, and experiencing no duress in finding what they want. Our definition of "effective" is that the user can navigate the website to find what they want as efficiently and easily as possible. HCI researchers have considered the best approaches for designing web pages since the early years of the internet (Comber, 1995). While the field has evolved considerably since then the goal of usability remains integral to HCI research. However, factors such as aesthetics and first impressions can be just as important to user satisfaction (Sørum et al., 2012). While there are proposed measures to evaluate these

factors objectively, a user's impression of an application's design can vary considerably depending on the individual. As a result, qualitative user studies are an important tool for HCI research (Blandford et al., 2016). Said qualitative studies are particularly effective for denoting areas for future work. As the Congressional Database is still undergoing development this research method proves useful for our purposes.

In this paper I will first explore the advantages and disadvantages of the sources that provide similar services to the Congressional Database. I will then outline and explain the design choices of both the database and website. Finally, I will explain the methodology and discuss the findings of the user study we conducted to evaluate the project.

Related Work

The utility of the Congressional Database is in most regards not unique to it. It is neither the only database for the United States Senate nor is it the only website that presents information on the topic. Furthermore, if those sources did not exist it would have been far more difficult to construct the Congressional Database. APIs like the ProPublica and Sunlight Congress have similar databases to our own, but neither directly connect to a website to allow users to interact with their data. An interested researcher that knows how to utilize APIs could likely find the information they need through these sources. However, this assumes the interested party knows exactly what specific information they are searching from the outset. These APIs output the requested information from queries and cannot be readily browsed. By building a website in conjunction with the database this allows for the information of the database to be easily navigable for less directed use while still allowing for specific queries. However, even as a website there are others that offer similar services. We will now take a closer look at the most relevant competitors.

GovTrack.us

The most prominent and similar competitor to what the Congressional Database offers is *GovTrack*. *GovTrack* is a website that presents information on both chambers of the United States Congress. Its focus is on allowing its users to "track" the activities of the current Congress, particularly the activities of certain congresspeople and the bills being proposed. It

allows users to create accounts, which makes selecting certain Congress members and bills to track the activities of easier. If a user navigates to a particular bill or member, they will find a large amount of information. In the case of a bill, it will generally include a timeline of the bill's progress, cosponsors of the bill, press statements related to the bill, and more. If a user were to look at a specific congressperson's page, they will find a very brief biography, an ideology score, enacted legislation, and more. Ultimately, *GovTrack* specializes in offering a wealth of information on specific Congress members or bills that a user takes interest in. The site is intuitive to use if a user is interested in the representative of their own home or the most recent contested bills. It becomes considerably more difficult to navigate if a user is hoping to find information outside of that.

As *GovTrack* specializes in offering large amounts of information on individual members and bills, its weakness shows when considering this sort of data in bulk. It is possible to browse the name of current members of Congress or bills in a list form, but figuring out how to do so would not be readily apparent to most users. Furthermore, these lists consist of small text and images that the user must continuously scroll through if they do not manually narrow down the search. *GovTrack* evidently does not focus on considering member or bill data in bulk and seem to have expended minimal effort in making such tasks easy to accomplish with their site. Furthermore, it is rather difficult to use the platform to search for data on the activities outside of the current session of Congress. For instance, members of Congress are organized by those who are in the current session of Congress and those who were seated in a previous session. If one wanted to figure out the makeup of the 116th Senate using *GovTrack* it would be a rather difficult task.

The Congressional Database does not currently offer as much in-depth information on specific bills or senators as *GovTrack* can. Instead, we offer significantly higher usability regarding navigating the website and engaging with large amounts of data. Whether a user wants to navigate through members, bills, roll calls, or committees they will find information presented in a card format that allows users to narrow down their search and to click on them for more specific information. This allows users to easily navigate large amounts of information more readily than *GovTrack* allows. We also make it simple to explore different sessions of Congress, as there is a selection form at the top of each page that allows users to change the session of Congress with ease.

Congress.gov

Another similar competitor to our services is Congress.gov. This is a government run website that provides information about the United States House of Representatives and Senate. The website primarily contains information on the legislation, nominations, and amendments proposed in Congress. The website also tracks other documents used in Congress such as communications and committee reports. The website has records on legislation circulating in Congress up until the 93rd Congress. One can find information on other documents such as congressional records as far back as the 77th Congress. Overall, it is the most comprehensive online source of legislation and other documents from Congress. Much of our data is scraped from this resource.

If one navigates to the page of a particular bill, they will see a card showing the "overview" of the bill and can navigate to find more specific information below. This specific information includes a summary, full text of the bill, cosponsors of said bill, and more. To the right of the overview there is a link to CBO estimates of the legislation and another link to view the full list of subjects associated with said bill. We believe this general format to be effective. As a result, we have implemented a similar layout when displaying information about issues on our website.

Figure 1. The top of the home page of Congress.gov as seen on desktop. While this website can provide users with a wealth of information, it may not be immediately apparent where to find what they are looking for. The home page is information dense with little visual guidance as to where the users should navigate first.

While Congress.gov offers readily digestible information on specific legislation, it is not always easy to locate the specific legislation one had in mind. As Figure 1 shows, the home page is dense and text heavy. Most of this text is of the same size with only the red headers having a larger size. There is no clear focal point to the page, which makes navigating the website without prior experience difficult. If one was interested in locating specific legislation, they would have to utilize the search bar found on the top of the page. This search function works well if the user knows the exact bill they are looking for, but they would need to navigate the dense advanced search options or filters otherwise.

The primary issue with Congress.gov is that the vast amount of information it holds comes at the expense of usability. The home page does little to visually differentiate the categories of data it links to. This website also limits its scope to congressional issues, documents, and members. Committees are only relevant in how they are assigned to a given bill. Additionally, if one wanted information on a particular roll call vote they are instead directed to Senate.gov. In comparison the Congressional Database provides a more streamlined experience for finding the relevant page to search for issues. Also, the Congressional Database contains

information about committees and roll calls on the website where Congress.gov relies on external links for that information.

Design

The Congressional Database is designed to have a similar structure on the backend and frontend. This design choice both supports scalability for the project and helps make the user experience on the frontend as straightforward as possible. The primary goals of our database design are consistency, accuracy, efficiency, and scalability. While our database currently only contains data on the United States Senate, its design could easily incorporate data on the United States House of Representatives as well. The frontend of the website is designed to preserve the intuitive layout and efficiency of the database. Beyond mirroring the database's goals the frontend design has the additional aims of pleasant aesthetics and ease of use. We will now explore the specifics of the Congressional Database's design.

The Database

The Congressional Database is a SQL relational database that is primarily organized around the *issues*, *committees*, *members*, and *rollCall* tables along with the relationships among them. We consider these four tables to represent the most important components of Congress. The issues table represents all documents that senators propose and vote on. The *committees* table contains data on the primary committees of the Senate. The *members* table represents individual senators. Finally, the *rollCall* table provides data on votes conducted during a given session of Congress. We will now do an overview of the particulars of each table structure.
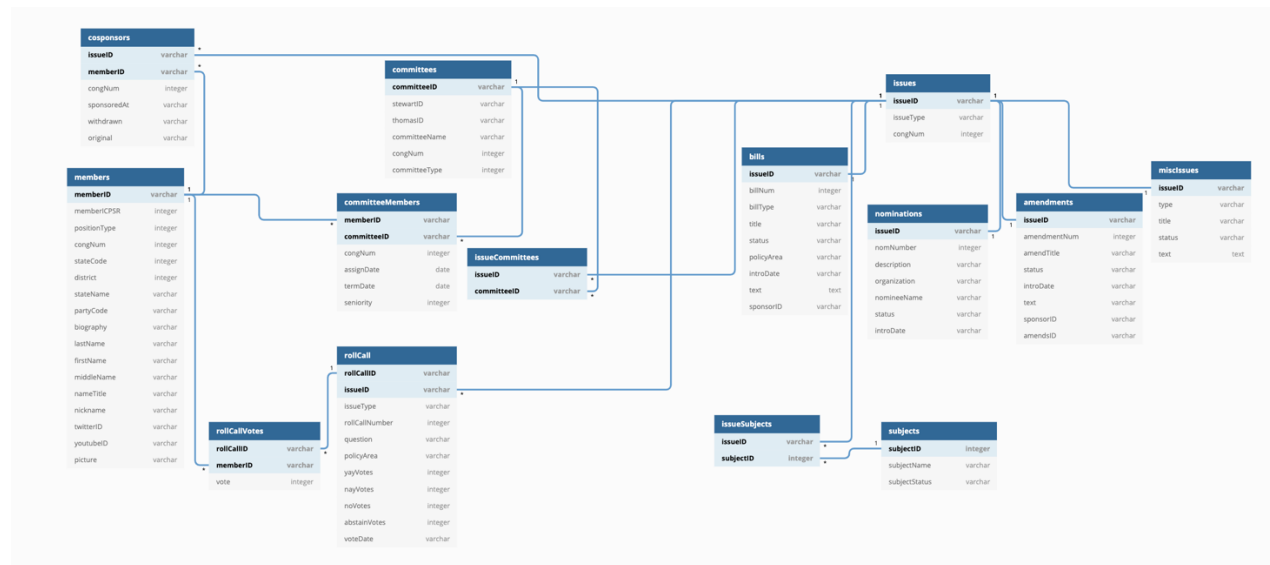
Figure 2. A bird's eye view of the database schema detailing the tables and the columns within them.

Issues Table

The *issues* table itself does not have too many columns. It simply holds a unique ID, its type, and the meeting of Congress it was put forward in. The bulk of the data associated with a particular issue is found in their subtype table. The subtype tables are *bills*, *amendments*, *nominations*, and *miscIssues*. The *bills* table corresponds to all legislation proposed by congresspeople. The *amendments* table contains amendments to bills (or other amendments) proposed by senators. The *nominations* table contains data on nominations the president put forward for the Senate to consider. Finally, *miscIssues* is a fallback table for issues voted on during roll call that do not cleanly fall into the other three categories. These miscellaneous issues largely consist of motions and treaty documents.

These subtypes are given distinct tables to avoid the large number of null entries that would populate the database otherwise. For instance, the *amendments* table requires a reference to the issue it is amending. This column is not directly applicable to any other type of issue and would be left null as a result. By giving each issue type its own table we can fashion the columns to fit their respective needs.

In addition to issues and their subtypes, we also have a table dedicated to the subjects of each bill. The Congressional Research Service (CRS) determines the primary subject for each bill, which is called the policy area. In addition to that they assign a variable amount of more

specific subjects such as "income tax deductions" or "animal and plant health." This information on specific subjects is stored in the *subjects* table. This *subjects* table holds the name of every subject, their ID, and whether it is a current or historical subject. Before 2009 the CRS used a much larger list of subjects, but shortened the list to the 1004 subjects currently used today (cite Congress.gov). The Congressional Database stores the subjects associated with any given bill in the *billSubjects* table. The *billSubjects* table presently has the most entries of any table in the database due to the large number of subjects associated with any given bill.

Committees Table

Committees play an integral role to shaping public policy. They have the power to block certain bills from every reaching the Senate floor and convincing the Senate to take up certain bills. Committee assignments are also an important component in understanding how particular senators influence policy and advance their careers (Marshall & Wolpe, 2018). The *committees* table holds the primary committees for a given session of Congress. It should be noted that committees have multiple entries for each session of Congress they exist in. While data on a particular committee will rarely change between sessions, the composition of their membership inevitably will.

Like the other tables in the database, the *committees* table currently focuses on the committees of the Senate. However, it also contains data on certain house committees, as a house committee may be assigned to a bill the Senate considers. For the purposes of this database, leadership positions such as the Majority Leader are also treated as committees. While these positions are not technically committees, they can be represented using the same columns.

The *committees* table is related to the *committeeMembers* and *issueCommittees* tables. The *committeeMembers* table contains data on the committee assignments for each senator. It is related to *members* table using the memberID key. Additionally, it holds information on the member's rank within their assigned committees. The *issueCommittees* table indicates which committees are assigned to a particular bill. It is related to the *issues* table through the issueID key. These two relationship tables can serve as a lens to understanding the committee's impact on policymaking.

Members Table

The *members* table holds the essential information on each senator including their party identification and basic biographical information. Besides the *committeeMembers* table *members* is related to the *rollCallVotes* and *cosponsors* table. The *rollCallVotes* table records the way in which a member voted on a specified roll call. It is also related to the *rollCall* table, which is connected to *rollCallVotes* using the rollCallID key. The *cosponsors* table stores the variable number of senators that chose to cosponsor a specified issue. The primary sponsor themselves is not stored in this table as "sponsorID" is a column in the *amendment* and *bills* page. The *cosponsor* table is related to *issues* with the "issueID" key.

RollCall Table

The roll call is arguably the most visible Senate activity. The votes of the senators determine whether a proposed bill will become a public law, assuming the president does not veto it. High profile nominations such as Supreme Court appointments also attract significant public interest. Furthermore, the voting behavior of senators is an important aspect of how the public and researchers understand a senator's ideology. The *rollCall* table holds data on all votes cast in a session of Congress regardless of their profile. Each entry holds data on the issue being voted on, the vote count, the outcome of the vote, and more.

We have added a column for the "policyArea" of the roll call. While only bills have a policy area column, we have approximated a policy area for other issue types. The policy area for amendments in a roll call is based off the bill they are amending. For nominations we use the organization the individuals are appointed to. For the rare miscellaneous issues we simply refer to the policy area as "Treaty" or "Miscellaneous." This column's existence is an example of how we designed our database to provide an easily navigable user experience on the frontend. The "policyArea" column is shown as part of the card made on the frontend for each roll call. This information helps to quickly convey what a particular roll call is about to a user navigating the page.

Website Design

In designing the website we opted for a minimalistic approach. We ensured that the text displayed on screen is strictly limited to what the users need to help navigate the website. It is only when a user selects a specific member, roll call, issue, or committee that they will see a large amount of text to convey the data associated with that entry.

To further support this minimalistic design, we gave the site a dark bluish gray background contrasted with white text. This color scheme is similar to the "dark mode" that many websites and apps have implemented in recent years. The "dark mode" is believed to be less straining for the eyes in dim light conditions. It also reduces power consumption. However, this benefit is only applicable to certain types of displays such as OLED (Pedersen et al., 2020). For our purposes the "dark mode" design further differentiates the Congressional Database from its closest competitors aesthetically while still maintaining an easily readable high-contrast color scheme. This color scheme is kept consistent in every page of the website to maintain visual continuity.

The pages of the website are organized in a similar manner to the database. The website's home page and header prominently display links to the four main pages of the website, which are the Members, Committees, Issues, and Roll Call pages. Each of these pages corresponds to one of the primary tables of the database. We will now explore the particulars of the designs of the home page and the four primary data pages.
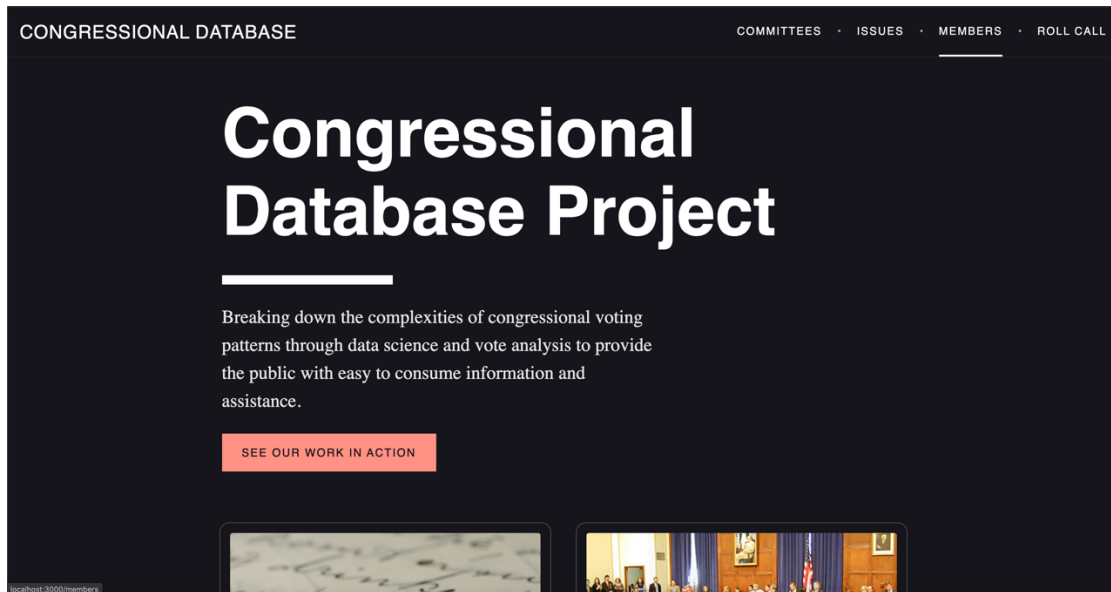
Home Page

Figure 3. The top of the home page users are greeted with upon first entering the website. Upon scrolling down there are four images with text below that direct the user to the respective data pages.
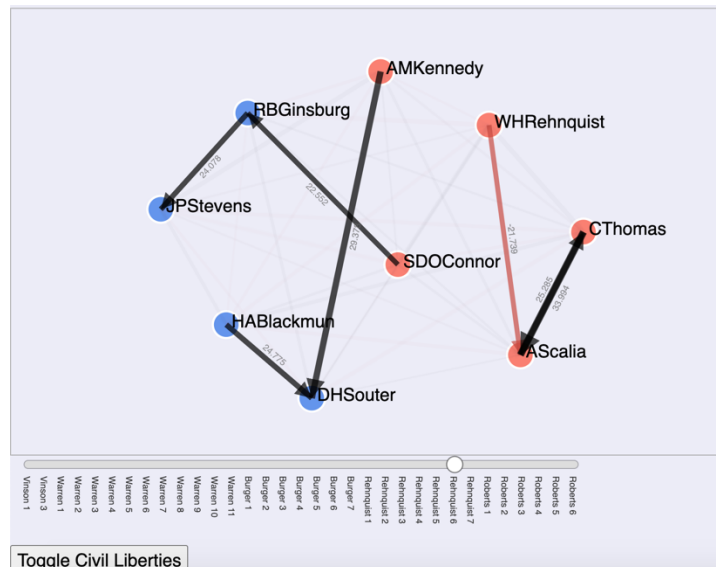

Figure 4. An example of the visualizations that will eventually populate our research page. Pictured here is an influence network of the United States Supreme Court, which allows the user to move individual nodes around to get a clearer view.

Upon entering the website, the user is shown the title of the Congressional Database along with a brief explanation of the website's purpose. This explanation is intended to quickly orient new users and help them understand how to navigate the website. Researchers have found

that the title and the most prominent text of a visualization are integral to its memorability (Borkin et al., 2016). Using these findings, we display the title and explanation text first in order to clearly convey the purpose of our web application.

Immediately below the explanation is a button that prompts the user to "see our work in action." This button will ultimately lead to our dedicated research page. The research page is unfortunately not fully implemented currently. If one clicks on that button now, they will be led to a page that shows an interactive influence network of the Supreme Court as seen in Figure 4. This visualization is an implementation of the Irfan and Ortiz model of linear influence games using behavioral data (Irfan & Ortiz, 2014). The visualization shows the sitting justices of each Supreme Court and the most significant positive and negative edges between them. Ultimately, this influence network will become one of the many visualizations that our website holds to help present our research findings.

Below the "see our work in action" button are four images with text below them corresponding to the Members, Committees, Issues, and Roll Call pages. Upon hovering over one of the images it will darken and text saying "view page" appears in the center. Clicking these images directs the user to the data page they selected. In addition to these links, they can also access the data pages using the header. This header appears on the top of every page of the site. This header, as seen in Figure 3 links back to the home page when they click Congressional Database on the left or the selected data page if they click one of the names on the right. Through this set up the user can readily find the page they are looking for in two different ways from the home page.
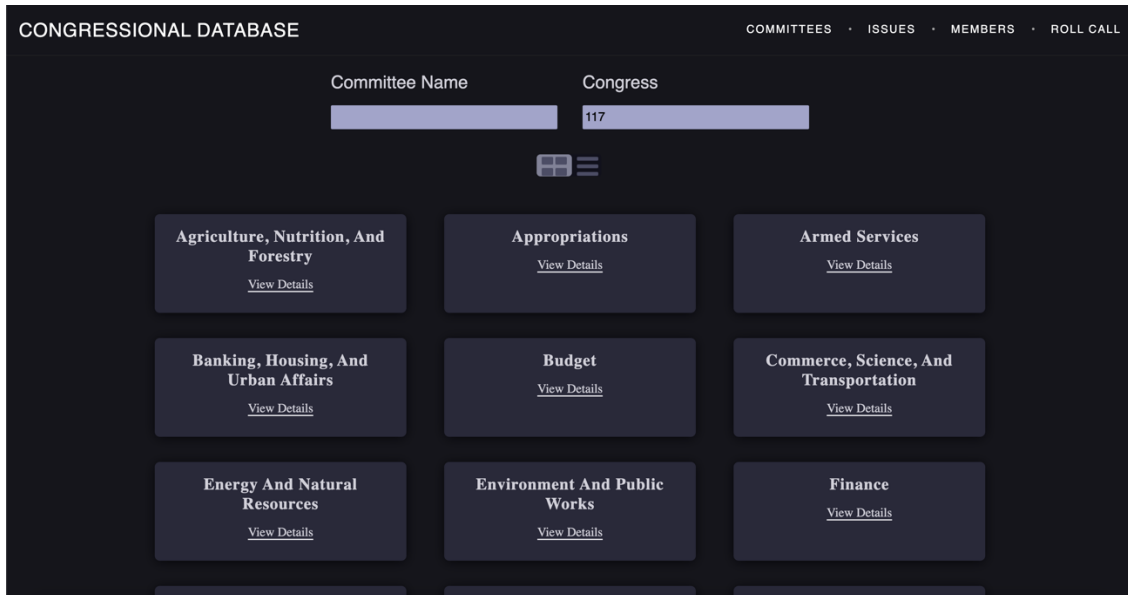
Committees Page

Figure 5. The design of the Committees page. Users can search for specific committees by entering a search into the "Committee Name" box and select a session of Congress using the dropdown menu.
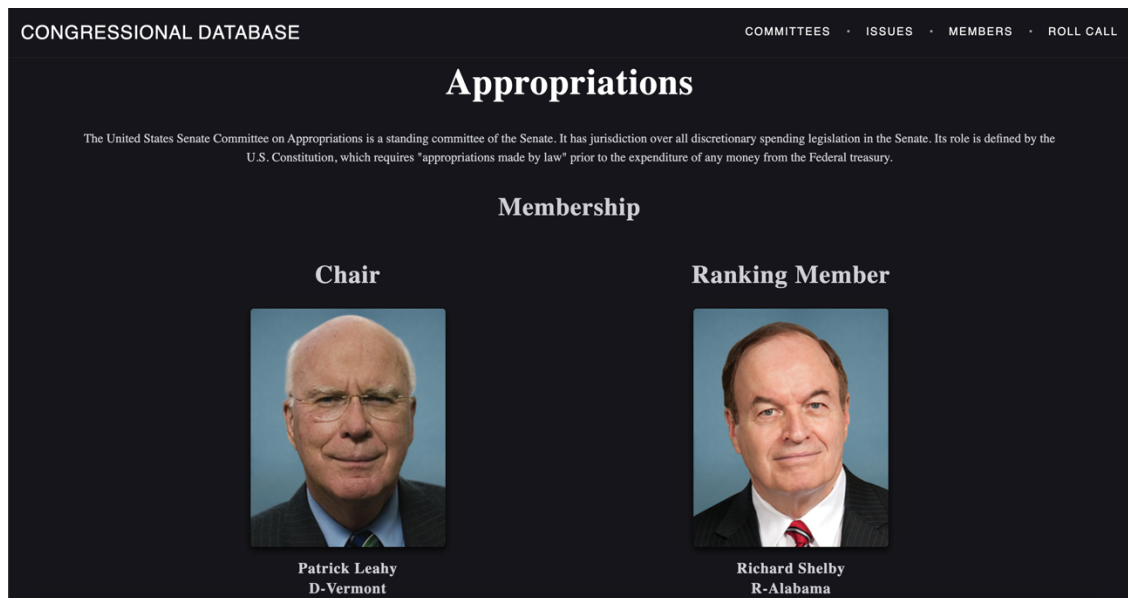

Figure 6. The top of the page for the Appropriations committee of the 117th Congress on desktop. Pictured here is a brief explanation of the committee's purpose along with the chair and ranking member.

The Committees page is designed with a simple card format. Every Senate committee for a given session of Congress is displayed in a card, which shows only the name of the committee

and a link to its page. As shown in Figure 5 there is a search bar above the cards where the users can input the name of any committee to quickly find it. To the right of the search bar is "Congress" which users can click to select from a drop-down menu to select a different session. The icons below that allow the user to switch between the cards view mode and a simple list mode, which could be easier for some to browse through.

Upon clicking the link to "View Details" the user navigates to the page associated with the committee they selected. As seen in Figure 6 the page summarizes the purpose of the committee and shows the membership for the specified year of Congress below. The committee chair and ranking member are shown at the top, and the rest of the senators are shown below ordered by rank and separated by party. The images of the chair and ranking member are 1.27 times larger than the images of the remaining members. This size discrepancy is intended to convey the importance of the committee chair and ranking member.
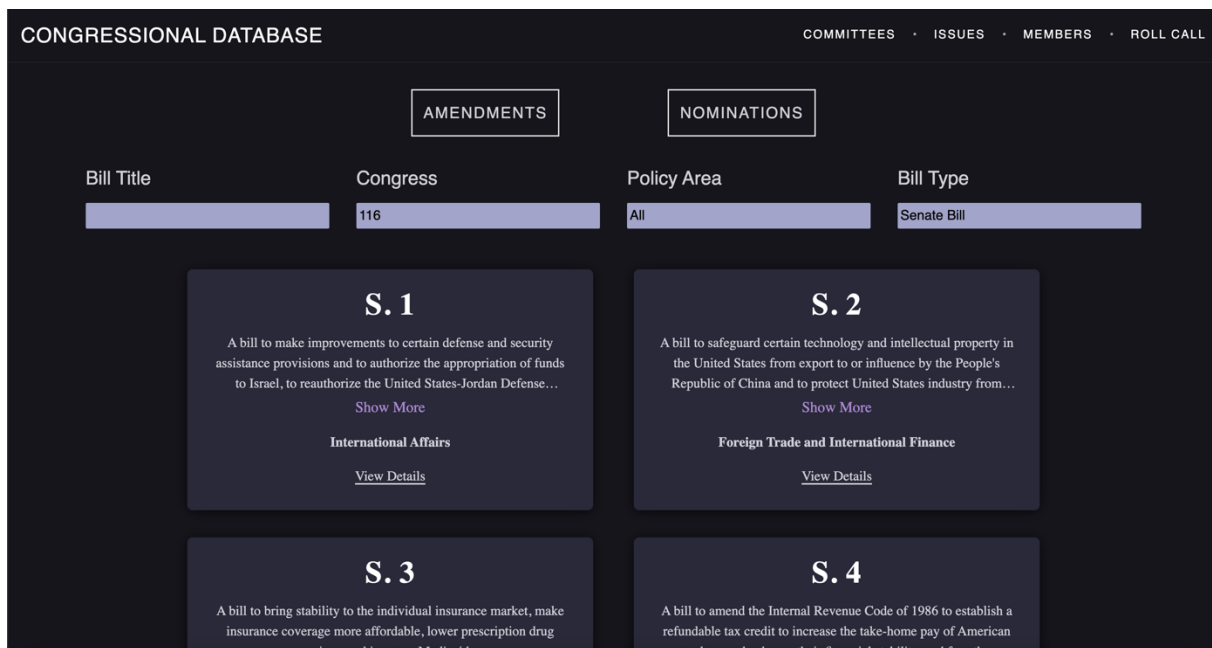
Issues Page

Figure 7. The top of the bills page for the 116th Congress. The amendments and nominations page have highly similar layouts. All employ the card format with variations in the text shown on the card depending on the issue. Amendments and nominations also do not have bill types, but nominations allows you to search by organization.

Figure 8. The top of the page showing the details for a specific bill. This includes the subjects associated with the bill, its full text, cosponsors, committees assigned to said bill, and any roll call votes.

The Issues page actually directs to three different pages corresponding to the primary subtypes of the *issues* table. We chose not to create a dedicated page for miscellaneous issues as they are far lower in number than the other three types. Presently we only have data for 32 different miscellaneous issues in total. In contrast, we currently have 102,909 rows in the *bills* table, 66,438 rows in the *amendments* table, and 37,817 rows in the *nominations* table. For our purposes we only deem the miscellaneous issues important when they are voted on by the senators.

All three of these pages use virtually the same layout shown in Figure 7, with minor differences due to the nature of the issue type involved. For instance, the bills page is the only one that allows the user to narrow their search by policy area. All Issues pages use a similar card layout to the Committees page, but the cards are larger to grant more space to convey the relevant information. Unlike committee names, the bill numbers themselves are not informative. So we needed additional information to help the user find the correct bill.

When the user clicks "View Details" they will then navigate to that issue's page as seen in Figure 8 to get more in-depth information on it. The design of these pages take inspiration from the Congress.gov format, particularly the tabs. These tabs allow the user to find their desired information about the bill without having to scroll through irrelevant information. The

individual pages for amendments and nominations have virtually the same layout with different subsections for the tabs. For instance, neither amendments nor nominations have a subjects tab. While the details of implementation vary depending on the type of data being presented, we attempt to keep the layout as standardized as possible.

Members Page
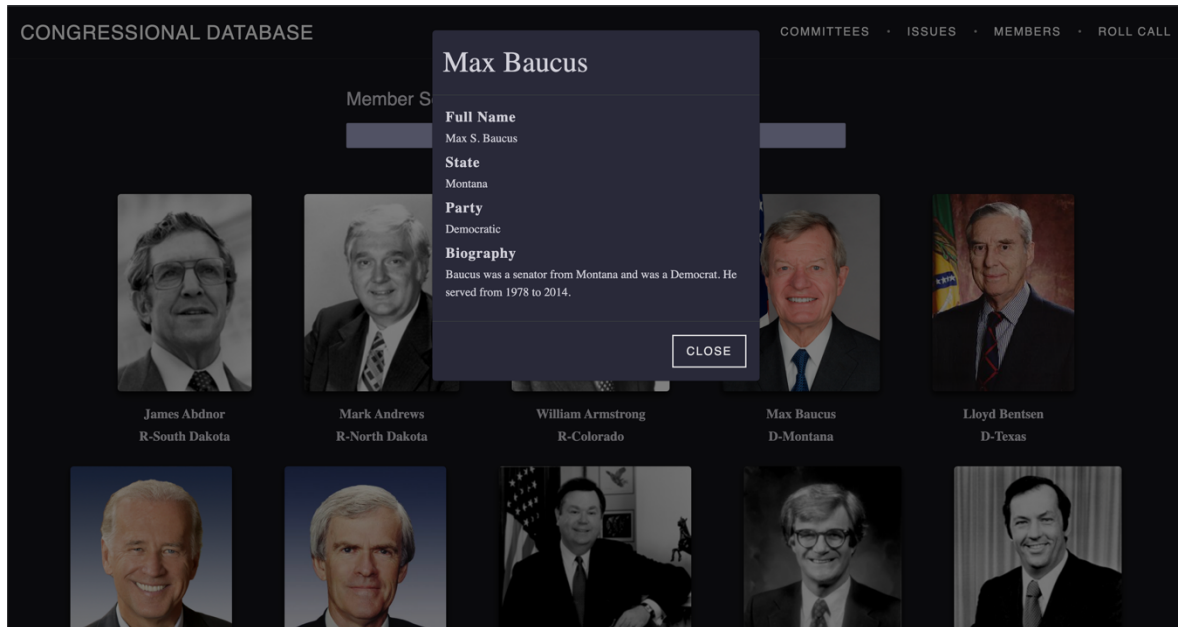
Figure 9. The Members page as seen when clicking on the image of a specific senator. Once can click on the close button or anywhere outside of the modal to return to scrolling through members.

The Members page is notable for using pictures instead of the card format seen in all other data pages. Unfortunately, there is presently no reliable method to assign unique and informative images to every bill passed in the Senate. Like the other pages users can search for the member's name and select different sessions of Congress from the drop down menu. The screenshot seen in Figure 9 displays the Senate delegation of the 99th Congress.

When a user hovers over a particular senator's image it will darken slightly, which indicates that it is clickable. Clicking on a senator causes a modal to appear giving more details about the specific senator as seen in Figure 9. For more recent senators we include links to their respective twitter accounts as well. We presently do not have dedicated pages for individual members. Since the data we want to convey about individual senators can be contained within the modal we deem unique member pages unnecessary.

Roll Call Page
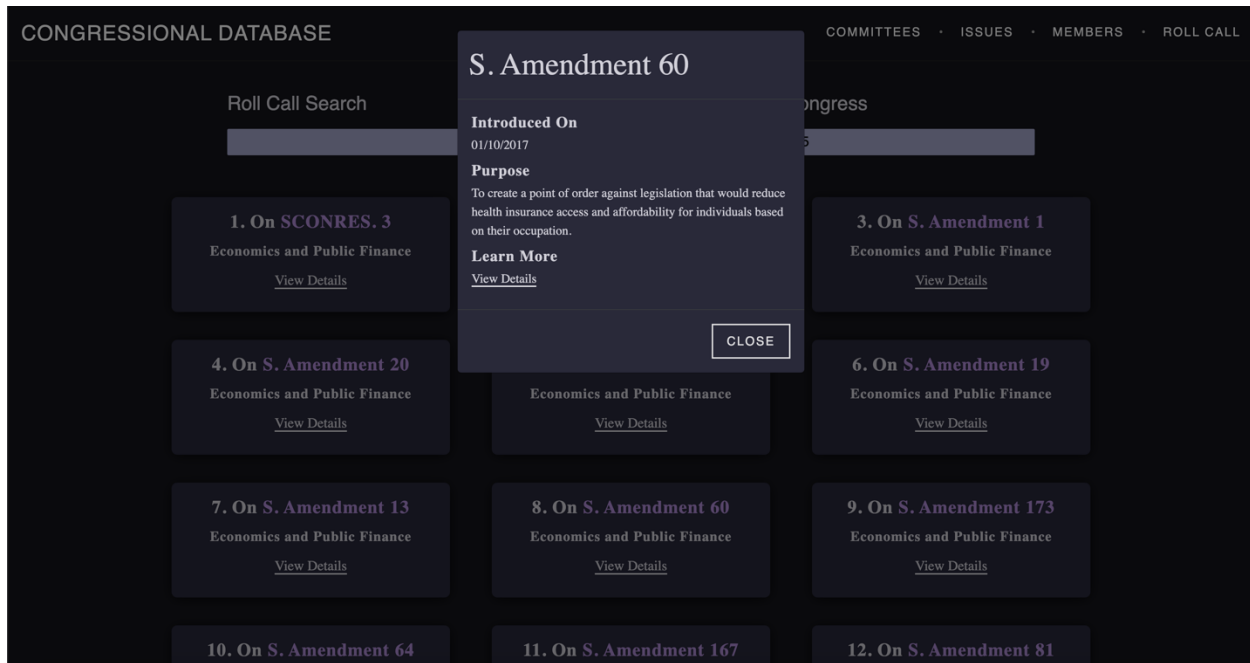
Figure 10. The Roll Call page upon clicking the name of a specific issue being voted on. A modal appears giving a brief overview of the issue and links to its page.

The Roll Call page employs a card format in a similar manner to the Committees and Issues page. The Roll Call page distinguishes itself by creating a modal for every issue the roll call votes on, which can be seen in Figure 10. We used purple text that darkens upon being hovered over to indicate a link to the modal. This choice implies the modal link is clickable. Additionally, this choice visually distinguishes the modal link from the white underlined text that most other links on the website are designed with.

Besides this feature the Roll Call page is formatted almost identically to the nominations page. Furthermore, upon clicking on "View Details" to navigate to a specific roll call the page format is nearly identical to that of a specific issue. The specific roll call page features a tab that summarizes the vote result and a tab that shows how each senator voted. Overall, the Roll Call page is designed to convey only as much information as is needed for users to find the specific roll calls they are interested in. We used this approach for the design of every page that displays content from our database on the website.

One could argue that the functionality of the Roll Call page could be incorporated into the Issues page. It is true that these tables are inextricably linked as a roll call is a vote on an

issue. We currently include a link to every roll call associated with a bill on that bill's page. However, solely relying on those links to convey the roll calls would have many pitfalls. Only a small portion of the thousands of issues put before the Senate each year are ever put to a vote. Solely including roll call as information in the Issues page would require additional visual indicators to show that said issue has roll calls associated with it. Furthermore, solely linking roll call votes to the issues that were voted on makes it hard to understand the timeline in which these roll calls were cast. Our Roll Call page shows every vote for a particular session of Congress in the order they were called.

Alternatively, one could remove all references to issues that were never put to a vote. It may be true that issues that were voted on attract more interest than those that do not. However, the issues that never get voted on can still provide interesting information. For example, one could look at the cosponsors of bills that may have little hope of passing at the time to understand the ideology of a senator. *S. 1129* of the 116th Congress would have established a "Medicare-for-all" national health insurance program if passed. This program was a popular topic for discussion in the 2020 Democratic debates and polls of the American public have found significant support for the proposal particularly among left-leaning individuals (Jones, 2020). This bill was never put to a vote in the 116th Congress, but the cosponsors data shows which senators are willing to publicly support the bill. Similar arguments could be made for the value of researching other issues that never received a roll call. The reasons a certain bill failed before it got put to a vote can be of interest as well. Minimalism is an important tool to help users navigate the website efficiently, but we want to encourage its use for a wide variety of inquiries.

Implementation

In this section we will explain how and where the Congressional Database obtained its data. Furthermore, we will discuss the technologies involved in creating the database and website itself.

Data Sources

We were able to make the Congressional Database as comprehensive as possible by utilizing multiple sources for data collection. By compiling data from multiple sources, we can

obtain a large range of historical data for every table. One challenge raised by this approach is that the data from these distinct sources are often formatted differently. However, regardless of the format of the data they largely contain the same information. As a result, they can be standardized as entries in our own database.

A prime example of how we obtain data from multiple sources can be seen in how we generate the *bills* table. Most information about each bill are derived from two sources depending on their age. Bills from the 93rd Congress up to the 114th Congress are sourced from the *Congressional Bills Project*. This website is no longer being updated, but it is still the best source available for data on historical bills. For the more recent bills (115th Congress through 117th Congress at the time this paper was written) we utilized the *@unitedstates* project. This source is constantly updating as new bills are introduced. Their code can be used to download data on every new bill and convert it to an easily usable json format.

Unfortunately, these two sources for our *bills* table contain only part of the data for each bill we are interested in. The rest needs to be obtained through web scraping. We obtain the full text of the bill along with subject data by scraping a particular bill's page in *Congress.gov*. Additionally, we get a more standardized version of the bill's status by scraping its page on *GovTrack.us*.

We obtain data on the other issue types in a similar manner. We obtain data on amendments from the most recent sessions of Congress using the *@unitedstates* project. For older amendments we resorted to scraping data from *Congress.gov* directly. Furthermore, unlike bills *Congress.gov* does not directly show the text of amendments. Instead, they often link to the page of the congressional record where the amendment's text can be found. Since the text of the amendments in the congressional record are often formatted the same way, we can obtain the relevant text using specially crafted regular expressions. Unfortunately, this method is not as consistent as scraping the bill text often is. This is especially true for amendments that amend another amendment. In these cases the amendment text is often prematurely cut. Data for nominations and miscellaneous issues are also obtained from *Congress.gov*.

Committees are another example where we need to use two different sources depending on the session of Congress we want data on. *Charles Stewart's Congressional Data Page* contains an excellent source for data on committees and membership from the 103rd Congress to

115th Congress. However, this source is no longer being updated. We use the *@unitedstates* project for committee data for the most recent sessions of Congress.

For older committee data before the 103rd Congress we create new entries based off of the names of committees assigned to each bill. The earlier Senate committees often use the same name, or serve the same purpose with a different name. In these cases much of our data from more recent sessions of Congress can be directly copied over. In rare cases where this is not possible we independently research the committee in question to add its data to our database. I personally wrote the brief descriptions of each committee's purpose by adapting the descriptions from the respective committee's website.

For roll calls we obtain the relevant data by scraping *Senate.gov*. This source has data ranging from the 101st Congress to the present. We are not currently aware of a reliable digitized source for roll call votes from earlier years. This source is reliable and comprehensive for the years of Congress it covers. However, scraping from this source is complicated by the need to have all the issues voted on in the database prior to scraping. These roll calls are what determine the miscellaneous issues that enter our database. Furthermore, this website often temporarily blocks our scraper when it is used too often within a short period of time.

Finally, for data on individual senators we utilize *VoteView* and *GovTrack.us* to obtain the relevant information. Incidentally, these sources have data that spans back much further in the past than our other sources. If we wanted to our database could contain information on every US legislator that existed since Congress was founded. However, we decided to keep our *members* table consistent with the other tables, which presently only have data spanning back to the 93rd Congress at the earliest. Additionally, after a certain point in the past we are no longer able to reliably obtain photos for every senator. Our web scraping for member data and that of other tables is accomplished using the Nokogiri library in ruby.

Technology Used

Our database is currently hosted on *MySQL*, which is one of the most popular and reliable SQL database systems. Since we are most likely going to use Heroku to publish our website, we may ultimately need to migrate our database to PostgreSQL. Both systems use SQL query language in the same manner most of the time, which means the required changes to the frontend would be minimal.

The Congressional Database website interacts with our database using *Next.js* within a *Node.js* runtime environment. *Next.js* is a framework that provides a hybrid server side and static page generation for React frontends. Our website must generate unique pages for specific issues, committees, and roll calls. Having to render every part of these pages upon request from the client would result in considerable delays. The primary advantage of using *Next.js* is that it allows the backend to prerender these pages so the client can be given the complete page on request.

This framework also provides other advantages for development such as simplifying the routing system. The framework automatically creates new routes based off the names of files within the pages folder. *Next.js* also provides a convenient method for dynamic routing of the unique issues, roll call, and committee pages. Finally, we appreciated how it accomplished certain complex tasks such as webpack configuration for us so we could focus on the content of the website.

The frontend uses a *React.js* framework, as *Next.js*'s usage encourages. *React.js* streamlines the integration of HTML and CSS with JavaScript, which helps streamline development. A key advantage of using *React.js* is its support for creating reusable components. Said components make our goal of making the entire website visually consistent significantly easier. Using components we can import and configure elements such as the search bar and modal into each page without having to rewrite large amounts of code each time. Furthermore, like *Next.js* and *Node.js* it is a JavaScript framework. Since all these frameworks use JavaScript they integrate together seamlessly and make it easier for future students to work on the project.

User Study

For the purposes of testing whether our design choices were conveyed effectively and identifying areas for improvement we conducted a user study. In this section we will explain the design of our user study.

Methods

The design of our user study takes inspiration from the research of Bateman et al. (2010) and Borkin et al. (2016). The Bateman et al. study was structured with two phases. In the first

phase participants were asked to look at multiple visualizations and were asked questions about the visualization's contents. This phase allowed participants to take as much time as necessary to answer the questions correctly. Afterwards they were asked to recall the contents of the visualization either shortly afterwards or 2-3 weeks afterwards. After the recall phase was complete, they were asked to complete a questionnaire about their personal preferences for visualizations. Borkin et al. employs a similar phase structure with greater emphasis on analyzing the eye tracking data. Our study employs a similar two phase structure, except it substitutes the recall phase with qualitative questions.

  The user study was conducted using a build version of the Congressional Database website on a laptop. Our study is divided into two phases: the tasks phase and the question phase. These phases occurred in immediate succession and took around 4 to 9 minutes to complete depending on the participant. A total of 5 individuals participated in the user study. This sample size is primarily a product of the circumstances of the pandemic. The contact restrictions in place at the time limited our options for reaching out to large numbers of potential participants. For each study we recorded the screen using QuickTime. We also recorded the participant's comments and responses using the iOS Voice Memos app. These audio recordings were then manually transcribed and promptly deleted.

  For the tasks phase of the user study participants were asked to navigate the website to find specific requested information. The participants were not given any guidance on how to locate the requested information besides reminders of what was being asked. For instance, if a participant found the correct committee but it was from the wrong session of Congress they would be reminded of the correct session.

  There was a total of four different tasks prepared for the participants, which correspond to the four primary data pages. Each participant was asked to perform two of the four available tasks. One randomly selected task from a pair we believe to be easier and one randomly selected task from a pair we believe to be more difficult. Participants were first prompted with one of the easier tasks, which they would find from the home page. The page the user landed on upon successful completion of the first task was used as the starting point for navigating to the requested information in the more difficult task. Below this paragraph is the exact wording of the tasks the participants were asked to perform. We considered the first two tasks to be the easier pair and the latter two tasks to be the more difficult pair.

- Find out who the chair of the Armed Service Committee of the 113th Congress is.
- Find out who the senators from Maine are in the 103rd Congress.
- Find the sponsor of Senate Bill 2021 in the 116th Congress.
- Find out which bill is voted on in the 140th roll call of the 115th Congress.

Upon successful completion of their two assigned tasks the participant was asked to respond to four open-ended qualitative questions about the design of the website. Beyond highlighting potential rooms for improvement with the website, the tasks phase was used to give the participants experience using the website. This allowed the participants to form an initial impression of the website that could be used to answer the qualitative questions. For every participant we asked the same four qualitative questions in the same order. The qualitative questions are worded as follows:

- What are the first few words that come to mind when you think of the design of this website?
- How easy or difficult do you find navigating this website?
- What do you think would help make this website easier for you to use?
- What would make this data more meaningful to you?

The study was concluded after the participant answered all four of these questions. After the conclusion of the study we reviewed the screen recordings of the participants navigating the website. We counted the number of mistakes the participant made in search of the requested information. For our purposes a mistake is defined as clicking on a component of the page that does not help the user get closer to the answer. For instance, a participant navigating to the Members page in search of a committee chair would be considered a mistake. Finally, we reviewed the transcripts of the participant responses to qualitative questions and assessed interesting ideas and points for improvement for the website.

Results

We have calculated the mean number of mistakes made by each participant for the respective tasks, which can be seen in Figure 11. The two participants that were assigned the first task were both able to locate the two senators without making a single mistake. Notably both

subjects utilized the Command+F search function to find the word Maine instead of the built in search bar on the page. For the second task the most common mistakes were selecting the correct committee in the wrong Congress and attempting to find the relevant information in the Members page.

For the third task one subject found the relevant information without a single mistake but did require multiple clarifications of the question. The other two subjects experienced somewhat more difficulty locating the relevant information. Notably, one subject took some time before they figured out that the Senate bills are located in the Issues page. With a mean number of mistakes of 4.5 the fourth task appeared to be the most difficult for the participants. Both subjects assigned this task were able to locate the 140th roll call with little issue but were confused as to how to find the bill said roll call voted on. This confusion resulted in the participants doubting they found the right page and backtracking unnecessarily.

The participant's responses to the four qualitative questions can be seen in Figures 12 through 15. These excerpts encapsulate each participant's full responses with filler words like "um" removed and light edits for sentence clarity.
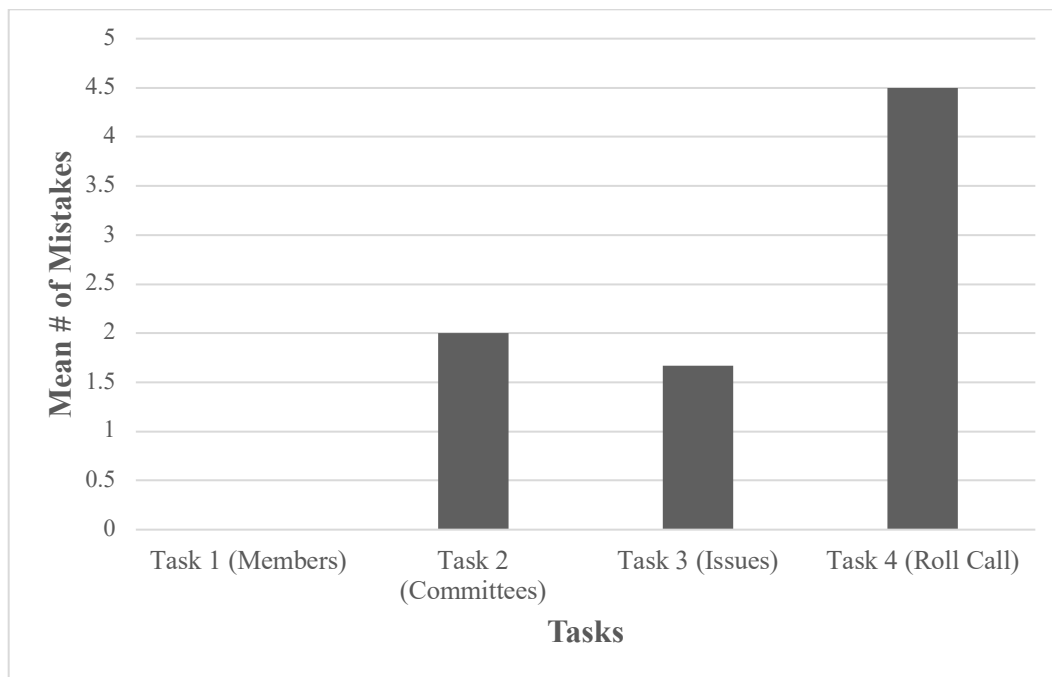

Figure 11. Mean number of mistakes recorded for each participant by task. N=2 for Task 1 and 4, and N=3 for Task 2 and 3.

Q1: What are the first few words that come to mind when you think of the design of this website?

- "It was very easy to scroll through … accessible … some of the titling was a little vague for me since I'm not really used to like, legal terms … it was a little like you have to know the terms to know what to click on … it was user friendly with the big buttons"
- "I think it looks professional … it's simplistic so it's easy to read and it's… informative?"
- "I think it's pretty well structured and organized it just takes one or two tries to figure out how it's structured and works, but otherwise it's fine."
- I'd say minimalistic … at the very front, it looks very straightforward and easy to understand … I do like how you divided these into like sectors you have the links to that up ahead too … in terms of user experience it seems pretty straightforward you got multiple ways to access links. … I guess one kind of inconsistency would be how you know it's like committees, issues, members, roll call it seems like these aren't like, consistent with it. I mean I would switch committees and issues.
- "I would say seems very organized. I would say streamlined also … I feel like the overarching umbrellas on them were actually very helpful like the issues, members, committees. Clean you know."

Q2: How easy or difficult do you find navigating this website?

- "I think it's helpful since- especially because it has the search things there that let you choose different congresses. I didn't notice it the first time around, but it's good."
- "I'd say overall it took me a hot second to understand where to find stuff initially. My initial impression would be to go to members first to see the specific person you were trying to find … but after you talked about the 113th Congress, Armed Services it seemed pretty straightforward."
- "Pretty easy, first one was definitely very easy. Cause I feel like you go to members and you kind of go from there. And for the other one it was just remembering the entire phrase that was ****ing me up. But once I knew what I was looking for then it was not super difficult … On an easy to hard scale I'd put it like a 2 out of 10 on the easiness side."
- "It was pretty easy … it was just one thing like the highlighting portions I just didn't know at the time. Other than that it was pretty easy to navigate and look through."
- "I found it pretty easy … I guess when I saw this I didn't recognize that it was just like alphabetical by senator rather than state … using the finding functions is a lot easier than just recognizing the order. Very easy to navigate altogether."

Q3: What do you think would help make this website easier for you to use?

- "It would be kind of nice if there were like, brief descriptions of each process. Like what the committees are maybe or how often roll call is done … Cause I imagine someone that works with this stuff a lot will find this as a really great reference that they can quickly turn to. But someone who doesn't know that much about Congress might find it harder to use."
- "I guess just design wise make it more colorful instead of just straight up black and white. But otherwise I see no issue with it structure wise… just aesthetics."
- "… if I didn't have the questions and prompts and stuff … maybe like an about? Cause I didn't really know what the website was about just at first. Maybe something like that … I don't think you need to explain how to use it because that's intuitive … just like what exactly it's for. It might've said that on the landing page, but I wasn't quite clear on that."
- "Maybe if I wanted to look at a particular bill you should have everything laid out when you click on it. Or just having the information laid out more specifically on a particular title or bill"
- "Based off of the questions you were asking it was really focused on navigating to which Congress. … In terms of hierarchy maybe if you had Congress as like, a higher hierarchy than the committee name then I would know from the very start I should look at the Congress first … and then I didn't think about the Congress until you gave me a hint. I guess one change could be bringing the Congress to a more significant location so people would target that first … I guess one minor detail would be having a dropdown bar icon right here so that people know it's interactive."

Q4: What would make this data more meaningful to you?
• "The data would be meaningful to me if I was interested in a particular policy or a particular law … I would be interested in like going on the database in what particular votes happened or a particular legal proceeding." • "In terms of data what I've seen so far it seems like kind of information but it's not very visual. Maybe some sort of mapping or graphing component would make it a lot more fun … I mean I don't even think it even needs to be interactive but just some sort of visual graphing something like that" • "I'm not really sure how people would use this information. OK I do see- it seems pretty straightforward. I like the fact that you're keeping things simple, like if you want more information you click on them. … I think you did a pretty good job with just laying out this stuff." • "I think it could be fun if you had like a recent news feature where recent bills would appear or what's happening right now kind of things. Cause I feel like that's also a very hard thing to follow." • "It would have to be what context I'm using it. Like if I need it for class or I'm just curious. I see nothing wrong with how it currently is."

Figures 12-15. Lightly edited excerpts from the transcripts of each participant's response to each question. Said edits were exclusively to cut down on filler words and to make them easier to read.

Limitations

This user study was conducted with only five participants, and only two to three participants were assigned any individual task. As a result, we cannot guarantee that the data on average mistakes performing each task is representative. The assignments for each task were given to the participants orally, and they often needed reminders for the specifics of the instructions. A portion of the mistakes recorded can be explained by the participants simply forgetting a detail of the assignment such as clicking on the right committee in the incorrect session of Congress. However, the primary purpose of the tasks phase is to introduce the participants to the website and highlight rooms for improvement in the design. The average number of mistakes per task may not be representative, but the specific reasons users made said mistakes are informative.

Relatedly, since each participant was only asked to perform two of the four possible tasks they did not see every page before they were asked the qualitative questions. We do not consider this to be a major issue due to the visual similarity of each page's design. Furthermore, if this

issue dramatically impacted their impression of the website, we would likely observe greater variation in the responses to the first and second qualitative questions.

Discussion and Future Work

Starting with our findings from the tasks phase we have located multiple areas for improvement to our design. The task our participants experienced the most difficulty with was the fourth task, which involved finding the bill associated with a specific roll call. Presently, the issue the roll call is associated with is highlighted with purple text, which can be clicked on to view more information about the specific bill. The purple-colored text alone failed to convey that it was a clickable component. This problem is also a result of using unfamiliar terminology. The participants were not aware that "S." means Senate bill. To remedy this issue we will consider additional visual cues to indicate that the issue titles are clickable and introducing keys to help guide users in understanding the naming conventions of issues.

Another major issue highlighted in the tasks phase is that the dropdown menu that allows users to change the session of Congress may not be immediately noticeable for some users. This is especially important for the Committees page, where there is minimal visual change when a user switches to a different Congress. The session of Congress is easily one of the most important columns for almost every table in the database. All unique IDs in the four primary tables of the Congressional Database incorporate the session of Congress. As is the case in the database, the session of Congress is integral to the organization of the website. To resolve this issue, we will consider ways to visually privilege the dropdown menu for sessions of Congress and add additional visual indicators when the page changes to a different session.

While there are many ways in which we can improve the website's design, the subject's responses to the qualitative questions are encouraging. This is particularly apparent in the responses to the first question as seen in Figure 12. The participants were not directly told what the purpose of the website was or the central goals of its design. We hoped that its purpose and design philosophy would be conveyed to the participants simply by letting them use the website.

Ultimately, many of our participants used similar words to describe the website as are found in our goals for its design. One participant called the website "minimalistic." Another referred to the website as "streamlined" and "very organized." A central goal of our design is to

make the experience as straightforward and easy to use as possible. All of our participants used variations of those terms to describe the website. Even those that described the website as having a learning curve described the website as easy to use once they figured it out.

The subject's responses to our second question are also largely encouraging. Three of the five participants led with the term "pretty easy" in describing navigating the site. The two participants that did not use the term "pretty easy" used the terms "helpful" and "straightforward" in their responses. However, some users expressed the presence of a learning curve in their use of the site. For instance, one participant said, "I'd say overall it took me a hot second to understand where to find stuff initially. My initial impression would be to go to members first to see the specific person you were trying to find." While they ultimately found the website 'pretty straightforward' once they figured it out, this response indicates that certain aspects of the website are not immediately intuitive. The first and second questions were designed to get the participants' overall impressions of the website. The primary takeaway from these responses is that our overall design approach was successfully conveyed, but there are still many ways in which we could make the site easier to use and understand.

The third and fourth questions were designed to prompt the participants on potential ideas for improvement to the website. A particularly important finding from these answers and previous ones is a need to familiarize users with the terminology of the website. Researchers interested in using our data would likely have little trouble navigating the terminology used on the website. However, as one of our subjects pointed out: "someone who doesn't know that much about Congress might find it harder to use." This idea is affirmed by the difficulties certain participants had figuring out how to navigate the website. The abbreviated terminology used to refer to Senate issues such as "S." and "PN" will certainly need some form of explanation. Even the names of the four primary data pages may not be initially meaningful to some users. Our website has explanations for the roles of specific committees but takes for granted that the user understands the role and importance of committees overall.

There are multiple potential avenues to address the user's potential lack of familiarity with congressional terminology. However, depending on how we implement additional guidance it could inconvenience users who are confident in the terminology. As mentioned earlier one possible solution is to implement keys on certain pages such as Issues to help users understand the terminology. While this feature could be helpful, it could be visually intrusive depending on

how much text is required to successfully explain the terms. Any additional explanations found on the website should be consistent with our minimalist design approach to the website. Too much explanatory text would inconvenience experienced users and potentially visually confuse new users as well. In keeping the keys consistent with the rest of the website's design we could either direct the user to a dedicated page for explanations or make the appearance of the key toggleable.

Relatedly, one participant mentioned that our website should have some form of explanation for its overall purpose. The text on our home page is intended to serve this purpose, but evidently was not an effective explanation for said individual. Beyond rewriting the text on the home page they suggested having an About page explaining the website's purpose. This idea was already in our plans prior to conducting the user study. They are a common feature on many websites to explain the website's purpose or to provide information about its developers. This participant's input serves as a reminder to make said About page a high priority.

One response to our third question has a less clear solution. One participant expressed a desire for the website to be more colorful and hence, more visually appealing. The visual appeal of the website is ultimately subjective, and it is worth noting that no other participant said something similar about the website. However, the color choice of the website is worth considering. Users may respond positively to websites that use saturated colors if it fits their cultural context. However, the most common colors used on web pages worldwide are whites, grays, blacks and dark blues (Kondratova & Goldfarb, 2007). These colors have the advantage of enabling contrasts between text and background and are unlikely to elicit strong negative reactions. However, it seems plausible that these safe color choices risk being unmemorable. Presently we are unlikely to make major changes to the background color of the website. However, it is certainly worth considering making the coloration of select components of the website appear more distinctive.

Another idea a participant offered to make the website more visually appealing is the implementation of "some sort of mapping or graphing component." This was already a goal for our research page, but there were no such plans for the rest of the website. Presently, the only source of imagery our data pages have are the photos of individual senators and the pictures on the home page. We are not aware of any pre-existing tool that could provide engaging visualizations for individual Senate bills for instance. However, it is an intriguing area for future

exploration. We have considered making use of our data of bill texts to conduct some form of text analysis. If this work were to materialize it could provide interesting opportunities to add visualizations as well. Like the idea of a news feed, these sorts of features will likely come down to the ambition of future students working on this project.

Conclusion

The Congressional Database is expected to be an ongoing project. The full extent of its ambitions is yet to be realized, and the user study in this thesis may very well inspire further aspirations. However, even if no further changes were made to the project it would stand as a comprehensive database that covers almost fifty years of Senate activity. While it would be difficult to get historical data beyond what we currently possess, our database can continue to grow with every new Congress as long as someone maintains it.

Through our user study we have identified many areas for improvement for the website. The internet, HCI research, and the needs of users are constantly evolving. As such, claiming that the website would benefit from certain changes is almost mundane. As an HCI project we take great interest in uncovering new ways to provide an effective user experience. However, the Congressional Database in its current state largely leaves its desired impression on users. The Congressional Database provides a solid foundation on which future work can be built upon in many possible directions.

References

Bateman, S., Mandryk, R., Gutwin, C., Genest, A., McDine, D., & Brooks, C. (2010). Useful junk?: the effects of visual embellishment on comprehension and memorability of charts.

Blandford, A., Furniss, D., & Makri, S. (2016). Qualitative HCI research: Going behind the scenes. Synthesis lectures on human-centered informatics, 9(1), 1-115.

Borkin, M. A., Bylinskii, Z., Kim, N. W., Bainbridge, C. M., Yeh, C. S., Borkin, D., . . . Oliva, A. (2016). Beyond Memorability: Visualization Recognition and Recall. *IEEE transactions on visualization and computer graphics, 22*(1), 519-528. doi:10.1109/TVCG.2015.2467732

Comber, T. (1995). *Building usable web pages: An HCI perspective.* Paper presented at the AusWeb95, innovation and diversity: The World Wide Web in Australia, Proceedings of AusWeb95, the first Australian World Wide Web Conference.

Congress.gov. (2021).

Hendler, C. (2009). Senate Goes XML. *Columbia Journalism Review.*

Irfan, M. T., & Ortiz, L. E. (2014). On influence, stable behavior, and the most influential individuals in networks: A game-theoretic approach. *Artificial Intelligence, 215*, 79-119.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information systems management, 29*(4), 258-268.

Jones, B. (2020). Increasing share of Americans favor a single government program to provide health care coverage. *Pew Research Center*.

Kondratova, I., & Goldfarb, I. (2007). Color Your Website: Use of Colors on the Web, Berlin, Heidelberg.

Lewis, J. B. Voteview.

Marshall, B. W., & Wolpe, B. C. (2018). *The committee: a study of policy, power, politics, and Obama's historic legislative agenda on Capitol Hill*. Ann Arbor: University of Michigan Press.

Pedersen, L. A., Einarsson, S. S., Rikheim, F. A., & Sandnes, F. E. (2020). *User Interfaces in Dark Mode During Daytime–Improved Productivity or Just Cool-Looking?* Paper presented at the International Conference on Human-Computer Interaction.

Sørum, H., Andersen, K. N., & Vatrapu, R. (2012). Public websites and human-computer interaction: an empirical study of measurement of website quality and user satisfaction.

Behaviour & information technology, 31(7), 697-706.
doi:10.1080/0144929X.2011.577191

Stewart, C. Charles Stewart's Congressional Data Page.

Tauberer, J. (2021). GovTrack.

Wilson, B., & Cong, C. (2021). Beyond the supply side: Use and impact of municipal open data in the US. *Telematics and Informatics, 58*, 101526.