Part 1

The Scoring System

We have used each of our metrics that we calculate for the websites to also give them scores; every metric in both the Automated and User Interaction sections carries a score between 0 and 10, with 10 representing the ideal performance and 0 representing the worst. As we have nine automated metrics and 3 user ones, this comes out to total scores out of 90 and 30 respectively for the two sections.

The scale for this scoring is decided somewhat arbitrarily; we have given equal weightage to every single metric we measure, though it can be argued that some of them would be much more important than others in real life. However, the actual importance each metric holds can't be easily calculated objectively, and as such we have decided to go with the simple way of equal weightage, as it still helps give a rough idea about the sites' relative performance.

Finally, note that some of the scores may be biased due to users' prior familiarity with the sites (like the BITS website) or due to the tasks for some websites being easier than others (though we did our best to make them similar in difficulty).

a) Automated Metrics

Metrics used

1. Calculating the average load time for the page to load
2. Availability of skip to main content button
3. Images with alt text
4. Number of broken links
5. Pages with multiple languages reflecting the switch
6. Handle unknown routes
7. Fraction of elements with an appropriate colour contrast
8. Pages having title descriptions
9. Pages limiting zooming and scaling

Average Load Time

Users generally expect websites to load quickly. A slow-loading website can lead to frustration, impatience, and a negative perception of the site. Users are more likely to engage with and return to websites that offer a fast and responsive experience. As such, we calculated the average load time for each website as our first criterion. We just let selenium method to load the website and use python's time library to calculate the load time.

We score this metric on a linear scale, with a loading time of 0 seconds corresponding to a score of 10, and a loading time of 5 seconds or higher corresponding to a score of 0.

Availability of Skip To Main Content Button

The "Skip to main content" button is a web accessibility feature that enhances the user experience for individuals who use screen readers or keyboard navigation. This button allows users to skip repetitive navigation links and jump directly to the main content of a web page. The implementation of this feature is done differently by different websites, so for this assignment, we are only testing if the button is provided by the target website. For this test we used generic xpath to check the presence of this element.
https://developer.mozilla.org/en-US/docs/Web/HTML/Element/main#skip_navigation

Sites get a score of 10 for this metric if such a button is found, and 0 otherwise.

Alternate Text in Images

Alt text, short for alternative text, is a descriptive attribute added to an HTML image element. It provides a textual description of the content and function of the image, making it accessible to users who may not be able to see the image due to various reasons, such as screen readers for people with visual impairments or slow internet connections that prevent image loading. For this test we collect all the image tags and look for non-empty alt attributes. Some websites also filled the alt attribute with unhelpful information; we classified them as not having alternate text as well.
https://developer.mozilla.org/en-US/docs/Web/API/HTMLImageElement/alt

This metric is scored as 10 multiplied by the fraction of images with (meaningful) alt text.

Number of Broken Links

Users expect to find relevant information quickly, and broken links hinder this process, leading to frustration and a potentially negative perception of the site. Broken links can undermine this confidence, making users hesitant to click on links and explore further. A higher number of broken links would make the website a game of Minesweeper. For this test we used selenium to get all link tags(a) and made requests to the href attribute and used the status code to classify if it is a broken link or not. We consider redirects as valid links too. If a url throws an error instead of sending anything back we consider it as an invalid link.

This metric is scored as 10 multiplied by the fraction of functional links.

Reflecting Changes in Language

Screen readers use the "lang" attribute in HTML to announce content in the appropriate language, improving the accessibility of the website for users with visual impairments. Translation services and browser translation tools can use the lang attribute to identify the

language of the content, enabling more accurate translations. For this test we visited the website in other languages and looked for the attribute in the top level.
https://developer.mozilla.org/en-US/docs/Web/HTML/Global_attributes/lang#accessibility

Sites get a score of 10 for this metric if the attribute is found, and 0 otherwise.

Handling unknown routes

Users may encounter unknown routes due to typos in the URL or outdated bookmarks. In such cases, a well-crafted error message can guide users on what went wrong and how to proceed. For this test we picked a random alphabet and tried to use it as a route and see how the websites behave(for example https://url/t where 't' is the random alphabet). The sites that throw a valid 404 are considered to be usable.

Sites get a score of 10 for this metric if a valid 404 is thrown, and 0 otherwise.

For the next three metrics, we acquired the data by using axe (which uses Selenium). This is a library that automates web accessibility testing by checking if various web accessibility guidelines are followed by a website or not. The axe script dumps data into a JSON which we parsed using **jq**.

Colour Contrast Ratios

Some people with low vision experience low contrast, meaning that there aren't very many bright or dark areas. Everything tends to appear about the same brightness, which makes it hard to distinguish outlines, borders, edges, and details. Text that is too close in luminance (brightness) to the background can be hard to read. Hence, all text elements must have sufficient contrast between text in the foreground and background colours behind them in accordance with WCAG 2 AA contrast ratio thresholds.
https://dequeuniversity.com/rules/axe/3.1/color-contrast?application=axeAPI
We got the number of elements with sufficient and insufficient colour contrast by checking the "passes" and "violations" array respectively.

This metric is scored as 10 multiplied by the fraction of components with acceptable contrast.

Documents must contain a title element to aid in navigation

Screen reader users use page titles to get an overview of the contents of the page. Navigating through pages can quickly become difficult and confusing for screen reader users if the pages are not marked with a title. The page title element is the first thing screen reader users hear when first loading a web page.

The title is the first thing that screen reader users hear when they arrive at a page. If there is no title or if the title is not descriptive and unique, screen reader users must read through the page to determine its contents and purpose.
https://dequeuniversity.com/rules/axe/3.1/document-title?application=axeAPI

We checked if a test with the id "document-title" exists in the "passes" array to determine if the website has a title description.

Sites get a score of 10 for this metric if the page has a title description, and 0 otherwise.

Zooming and scaling must not be disabled

The user-scalable="no" parameter inside the content attribute of the <meta name="viewport"> element disables zooming on a page. The maximum-scale parameter limits the amount the user can zoom. This is problematic for people with low vision who rely on screen magnifiers to properly see the contents of a web page.

Users with partial vision or low vision often choose to enlarge the fonts on their browser to make text on the web easier to read. The browser's viewport focus is everything visible in the browser window at a given moment. Maximising the browser to full size on a high-resolution monitor creates a large viewport focus area and may include the entire web page.
https://dequeuniversity.com/rules/axe/3.1/meta-viewport?application=axeAPI
We checked if a test with the id "meta-viewport" exists in the "passes" array to determine if the website supports zooming and scaling.

Sites get a score of 10 for this metric if unrestricted zooming is enabled, and 0 otherwise.

b) Metrics with User Interaction

Methodology

Participants were all informed about the nature and purpose of the study, as well as what details of theirs would be recorded. They were made to fill a demographic questionnaire, after which they were asked to visit the websites linked and complete the given tasks. The tasks and links were both part of the questionnaire provided, and there was no intervention from the interviewers during task completion. The users were told that they were given 5 minutes for each website to perform the task and they were also informed before the test that they have to option to not continue if they wish.

All users went through the tasks on the same laptop, in incognito mode (to avoid one session affecting others by stored cookies, purple links that are already visited, etc.), as well as with any browser extensions turned of except for selenium IDE extension. All experiments were conducted over the college LAN to ensure a uniform experience.

Metrics used

1. Number of Clicks
2. Time to Finish Task
3. Number of back button clicks

Number of Clicks

Needing a lot of clicks to find some information on a website generally indicates deeply nested hierarchies, which can make it hard for users to keep track of what page they are on and figure out where they need to go. Thus, we measure the number of clicks participants took to finish their tasks and penalise the website accordingly. Note that this does not include the number of times the back button was clicked.

As our minimum and maximum value recorded for this metric fall close to 0 and 100 respectively, the score for this metric is a linear scale with value 10 for 0 clicks and 0 for 100 clicks.

Time To Finish Task

This metric is self-explanatory; the less time it takes for a user to complete the given task, the smoother and more pleasant their experience is. We kept an upper limit of 5 minutes per task, giving a failing score to any website where the user couldn't finish their task within the time limit.

The score for this metric is a linear scale with value 10 at 0 seconds at 0 at 300 seconds (timeout).

Number of Back Button Clicks

Having to press the back button at all in a website is an especially potent indicator of confusing or misleading design, as it means the user had accidentally gone to an incorrect page (or at least perceived the page they had arrived at as incorrect) and had to backtrack. All our tasks were designed to be doable without a single back button click being necessary, and as such these clicks are penalised much more than the regular clicks in our first metric.

As our minimum and maximum value recorded for this metric fall close to 0 and 10 respectively, the score for this metric is just 10 - (number of back button clicks)

Note:
- In case a task is failed either due to timeout or because the user gives up/does not wish to continue, the website is given a score of 0 for all three user interaction metrics for that user.
- The final score for the user interaction metrics is the average score attained by that website over all users.