What is the CS410 Digital Library?

The CS410 Digital Library is a Chrome extension and website with the goal of helping the scaling of learning in online classrooms. It supports contextual question-asking and digital library creation. Online information is curated via a Chrome extension and is searchable by a website. Below, we explain the installation process and use cases in detail.

Why is this needed in CS410?

CS410 has nearly 400 students registered for the course. While a course of this size makes it difficult for the course staff to meet the learning needs of each individual student, the size also provides a great opportunity for us to leverage the benefit of collaborative learning. We can already achieve collaboration to some extent by having you help each other answer questions posted on Campuswire, but we hope to do more by offering a new tool: a work-in-progress Chrome extension and website that currently supports contextual question-asking and digital library creation. As the semester progresses, we plan to add more real-time benefits to using the platform, such as providing you with helpful links upon opening the extension. We believe that these benefits can help us efficiently and collaboratively scale learning in online classrooms and improve the learning experience for all of you.

The CS410 Digital Library will be used in lieu of LiveDataLab for MP 2.1 and MP 2.2. For MP 2.1, you will use the extension to collect parts of web pages that you have found helpful while studying the course material. And for MP 2.2, you will use the website to perform some basic searches attempting to retrieve your submitted content, and you will submit relevance judgements over the returned results.

Outside of the required use of the CS410 Digital Library for MP 2.1 and MP 2.2, we strongly recommend that you use the extension and search engine to explore all kinds of ways of leveraging it to facilitate your own learning or help others learn. We also encourage students to propose course projects that extend the platform beyond its current functionality. As it is an innovative tool that can improve our teaching, we would greatly appreciate it if you could all take the time to download it and try it out

How to use the CS410 Digital Library?

To access the Chrome extension or website, you must first be using the Illinois VPN. Before proceeding, please set up and activate the VPN on your local computer by following the directions outlined here.

Chrome Extension

Installation Instructions

The Chrome extension is not yet in the web store, but it can be downloaded from here. After downloading, please follow the instructions below:

1. Decompress the extension folder, and move it somewhere permanent.
    a. Note that deleting or moving the decompressed extension folder after completing these steps will result in the extension not working in the browser.
2. In your Chrome browser, paste chrome://extensions/ into the address bar and navigate to the page.
3. Toggle on "Developer Mode" (at the top right of the screen).
4. Click "Load Unpacked", and select the decompressed extension folder.
5. Toggle off “Developer Mode”
6. Click the "Extensions" icon in the Chrome toolbar (puzzle at top right of toolbar).
7. Pin "DL Extension".
8. You can now use the extension!
    a. Note that it will not work on the chrome://extensions/ page
    b. When creating an account, please use your university email as the email and your NetID as the username (important for grading!)

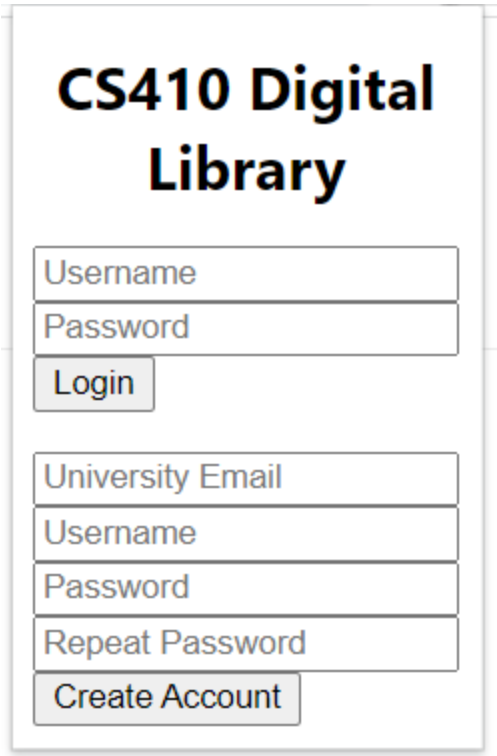
Update Instructions

Over the semester, we may make some changes to the extension UI (bug fixes, additional features, etc.) If and when we do, we will make an announcement on Campuswire and provide the link to download the updated version.

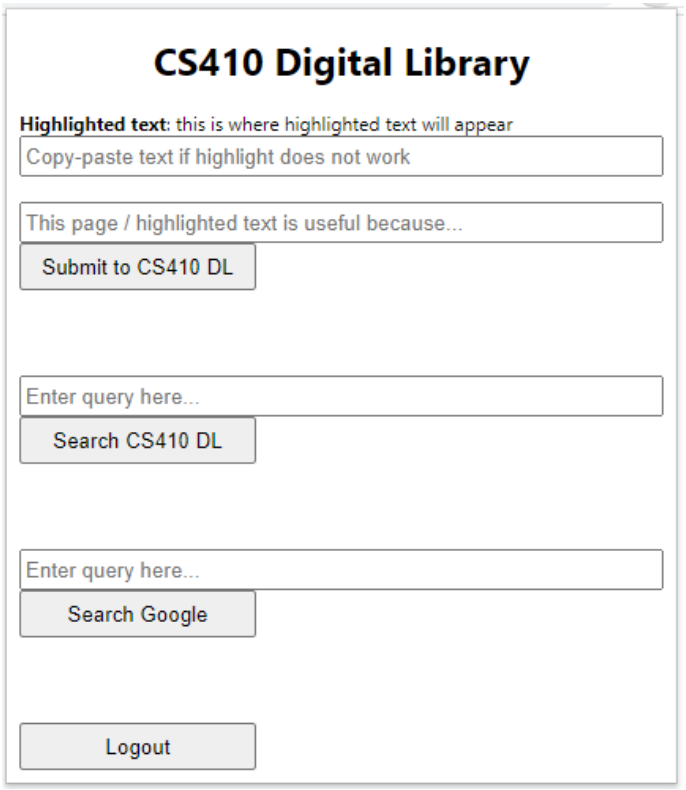
To update your extension locally, download the updated compressed folder, extract it, and move the folder to the same location (replacing your existing uncompressed extension folder). Then, the next time you use the extension in-browser, it should be the updated version.

Functionality

After opening the extension, you should see something like the following:



Please create an account (**make your username your NetID and use your university emai**l). And make sure to write down your password!. After successfully creating an account, you'll see a page like this (depending on whether or not you have highlighted any text in your browser tab):



The functionality of each button is described below.

Submit to CS410 DL

Clicking this button will save the current URL (i.e., the page that you opened the extension on), your short explanation of relevance, and the highlighted text (if any) to the backend database. Here, the idea is to create a “digital library” or collection of webpages that are useful for students (try not to save any personal or private information). The submitted pages should be immediately available to search on the CS 410 Search Engine website (described in the section titled “CS410 Search Engine” below).

Note that the extension will not send the webpage directly from your browser, rather it sends the URL highlighted text, and the explanation that you provide. This means that you are able to submit snippets of pages that are only available to CS410 students (e.g., a Campuswire or Coursera page).

For some webpages, highlighting the text will not work. So, we provide an optional field right below the highlighted text display that reads “Copy-paste text if highlight does not work”. By copying text from a webpage and pasting it here, you can submit the snippet as is described above.

Examples

You find a YouTube video that helps explain the intuition behind the retrieval formula BM25. While on the page, you open the extension, type in “intuition behind bm25”, and click “Submit to CS410 DL”

You read a blog post that walks through the basics of the vector space model, and there is a particular paragraph that helped you understand the construction process. While on the page, you highlight the paragraph, open the extension, type “vector space construction process”, and click “Submit to CS410 DL”.

More generally, you can submit anything: papers, textbook pages, Reddit posts, etc. and explain why you found it helpful.

Search CS410 DL

Using this button will use the query, current website URL, and highlighted text to search the CS410 Digital Library (which contains all of the submitted content described in the previous section). The button will open a new tab and send the query to the CS410 Search Engine (described in the section “CS410 Search Engine” below).

Examples

You are reading a Wikipedia article on information retrieval, and are curious about how a concept relates to CS410. So you highlight the paragraph, write the concept in the search bar, and click “Search CS410 DL” to see how other students have curated information relating to this concept.

Search Google

Alternatively, you can use the “Search Google” button, to search Google just like you would from https://google.com. After entering a query, clicking the search button will open a new tab with your search query entered on Google. We include this functionality so that you can get instant answers to simple questions. When you perform a search in this manner, the backend records the current URL (i.e., the page that you opened the extension on), the highlighted text (if any, or the copy-pasted text), and your typed query. So, if you are about to perform a Google search with respect to a specific paragraph on the webpage, we ask that you first highlight it, then enter the search query via the extension.

Examples

You are watching a Coursera lecture and come across the term “syntagmatic”. The term is new to you, so while on the lecture page, you open the extension, type “syntagmatic definition”, and click “Search Google”.

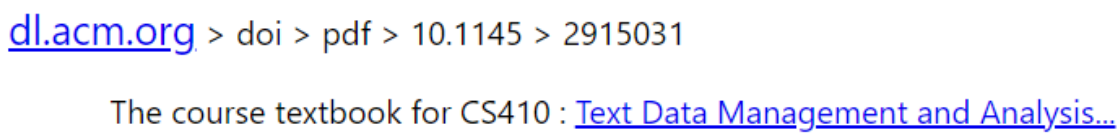
Chaining Together Google Searches and Submissions

With the functionality described above, you can easily save answers to questions. After searching Google from a highlighted context or webpage and browsing the results, you may find the answer to your question. Then, you can highlight the answer, open the extension, add a description like “helped answer X”, and click “Submit to CS410”.

CS410 Search Engine

The CS410 Search Engine can also be accessed directly via this link. After logging in (with the same credentials as the Chrome extension), you can enter basic search queries and see the curated information submitted by all of the students. The search results are grouped by general URL, and each URL will have one or more links to highlighted portions of the page, along with the submitted explanation for why the page/highlighted text is useful.

An example search result is displayed below:



The first line provides the hyperlink to the webpage “https://dl.acm.org/doi/pdf/10.1145/2915031”, and the second, indented line is a user submitted “explanation : highlighted text”. The link in the second line goes directly to the highlighted text. If multiple students submit contexts on the same web page, then there will be multiple indented lines below the first line, each displaying the submitted explanation / highlighted text.

You can also click “My Submissions” at the top of the page to see all of the content that you have submitted. Moreover, using this page, you can delete any of your submissions (in case you would like to edit, just delete and re-submit).

Frequently Asked Questions

How can I report a bug, give a suggestion, or discuss a concern?

Please post on Campuswire or email Kevin Ros at kjros2@illinois.edu.

What should I do if the highlight doesn’t register?

For some pages, your highlighted text will not be captured by the extension. You can tell that this is the case when the “Highlighted text:” area doesn’t display any text. If you want to save the highlighted text in the context or ask a question about the highlighted text, then first copy-paste it into the field right below the highlighted text display.

How can I submit a specific PDF page?

You can add a fragment in the URL. For example, if you would like to submit page 9 of this PDF, you can add “#page=9” to the end of the URL. That results in this link.

What should I do for specific YouTube video times?

If you want to ask a question at a specific time, you can change the webpage URL to include the video time (right click video → copy URL at current time → paste into browser), and re-open the extension.

What should I do for Coursera videos?

You can highlight the transcript and open the extension, and it should register the highlighted text.

Overview of CS410 MP2

MP2 is a 4-part assignment in which you will get familiar with building and evaluating Search Engines. Following our philosophy of promoting collaborative learning, we have designed this comprehensive MP with the goal of not only teaching you useful skills of building and evaluating search engines, but also allowing all of you to leverage your MP2 work to collaboratively build a useful Digital Library (DL) search engine for CS410, which you all can then use to enrich your learning experience. You will be able to use the CS410 DL search engine to easily locate course information. For example, you can search with a query like "syllabus" to find the Coursera weekly modules or using the query "textbook" to directly find the textbook used for CS410, essentially by using the CS410 DL as a convenient bookmark search tool. Alternatively, you may also find an explanation of concepts covered in CS410 with a query such as "precision" or "explanation of information retrieval". You may also find useful Youtube videos using the CS410 DL search engine. The actual content the search engine supports would depend on what you all would save to the DL and thus will naturally evolve to include increasingly more content as we all use it over time in the course. The search engine you will build will also serve as a basis for you to design your course projects, which can then further improve the utility of such a CS410 DL system, providing more benefit to all of us as well as many students of CS410 in the future.

Specifically, the four parts of MP will systematically expose you to four major tasks involved in building and evaluating a search engine:

1. MP2.1 covers content acquisition and creation of a collection of searchable documents. The outcome of MP2.1 is a collection of documents relevant to CS410.
2. MP2.2 covers annotation of the collection created in MP2.1 to create a set of queries and the corresponding relevance judgments. The outcome of MP2.2 is a test set that can be used to quantitatively evaluate a search engine algorithm.
3. MP2.3 Implementation and evaluation of a baseline ranking algorithm, where you will be asked to complete the implementation of a state of the art ranking algorithm using the MeTA toolkit and evaluate it using the test set you will build from MP2.2.
4. MP2.4 Improvement of ranking algorithms, where you will have an opportunity to freely explore any ideas to vary/improve over the baseline ranking algorithm(s). A leaderboard-based competition will be held to encourage you to propose new ideas to design an effective ranking algorithm for ranking CS410 content pages. Your new algorithms can then be used to power the CS410 DL search engine and potentially improve its utility for all of us.

To enable such a massive collaborative MP, we will leverage an existing baseline CS410 Digital Library system that our TA, Kevin Ros, has built, which consists of a search engine and a browser (Chrome) extension. The Chrome extension enables you to easily save any pages from the Web that are potentially relevant/useful to CS410 with an explanation of how it is relevant or useful. It also enables you to conveniently take notes from any webpage by highlighting the

content and saving it to the CS410 DL. The search engine enables you to search over the collection of the saved pages made by all the students in CS410 and has the ability of updating the collection in real time, meaning that once you save a page, it would be immediately searchable via this search engine.

With this baseline CS410DL system in place, you will be able to finish MP2.1 and MP2.2 conveniently without spending any extra time for submission to Coursera or LiveDataLab. Specifically, the baseline CS410DL system also supports user accounts and can log user activities, thus grading can be done automatically based on the pages you have saved and relevance judgments you have made via the system. This means that you do not need to make any "conventional" submission specifically for MP2.1 or MP2.2; you just need to follow the instructions to use the system to save useful content and make relevance judgments for some sample queries that you are interested in.

MP2.3 and MP2.4 will be done using the MeTA toolkit and LiveDataLab in the same way as you have done for MP1. More details will be provided in the description of each of them.

MP2.1

Your task for MP2.1 is to use the baseline CS410DL system (briefly described above) to save a certain number of web pages so as to help create a large collection of relevant online documents to CS410 content for building a useful CS410 Digital Library. More information about this baseline CS410DL system (including installation instructions and the URL to the search engine) can be found in this Google document.

Specifically, please follow the instructions below to complete MP2.1:

1. Follow the extension installation instructions in the Google document.
2. Become familiar with the extension, the website, and the general functionality.
3. Create an account using your NetID and university email. Please make sure to write down your password! (The system at this point cannot help you reset your password.) Using the correct NetID / email is essential for grading.
4. Use the "Submit to CS410 DL" functionality to record at least 15 web pages that you found useful while browsing material for CS410 or relevant content on the Web. This could be lecture points, papers, campuswire posts, videos, etc. We encourage you to explore multiple uses of the CS410DL system, ideally spend a few minutes to explore each of the following categories:
    a. Bookmark some useful course information pages (e.g., some pages/lectures on Coursera or some useful posts on campuswire)
    b. Save some relevant research papers or technical articles related to CS410 (e.g., you can do a search in Google Scholar with any interesting keywords); please make sure to highlight a lot of relevant content in those articles to enable CS410DL to match them with a query (think about using CS410DL to conveniently take notes from any paper or technical article and share your notes

with your classmates)

c. Find some interesting relevant videos on Youtube; again, you can search in Youtube with interesting keywords
d. Find some relevant development resources related to CS410 such as useful toolkits or online tools; you can search with appropriate keywords in Google
e. Find some pages about startups using technologies related to CS410 (e.g., using a query like "text mining startups" to search in Google).

In general, please try to provide an informative description of any saved page and/or highlight relevant content whenever possible; the extra text helps CS410DL work better as a search engine since at this point, it only matches a query with the text you typed in or highlighted (this enables high relevance whenever there is a match).

Your submitted content should be available immediately for everyone in CS410 to search for on the search engine website. And grading for the assignment will be updated on Coursera once the assignment is due.

MP2.1 should be completed by **Sept 11, 2022 at 11:59 pm (CDT).**