

# CS 410 Clip 'n Search

Software usage tutorial presentation


# Topics covered

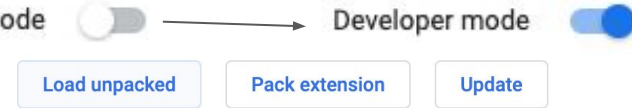
- (1) sufficient instructions on how to install the software if applicable
- (2) sufficient instructions on how to use the software
- (3) at least one example of use case so as to allow a grader to use the provided use case to test the software.

# Installing the Software (these instructions are basically the same as the DL Extension installation instructions except for link to extension)

Installation Instructions [download, decompress, load into browser]

The Chrome Extension can be downloaded from [here](https://github.com/brysonli12/CourseProject/tree/main/extension) (410clip\_extension.zip file in <https://github.com/brysonli12/CourseProject/tree/main/extension>) . 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).  The image shows the 'Developer mode' toggle switch in the top right corner of the Chrome browser's 'chrome://extensions/' page. The switch is currently turned on, indicated by a blue slider and the text 'Developer mode'.
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 "CS 410 Project: Clip 'n Search".
8. You can now use the extension!
  - a. Note that it will not work on the `chrome://extensions/` page



# Installing the Software

<https://github.com/brysonli12/CourseProject/tree/main/extension>

main CourseProject / extension /

Go to file

This branch is 7 commits ahead of CS410Assignments:main.



brysonli12 add extension zip file

3686fef 35 minutes ago History

..



410clip\_extension

Add files via upload

36 minutes ago



410clip\_extension.zip

add extension zip file

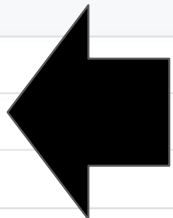
35 minutes ago



README.md

Rename index.md to README.md

10 days ago





README.md


Extension zip file and code will be put in this directory

# Installing the Software


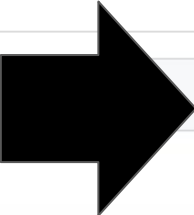
[https://github.com/brysonli12/CourseProject/blob/main/extension/410clip\\_extension.zip](https://github.com/brysonli12/CourseProject/blob/main/extension/410clip_extension.zip)

 main ▾ **CourseProject** / [extension](#) / 410clip\_extension.zip Go to file ...

 **brysonli12** add extension zip file Latest commit 3686fef 36 minutes ago History

 1 contributor

47.2 KB View raw

# How to Use the Software

Extension

Webpage <http://brysonli.pythonanywhere.com/>

# Extension

## Saving an article

## Searching using BM25

### Clip 'n Search (CS 410)

**Page Title:** Data Mining: What it is and why it matters | SAS

**Highlighted text:** Data mining is the process of finding anomalies, patterns and correlations within large data sets to predict outcomes. Using a broad range of techniques, you can use this information to increase revenues, cut costs, improve customer relationships, reduce risks and more.

Copy-paste text if highlight does not work

Introduction to data mining and its importance

Submit to Clip 'n Search

Enter query here...

Search Clip 'n Search database

Credit goes to CS 410 TAs for originally developing this extension (and I took the relevant parts and slightly modified/extended it [made a simple version for this final project], the server is something I did by myself). Additional credits on github/about section of website.

---

# Extension

Saving an article

Searching using BM25

## Clip 'n Search (CS 410)

**Page Title:** CS 410 Project Tutorial - Google Slides

**Highlighted text:**

Copy-paste text if highlight does not work

This page / highlighted text is useful because...

Submit to Clip 'n Search

data mining

Search Clip 'n Search database

Credit goes to CS 410 TAs for originally developing this extension (and I took the relevant parts and slightly modified/extended it [made a simple version for this final project], the server is something I did by myself). Additional credits on github/about section of website.



# Webpage

1. search
2. get extracted topics
3. view about

<http://brysonli.pythonanywhere.com/>

 [Clip 'n Search](#) [Topics](#) [About](#)

Search using BM25:

## Welcome to Clip 'n Search

**On this site you can:**

- Search the database of articles using BM25 (see above)
- View extracted [topics](#) from the current database of articles

Webpage



Clip 'n Search

Topics

About

Search using BM25:

1. search **Topics Extracted**
2. get extracted **from current article database (in no particular order)**
3. view about

<http://bryson>

#### Topic 1

ranking to query search information retrieval results engines documents an criterion rank terms result users queries list relevant fundamental ir

#### Topic 2

clustering between task algorithms with many consequently we has attention analysis precisely approached performance implications 108 heterogeneous metrics received methods

#### Topic 3

tf idf to document frequency text retrieval term word for information or by which number inverse weighting corpus based how

#### Topic 4

clustering task world defined between data as real complex on this approach pitfalls comprehensive much parameters classes choice datasets properties

# Webp

## Clip 'n Search Topics About

## Clip 'n Search Topics About

Search using BM25:

1. sea

2. get

3. view

### Results for "search"

#### Result 1 : Ranking (information retrieval) - Wikipedia ~ Score: 1.3252889191812833

[https://en.wikipedia.org/wiki/Ranking\\_\(information\\_retrieval\)](https://en.wikipedia.org/wiki/Ranking_(information_retrieval))

It is an introduction to ranking. - Ranking of query is one of the fundamental problems in information retrieval (IR).[1] the scientific/engineering discipline behind search engines.[2] Given a query q and a collection D of documents that match the query, the problem is to rank, that is, sort, the documents in D according to some criterion so that the "best" results appear early in the result list displayed to the user. Ranking in terms of information retrieval is an important concept in compute...

#### Result 2 : Okapi BM25 - Wikipedia ~ Score: 1.131281525292409

[https://en.wikipedia.org/wiki/Okapi\\_BM25](https://en.wikipedia.org/wiki/Okapi_BM25)

Explains Okapi BM25 - In information retrieval, Okapi BM25 (BM is an abbreviation of best matching) is a ranking function used by search engines to estimate the relevance of documents to a given search query. It is based on the probabilistic retrieval framework developed in the 1970s and 1980s by Stephen E. Robertson, Karen Spärck Jones, and others. The name of the actual ranking function is BM25. The fuller name, Okapi BM25, includes the name of the first system to use it, which was the Okapi in...

#### Result 3 : Machine learning, explained | MIT Sloan ~ Score: 0.0

<https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained>

Introduction to machine learning, the history of and what it is including types and applications. ~ Machine learning is a subfield of artificial intelligence, which is broadly defined as the capability of a machine to imitate intelligent human behavior. Artificial intelligence systems are used to perform complex tasks in a way that is similar to how humans solve problems. The goal of AI is to create computer models that exhibit "intelligent behaviors" like humans, according to Boris Katz, a princ...

<http://br>

# Use Case

1. Have the extension installed
2. pick a website about a cs 410 topic, let's say "Latent Dirichlet allocation" and navigate to it
3. highlight a portion of the article that is a good representation of most of the topics of the article (or whatever you want to focus on)
4. open the extension
  - a. copy+paste the article portion if it didn't work.
  - b. type something about the article, click on submit. "Successful" should pop up.
  - c. **search for** "latent ranking"
  - d. You will be redirected to webpage where you can search for more articles.
    - i. <http://brysonli.pythonanywhere.com/>
  - e. You will be able to see a list of article excerpts/ some basic info that was submitted
5. Click on the "topics" link in the navigation bar to view some extracted topics