

Proposal for Project - Citation Format Generator for CUHK(SZ)ers

Background & Problem

Every time we are required to write an ENG research paper or a reflective journal or term paper for our general courses, how to do citation is always a big headache. Since there are different types of citation sources and format styles, it is hard for us to remember them all and follow the corresponding regulations.

There indeed exist many online citation machines to meet our demand, but they have many disadvantages. First, citation machines in our own country tends to be simple and incomplete, which may not include the majority of citation sources and format styles. Second, as for foreign citation machines, it is hard for us to open their sites due to low broadband or a lack of VPN. Besides, we may do Chinese and English citations simultaneously when dealing with our papers, which is not achievable for a single citation machine in our country or the others. Furthermore, as current citation machines are set online, it can be easily influenced by other incidental factors. If we design our citation machine to be offline, then we can easily utilize them whenever and wherever we want, regardless of the condition of Internet.

Solution

We want to design an offline version of Citation Format Converter to help CUHK(SZ) students who struggle in the citation work. In our widget, we serve the choice of APA / MLA format, and Book / Article / Website / Journal / Magazine source as well as the corresponding citation guidance. Then we provide the specific blocks for user to enter the source's information. Our widget will generate the "reference" and "citation within paper" for users.

Benefits

1. The offline version could save students' time while using it since it is independent with the network speed.
2. We combine APA and MLA format styles in the same widget, which meets all the demands of CUHK(SZ) students who are struggling with their papers of English or General courses.
3. We distinguish different format styles between Chinese and English, which meets the demands of CUHK(SZ) students who cite both Chinese and English quotations.
4. The widget will provide simple citation example before the user choose the format and source. Thus, the widget could also be used as "reference handbook". Meanwhile, it provides several website links which give detailed tutorials for each citation type.
5. We divide each source into detail. For example, when the user choose APA format and Book source, we then generate corresponding in-text citation as well as the reference.

Functionalities

End Product: Paper

Considering our school students dealing with the paper mostly, we only support this end product.

APA / MLA Guidance:

This serves as our tutorials for every certain format style (APA / MLA) and every certain reference source type (Book / Magazine/ Journal / Website / Newspaper). It will show up Chinese and English versions for each. Meanwhile, we add a website link for every title. The user can right-click it, copy the link address and paste on the safari to open it. Then it will show full and detailed information corresponding to that type. For example, for the type APA — Within a Paper — Book, we show like below:

[APA -- Within a Paper -- Book](#)

Last name, First initial.Middle initial. (Year). Book title. Place: Publisher.

姓名. (年份). 书名. 出版地点: 出版社.

Remark:

1) The first line is the title with a copy link. The second and third lines are Chinese and English versions where every part has been replaced by the part name.

2) The website contains complete information for that type. For example, for APA -- Within a Paper — Book type, it contains detailed categories. See below:

APA -- Within a Paper — Book,

APA -- Within a Paper -- Book — Chapter,

APA -- Within a Paper -- Book -- Corporate or Government Author,

APA -- Within a Paper -- Book -- Dissertation Or Thesis,

APA -- Within a Paper -- Book -- Edited Book,

APA -- Within a Paper -- Book -- Encyclopedia Article — Signed,

APA -- Within a Paper -- Book -- Encyclopedia Article -- Signed — Electronic,

APA -- Within a Paper -- Book -- More than 7 Authors.

However, limited by time, our widget can only support the basic format for a certain type (such as APA -- Within a Paper — Book).

At the naviguide page, we let the user to choose the source type from the following five:

Book / Article / Newspaper / Journal / Magazine / Website.

Once a certain type is choosed, it then immediately shifted to the book page and let the user to input some required information and generate the general formats (In-text Citation and Reference). Take Book as example, the user are required to input the book title, place: publisher, contributors, published year and page.

However, there are several notes here:

1. Editor / Translator Checkbox:

Since the contributor includes the author(s), the editor and the translator, we set different input bars for them. In order to make it simple, we first hide the editor and translator bars as defaulted because most reference books have author(s) while do not have editor and translator. When the reference book the user used have editor or translator, he/she can click the editor or translator checkbox and therefore the hidden bars will show up.

2. Multiple-author, Mutiple-editor, Mutiple-translator Checkbox:

There can be also multiplied authors. The format under this constraint can be different from the basic one. Considering the simplification of inputting and doing later data processing, we set a multiple-author checkbox. If the user clicked it for there are multiple authors, we will pop up a window to remind the user to input the authors' name separated by commas. So as multiple editors and multiple translators.

3. English Checkbox:

Because the Chinese and English characters should be processed differently (otherwise it will turn out to be messy codes), we add a English checkbox, which means the user have to click it when the reference book is an English book.

4. Clear Button:

If the user find the input information incorrect, he/she don't need to delete one by one. He/she can click on the clear button and it will return to the book page without any input.

5. Get Citation Button:

We provide the formats (within Paper and within References) when clicking on the get citation button.

You can choose the format style at the right side. The default version is for APA format style. You can choose to convert to MLA. The generated citations will show up at bottom in the text lines, with some necessary parts italic.

Open Reference File from Zhiwang or Duxiu

For Journal / Newspaper / Magazine, the user have another choice, which he / she don't need to input all the information. As we have investigated, many students prefer to search Chinese papers on 知网(Zhiwang) and 独秀(Duxiu). Therefore, we include a function which can open up a reference txt file from Zhiwang or Duxiu, and then we evoke inside functions in our .cpp/.h files to parse the information in the txt file and generate a correct citation. Since there are several types of reference txt files, we only support the Endnote document export format. Click on the "open reference file" button, you can add document export format txt file. If the type are not corresponded (e.g. You open a magazine format txt file in the book page), a window containing the reminder will pop up.

Search in our Database

As for the Book type, we also include a new function for the user to input incomplete book information and we search in our offline light SQL database. The SQL database includes the basic information of a large amount of Chinese books as well as several English books. Those are got from the Internet, mainly from 豆瓣(Douban). The total number of the books is approximately 20,000+, with 6M size. For example, if the user input "超越死亡" in the book title of the book page, then the widget will automatically filled all the information bars. Then the user just need to click get citation, and all the things can be done at a second. However, due to the limitation of our database, it is possible we cannot find any information related to the user's input.

Back and Close-up

Every single page (not the navigation page) have a back button so that the user can turn back to navigation page to choose another source type and generate next citation. Or the user simply close the widget by clicking on the cross on the up-left corner.

Implementation and Difficulties

GUI Impletation:

The main framework used pointer in C++. Considering the user may not quit from the navigaiton page - close up the widget by clicking the cross on each source type page or just force the widget to quit, we use the pointer to avoid storing in the stack and popping up error (something related to memory allocation) while quitting.

We used object-oriented programming a little while implementing the relationship between the navigation page and the specific source type pages. However, the inheritance and polymorphism were not fulfilled. After finishing the project, we do think it can be simplified if using inheritance since the structure of each page are similar to some extent. We can use virtual methods in the navigaiton page and give detailed and specific implementation in each source type page.

As for the generated citation, it includes the formal font as well as italic font. However, if we deal with QString (the string type in Qt supporting ui interface), it cannot change its attribute to italics. Meanwhile, if we use QTextEdit, it is still a problem that we cannot change the font partly. Therefore, we devide the processed data into two parts and combined them together using `vector<String>` collection.

Due to the limitation of Qt, the user can adjust the window size yet the window cannot largen or shrink dynamically. Therefore, it is possible that the window may be too small to show all the information. We have considered to lock the operation of changing the window size, however, when we run on Mac, it still allows the user to do so depite the code to prohibit this operation. (Though we do not know whether Windows allow it or not.)

Due to the limitation of Qt, it UI interface is imperfect for the contributor part. Because the ideal interface we want is: when we hide the editor and translator bars and texts, we also want the places they occupied can be automatically filled with below parts. But the final interface turned out to be successfully hidden yet just become blank with other parts remained at their original places. We decided to grid layout to solve this problem. But the disadvantage of grid layout is that we cannot overlap each UI component which largely lower the flexibility of our

interface, making it really ugly.

Database Implementation:

When doing the database connection, there are also several difficulties.

Firstly, it is very hard to find resource. APIs from companies (such as Baidu Xueshu, Amazon, Douban, Google Scholar, Dangdang, etc.) were closed or not provided or not free though we write many emails to them. Meanwhile, those web crawlers are slow and hard to use. Meanwhile, as we all are new to database, the building and connection have failed many times because of encoding error. Especially for Chinese character case, since SQLite is UTF-8 supported only, we have to handle it specially.

Furthermore, intergrating QtSQL and making it run on other devices (such as Mac) is also a headache. It has to devide the situation into different conditions and consummate the coding.

Fail to Connect Online Search:

We actually want to do online citation generator at the mean time, yet failed.

The goal we want to achieve is that we search the incomplete input information online and return back the correct citation. This can breakthrough the limitaion of offline database we built since the database has limited book information and few English books. However, the APIs from companies (such as Baidu Xueshu, Amazon, Douban, Google Scholar, Dangdang, etc.) were closed or not provided or not free. We tried to contact with our school library and they cannot provide such kind of API for us to search because they subscribe Endnote yearly and Endnote can provide all the information.

We have also considered to implement this function by making a simple http request. However, the http request tends to have access to the original html code of the website. We can only retrieve the webtitle, author and time by recognizing the labels in the html code. But it may not be accurate since it can grab any title, name and time it first encountered. As there maybe more than two titles/names/time, it may return back wrong information. This is what we don't want to see. Therefore, we discarded this idea and decided not to connect online search.

Assignment of responsibility

Ye Haolin and Yu Jiahan are responsible for the GUI design and data parsing. Ye Haolin implemented the parsing functions mostly, especially for the special case for Chinese. Yu Jiahan designed the whole UI interface mostly and define each buttons function as well as the jump terms.

Min Tian are responsible for building a light SQL database and loading the book information (20,000+ books, 6M) into the built database. Meanwhile, the database connection and necessary supporting functions are implemented by Min Tian.

The contribution percentage of our team is approximately:

Ye Haolin 40%

Yu Jiahan 30%

Min Tian 30%

Declaration:

Except for the usage of STL and stanford library in C++, we three do all the coding. We do not use outside libraries as well as others' code for our implementation.

Members of Project Team

Ye Haolin (116010266)

Yu Jiahan (116010274)

Min Tian (116010168)