

# EE101 Final Project Report

Yifu Chen, Jialong Guo , Ziliang Guo, Aofan Jiang

2019 年 6 月 22 日



# 目录

|  |           |
|--|-----------|
| <b>1 Overview</b>                                | <b>1</b>  |
| 1.1 Project File Tree . . . . .                  | 1         |
| 1.2 Develop Environment . . . . .                | 1         |
| <b>2 Front-end</b>                               | <b>2</b>  |
| 2.1 Overview . . . . .                           | 2         |
| 2.1.1 Basic Idea . . . . .                       | 2         |
| 2.1.2 Process of development . . . . .           | 2         |
| 2.2 Index.php . . . . .                          | 3         |
| 2.2.1 Logo . . . . .                             | 3         |
| 2.2.2 Greeting Words . . . . .                   | 3         |
| 2.2.3 Search box & buttons . . . . .             | 3         |
| 2.2.4 Copyright & Related Websites . . . . .     | 6         |
| 2.3 Search.php . . . . .                         | 7         |
| 2.3.1 Navigation Bar . . . . .                   | 7         |
| 2.3.2 Body Part & Text . . . . .                 | 7         |
| 2.3.3 Paper Turning & "Jump to" Button . . . . . | 8         |
| 2.4 Conference.php . . . . .                     | 8         |
| 2.5 Author.php . . . . .                         | 9         |
| 2.6 Title.php . . . . .                          | 10        |
| 2.7 Acknowledgment . . . . .                     | 11        |
| <b>3 Graphs</b>                                  | <b>12</b> |
| 3.1 Overview . . . . .                           | 12        |
| 3.2 Design . . . . .                             | 12        |
| 3.3 Searching Data . . . . .                     | 13        |
| 3.3.1 From Mysql to Solr . . . . .               | 13        |
| 3.3.2 Searching in Solr . . . . .                | 14        |
| 3.4 Formatting Data . . . . .                    | 14        |
| 3.5 Drawing Graph . . . . .                      | 14        |
| 3.6 Fruits' Display . . . . .                    | 14        |
| <b>4 Overview</b>                                | <b>16</b> |
| <b>5 Keyword Highlighting</b>                    | <b>17</b> |

|          |  |           |
|----------|--|-----------|
| <b>6</b> | <b>Hyperlinks</b>                      | <b>18</b> |
| 6.1      | Hyperlink of each title . . . . .      | 18        |
| 6.2      | Hyperlink of each conference . . . . . | 19        |
| 6.3      | others . . . . .                       | 19        |
| <b>7</b> | <b>Code optimization</b>               | <b>20</b> |
| 7.1      | Solr . . . . .                         | 20        |
| 7.2      | Mysql . . . . .                        | 20        |
| <b>8</b> | <b>Contact and Open source</b>         | <b>21</b> |

# **1 Overview**

## **1.1 Project File Tree**

## **1.2 Develop Environment**

## 2 Front-end

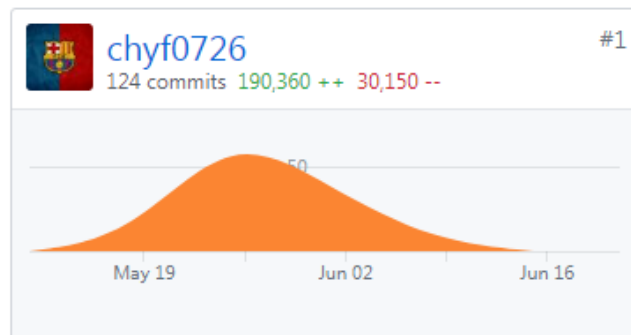
written by Yifu Chen

### 2.1 Overview

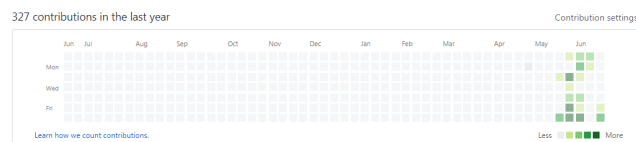
#### 2.1.1 Basic Idea

I was in charge of the front-end part.I mainly used CSS and BOOTSTRAP to beuify our websitek.In my opinion,I hope my websites be plain and straight-forward,so I did not decorate our websites deliberately,and this my idea of designing the layout fo our websites.

#### 2.1.2 Process of development

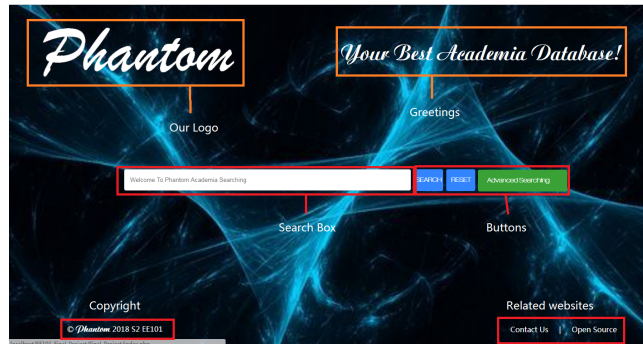


We took the advantage of *Github* to promote our cooperation.During the period, I wrote down about 200,000 lines of codes which was mostly written independently,and I was the number 1 contributor in our team.



I started my job on 26th May and ended on 8th June and I will introduce my job in the following part.

## 2.2 Index.php



Our index page is shown above. First of all, I would like to introduce the process of my designing our home page. At first, there were three search boxes in the page. The search box "Author", "Title" and "conference", and the layout of the index.php was settled. Then, we decided to use multi-searching. Therefore I cut down the number of search boxes into one. Finally, I polished the index.php and the page became what you can see now.

The elements the index page consists of was shown in the graph above, and I am going to introduce every part respectively.

By the way, the favicon of our websites was *Raffaello's The School of Athens*.

### 2.2.1 Logo

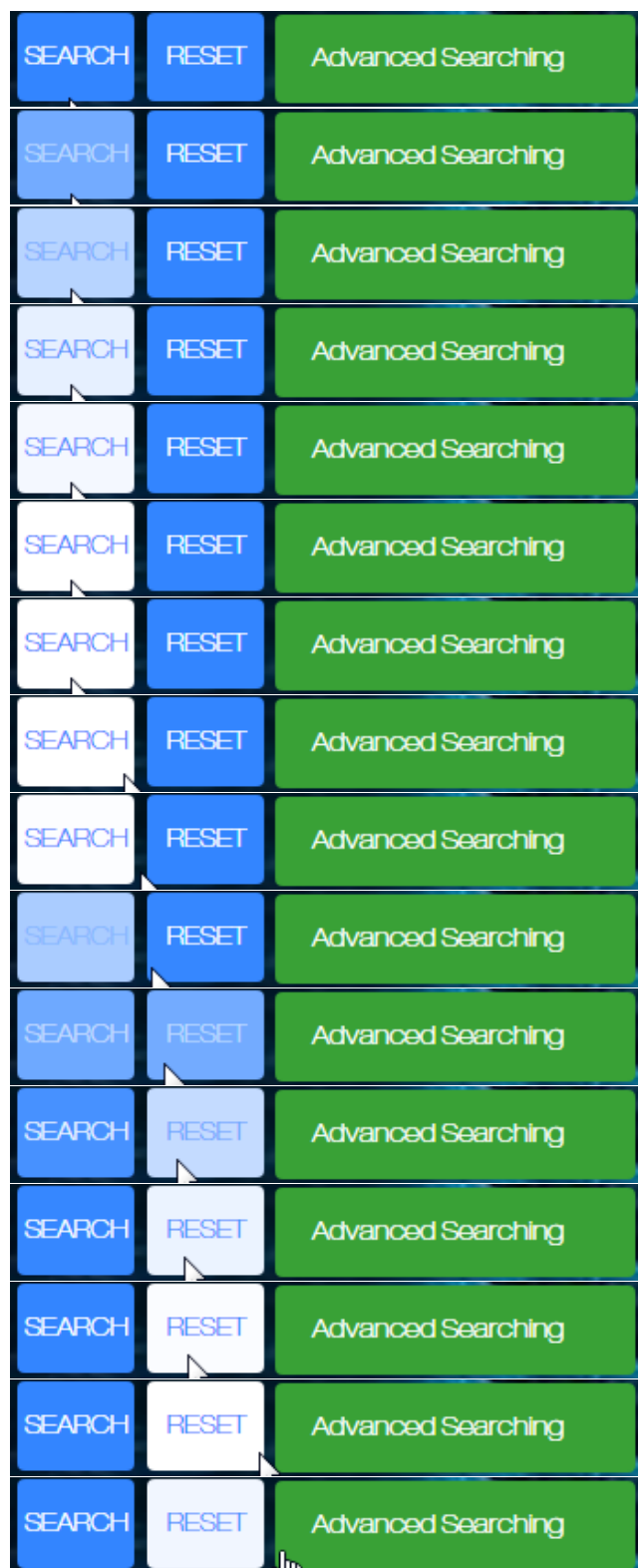
The name of our search engine came from *The Phantom of the opera*, one of my favorite films, and I hope that the speed of our searching engine can be as fast as a phantom. It took me so much time to find an appropriate font to display our logo. Finally I found out a font "书体坊兰亭体". However, the words could not be shown in terms of vector graph, which meant that the edge of the words were not smooth, and that is a defect of our logo.

### 2.2.2 Greeting Words

To display the greeting words, I also spent plenty of time to find an appropriate font. Finally, I found a font named "ChannelSlanted2" to present the greeting words.

### 2.2.3 Search box & buttons

The search box and the buttons are the key part of the page, and I beautified the buttons, and the result was shown in the following pictures.









#### 2.2.4 Copyright & Related Websites

These two parts were mainly written by Ziliang Guo and Aofan Jiang and Hyperlinks were set on the texts "Contact Us" and "Open source".

## 2.3 Search.php

I am going to introduce every part of search.php respectively.

Phantom navigation bar Welcome To Phantom Academia Searching Search

Multi Field Search: jiawei han

body part

| Title  | Authors                                      | Conference |
|--|--|------------|
| mining heterogeneous information networks  | jiawei han                                   | SIGKDD     |
| mining frequent patterns by pattern growth methodology and implications                            | jiawei han, jian pei                         | SIGKDD     |
| mining heterogeneous information networks a structural analysis approach                           | yiyou su, jiawei han                         | SIGKDD     |
| closegraph mining closed frequent graph patterns   | xifeng yan, jiawei han                       | SIGKDD     |
| dynamic generation and refinement of concept hierarchies for knowledge discovery in databases      | jiawei han, yongjian fu                      | SIGKDD     |
| chinese-japanese cross language information retrieval a han character based approach               | yui matsumoto, md maruf hasan                | ACL        |
| resource and knowledge discovery in global information systems a preliminary design and experiment | osmar r. zaslav, jiawei han                  | SIGKDD     |
| on trivial solution and scale transfer problems in graph regularized nmi                           | chris ding, jiawei han, quanquan gu          | UICAI      |
| advances of the oblearn system for knowledge discovery in large databases                          | simon tang, jiawei han, yongjian fu          | UICAI      |
| classifying large data sets using svms with hierarchical clusters                                  | jiong yang, jiawei han, hwanqiao yu          | SIGKDD     |
| an efficient multi-relational naive bayesian classifier based on semantic relationship graph       | xiaohu yin, jiawei han, hongyan liu          | SIGKDD     |
| collective topic modeling for heterogeneous networks   | jiawei han, bo zhao, hongbo deng             | SIGIR      |
| tensor space model for document analysis   | jiawei han, deng cai, xianlei he             | SIGIR      |
| clustering moving objects  | yilan li, jiong yang, jiawei han             | SIGKDD     |
| metarule guided mining of multi dimensional association rules using data cubes                     | micheline kamber, jiawei han, jennifer chang | SIGKDD     |
| robust tensor decomposition with gross corruption  | huan guo, jiawei han, quanquan gu            | NIPS       |
| trust analysis with clustering   | marish guo, jiawei han, yiyou su             | WWW        |
| mining event periodicity from incomplete observations  | jingqiang wang, zhenhui li, jiawei han       | SIGKDD     |
| ranking based clustering of heterogeneous information networks with star network schema            | jiawei han, yiyou su, yintao yu              | SIGKDD     |
| ranking based classification of heterogeneous information networks                                 | marina danielovskiy, jiawei han, ming ji     | SIGKDD     |
| building enriched web page representations using link paths  | tim werninger, jiawei han, chengxiang zhai   | WWW        |
| parallel mining of closed sequential patterns  | jiawei han, david padua, shengnan cong       | SIGKDD     |
| spectral regression for efficient regularized subspace learning                                    | jiawei han, deng cai, xianlei he             | ICCV       |
| sparse projections over graph  | jiawei han, deng cai, xianlei he             | AAAI       |
| closest searching for the best strategies for mining frequent closed itemsets                      | jianying wang, jiawei han, jian pei          | SIGKDD     |

Found 401 results. Each page: 25 items. Altogether: 17 pages.

Text

Phantom

Page turning 1 2 3 4 5 6 7 8 9 10 Next

\*Jump to\* Button

Jump to: Go

### 2.3.1 Navigation Bar

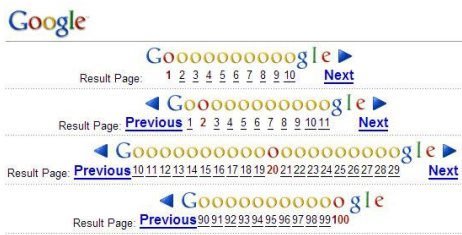
I used css and bootstrap to construct the navigation bar. I thought there was no need to build a that complicated navigation bar, so I just put our logo in the top left corner of the screen and set it with a hyperlink to the homepage, and in the top right corner of the screen was a search box to conduct multi-search.

### 2.3.2 Body Part & Text

I selected the font "Regencie" to beautify the table and "书体坊赵九江钢笔楷书" to beautify the texts.

2.3.3 Paper Turning & "Jump to" Button

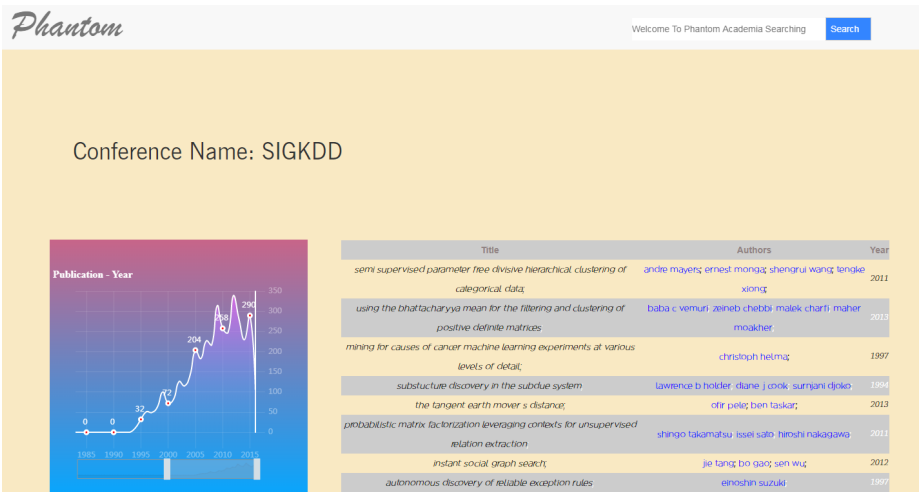
Our idea to design the pattern of the Paper Turning is to imitate *Google's* pattern.

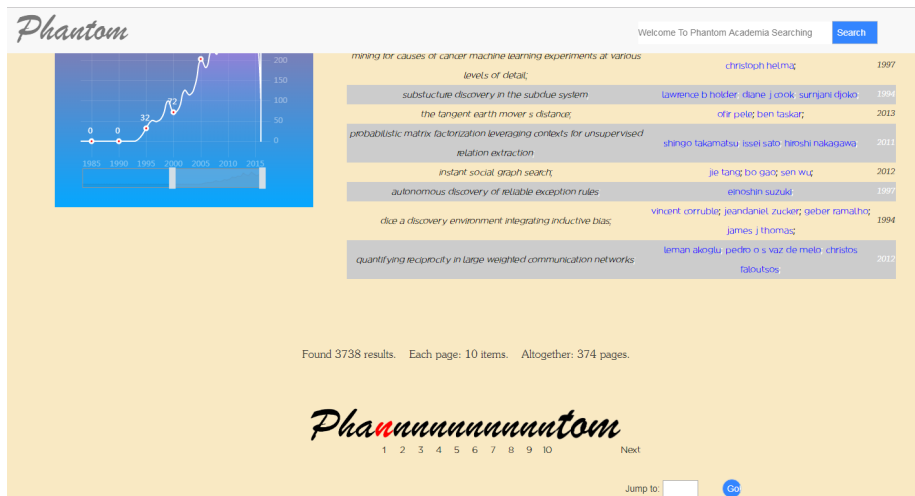


Therefore,I photoshopped some pictures and realized this idea.

I beautfied the "Jump to" box,and the button "Go!" was modified by Aofan Jiang.

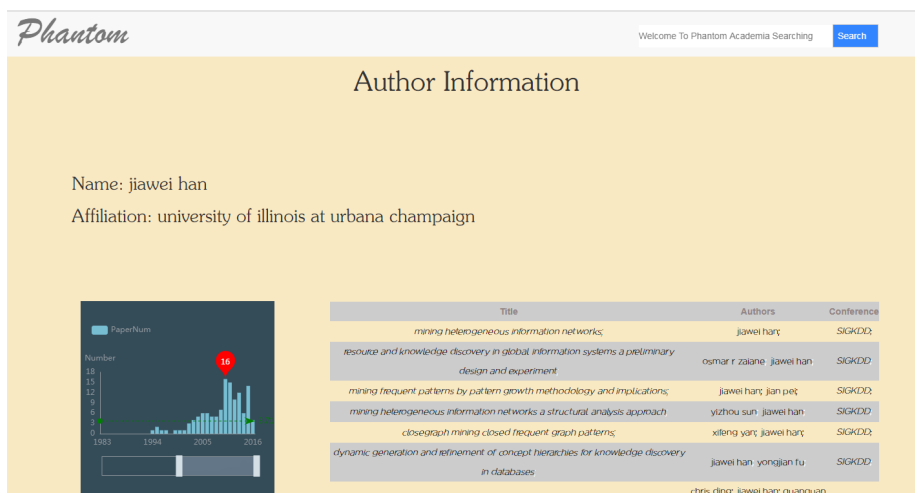
2.4 Conference.php

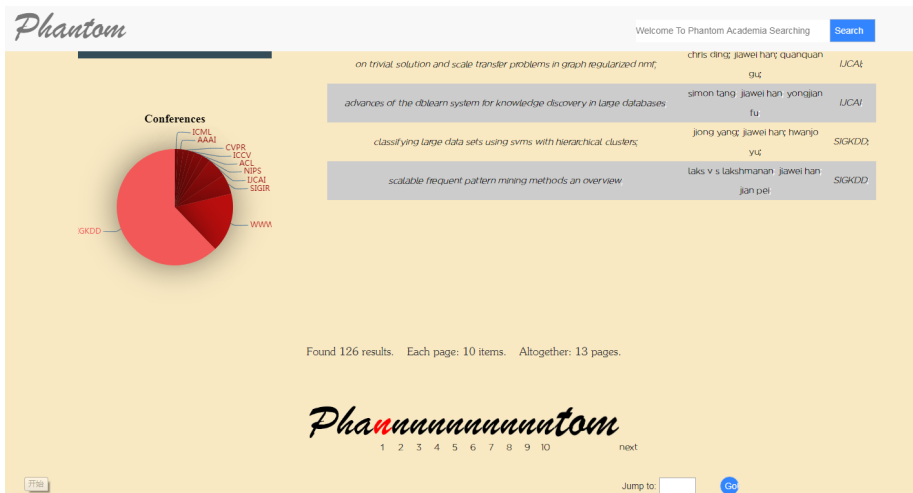




The beautification of Conference.php was similar to the search.php. Therefore I will not go into details here.

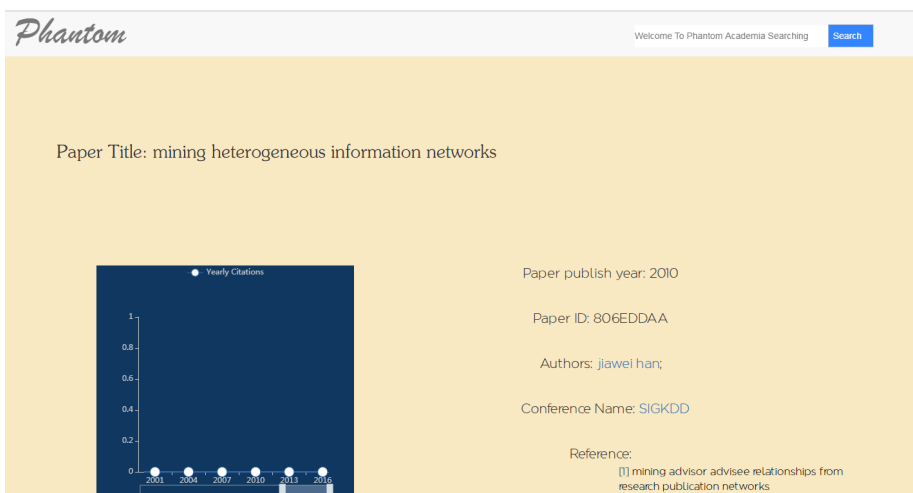
## 2.5 Author.php

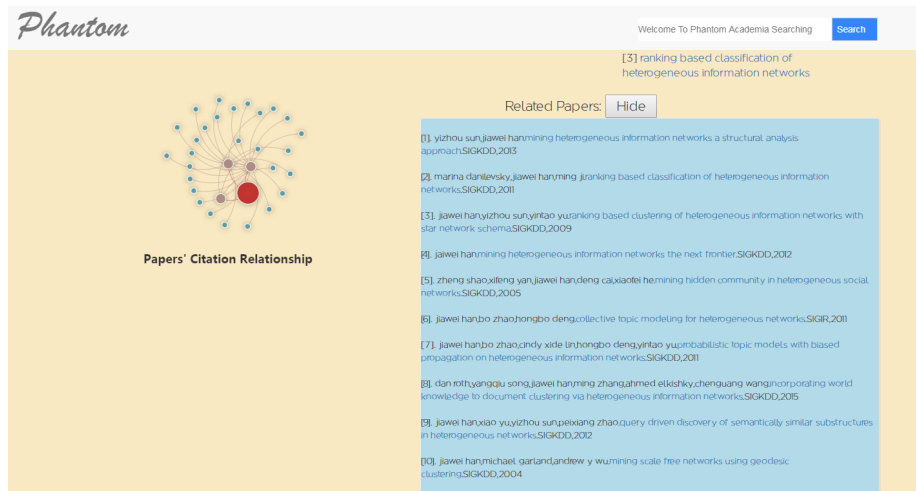




The beautification of the page author was similiar to the search.php. Therefore I will not go into details here as well.

## 2.6 Title.php





I set the texts of the website in a conspicuous position, and the all the texts in the page were shown in selected fonts. I beautified the "Show" button as well.

## 2.7 Acknowledgment

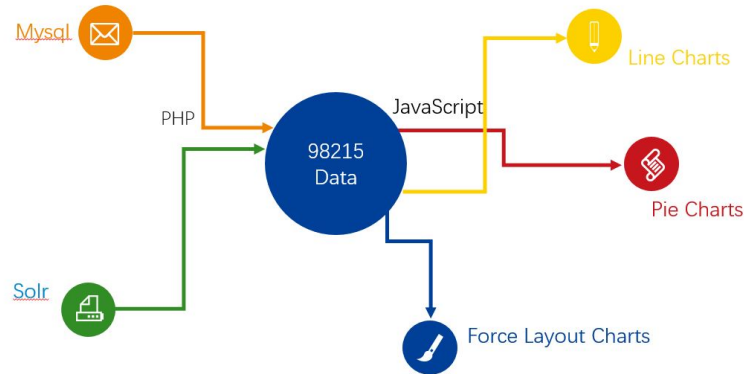
I would like to talk about what I learnt after taking the course *Introduction to Engineering for Electronic Information (B)*. Well, I have to say that this course is a little challenging. The 3 labs and the final projects all covered the fields that I was not familiar with or even never heard about. All these tasks needed me to solve tons of problems and many of them made me mad. However, I got through all these problems with my own effort and the help from my roommates, my classmates and TAs. So here I want to say "Thank You" to all that helped me. I especially want to express my gratitude to TA *Hui Xu*. It was his help and patience that benefited me a lot. I remembered that I met a problem in lab3, and I spent about 4 days solving the problem, but it was still unsolved. Then I turned to TA *Xun*, and we solved the problem together and finally worked it out.

Then I would like to talk about my work in the final project. As my roommate took the compulsory parts of the final projects, I "had to" do the beautification. I do not think beautification is a job of difficulty, but it required patience. I usually spent plenty of time searching for a awesome fonts and racked my brains to think how can do the webpage better, but when writing this report, I really found it difficult to present my job, for what I have done are only about the layout of the website and some trivial details. I cannot write something like "I used xxx to quicken the search" or "I modified my algorithm with this bright idea"

## 3 Graphs

[Start] — By Jialong Guo 518030910272 —

### 3.1 Overview



When it comes to our graph part, it can be chiefly divided into three parts:

The first of which is to get all the data we need in order to draw a graph.

The second part of which is to format all the data into the form that the corresponding graph is needed.

The last part of which is just to use specific javascript library to create graphs and make them beautiful.

### 3.2 Design

In author page' s left part, I add a column graph, which displays the number of publication the author has yearly and shows the trend of his/her academic activity to give users a clear impression about the author' s research fruit and the frequency of his/her delivery.

In author page' s left part, I also add a pie graph, which showcases how many papers the author publishes in a conference which brings a clear view on the author is mostly connected with what conference.

In paper page's upper part, there' s a line-collumn graph showing its yearly



citations, from which we can gain a insight into the popularity of its research field.

In paper page's lower part, I use a force directed graph to display the relations between similar papers. It gives a convience way to find related papers and messages.

In conference page, a line graph of its yearly amount of papers may reveal its academic influence.

### 3.3 Searching Data

In this subsection I mainly discuss the first part of graph drawing, get data from database. Since diverse graphs need diverse data, here I just demonstrate a specific example of getting data, which can mostly stand for my means of searching data.

#### 3.3.1 From Mysql to Solr

Taking time cost into consideration, we may need to import data into solr from mysql in advance, for searching from solr will spend more time than from mysql. Thereby, it is required that we write the referenceID of reference papers into solr's schema, which allows us to search the reference papers of a paper just from solr when needed.

The final data we put in solr is formed in this way:

```
with codecs.open(FP_out, 'w', 'utf-8-sig') as f:
    data = {"PaperID": result[0][0],
            "Title": result[0][1],
            "Authors'ID": [result[0][2]],
            "Authors'Name": [result[0][3]],
            "ConferenceID": result[0][4],
            "ConferenceName": result[0][5],
            "Year": result[0][6],
            "ReferenceID": [result[0][7]]}
    # "AffiliationID": [result[0][8]]}
    for i in range(1, len(result)):
        out_print = False
        if result[i][0] == data["PaperID"]:
            if (result[i][2] not in data["Authors'ID"]):
                data["Authors'ID"].append(result[i][2])
            if (result[i][3] not in data["Authors'Name"]):
                data["Authors'Name"].append(result[i][3])
            if (result[i][7] not in data["ReferenceID"]):
                data["ReferenceID"].append(result[i][7])
```

You can refer to the codes attached for detailed information.

### 3.3.2 Searching in Solr

The first step is getting value form user's input, then create the url link to search in solr. After this step, we can get a .json file with the result.

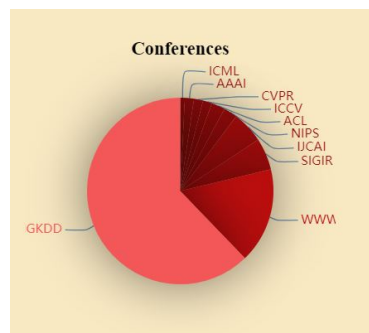
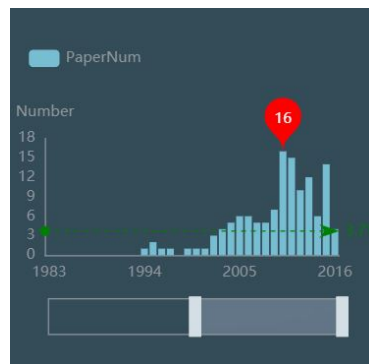
## 3.4 Formatting Data

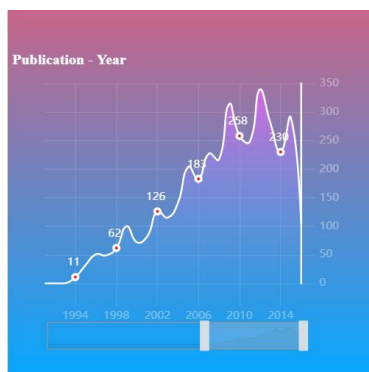
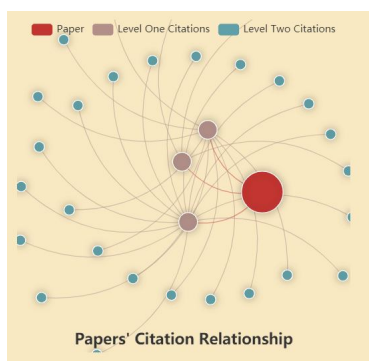
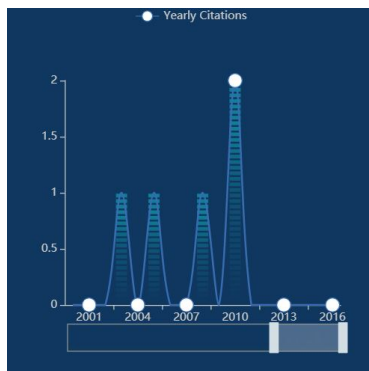
In this section, we mainly discuss how to change the search result into suitable formation of drawing a graph.

## 3.5 Drawing Graph

In the section, the process of drawing with echarts is mainly discussed.

## 3.6 Fruits' Display





[End] — By Ziliang Guo 518030910273 —

## 4 Overview

*[Start] — By Ziliang Guo 518030910273 —*

(1) I took the initiative that we take full advantage of Github to accelerate our project. I also create a document to take notes of the problems we met and the solutions.

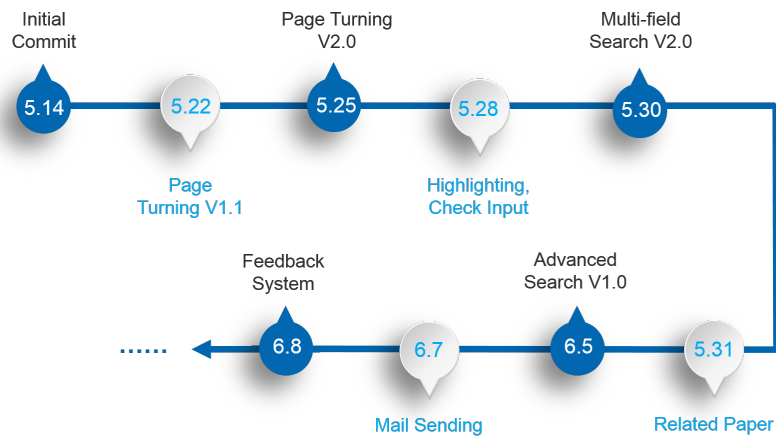
(2) Actually, I wrote the manual and uploaded my Lab 01 - 03 codes to unify the databases.

(3) I mainly focus on the back-end development.

(4) Of all my codes, I wanna highlight that approximately 85% are mainly created independently. For the remaining codes, modification is applied, with reference to some online blogs.

(5) Meanwhile, during my coding, I always remember to leave interfaces for my collaborators.

(6) As is vividly depicted in the timeline graph, I realized and improved different sections separately, in other words, term by term. Of course, my constant improvements are shown.



## 5 Keyword Highlighting

I adopted the “hl” settings of Solr. It is somehow very simple. Just echo the corresponding urls will do.

However, please notice that, for multivalued fields such as Authors\_Name, only the highlighted part is returned. So I made judgements in such special cases.

Codes:

```
$url = "http://localhost:8983/solr/
lab02/select?indent=on&q=Title:". $query. "
^1+OR+Authors_Name:". $query. "^0.7+OR+ConferenceName:
". $query. "^0.5&start=" . ($page_limit*($page-1)). "
&rows=". $page_limit. "&wt=json&hl=on&hl.fl=Title,Auth
ors_Name,ConferenceName&hl.simple.post=<%2Fb><%2Ffont>&h
l.simple.pre=<font%20color%3D%23FF0000><b>" ;
```

*[End] — By Ziliang Guo 518030910273 —*

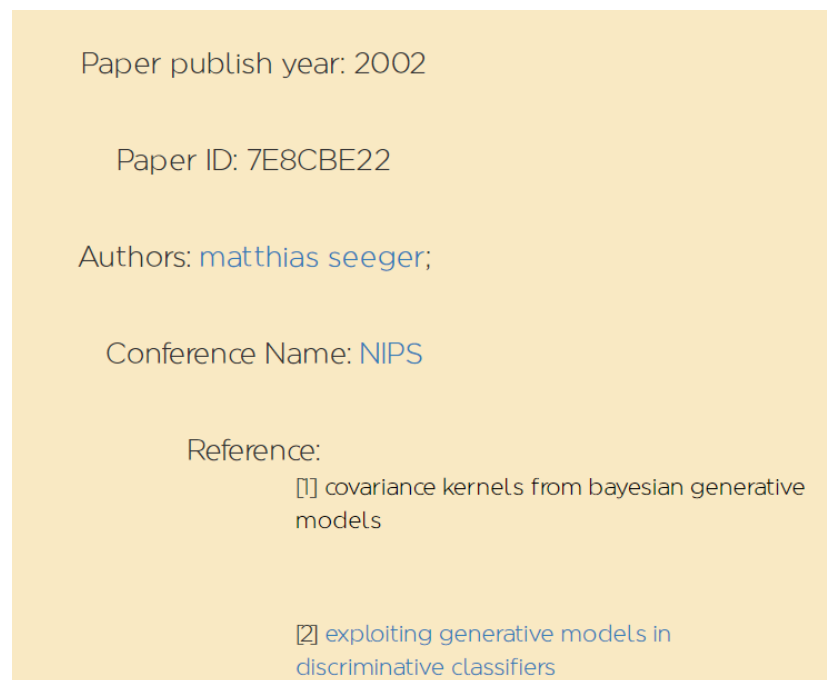
## 6 Hyperlinks

— *By Aofan Jiang 518030910275* —

As required in the final project, I add the hyperlink of each title, author and conference. So the users can click to get more information about the result they want to search for.

### 6.1 Hyperlink of each title

Different from the table about the information of key words, the title page is simply a series of main information about the corresponding paper. It includes the paper ID, the publish year, the authors, conference name, and reference papers. This can make the whole page clearer, just as the following picture.



It might be ignored that the table of reference paper title of each paper is given in the original data. Although it is not required in the lab two, I still add it as a field in Solr to make sure that more detailed information can be gotten by the users. This explains the existence of references in the paper page.

What's more, to make a user-friendly website, you can see in the following picture that at the top of website, the link information about the title. It's also

clear and easy to find the information. It' s linked by the symbol of addition.

localhost/EE101-Final\_Project/Final\_Project/title.php?title=learning+to+learn+with+the+informative+vector+machine

On the comparison, you can see even in Baidu NetDisk that the link is not so comfortable as ours. They are all connected with %2 while our website link are connected with + . Since all the information is imported in solr, the searching speed is quite fast with no delay.

=Auth%20Login%20Success&&bduss=&ssnerror=0&traceid=#/all?vmode=list&path=%2F我的资源%2F冰与火：权力的游戏1-6

## 6.2 Hyperlink of each conference

This page is also mainly an information table like the main searching page. It includes each paper' s title, authors published on the given conference name. However, at the last column of table, the original conference name is replaced by the publish year of each paper published in this conference.

| Title   | Authors  | Year |
|---|--|------|
| <i>von mises fisher clustering models;</i>  | siddharth gopal; yiming yang;  | 2014 |
| <i>mixed membership matrix factorization</i>  | lester mackey; michael i jordan; david j weiss;                            | 2010 |
| <i>a randomized anova procedure for comparing performance curves;</i>                               | michael atighetchi; justus plater; xiaoqin zhang; paul r cohen;            | 1998 |
| <i>path normalcy analysis using nearest neighbor outlier detection;</i>                             | muthukumaran chandrasekaran; david lupar; khaled rasheed; hamid r arabnia; | 2008 |
| <i>distributed stochastic gradient mcmc;</i>  | babak shahbaba; max welling; sungjin ahn;                                  | 2014 |
| <i>regularization of neural networks using dropconnect;</i>   | li wan; sixin zhang; rob fergus; yann le cun; matthew d zeller;            | 2013 |
| <i>automated cephalometric landmark localization using sparse shape and appearance models;</i>      | dirk vandermeulen; johannes keustermans; paul suetens; dirk smeets;        | 2011 |
| <i>semi supervised learning through principal directions estimation</i>                             | bernhard scholkopf; olivier chapelle; jason weston;                        | 2003 |
| <i>forgetting counts constant memory inference for a dependent hierarchical pitman yor process;</i> | nicholas bartlett; david pfau; frank wood;                                 | 2010 |

## 6.3 others

Even at different pages among paper, author, conferences. Almost all the items are hyperlinked. So it means you can jump to different pages in any given page. Here comes an example of authors' information

| Title   | Authors                              | Conference |
|---|--------------------------------------|------------|
| mining heterogeneous information networks;  | jiawei han;                          | SIGKDD;    |
| resource and knowledge discovery in global information systems a preliminary design and experiment; | osmar r zaiane; jiawei han;          | SIGKDD;    |
| mining frequent patterns by pattern growth methodology and implications;                            | jiawei han; jian pei;                | SIGKDD;    |
| mining heterogeneous information networks a structural analysis approach                            | yizhou sun; jiawei han;              | SIGKDD;    |
| closegraph mining closed frequent graph patterns;   | xifeng yan; jiawei han;              | SIGKDD;    |
| dynamic generation and refinement of concept hierarchies for knowledge discovery in databases;      | jiawei han; yongjian fu;             | SIGKDD;    |
| on trivial solution and scale transfer problems in graph regularized nmf;                           | chris ding; jiawei han; quanquan gu; | IJCA;      |

## 7 Code optimization

### 7.1 Solr

By testing, I find that the searching speed by solr is faster than mysql. So all the places that can be used in solr are changed from mysql.

### 7.2 Mysql

When using mysql, there are still some methods to improve the speed of program.

If the searching result is only one piece. For instance, the publish year, the affiliation and the conference and so on. We can limit the searching result numbers by add the code “LIMIT 1 ” . As a result, when mysql get one information about the result, it will stop the searching process, which is a great improvement in the speed of a program.

```
$result = mysqli_query($link, "SELECT AuthorName from authors where AuthorID='$author_id' limit 1");
```

Instead of using inner join, try to use more simple searching and output the result together is faster.

```
$affi_id_name_result = mysqli_query($link, "SELECT affiliations.AffiliationID, affiliations.AffiliationName from (select AffiliationID, count(*) as cnt from paper_author_affiliation where AuthorID='$author_id' and AffiliationID is not null group by AffiliationID order by cnt desc) as tmp inner join affiliations on tmp.AffiliationID = affiliations.AffiliationID");
```

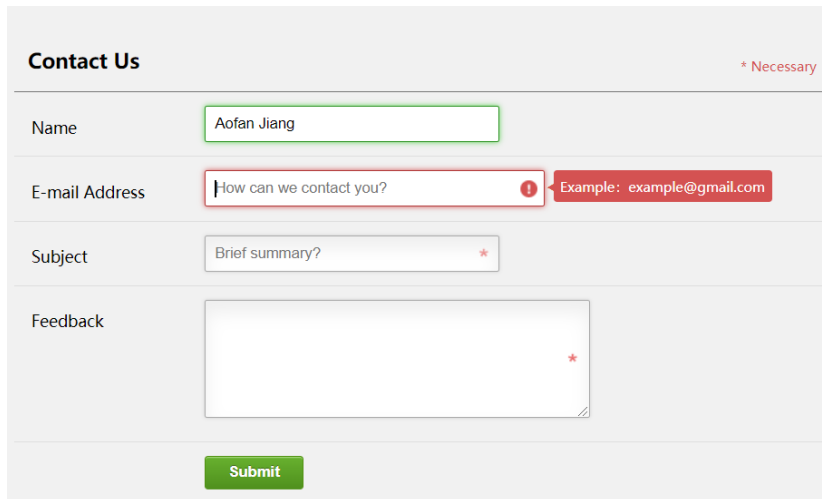
```
$affi_id_row=mysqli_fetch_row(mysqli_query($link,"SELECT AffiliationID, count(*) AS count FROM paper_author_affiliation where AuthorID=$author_id GROUP BY AffiliationID ORDER BY count DESC LIMIT 1"));
$affi_id=$affi_id_row[0];
$affi_id_name_result=mysqli_query($link,"SELECT AffiliationName from affiliations where AffiliationID='$affi_id'");
```

As for the code in php, we can choose to use mysqli\_fetch\_row rather than mysqli\_fetch\_array since the information class is of each column is known to us.



## 8 Contact and Open source

At the home page, you can find the word “contact us ” . After clicking it, you will jump to a new page with a sophisticated chart. You can input your information and your information, we will get your feedback after you clicking the submit button.



The image shows a 'Contact Us' form with the following fields and elements:

- Title:** 'Contact Us' with a red asterisk and the text '\* Necessary'.
- Name:** A text input field containing 'Aofan Jiang'.
- E-mail Address:** A text input field containing 'How can we contact you?'. To its right is a red box with a white exclamation mark icon and the text 'Example: example@gmail.com'.
- Subject:** A text input field containing 'Brief summary?' with a red asterisk to its right.
- Feedback:** A large text area with a red asterisk to its right.
- Submit:** A green button labeled 'Submit'.

Also, you can find the word ” open source ” . After clicking it, you will jump to our project on Github website. At here, you can check all the detailed codes and corresponding documents.

SJTU 2019 S1 EE101 Final Project

|  |                              |             |                |
|--|------------------------------|-------------|----------------|
| 312 commits  | 2 branches                   | 0 releases  | 4 contributors |
| Branch: master   |                              |             |                |
| Create new file Upload files Find File Clone or download |                              |             |                |
| Sjtugil 图片美化 Latest commit ecbe35 12 days ago            |                              |             |                |
| Documents  | cyf: advanced_search的解决方案及原因 | 12 days ago |                |
| Final_Project  | 图片美化                         | 12 days ago |                |
| lab01  | 将ReferenceID导入solr           | 16 days ago |                |
| lab02  | cyf                          | 13 days ago |                |
| lab03  | Add files via upload         | last month  |                |
| README.md  | Initial commit               | last month  |                |
| github-stats.py  | Bug Fix & Stats              | 21 days ago |                |
| question.txt   | cyf                          | 21 days ago |                |
| relations.gexf   | title 最后一个图                  | 13 days ago |                |
| test.txt   | Update test.txt              | last month  |                |
| 项目进度.txt   | cyf                          | 19 days ago |                |