

Interactive web-based system for examining and comparing academic influence for Computing and Information School

INFSCI 2950 - Independent Study

Zijing Gao
ZIG13@pitt.edu

Supervisor: Dr. Peter Brusilovsky

I. Project Introduction

The project is a Web-based system for examining and comparing research performance of Computing and Information Schools, which stores research performance data of several peer schools. The system provides interface for maintenance staff to review and edit this information since it could change. It also compares research performance data of several peer schools from several aspects and generate dynamic graphs.

This system allows SCI administration and faculty to compare the performance of SCI in research with performance of peer schools.

II. Data preparation

1. Dataset description

The dataset is crawled from google scholar with given crawling script from my supervisor Dr.Peter Brusilovsky. The dataset size is 277.

2. Dataset preparation

I did some data pre-processing on the original dataset.

1) Choose required features of data. This project chooses 10 features of faculties, including name, gender, link, citation, h index, i10 index, first publication year, school, department and job.

2) Drop unneeded features.

3) Organizing selected feature. In this process I did work to organize the data into a uniform and usable format for importing to the database later.

4) Combing the data files of each peer school into one dataset and added school feature.

III. Database Design

1. Environment

This system uses MySQL on XAMPP php admin and uses sql for query.

2. Data Dictionary

Faculty

Field Name	Type	Field Size	Required?	Desicription
f_id	INT	11	y	Unique number for each faculty
name	VARCHAR	500	n	Name of the faculty

gender	VARCHAR	100	n	Gender of the faculty
link	VARCHAR	1000	n	Home page link of the faculty
citation	INT	11	n	Citation number of the faculty
h_index	INT	11	n	H index of the faculty
i10_index	INT	11	n	I10 index of the faculty
first_pub	INT	11	n	The first year of publication of the faculty

School

Field Name	Type	Field Size	Required?	Desicription
s-id	INT	11	y	Unique number for each school
title	INT	500	n	Name of the school

Department

Field Name	Type	Field Size	Required?	Desicription
d-id	INT	11	y	Unique number for each department
title	INT	500	n	Name of the department

Job

Field Name	Type	Field Size	Required?	Desicription
j-id	INT	11	y	Unique number for each job
title	INT	500	n	Name of the job

Faculty-School

Field Name	Type	Field Size	Required?	Desicription
fs-id	INT	11	y	Unique number for each relationship between faculty and school
f-id	INT	11	n	Unique number of the faculty
s-id	INT	11	n	Unique number of the school

Faculty-Department

Field Name	Type	Field Size	Required?	Desicription
fd-id	INT	11	y	Unique number for each relationship between faculty and department
f-id	INT	11	n	Unique number of the faculty
d-id	INT	11	n	Unique number of the department

Faculty-Job

Field Name	Type	Field Size	Required?	Desicription
fj-id	INT	11	y	Unique number for each relationship between faculty and job
f-id	INT	11	n	Unique number of the faculty
j-id	INT	11	n	Unique number of the job

IV. System development

1. Technology Description

This system uses HTML, CSS, Javascript and jQuery with Bootstrap for front end development. The backend of the system uses PHP and PDO connecting to MySQL database.

The development environment is XAMPP Server.

2. Interface and Function Description

1) Home Page

As shown in figure1, this is the home page of the website. It is designed in a simple and clean style. The button “Go Compare” can navigate to the main function page of the website. Figure 2 shows the content when the user scroll down on the home page. There are description and shortcuts to the four sections of the website: Compare, Faculty, School and Department.

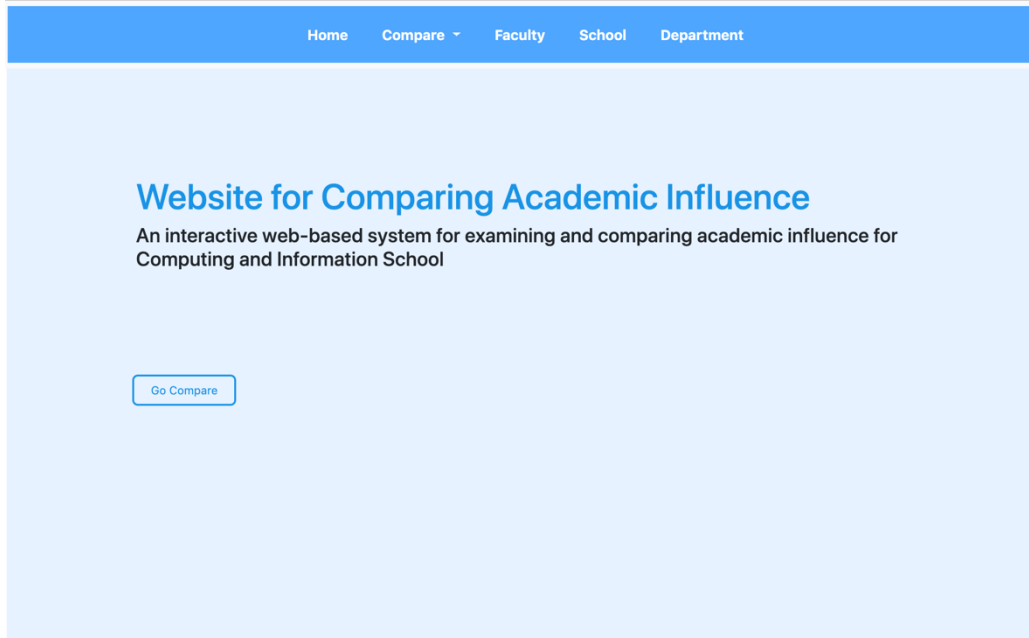


Figure1 Home Page

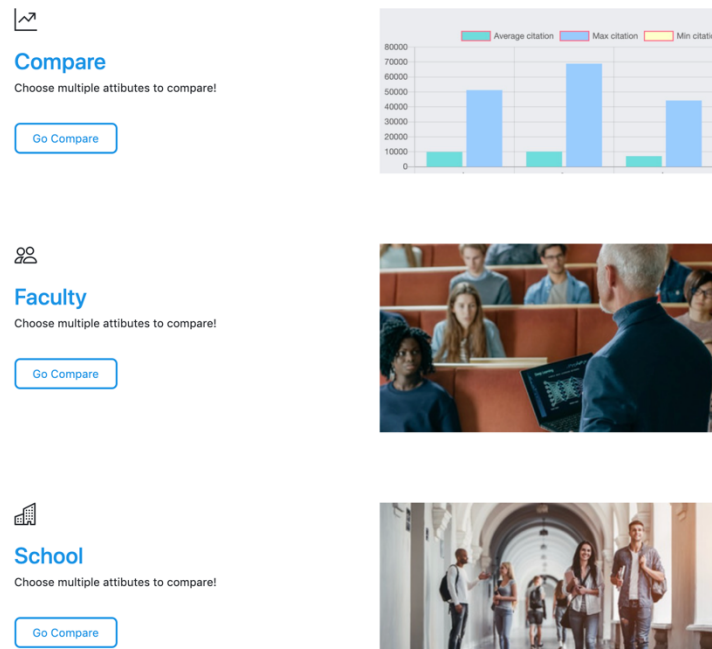


Figure2 Home Page Part2

2) Compare Page

The Compare page provides main functions of the website - examining and comparing research performance of Computing and Information Schools. In this section, the research performance data including comparing citation number, h index, i10 index and first year of publication of faculties are compared by two aspects, school and department. This section also generates dynamic graphs for

user to have a more visual comparison. The following two figures are two comparison examples. Figure3 is the comparing output graph and data list for average and min citation of faculties divided by schools. Figure4 is for average h index and i10 index of faculties divided by departments.

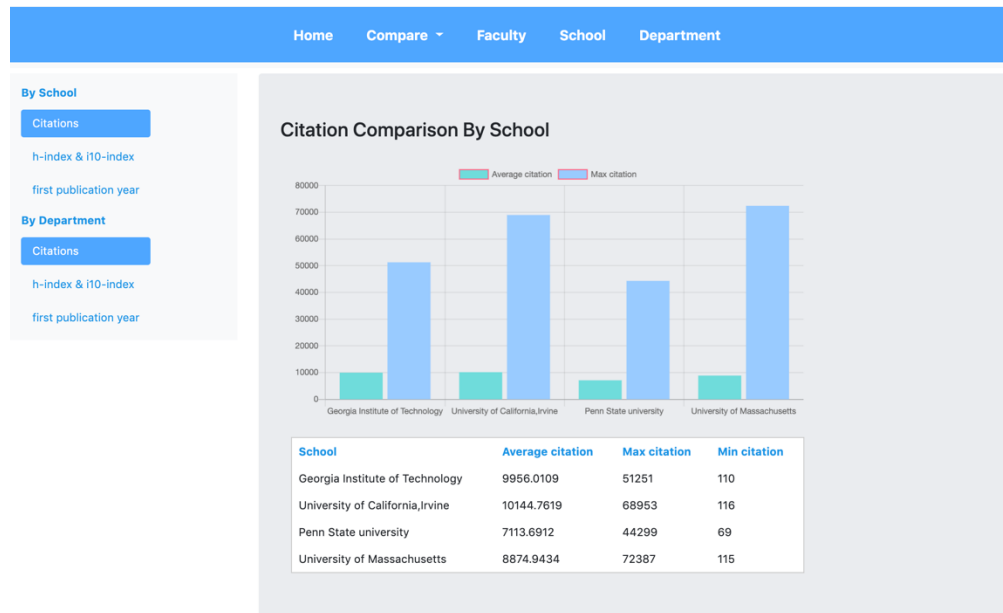


Figure3 Compare Page – Compare by School

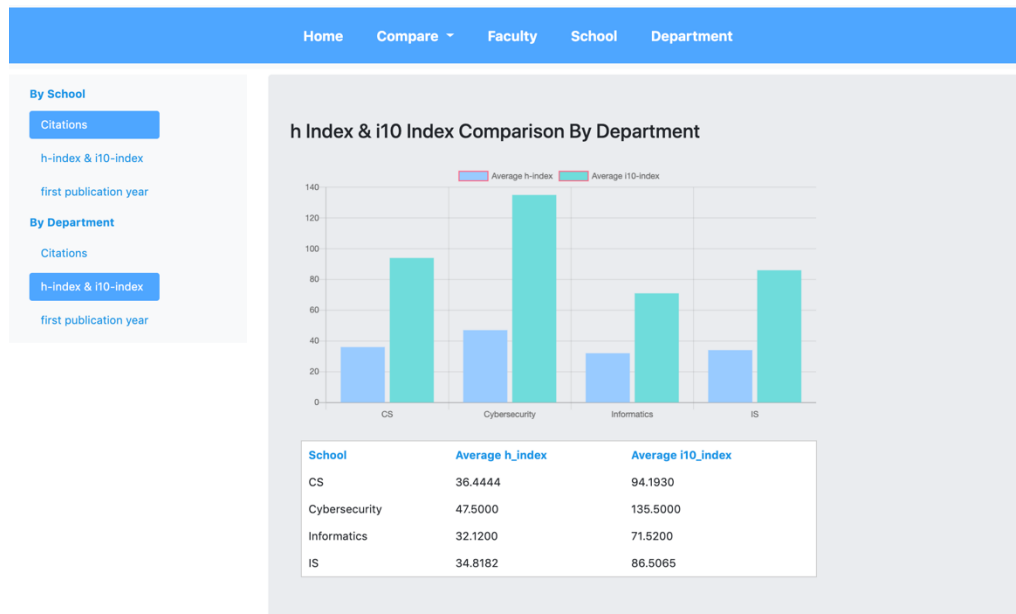


Figure4 Compare Page – Compare by Department

3) Faculty Page

The faculty page provides function of analyzing data distribution in several aspects, including gender, school, department, citation, h index and i10 index.

Figure5 shows the output graph and data list of department distribution. Figure6 shows the output graph of citation numbers of faculties distribution.

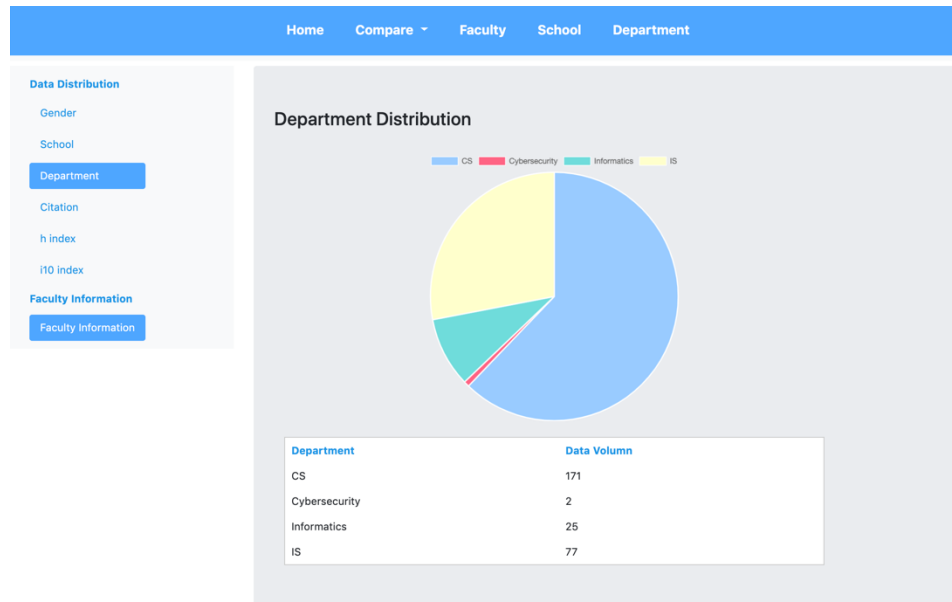


Figure5 Faculty Page – Department Distribution

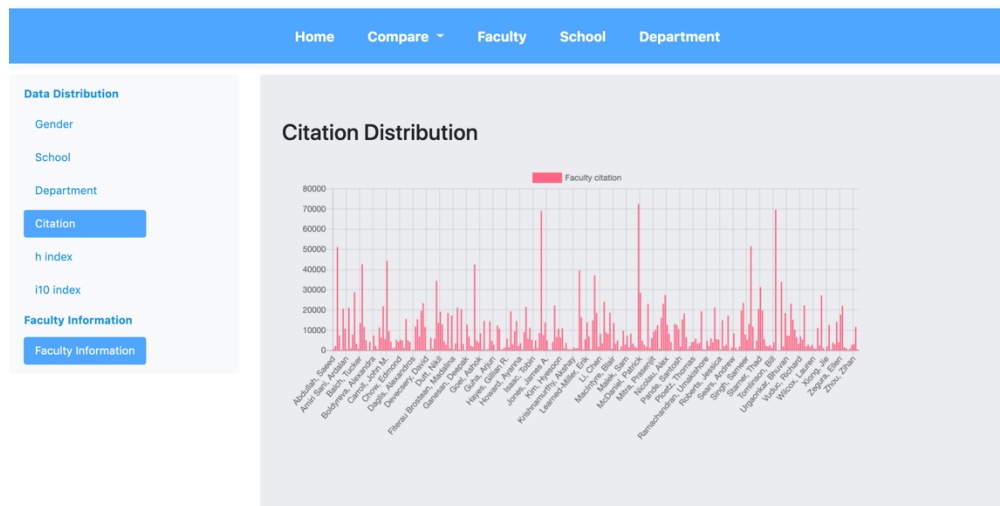
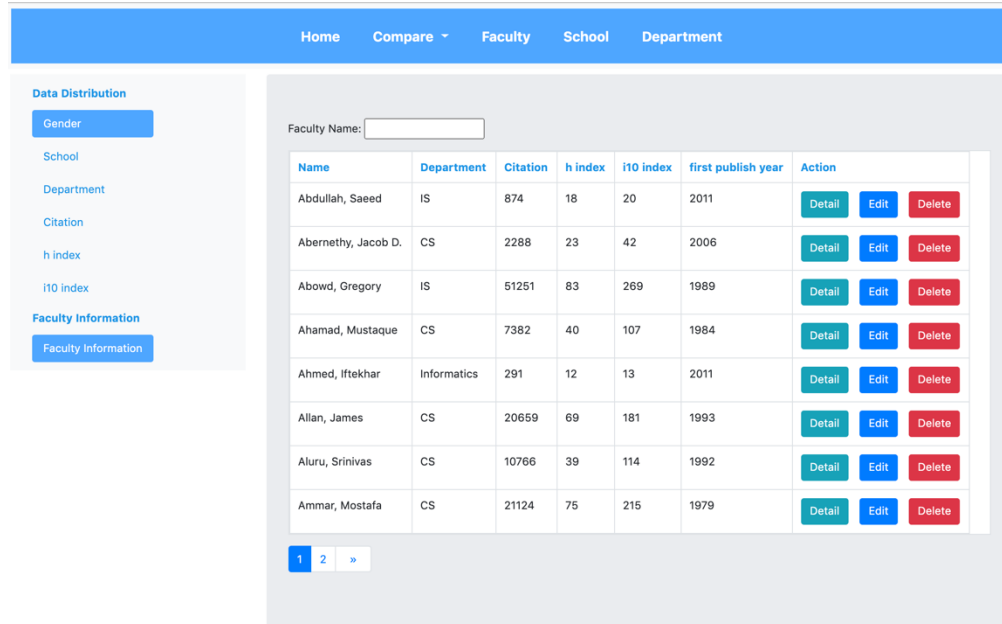


Figure6 Faculty Page – Citation Distribution

The faculty page also supports faculty information browsing and editing as shown in figure7. It also supports fuzzy query of faculty name. As shown in figure8, if the user input “James”, all the faculty information with a name including “James” in it will be displayed.

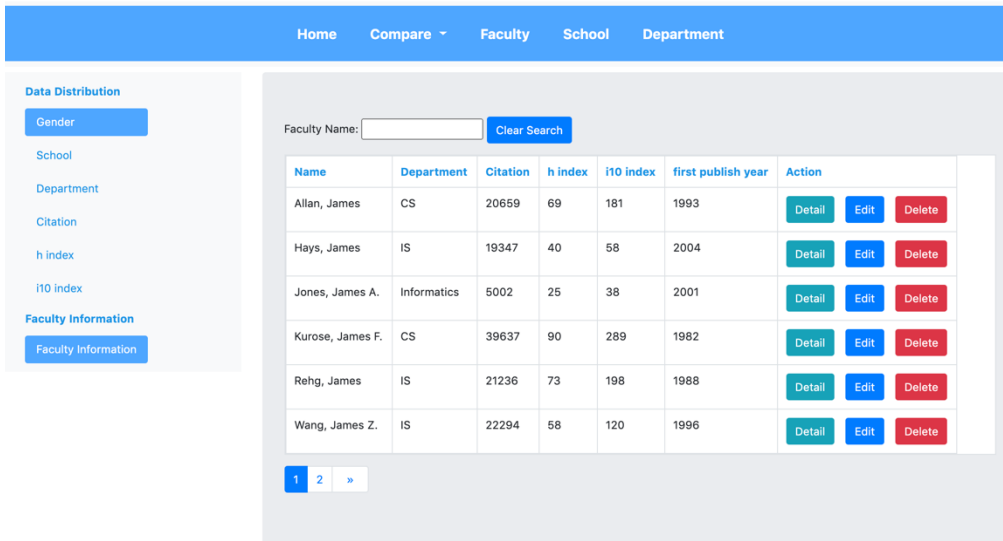


The screenshot shows the Faculty Page interface. On the left, there is a sidebar with 'Data Distribution' filters (Gender, School, Department, Citation, h index, i10 index) and 'Faculty Information' (Faculty Information). The main content area has a 'Faculty Name' search bar. Below it is a table with the following data:

Name	Department	Citation	h index	i10 index	first publish year	Action
Abdullah, Saeed	IS	874	18	20	2011	Detail Edit Delete
Abernethy, Jacob D.	CS	2288	23	42	2006	Detail Edit Delete
Abowd, Gregory	IS	51251	83	269	1989	Detail Edit Delete
Ahamad, Mustaque	CS	7382	40	107	1984	Detail Edit Delete
Ahmed, Iftekhar	Informatics	291	12	13	2011	Detail Edit Delete
Allan, James	CS	20659	69	181	1993	Detail Edit Delete
Aluru, Srinivas	CS	10766	39	114	1992	Detail Edit Delete
Ammar, Mostafa	CS	21124	75	215	1979	Detail Edit Delete

At the bottom of the table, there are pagination controls showing '1 2 »'.

Figure7 Faculty Page – faculty information



The screenshot shows the Faculty Page interface with a search filter applied. The 'Faculty Name' search bar contains the text 'James' and a 'Clear Search' button. The table below displays the search results:

Name	Department	Citation	h index	i10 index	first publish year	Action
Allan, James	CS	20659	69	181	1993	Detail Edit Delete
Hays, James	IS	19347	40	58	2004	Detail Edit Delete
Jones, James A.	Informatics	5002	25	38	2001	Detail Edit Delete
Kurose, James F.	CS	39637	90	289	1982	Detail Edit Delete
Rehg, James	IS	21236	73	198	1988	Detail Edit Delete
Wang, James Z.	IS	22294	58	120	1996	Detail Edit Delete

At the bottom of the table, there are pagination controls showing '1 2 »'.

Figure8 Faculty Page – faculty search

This section also provides functions of reading detail of the information, editing and delete. Figure9 and figure10 are examples of reading detail and editing.

Read Records

Name	Abdullah, Saeed
gender	m
link	https://scholar.google.com/citations?user=UK1Z4wYAAAAJ&hl=en&oi=ao
school	Penn State university
department	IS
job	Assistant
citation	874
h_index	18
i10_index	20
first_pub	2011
	Back to records

Figure9 Faculty Page – Read Detail

Update Records

Name	Abdullah, Saeed
gender	m
citation	874
h_index	18
i10_index	20
first_pub	2011
	Save Changes Back to records

Figure10 Faculty Page – Update

V. Limitation and Future work

1. With limited time and ability, the functions of this system are relatively basic. It can only do comparison from a few aspects.
2. The output graph only includes bar chart and pie chart, which can be displayed in more kinds of charts.
3. The interface design only achieve clean and simple, which is not user friendly and appealing enough.
4. The database only contain a small size of data set.
5. The website does not achieve crawling data from google scholar automatically.