

Developing an Objective Measure for Research Publications

Name: Shubhi Jain
Roll number: 2016267
BTP Track: Engineering Track
BTP Advisor: Dr. Pushpendra Singh

BTP report submitted in partial fulfillment of the requirements for the Degree of B.Tech. in
Computer Science Engineering on **5th June,2020.**



INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY **DELHI**

Student's Declaration

I hereby declare that the work presented in the report entitled "**Developing an objective measure for research publications**" submitted by me for the partial fulfillment of the requirements for the degree of *Bachelor of Technology in Computer Science & Engineering* at Indraprastha Institute of Information Technology, Delhi, is an authentic record of my work carried out under guidance of **Dr. Pushpendra Singh**. Due acknowledgements have been given in the report to all material used. This work has not been submitted anywhere else for the reward of any other degree.

.....
Shubhi Jain (2016267)

Place & Date:

Certificate

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

.....
Dr. Pushpendra Singh

Place & Date:

Abstract

Using the publicly available data-sets and the surveys, we aim to design a web-portal that would be able to provide insights into venues that are considered important for research as well as information of researchers publishing in such venues. We believe that our portal would be able to help students and researchers alike in understanding the current state of research in Indian research vis a vis International Universities. We also aim to how the quality of a research publication is evaluated, how the researchers decide the venue for their publications and how they assess publications of others.

Keywords: DBLP, CSRankings, CORE Rankings, GRIN Ranking, MongoDB, Node.js.

Acknowledgement

I owe my deep gratitude to our project advisor Dr. Pushpendra Singh, who took keen interest on our project work and guided us all along by providing all the necessary information for developing a good system. I would also like to thank my project partners, Lakshya and Kaustav without whom the project wouldn't have been possible.

Contents

1	Introduction	1
1.1	Related Work	1
1.2	Data Set	2
2	DBLP	3
2.1	Introduction	3
2.2	DBLP Records	3
2.3	DBLP Persons	5
2.3.1	DBLP Person Records	5
3	CORE and GRIN Ranking	6
3.1	CORE Ranking	6
3.2	GRIN Ranking	7
4	Previous Work	8
5	Current Work	9
5.1	Database	9
5.1.1	DBLP Data	9
5.1.2	CS Ranking Data	13
5.1.3	CORE Ranking Data	13
5.1.4	GRIN Ranking Data	13
5.2	Web-Portal	14
5.3	Survey Results	17
5.3.1	Student Survey	17
5.3.2	Faculty Survey	20
6	Future Work	24

Chapter 1

Introduction

Rankings are intensely popular and influential. While we might wish for a world without rankings, wishing will not make rankings go away.

Given this state of affairs, it makes sense to aim for a ranking system, for institutes and authors, that is meaningful and transparent. The most influential rankings right now are entirely reputation-based and relies on surveys sent to department heads and directors of graduate studies.

Our work is entirely metric-based and mainly focused on understanding the current standing of Indian Institutes vis a vis International Universities. We also work on understanding how does the students and professors of the Indian Universities follow the path of research and compare it to the way done in International Universities.

We are not only going to use our metric to just rank institutes but will also focus on ranking the researchers in each institute by taking into account the ranking system of various other websites like GRIN Ranking [5], CORE Ranking [4] and CS Ranking [3].

The user target that we have for our work are the students and the researchers. Through our web-portal we aim to provide insights into venues that are considered important for research as well as information of researchers publishing in such venues.

1.1 Related Work

There are many websites like QSRankings [1], CSRankings, etc which have their own metrics to rank institutes.

QSRankings use a compilation of 6 metrics [2] to rank institutes - Academic Reputation, Employer Reputation, Faculty/Student Ratio, Citations per faculty, International Faculty Ratio, International Student Ratio - with different weights being assigned to each of the metrics to calculate the rankings of the institute.

CSRankings on the other hand weighs departments by their presence at the most prestigious publication venues. CSRankings argue that although incorporating citation count can be tempt-

ing, it has been repeatedly argued that citation count can be subjective to interpretation and may also include the effect of "citation cartels".

Some other ranking websites have many more metrics on the basis of which they rank institutes. Some of the metrics like, lab count, number of people per system, etc. are obsolete since they don't portray the real picture of the research being done in the institute. Nowadays, most of the students have their own systems. Hence, the metric based upon the system count provided by an institute is flawed. Similarly, other metrics may not be suitable for ranking of an institute.

Although CSRankings seems to be the best in terms of the ranking metric being used, it has a few shortcomings like, it ranks colleges only in the Computer Science domain and it doesn't include rankings of individual researchers.

1.2 Data Set

The data-set that we have used are four different publicly available data.

1. DBLP Data [6]
2. CS Ranking
3. CORE Ranking
4. GRIN Ranking

We have used the DBLP data for extracting information about the authors like all the articles of the authors, affiliation of the author, all the venues where the author has published, etc. We have also extracted the information for each institute with the information we found on the authors.

We made our metric for ranking authors and institutes based on the ranking metric of the rest of the data-sets.

Chapter 2

DBLP

The DBLP computer science bibliography is the on-line reference for bibliographic information on major computer science publications.

2.1 Introduction

For computer science researchers the DBLP web site is a popular tool to trace the work of colleagues and to retrieve bibliographic details when composing the lists of references for new papers. Ranking and profiling of persons, institutions, journals, or conferences is another usage of DBLP.

The DBLP data may be downloaded. The bibliographic records are contained in a huge XML file. It is easy to derive several graphs like the bipartite person-publication graph, the person-journal or person-conference graphs, or the coauthor graph, etc.

The main advantages are the free availability and the inclusion of many conference proceedings which play an essential role for many branches of computer science and are poorly covered by other bibliographic data bases. On the other hand, the main disadvantages of DBLP are the lacking citation information and the varying coverage for different sub fields of computer science.

2.2 DBLP Records

The file dblp.xml contains all bibliographic records which make DBLP. It is accompanied by the data type definition file dblp.dtd.

The XML root element <dblp> contains a long sequence of bibliographic records. The DTD lists several elements to be used as a bibliographic record:

```
<!ELEMENT dblp (article—inproceedings— proceedings—book—incollection—  
phdthesis—mastersthesis—www)* >
```


These tags correspond to the entry types used in BibTEX. DBLP records may be understood as “BibTEX records in XML syntax + ϵ ”:

```
<article key="journals/cacm/Szalay08" mdate="2008-11-03">
<author>Alexander S. Szalay</author>
<title>Jim Gray, astronomer.</title>
<pages>58-65</pages>
<year>2008</year>
<volume>51</volume>
<journal>Commun. ACM</journal>
<number>11</number>
<ee>http://doi.acm.org/10.1145/1400214.1400231</ee>
<url>db/journals/cacm/cacm51.html#Szalay08</url>
</article>
```

This record describes an article from CACM as follows:

1. The enclosing **article** element has two attributes :
 - a. **key** is the unique key of the record. The most important sub-trees in the key namespace are `conf/*` for conference or workshop papers and `journals/*` for articles which are published in journals, transactions, magazines, or newsletters.
 - b. **mdate** is the date of the last modification of this record. The format complies with ISO 8601, i.e. YYYY-MM-DD.
2. An **author** element is defined separately for each author that contains the name of the author. The order of the `textbfauthor` elements inside a record is significant, it should be the same as on the head of the paper. If the author of a publication is unknown, the DBLP record does not contain any `textbfauthor` element.
3. The **title** element is only element which has to exist in every DBLP publication record. It may contain **sub** elements for subscripts, **sup** elements for superscripts, **i** elements for italics, and **tt** for typewriter text style
4. The **pages** element is written for page numbering in "from-to" way. In case, the number of the last page is unknown, `from-` is written while there is a single paged paper, the page number is written without hyphen.
5. The **year** element should always contain a four digit number to be interpreted according to the Gregorian calendar
6. A DBLP record may contain up to two URLs in the fields **url** and **ee**. Both URLs may be either global or local. A global URL is standard internet URL, it always starts with protocol specification of the form "letter+ :" (**http:**, **ftp:**, ...). If the **url** or **ee** contents does not start with protocol name followed by a colon, it is a local URL pointing inside the DBLP web site. To get valid URL, you simply have to add a base URL of a DBLP server as a prefix.

2.3 DBLP Persons

URL stability is regarded as a very important virtue to make a service reputable. Initially, when DBLP was started in 1993, it was decided that each person should have her/his own **"DBLP author page"** but in today's time there are multiple authors with the same name. So, in these cases, a small unique number (unique within the set of the people having the same names) is appended after the name of the person so as to distinguish them. Although the problem of finding different people of the same name is difficult, small heuristics used by DBLP proved to be useful.

2.3.1 DBLP Person Records

Normally, DBLP people have their own websites as home pages. Overtime, it became obvious to add these links from DBLP author pages to personal home pages. These are added as a part 'www' tags.

```
<www key="homepages/h/AlonYHalevy" ...>
<author>Alon Y. Halevy</author>
<author>Alon Y. Levy</author>
<title>Home Page</title>
<url>http://alanhalevy.googlepages.com/</url>
</www>
```

Chapter 3

CORE and GRIN Ranking

Conference papers and the venue where they are published are important to researchers. With this, there comes the need to know the best ones, hence we need a ranking system. There are several websites with their own metric system to rank the conference venues. We have selected two of the most known ones, CORE Ranking and GRIN Ranking.

3.1 CORE Ranking

The Computing Research and Education Association of Australasia, CORE, is an association of university departments of computer science in Australia and New Zealand.

CORE determines the conference rankings on the basis of a mix of indicators, including citation rates, paper submission and acceptance rate. Apart from this, the visibility and research track record of the key people hosting the conference and managing its technical program also plays an important role in the conference rankings.

CORE assigns conferences to one of the following categories of ranks:

A*	Flagship conference, a leading venue in a discipline area.
A	Excellent conference, and highly respected in a discipline area.
B	Good conference, and well regarded in a discipline area.
C	Other ranked conference venues that meet minimum standards.
Australasian	A conference for which the audience is primarily Australians and New Zealanders.
National	A conference which is run primarily in a single country, with Chairs from that country, and which is not sufficiently well known to be ranked. (Papers and PC may be international).
Regional	Similar to National but may cover a region crossing national borders.
Unranked	A conference for which no ranking decision has been made.

3.2 GRIN Ranking

GRIN Ranking portal is an initiative sponsored by GII (Group of Italian Professors of Computer Engineering), GRIN (Group of Italian Professors of Computer Science), and SCIE (Spanish Computer-Science Society).

The rankings are generated by a joint committee of GII, GRIN and SCIE by using an automatic algorithm based on well-known, existing international classifications. This automatically-generated rating is updated periodically, usually every two years. The rating algorithm brings together and tries to unify the base data sources : The CORE 2018 Conference Rating, Microsoft Academic, and LiveSHINE.

The conferences are classified mainly into four tiers or in the following decreasing order :**A++**, **A+**, **A**, **A-**, **B**, **B-**, **C**.

Tier 1	A++ , A+	Top Notch conferences.
Tier 2	A , A-	Very high-quality conferences.
Tier 3	B , B-	Good quality conferences.
Tier 4	Work in progress	Work in progress.

Chapter 4

Previous Work

The code for CSRanking is available online so we started by reading the code for CSRanking and exploring it. Along with that we researched for different databases that we could have used. The databases that we initially were aware of were ArangoDB, MongoDB and Neo4J. After researching further about these databases, we decided to go with MongoDB.

Although ArangoDB was released after MongoDB and Neo4J, it did not have as many language support as MongoDB. Since this database is quite new and not very popular there were very less references and tutorials available. Also, there were very less blogs which resolved the issues arising in ArangoDB.

On the other hand, Neo4J is the oldest database among the three, the queries here are too long as they are nested. For example, for any node there were multiple node and which in turn had multiple nodes relation that made it difficult to run even simple queries.

In the xml file of the DBLP data, the data is given in terms of articles, journals, and persons. We converted xml to json format for using it in MongoDB.

In MongoDB, we made documents (corresponding to a record in MySQL) and collections (corresponding to a table in MySQL) for authors of the top 10 Indian Institutes with affiliation. We also made document and collection of articles of each author by querying the person page API given in the DBLP data.

We decided some of the queries as follows:

1. Information of an author
2. Papers published at a conference sorted year-wise
3. Papers published by an institute sorted year-wise with conference names
4. Papers published by an author in a particular conference sorted year-wise

Other than this, we have decided to use Django for the web-application part.

Chapter 5

Current Work

5.1 Database

This semester we started by revamping of the data set prepared in the previous semester due to inconsistencies found in the data-set. We have now divided the data into four parts :

- DBLP
- CS Ranking
- CORE Ranking
- GRIN Ranking

The data has redundancy but it was done in order to get a faster query response. We did not limit the size of the data but worked more on the optimization of user experience. In case, we had removed the redundancy then since MongoDB does not have inner and outer joins and only a lookup function, it was taking a lot of unnecessary time which would bring lag to the web-portal whenever the user would add/subtract filters.

5.1.1 DBLP Data

The revamping of data was required as we had first extracted data from the .xml file of the DBLP and when we made an API call, we found that the data here was different than the one we extracted. Apart from this, the tags used were also different in the data by API call, so we had to re-write the code for extracting the data.

In the API call query, whenever we search for anything like an author or a venue, it would search the name partially. For example : if we search for "Pushpendra Singh" using the API call, it would give back the result with all the authors name which has the query name in it like "Pushpendrai Singh Bharti", "Pushpendra Bahadur Singh" and "Pushpendra Singh 001". So, to overcome this, whenever we made an API call for any author, we would go through each

article of the authors and pick those articles which has the queried author as one of the authors of the article.

We have formatted the data in JSON format with the following keys:

- **Author Data** : The Author data has the following keys :

Keys	Description
Affiliation	Institute
Alias	Alias of the author
Author	author name
Homepage	link to author's page on his/her respective institute page
url	DBLP url page of the author.
Papers	All the articles published by the author
Venues	All the conferences that the author has published in
Graph	Year-Wise count of the number of paper published by the author.

Example: The following data contains all the keys except Alias as mentioned above but it doesn't contain the whole data as the number of venues and papers were a bit large.

```
{
  "affiliation": "IIIT Delhi",
  "author": [
    "Rajiv Ratn Shah"
  ],
  "homepage": "https://old.iiitd.ac.in/people/adjunct-visiting",
  "url": "https://dblp.org/pid/134/3502",
  "scholarid": "WACHZv4AAAAJ",
  "papers": [
    "conf/aaai/SalikAKSJZ19",
    "conf/www/JangidSSZ18"
  ],
  "venues": [
    "WWW",
    "AAAI",
  ],
  "graph": {
    "2019": 32,
    "2018": 20,
  }
}
```

- **Articles Data** : The Article data has the following keys :

Keys	Description
Author	List of co-authors
Title	Title of the article
Venue	Venue where the article was published
Pages	Number of pages of the article.
Year	Publishing Year
Type	Either Conference and Workshop Papers or Journal Articles
Key	Unique key of the article
doi	The DOI of the document.
ee	electronic resource (e.g., a web page provided by the publisher, or a pdf document).
url	Link to the DBLP page
Affiliations	List of all the institutes of the co-authors

Example: The following example contains all the keys as mentioned above.

```
{
  "authors": {
    "author": [
      {
        "@pid": "66/4913",
        "text": "Dheeraj Sanghi"
      },
      {
        "@pid": "j/PankajJalote",
        "text": "Pankaj Jalote"
      },
      {
        "@pid": "97/5792",
        "text": "Puneet Agarwal"
      }
    ]
  },
  "title": "Using Proximity Information for Load Balancing in Geographically
           Distributed Web Server Systems.",
  "venue": "EurAsia-ICT",
  "pages": "659-666",
  "year": "2002",
  "type": "Conference and Workshop Papers",
  "key": "conf/eurasiaict/SanghiJA02",
  "doi": "10.1007/3-540-36087-5_77",
  "ee": "https://doi.org/10.1007/3-540-36087-5_77",
```



```

"url": "https://dblp.org/rec/conf/eurasiaict/SanghiJA02",
"affiliations": [
  "IIIT Delhi",
  "IIT Kanpur"
]
}

```

- **Institute** : The Institute data has the following keys :

Keys	Description
Name	Institute Name
Auth	List of authors in the Institute
Graph	Year-Wise count of the number of paper published by the authors in the Institute.

Example: The following example contains all the keys as mentioned above but it doesn't contain the whole data as the number of papers and authors were a bit large.

```

{
  "name": "IIIT Delhi",
  "authors": [
    "Mayank Vatsa",
    "Pankaj Jalote",
    "Pushpendra Singh 0001",
    "Rajiv Ratn Shah",
  ],
  "papers": [
    "journals/imwut/AroraMGKMP19",
    "journals/imwut/MukherjeeNVSP19",
    "conf/chi/AroraMSP19",
    "conf/chi/AroraSMJSP19",
  ],
  "graph": {
    "2018": 133,
    "2017": 99,
    "2019": 153,
    "2020": 35
  }
}

```

5.1.2 CS Ranking Data

The data for CS Ranking is available online. So, we extracted the faculty affiliations of the selected Institutes and mapping of fields-conferences from the CS Ranking data-set. We removed all those conferences which did not have a mapping that in turn removed all the articles which were published in the removed conferences. If all articles of an author were removed, that means CS Ranking does not have the information on that author. Hence, we remove the author too.

The CS Ranking data has all the same data-set as the DBLP just with a few changes which are as follows :

- Addition of fields to the author and institute data.
- Deletion of authors whose data is not available on CS Ranking from author and institute data-sets.
- Deletion of articles whose conference-field mapping is not given on CS Ranking from all the three data-sets.

5.1.3 CORE Ranking Data

Similar, to the CS Ranking data, CORE Ranking also provides field-conference mapping along with the ratings of each conference. We add the rankings key to the data of institute and author with the value as all the ratings from the range of ratings (A*,A,B,C, Australasian, National and Regional) such that the the author has published at least one paper in any conference with that ranking.

For example: if author 'a' has published in papers in 4 conferences and 2 of the conferences are classified as A*, 1 as A and 1 as B then we add rankings : [A*,A,B] to the author data of 'a' and the institute data to which the author is affiliated to.

We have evaluated the rankings of each institute and author among themselves, by taking into account the number of papers published in each of the classified conferences. The priority order is as follows: A*,A,B,C, Australasian and Others. Here, we have added Regional and National Ranking to others. We give higher rank to any author/institute which has published more papers in A*, then A and so on.

5.1.4 GRIN Ranking Data

Similar to the above Ranking data-sets, GRIN ranking only provides the ranking of conferences and not the fields. So, as we added and evaluated ranking in the CORE Ranking Data, we did the same here with the following priority order : A++, A+, A, A-, B, B- and Unranked.

5.2 Web-Portal

We started creating the web-portal using Django but as we went forward to integrate with MongoDB, the integration was not as smooth. So we switched the web-framework in use to Node.js as we were getting much smoother integration with MongoDB.

So, now we built our web-portal using HTML, CSS and JavaScript at the front end along with Node.js and MongoDB with Handle Bar as the templating engine, at the back end.

We have used the following in node.js:

- **Express** : It is a web-framework used for routing and allows to dynamically render HTML Pages based on passing arguments to templates.
- **Mongoose** : It acts as a front end to MongoDB and allows node.js access data from MongoDB
- **Handle Bars** : It is a templating engine that manipulates the HTML code from the server side using the server side code.
- **Ajax (Asynchronous JavaScript and XML)** : It allows web pages to be updated asynchronously by exchanging small amounts of data with the server using the GET and POST requests.

Apart from this, we used PyMongo in python to insert data into the MongoDB database.

Our web-portal has the following functionalities :

The screenshot shows the 'ResearchMetric' web portal. On the left, there is a sidebar with a 'CS Rankings' dropdown menu and input fields for 'Author', 'Institute', 'Venue', and 'Field'. Below these is a 'Ranking' button and a 'Submit' button. The main content area displays a table of authors and their affiliations. The table has four columns: 'Name', 'Affiliation', 'Fields', and 'Adj. Count'. The data is as follows:

Name	Affiliation	Fields	Adj. Count
Aditya Kanade	IISc Bangalore	AI, PL, Embedded, SE, Logic	3.250
Ambedkar Dukkupati	IISc Bangalore	AI, ML, Vision	3.917
Anand Louis	IISc Bangalore	Theory	1.583
Arindam Khan 0001	IISc Bangalore	Theory	1.000
Arkaprava Basu	IISc Bangalore	Arch, HPC	2.335
Arpita Patra	IISc Bangalore	Security, Crypto	1.583

Figure 5.1: Home Page : CS Ranking Data

- Allows user to choose which data he/she wants to see from all four categories of data-sets (DBLP, CSRanking, CORE Ranking, GRIN Ranking).
- Allows user to choose between the author and institute data of any of the above mentioned data sets and each data-set provides the following information:

DBLP	Author	Name, Affiliation, Adjusted Count.
	Institute	Name, Author Count, Venue Count, Adjusted Count.
CS Ranking	Author	Name, Affiliation, Adjusted Count.
	Institute	Name, Author Count, Venue Count, Adjusted Count.
CORE Ranking	Author	Name, Affiliation, Fields, Conference Ranks, Rating.
	Institute	Name, Author Count, Venue Count, Fields, Conference Ranks, Rating.
GRIN Ranking	Author	Name, Affiliation, Conference Ranks, Rating.
	Institute	Name, Author Count, Venue Count, Conference Ranks, Rating.

- The filters that are available are: *Author, Institute, Venue, Field and Ranking*. Our web-portal allows user to add filters according to their convenience but the user cannot add
 - Ranking and Field filter with DBLP data.
 - Ranking filter with CS Ranking data.
 - Field with GRIN Ranking data.

On the other hand, the user can use all the filters at once with the CORE Ranking data.

- Each row is clickable in both the author and institute data that redirects the user to the selected author or institute page which also has the functionality to let user choose the data-set.
 - The author page (figure 5.2) provides information about the author, all the articles that the author published linked to the DBLP page of the article, the URL of the author page on the official web page, the URL of the author’s google scholar page, all the other authors with whom the author has worked with, and an year-wise graph of number of publication of each year.
 - The institute page (figure 5.3) provides information about the institute, all the authors of the institute linked to the author page, all the publications that the author from the institute have published linked to their DBLP page and an year-wise graph of number of publication of each year.
- The year-wise graph in the author page also has the functionality of choose the range of year.

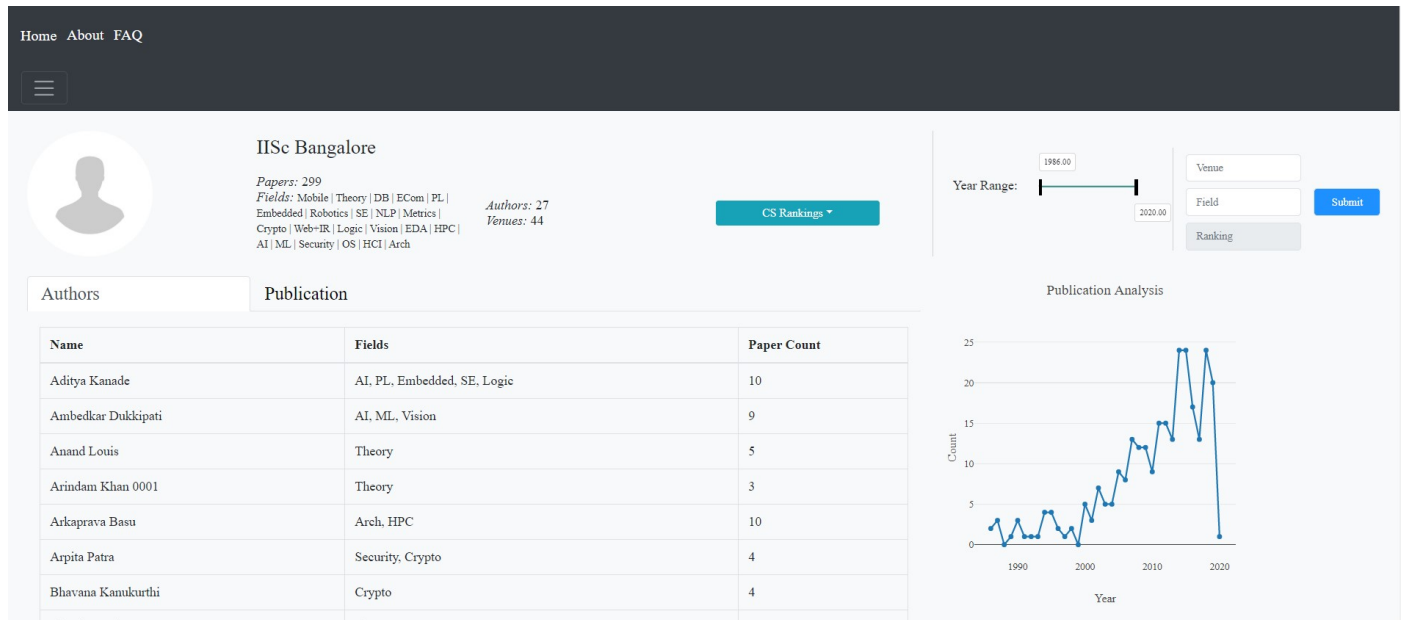


Figure 5.2: Institute Page

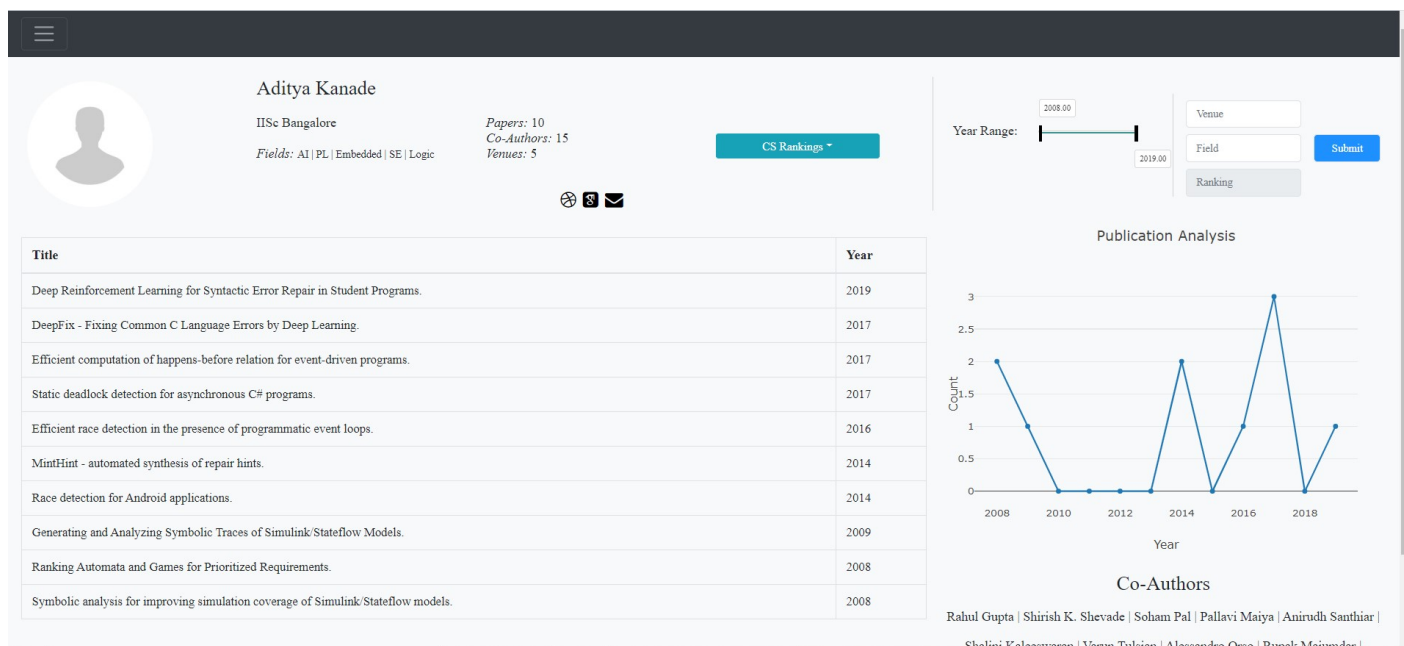


Figure 5.3: Author Page

5.3 Survey Results

We conducted a survey to know the insights of the faculty and Ph.D student on the way they do research and how they evaluate it. We launched two survey one for faculty and one for the Ph.D students. Both the surveys are still open.

5.3.1 Student Survey

We have received 22 results till date from the Ph.D students. One of the Questions of the survey was "According to you, what is good quality research and how do you assess it?". So for this question we received various sort of answers. Some of the answers had the following in common:

- Research should make some contribution to the community
- Research should have been recognised at some prestigious conference.
- Good quality research must have a crisp explanation of following questions — what is a problem statement, the motivation behind the work, research gap, strong evaluation of the hypothesis (in varied test scenarios), overall takeaway, broader impact, and last but not the least, what are the limitations.
- Base of new ideas for other researchers to build upon.
- Novelty of the method used for research

Rest of the questions had the following result :

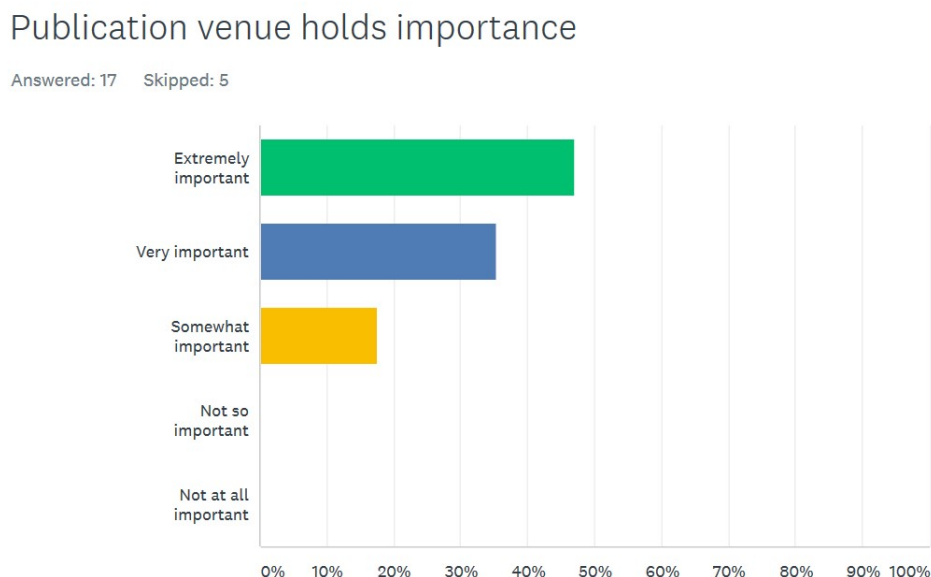


Figure 5.4

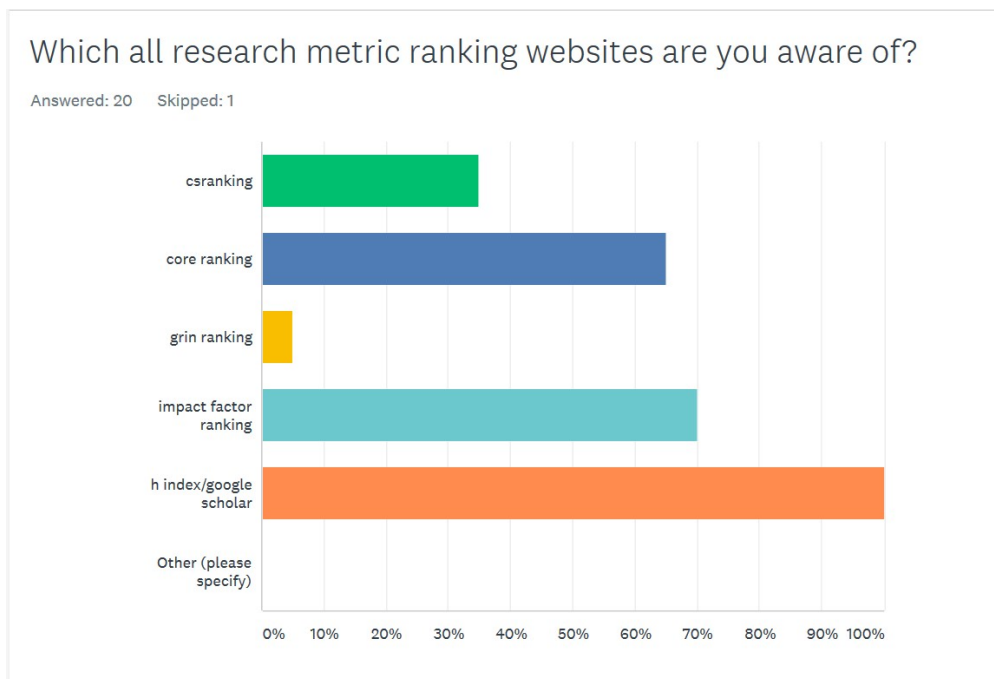


Figure 5.5

How do you rank an institute with respect to research?

Answered: 21 Skipped: 0

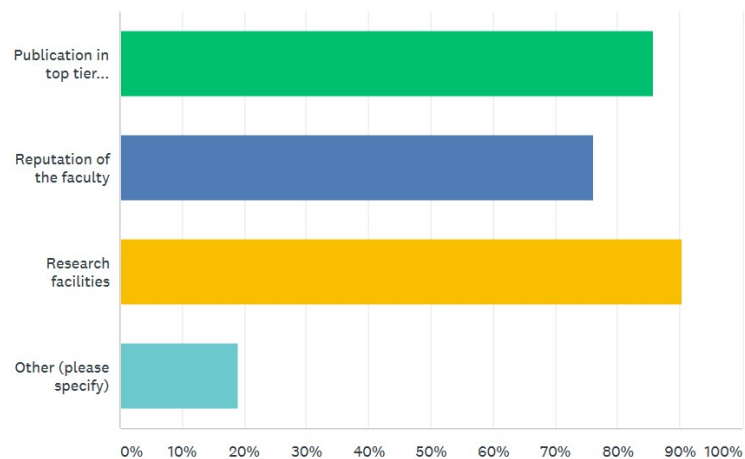


Figure 5.6

In the above question (figure 5.6), the other answers were as follows :

- Freedom to choose your own subject of study.
- Area of research, recent publication in the field
- Projects
- New and upcoming venues where top people publish

How do you choose a faculty member to work with?

Answered: 22 Skipped: 0

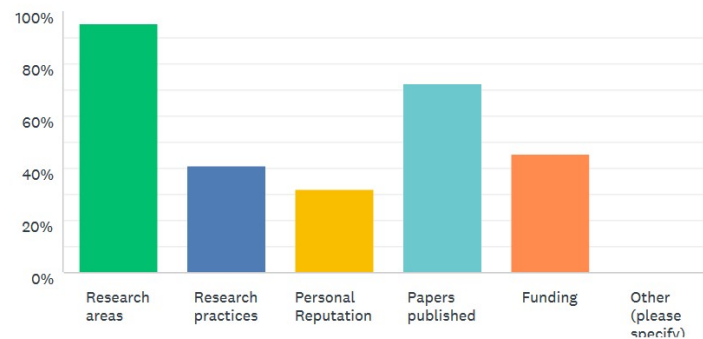


Figure 5.7

If you are a PhD student, then how many papers you suppose you should have published by the end of your PhD program?

Answered: 22 Skipped: 0

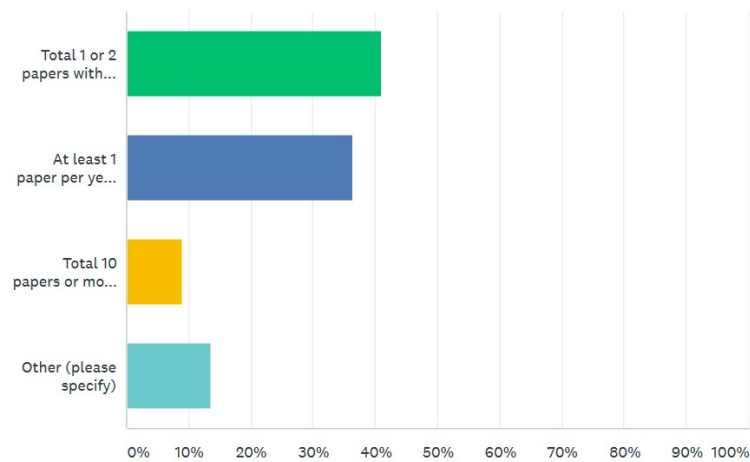


Figure 5.8

In the above question (figure 5.8) , the options given were as follows:

1. Total 1 or 2 papers with very high quality papers in your whole PhD program
2. At least 1 paper per year during your PhD program
3. Total 10 papers or more by the end of your PhD program

And the other answers were :

- A mix 1-2 high and 2 average quality papers
- The number of papers depends on the field in which a person is working, lab setup, journal vs conference, venues, grants, etc

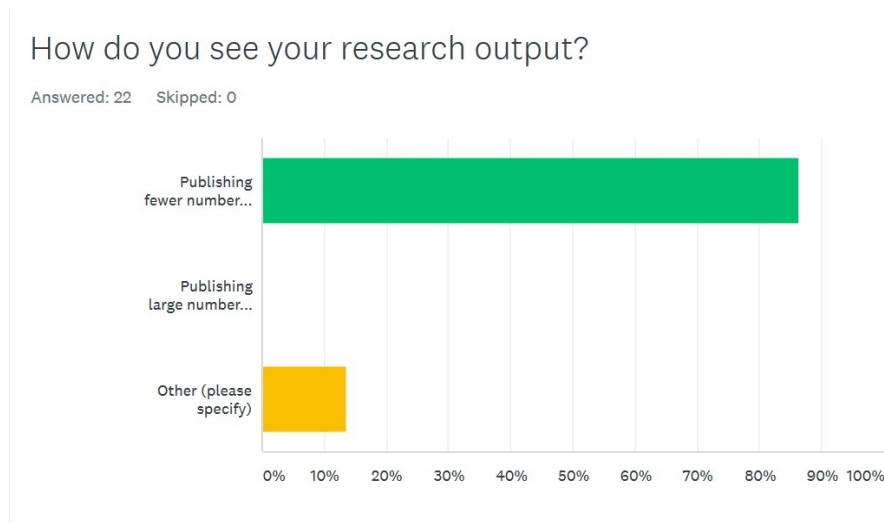


Figure 5.9

In the above question (figure 5.9), the options given were as follows:

1. Publishing fewer number of papers with very high quality
2. Publishing large number of papers with average quality

And the other answers were mainly a mix of both as both the type of papers require different type of skill sets.

5.3.2 Faculty Survey

We have received 8 results till date from the Faculty. Again, the same question was asked in the faculty survey too "According to you, what is good quality research and how do you assess it?". Some of the answers had the following in common:

- Research should well designed and replicable
- Research advances the state of the art.
- Venue where the research is published
- Research should have real applications

Another question that we asked in the survey was "What do you expect from PhD student, assuming a 5yr PhD program, in terms of his/her research output?", and we received the following results :

- Ability to ask good fundamental questions
- Top-tier publications in their field of study

- Central focus should be solving problems of our society and not publishing. Publications are an obvious consequence.
- A good hold on the subject of research.

The results for the rest of the questions are as follows :

The publication venue holds importance.

Answered: 7 Skipped: 1

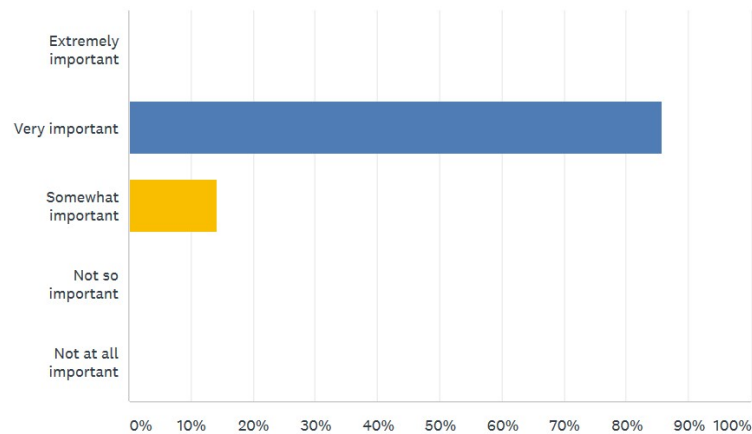


Figure 5.10

On what basis, do you judge a faculty's research papers?

Answered: 8 Skipped: 0

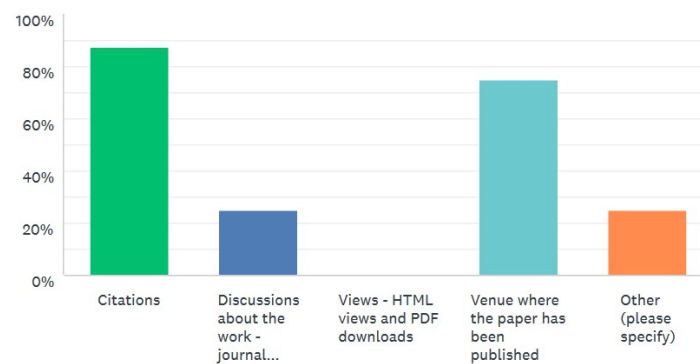


Figure 5.11

In the above question (figure 5.11), the result of the other options are as follows:

- How fundamental and original are the research questions. Rigor in analysis. Often original research is hard to publish due to various biases of the editors/journals/etc.
- The logic of the paper's argument

How do you evaluate a new faculty member?

Answered: 8 Skipped: 0

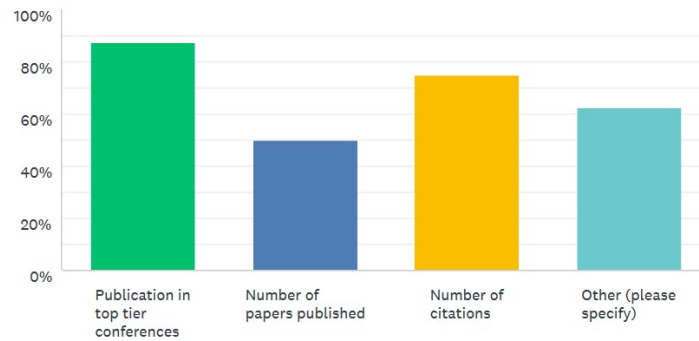


Figure 5.12

In the above question (figure 5.12), the other answer results are as follows:

- H-index and Impact factor of journals
- The candidate should be able convey command on their specialized subject matter. Publications are a result of many factors, which do not necessarily reflect ability to drive creation of knowledge.
- Suitability for teaching, exploring new research direction, guiding students

On what basis do you choose a conference/journal to publish your papers?

Answered: 8 Skipped: 0

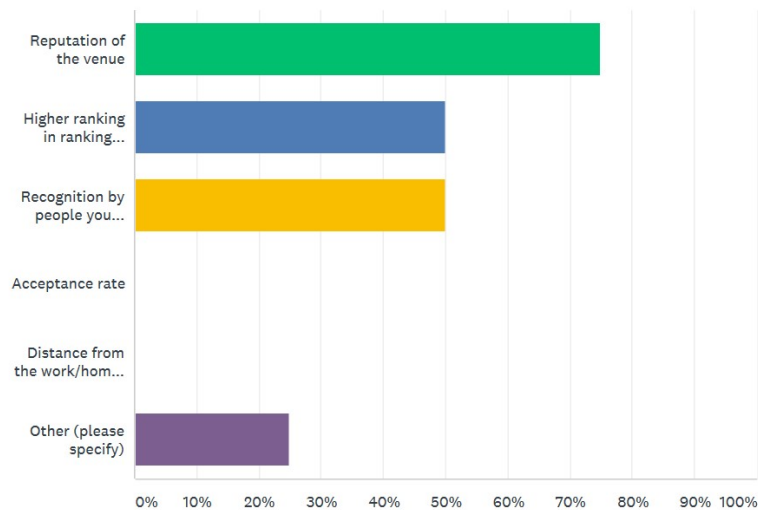


Figure 5.13

In the above question (figure 5.13), the options given were as follows:

1. Reputation of the venue
2. Higher ranking in ranking matrix like cs ranking,core ranking
3. Recognition by people you interact on a day-to-day basis
4. Acceptance rate
5. Distance from the work/home location

And the other answers were Impact factor of Journal, Travel difficulty and Publication costs.

This was the result for the survey till now and we are still awaiting more responses to get a better analysis of the data.

Chapter 6

Future Work

The next steps would be the following -

1. Incorporate field analysis in our Institute and Author data
2. Writing the FAQ for our web-page.

We also aim to expand our data-set from only 10 institutes to cover all the Indian Research Institutes.

In conclusion, we believe that there is a lot of scope of improvement in the web-portal, which we aim to do via user-feedback.

Bibliography

- [1] <https://www.topuniversities.com/university-rankings/world-university-rankings/2020>
- [2] <https://www.topuniversities.com/qs-world-university-rankings/methodology>
- [3] <http://csrankings.org/index?all>
- [4] <http://portal.core.edu.au/conf-ranks/>
- [5] <http://gii-grin-scie-rating.scie.es/>
- [6] <https://dblp.org/>