

DBLP Records and Entries for Key Computer Science Conferences

Swati Agarwal, Nitish Mittal, Rohan Katyal, Ashish Sureka, Denzil Correa

The dataset "DBLP-CSR.zip" is derived from September 17, 2015 snapshot of dblp bibliography database. It contains the last 16 years (2000 – 2015) of publications records of 81 Computer Science Research conferences used for a study conducted in our paper *Women in Computer Science Research- What is Bibliography Data Telling Us?* published in ACM SIGCAS Computers and Society Newsletter, Volume 46, Issue 1, February 2016. Link to the Newsletter Archive: <http://dl.acm.org/citation.cfm?id=J198>

Importing the Data in MySQL

```
mysql -u root -p;  
enter your password  
create database dblp;  
use dblp;  
source filename.sql
```

Dataset Description

The dataset contains 7 .sql files primarily named as affiliation_coord.sql, affiliation.sql, author_gender.sql, authors.sql, editor_gender.sql, editor.sql and main.sql.

The affiliation_coord.sql, affiliation.sql, authors.sql, editor.sql files create the tables with same name. While main.sql, editor_gender.sql and author_gender.sql create tables with the names general, genedit and genauth_old respectively.

Followings are the list and description of all attributes used in the dataset. Same attributes used in different tables are listed only once.

1. Table- general

- k- unique id of each article- primary key in the table
- year- the year of publication
- conf- Abbreviation for conference name
- crossref- cross reference link to all articles published in a conference in a year
- cs, de, se, th- a binary attribute denoting if a conference belongs to these domains (Computer Science, Data Engineering, Software Engineering, Theory)
- publisher- Name of the conference publisher
- link- unique DOI link to the article that re-directs to conference publisher page.

2. Table- authors

- pos- position of author in the paper. 0 denotes first author
- name- unique name of author in dblp dataset
- gender- gender of authors. Hyphen (-) denotes that gender was not determined. Please refer to the paper for more details.
- prob- probability of a name to be M, F, -.

3. Table- editors

- k- foreign key for crossref attribute in general table
- pos- position of editor in conference. 0 denotes the first editor.

4. Tables- genauth_old and genedit

- contain the records of gender information of authors and editors- derived from authors and editors tables and have no new attribute.

5. Table- affiliation

- affil- affiliation record of each author publishing in the 81 conferences mentioned above.
- year- year of publication

6. Table- affiliation_coord

country- country name of the author extracted from affiliation information country_code- code to be used for map (USA, NY, IN) lat, lng- geocode (latitude and longitude) information of affiliation.

Contact Information

If you have questions or suggestions regarding this document, please contact Swati Agarwal at 'swatia@iiitd.ac.in'.