

DBLP Publications Records and ACM Metadata for SIGWEB Conferences

Swati Agarwal, Nitish Mittal, Ashish Sureka

The dataset "DBLP-SIGWEB.zip" is derived from September 17, 2015 snapshot of dblp bibliography database. It contains all publications and authors records (available in dblp data) of 7 ACM SIGWEB conferences:

- ACM conference on Hypertext and social media (HT)
- Joint Conference on Digital Libraries (DL)
- Document Engineering (DocEng)
- Web Science (WebSci)
- Conference on Information and Knowledge and Management (CIKM)
- Conference on Web Science and Data Mining (WSDM)
- User Modeling, Adaptation and Personalization (UMAP) Conference

The dataset also contains the authors, chairs, affiliations and additional metadata information of conferences that are published in ACM digital library.

Importing the Data in MySQL

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

Dataset Description

The dataset contains dblp-sigweb.sql file that creates 15 tables in mysql primarily named as affiliations_info, author_tags, authors_affiliation_history, authors_info, cited_by, colleagues, concepts, conferences, editors, general_chairs, paper_authors, paper_references, papers, program_chairs, and acceptance_rates.

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

1. Table- papers

- dblp_key- unique id of each publication in dblp database
- crossref- unique id of each conference in dblp database
- doi- unique doi url to publisher page
- paper_id- unique id of each article in acm digital library (DL)
- cite_count- number of citations for each article calculated for the papers published in acm DL
- pages- number of pages for each article in conference proceedings
- conf_id- unique id of each conference in acm DL
- funding- funding source information of article. NULL- if no funding source available

2. Table- paper_authors

- author_id- unique id of an author in acm DL
- affiliation- affiliation information of author for associated article

3. Table- concepts

- concept- concepts in an article- tagged by ACM

4. Table- author_tags

- author_tag- Keywords/tags provided by authors

5. cited_by

- paper_id- acm DL id of article A to be cited
- cite_id- unique id of article that has cited article A

6. paper_references

- refer_id- unique id of the articles (published in sigweb conferences) cited in article A.

7. Table- conferences

- dblp_key- unique id of each conference in dblp database
- year- year of the conference
- publisher- publisher name of each conference (ACM, Springer, IEEE etc.)
- title- full name of the conference proceeding
- doi- unique doi url to the conference publisher page
- conf_id- unique id of each conference in acm DL

8. Table- general_chairs, program_chairs, editors

- author_id- unique id of author (as general chair, program chair or editor of conference conf_id)
- affiliation- affiliation of author

9. authors_affiliation_history, colleagues

- author_id- unique id of author A in ACM DL
- position- index of affiliation- starts from 0
- affiliation- lists all affiliations of an author
- colleague_id- lists acm IDs of all authors publishing papers in ACM co-authored with A.

10. authors_info

- author_name- full name of author acquired from ACM publisher page
- year_first- year of first article publication in ACM
- year_last- year of recent article publication in ACM
- pub_count- total number of publications in ACM DL
- cite_count- total number of citations mentioned in ACM publications
- avg_cite- average number of citation in ACM publications

11. affiliations_info

- affiliation- name of the affiliation
- affiliation_type- type of affiliation (Industry, Academic Institution)
- city, state, country- geographical location of affiliation
- lat, lng- geocodes (latitude and longitude) of affiliation

12. Table- acceptance_rate

- conf_id- unique id of conference (acm id)
- dblp_key- unique id of a conference in dblp database
- year- year of conference
- submitted- number of submission received in conference X in year Y
- accepted- number of accepted papers in conference X in year Y
- rate- acceptance rate of conference X in year Y.

Contact Information

If you have questions or suggestions regarding this document, please contact Swati Agarwal at 'swatia[at]iiitd[dot]ac[dot]in' or Nitish Mittal at 'nitishmittal94[at]gmail[dot]com'.