# Mô tả cấu trúc của cơ sở dữ liệu FaceDBLP

Tính đến tháng 1/2010 DBLP chứa thông tin của 1,3 triệu bài báo, trong lĩnh vực công nghệ thông tin.DBLP không dùng một hệ quản trị cơ sở dữ liệu nào để lưu dữ liệu mà dữ liệu được ghi trong 125000 files. Dữ liệu của DBLP được xuất ra các dạng CDF, XML và SQL, người phát triển có thể download các file dữ liệu này từ trên web của chương trình.

Sau đây là cấu trúc bảng SQL của DBLP được Tiến sĩ [Jörg Diederich](http://www.l3s.de/%7Ediederich) xây dựng lên từ file XML của DBLP. Dữ liệu này được sử dụng trong hệ thống tìm kiếm Faceted DBLP [[1]](#footnote-1) và được cập nhật mỗi tuần một lần bằng cách sử dụng một đoạn script đọc dữ liệu từ file XML.

Dữ liệu được lưu trữ trong 3 bảng:

* dblp\_pub\_new: lưu thông tin bài báo.

Thông tin của bài báo bao gồm tựa đề bài báo, năm xuất bản, số trang, tên nhà xuất bản, và một số thông tin định danh bài báo trong file XML của DBLP được giải thích chi tiết trong phần mô tả phía dưới. Trong cơ sở dữ liệu này mỗi bài báo có một id riêng được dùng chung cho các bảng có liên quan đến bài báo.

* dblp\_author\_ref\_new: lưu thông tin về tác giả bài báo.

Thông tin bao gồm tên tác giả, tác giả có phải là một người biên tập (editor) hay không. Trong bảng dữ liệu này, những tác giả viết cùng một bài báo thì có id giống nhau và giống id tương ứng của bài báo cótrong bảng dblp\_pub\_new.

* dblp\_ref\_new: lưu thông tin về các tham chiếu (reference) giữa các bài báo.

Bài báo có id tương ứng trong bảng dblp\_pub\_new được tham chiếu bởi những bài báo nào được xác định bằng khóa dblp\_key.



Chi tiết về các trường trong các bảng được giải thích như sau :

**Bảng** **dblp\_pub\_new**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  | | --- | --- | --- | --- | |  | **Trường (Field)** | **Kiểu**  **(Type)** | **Chú thích**  **(Comment)** | | Khóa chính  PRIMARY | id | int(8) | Khóa chính. | |  | dblp\_key | varchar(150) | Khóa trong file XML dữ liệu DBLP – Cấu trúc: tên loại tài liệu (book, conf)/ nhà xuất bản/tên tác giả đầu tiên trong tài liệu và năm công bố. | |  | title | longtext | Tựa đề của tài liệu. | |  | source | varchar(150) | Tên của hội nghị nơi bài viết được công bố: tên hội nghị, tạp chí, … | |  | source\_id | varchar(50) | Tham chiếu đến các nguồn xuất bản (phần đầu tiên của dblp\_key) - cấu trúc: tên tạp chí/tênviết tắt hội nghị + năm diễn ra hội nghị. | |  | series | varchar(100) | Series của tài liệu (chỉ áp dụng với sách và proceeding). | |  | year | int(4) unsigned | Năm xuât bản của tài liệu. | |  | type | varchar(20) | Thể loại của tài liệu ví dụ article, proceedings, … | |  | volume | varchar(50) | Tập của nơi phát hành tài liệu. (tài liệu thuộc tập mấy trong cuốn phát hành ) | |  | number | varchar(20) | Số tập của nơi phát hành tài liệu. (nơi phát hành có bao nhiêu tập ) | |  | month | varchar(30) | Tháng tài liệu được xuất bản. | |  | pages | varchar(100) | Tài liệu thuộc trang bao nhiêu trong cuốn xuất bản. | |  | ee | varchar(200) | Địa chỉ URL tới bản điện tử của tài liệu. | |  | ee\_PDF | varchar(200) | Địa chỉ URL tới bản PDF của tài liệu. | |  | url | varchar(150) | Địa chỉ của tài liệu trong dữ liệu của DBLP (bắt đầu bằng db/ ...). | |  | publisher | varchar(250) | Tên của nhà xuất bản ; tên trường đối với tài liệu là luận văn; hoặc trang chủ nơi xuất bản. | |  | Isbn | varchar(25) | *International Standard Book Number* - mã số tiêu chuẩn quốc tế có tính chất thương mại duy nhất để xác định một quyển sách. | |  | crossref | varchar(50) | Tham chiếu chéo đến một tài liệu khác. Các tài liệu trong cùng một hội nghị, cùng một năm thì có crossref giống nhau. | |  | titleSignature | varchar(255) | Tựa đề của tài liệu không bao gồm khoảng trắng và các ký tự đặc biệt. | |  | doi | varchar(255) | digital object identifier – cung cấp thông tin giúp người dùng có thể tìm được tài liệu trên Internet. | |  | mdate | Date | Lần cuối cùng chỉnh sửa thông tin tài liệu. | | | | |  |  |  |
|  | | | |  |  |  |
|  | | | |  |  |  |
| |  |  |  |  | | --- | --- | --- | --- | |  | **Trường (Field)** | **Kiểu**  **(Type)** | **Chú thích**  **(Comment)** | | PRIMARY | id | int(8) | Khóa tương ứng với id trong bảng dblp\_pub\_new. | | PRIMARY | author | varchar(70) | Tên của tác giả. | |  | editor | int(1) | Giá trị trả về giá trị là đúng khi tác giả cũng là một người biên tập (editor). | |  | author\_num | int(3) | Số thứ tự của tác giả (tương ứng trong file gốc XML) . Một bài báo có 5 tác giả thì số tương ứng bắt đầu từ 0, tác giả có số tương ứng như thế nào thì có author\_num tương tự vậy. |   **Bảng dblp\_author\_ref\_new**  **BảngTable dblp\_ ref\_new**   |  |  |  |  | | --- | --- | --- | --- | |  | **Trường**  **(Field)** | **Kiểu**  **(Type)** | **Comment** | | PRIMARY | id | int(8) | Khóa tương ứng với id trong bảng dblp\_pub\_new | |  | ref\_id | varchar(150) | Khóa dblp\_key của những bài báo được trích dẫn. | | | | |  |  |  |
|  |  |  |

1. <http://dblp.l3s.de/?q=&newQuery=yes&resTableName=query_result0n7KsQ> [↑](#footnote-ref-1)