**Mô tả cấu trúc của cơ sở dữ liệu DBLP**

Tính đến tháng 1/2010 DBLP chứa thông tin của 1,3 triệu bài báo trong lĩnh vực khoa học máy tính được thu thập từ các thư viện số, các hội nghị à các tạp chí.

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.



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. | |  | 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. | |  | 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. | |  | 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) |  | |  | 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 của tác giả trong file XML từ dblp. |   **Bảng dblp\_author\_ref\_new** | | | |  |  |  |
|  |  |  |
| Như vậy trong cấu trúc bảng không chứa dữ liệu về tóm tắt của bài báo (abstract) cũng như chủ đề của bài báo đề cập đến.  Từ cấu trúc này nhóm bổ xung thêm vào cấu trúc những bảng sau để bảng có thể lưu được các thông tin về chủ đề cũng như tóm tắt của bài báo. |  |  |
|  |  |  |
|  |  |  |

**Bảng dbsa\_sbj\_new**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Trường (Field)** | **Kiểu**  **(Type)** | **Chú thích**  **(Comment)** |
| PRIMARY | id | int(8) | Khóa chính của chủ đề |
|  | subj\_name | varchar(150) | Tên của chủ đề. |

**Bảng dbsa\_pub\_new**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Trường (Field)** | **Kiểu**  **(Type)** | **Chú thích**  **(Comment)** |
| PRIMARY | id | int(8) | Khóa ngoại của dblp\_pub\_new |
|  | sbj\_id | int(8) | Khóa ngoại của dbsa\_sbj\_new |
|  | abstract | longtext | Tóm tắt của bài báo |
|  | keyphrase | Varchar(150) | Các từ khóa trong bài báo |