**ACM (Association for Computing Machinery )[[1]](#footnote-2)**

ACM cung cấp một thư viện số mà tại đó người dùng có thể tìm kiếm các bài báo được công bố bởi ACM thông qua The ACM Digital Library, hoặc người dùng cũng thể tìm các bài báo,tạp trí ở các hội nghị khác thông qua The Guide.

Đối với The ACM Digital Library  chúng ta có thể tìm kiếm bằng một từ khóa bao gồm :

+ Thông tin về tác giả bài báo ([*Names*](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=3&dim=4)*,* [*Institutions*](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=3&dim=3208)*,* [*Authors*](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=3&dim=2)*,* [*Editors*](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=3&dim=3)*,* [*Reviewers*](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=3&dim=3209)*. )*

*+ Thông tin về nơi công bố (*[Publication Year](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=6&dim=3221), [Publication Names](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=6&dim=1), [ACM Publications](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=6&dim=3224), [All publications](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=6&dim=3225), [Content Formats](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=6&dim=3210), [Publishers](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=6&dim=3213)*).*

*+ Thông tin về hội nghị (*[Sponsors](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=5&dim=3207), [Events](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=5&dim=3198), [Proceeding Series](http://portal.acm.org/results.cfm?coll=portal&CFID=100715260&CFTOKEN=11391884&query=datamining&dl=ACM&dimgroup=5&dim=3206). *)*

Thông tin lấy được từ bài báo bao gồm :

* abstracts
* citings (where the paper has been referenced by other papers)
* references (by the paper to other papers)
* index terms from ACM's [Computing Classification System (CCS)](http://www.acm.org/class/)
* reviews from ACM's Computing Reviews.

Người dùng không được download tài liệu về nếu không có tài khỏan.

Các bài báo được phân chia chuyên đề theo khung phân loại được định nghĩa bởi ACM, khung phân loại dạng cây và được định nghĩa và bổ xung bởi người dùng . Khi một bài báo đưa lên thư viện số người dùng sẽ dựa vào hướng dẫn của ACM để chọn đúng chủ đề cho bài được đăng lên. Khi bài báo được submit lene trên trang ACM sẽ có một đội ngũ review bài của người đăng.

**CCS classification tree** [[2]](#footnote-3).

**CiteSeer [[3]](#footnote-4)**

Là một thư viện số mà tài liệu được thư viện cung cấp chủ yếu là về lĩnh vực máy tính. Thư viện số này dùng hệ thống Autonomous Citation Indexing (ACI) để đánh chỉ muc và tím kiếm tài liệu từ đó tạo cơ sở để người dùng có thể tìm kiếm được các bài báo. Khi ngườ dùng tra cứu thông tin về một bài báo thì hệ thống sẽ trả về các thông tin sau: Link download, các thông tin metadata (abtract, title, year).

======================24/8=========================

Tìm hiểu về DBLP (Digital Bibliography & Library Project) :

Là một website chứa thư mục các bài báo khoa học phát triển bởi trường đại học [Universität Trier](http://en.wikipedia.org/wiki/University_of_Trier) của [Germany](http://en.wikipedia.org/wiki/Germany). Ban đầu trang web chỉ tập trung vào cung cấp các thông tin về lĩnh vực DataBase systems và Logic Programming, sau đó hệ thống được mở rộng sang các mục khác trong khoa học máy tính. Tính đến tháng 1/2010 DBLP chứa 1,3 triệu bài báo trong lĩnh vực khoa học máy tính được thu thập từ các thư viện số, các hội nghị và 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 .

Ban dầu DBLP là một tập các danh sách các đề mục – mục lục (tables of contents –TOCs ) của các hội nghị và tạp trí trong lĩnh vực dữ liệu và lập trình logic . Các file TOCs được nhập bằng tay và theo định dạng HTML được dẫn link tới trang giới thiệu của trang đó .

Bước tiếp theo hệ thống sẽ pare file TOCs để xây dựng một "author pages" trong đó các giới thiệu các

Stefano Ceri, [Marco Brambilla](http://www.informatik.uni-trier.de/%7Eley/db/indices/a-tree/b/Brambilla:Marco.html): Search Computing: Challenges and Directions [outcome of the first SeCO Workshop on Search Computing Challenges and Directions, Como, Italy, June 17-19, 2009] [Springer 2010](http://www.informatik.uni-trier.de/%7Eley/db/conf/seco/seco2009.html)

đồng tác giả cách part như sau :

* Part file TOC để nhận diện các thông tin volume, number, author, title và page fields . lưu trong file TOC out bằng định dạng refer format.
* Dùng chương trình *mkauthors đọc dữ liệu đưa vào để so sánh với tập dữ liệu tên tác giả*

[*http://www.informatik.uni-trier.de/~ley/db/indices/AUTHORS*](http://www.informatik.uni-trier.de/~ley/db/indices/AUTHORS)

* *đề từ đó xây ựng page author*

*mkauthors* program và search engine thực hiện việc link tới cái author page của tác giả nếu nếu có trong cơ sở dữ liệu

*tiền xử lý dùng tage của* *mkhtml*

1. <http://portal.acm.org/portal.cfm> [↑](#footnote-ref-2)
2. <http://www.acm.org/about/class/ccs98-html> [↑](#footnote-ref-3)
3. <http://citeseerx.ist.psu.edu/> [↑](#footnote-ref-4)