

عنوان فارسی مقاله :

داده کاوی موسیقی برای پخش موسیقی الکترونیکی

عنوان انگلیسی مقاله :

Musical Data Mining for Electronic Music Distribution

François Pachet, Gert Westermann, Damien Laigre  
Sony CSL-Paris, 6, rue Amyot, 75005 Paris, France  
[pachet@csl.sony.fr](mailto:pachet@csl.sony.fr)

توجه !

این فایل تنها قسمتی از ترجمه میباشد.



برای تهیه مقاله ترجمه شده کامل با فرمت ورد (قابل ویرایش) همراه با نسخه انگلیسی

مقاله، [اینجا](#) کلیک نمایید.

### 3 Extracting Similarities

The automated extraction of similarities based on co-occurrence analysis requires three main steps: 1) information gathering, 2) music title and artist identification, and 3) co-occurrence analysis per se.

#### 3.1 Information gathering

Web robots were implemented to automatically query the web servers containing appropriate information. The output of this phase is a collection of text files, each file representing either an album, a radio program, or any document containing at least two music titles, said to be co-occurring, as illustrated in Figure 1, 2 and 3.

The technique consists in first identifying a sample database. We have conducted experiments with databases of various sizes as described below. For every pair of titles in the database we perform a query in each of the data sources, to look for documents containing both items. In the case of radio programs, this query is slightly modified to ensure that the two titles are actually neighbours in the play list. We assume that co-occurrence is a symmetrical function so there is a total of  $n(n-1)/2$  queries to perform for a database of size  $n$ .



توجه!

این فایل تنها قسمتی از ترجمه میباشد.

برای تهیه مقاله ترجمه شده کامل با فرمت ورد (قابل ویرایش) همراه با نسخه انگلیسی مقاله، [اینجا](#) کلیک نمایید.

همچنین برای مشاهده سایر مقالات این رشته [اینجا](#) کلیک نمایید.

**3. استخراج تشابهات**  
استخراج خودکار تشابهات بر مبنای تحلیل هم رخدادی نیازمند سه مرحله اصلی می باشد: 1) گردآوری اطلاعات، 2) شناسایی هنرمند و عنوان موسیقی و 3) تحلیل هم رخدادی

**1.3 گردآوری اطلاعات**  
برای جستجو و پرس و جوی خودکار از سرورهای وب حاوی اطلاعات مناسب، روبات های وب به معرض اجرا درآمدند. خروجی این فاز کلکسیون فایل های متنی است که هر فایل معرف یک آلبوم، یک برنامه رادیویی یا اسناد حاوی حداقل دو عنوان موسیقی است که همزمان باهم رخ می دهند، این مسئله در اشکال 1، 2 و 3 نشان داده شده است .

اولین مرحله اجرای این تکنیک شناسایی پایگاه داده نمونه می باشد. در اینجا آزمایشاتی با پایگاههای داده مختلف انجام داده ایم که در قسمت ذیل توصیف شده است. برای هر جفت عنوان در پایگاه داده، در هر یک از منابع داده دست به پرس و جو می زنیم ، هدف از این کار جستجوی اسناد حاوی هر دو آیت می باشد. در مورد برنامه های رادیویی، این پرس و جو تا حدی اصلاح شده است که هدف از این کار اطمینان از همسایه بودن دو عنوان در لیست پخش می باشد. فرض می کنیم هم رخدادی تابعی متقارن می باشد، بنابراین کلاً  $n(n-1)/2$  پرس و جو برای پایگاه داده ای با اندازه  $n$  اجرا می گردد.