Data Modeling and Databases: Project Phase 1

Due on Thursday, October 1, 2015

Qiang Qu

Alexey Chernyshov, Vladislav Kravchenko, Murad Magomedov

Contents

1 Phase 1. Design and Implement Relational Model.

1.1 Introducing

There are many repositories containing research article information, but not all of them provide interfaces to access this information. Thus we decided to use DBLP. One of its interfaces is (https://aminer.org/billboard/citation), which provides such additional information as cites and abstracts for articles. This information further will be used in ranking methods.

1.2 ER-Model

After studying the documentation we agreed on next entities for our ER-model:

- Article
- Author
- Keyword

We will extract Keywords from the abstracts in order to search for related articles. We designed this ER-Model:

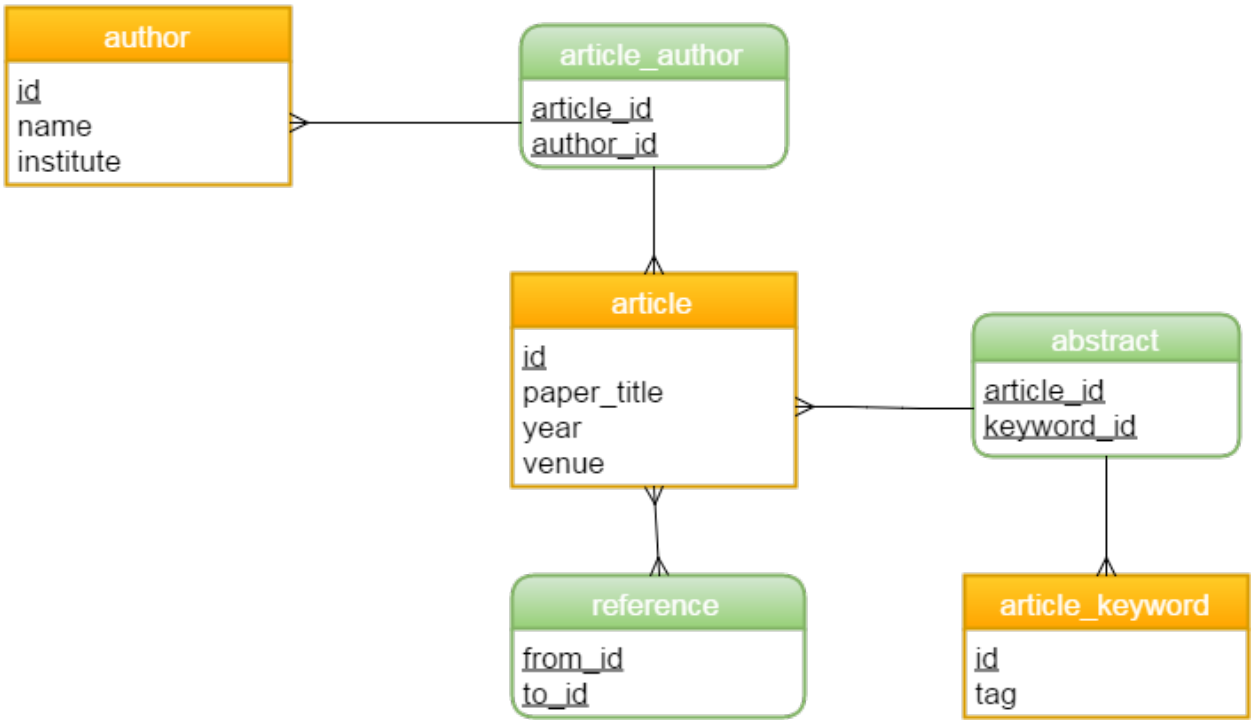

Figure 1: ER-Model diagram.

Our ER-Model was transformed into Relations:

- article: {[id: INTEGER, paper_title: STRING, venue: STRING, year: INTEGER]}
- author: {[id: INTEGER, name: STRING, institute: STRING]}
- keyword: {[id: INTEGER, tag: STRING]}
- article_author: {[id_article: INTEGER, id_author: INTEGER]}
- article_keword: {[id_article: INTEGER, id_keyword: INTEGER]}
- reference: {[from_id: INTEGER, to_id: INTEGER]}

1.3 Functionality

We designed the relations in an existing database management system (DBMS) by creating physical tables and their relationships. For our DBMS we have chosen the PostgreSQL. We imported 1.632.442 publications that we parsed using a Python script from an existing publication repository (DBLP). In addition, we created the SQL-query files that perform such operations:

- Select
- Update
- Delete
- Insert

You can search the publication based on author name, publication year, venue(conference/journal name), title, keyword, institution.
To search for the related articles you can use Keywords and References (to
from).
You can sort the publications based on such ranking methods:

- based on the number of times the Authors were cited. (In case of a new publication with no references to, but with popular authors)
- based on the number of times the Article was cited. (Popular publications)

SQL-query files can be found in attachment.

References

[1] `http://dblp.uni-trier.de/` DBLP

[2] `https://aminer.org/billboard/citation` aminer.org

[3] `http://dblp.uni-trier.de/faq/dblpxml[1].pdf` DBLP Some Lessons Learned by Michael Ley

[4] `http://dblp.uni-trier.de/xml/docu/dblpxmlreq.pdf` DBLP XML Requests