UV Distributed Information Management
Summer semester 2022

Assignment 2 – Addendum

This document contains additional remarks regarding some unclear aspects of the solution to Assignment 2.

Commands and queries are wrapped in a framed listing environment which also specifies the used command-line tool at the beginning of the title (separated by a dash –). The following listing shows an example that executes the command `ls` in a Linux `terminal` (explanatory comments are shown in gray and are not part of the command):

> `terminal` – Show directories.
> ```
> 1  ls     # this is a comment (i.e., not part of the command)
> ```

SQL Statement for Q3

Since there are different ways to modify query Q3 (as required to solve Assignment 2 properly), we provide the equivalent SQL statement for the original query Q3 (as stated below).

> **Query Q3:** mongo(.exe) – Join documents of the arXiv collection and the DBLP collection that have the same title (i.e., find identical publications that appear in both collections).
>
> ```
> 1  assignment2> db.arxiv.aggregate({
> 2  "$lookup": {
> 3    "from": "dblp",
> 4    "localField": "title",
> 5    "foreignField": "title",
> 6    "as": "arxivdblp"
> 7  }
> 8  }).pretty()
> ```

Remark: We use PostgreSQL as relational database system to execute the equivalent SQL queries, and we assume that two relations (i.e., tables) named "arxiv" and "dblp" exist, each of which has 2 columns, namely "title" and "authors".

In order to map the "as" part of the MQL query to an equivalent SQL query, we can use the `SELECT INTO` statement [1]. This statement will store the resulting tuples into a variable named "arxivdblp", which can in turn be used as a relation (cf. subsequent `SELECT` query). The commands are summarized in the following listing:

---

[1] https://www.postgresql.org/docs/current/sql-selectinto.html

> **Query Q3 using SQL:** psql(.exe) – Join tuples of the arXiv relation and the DBLP relation
> that have the same title (i.e., find identical publications that appear in both relations).

```
postgres=# SELECT arxiv.title, arxiv.authors, dblp.title, dblp.authors
           INTO arxivdblp
           FROM arxiv LEFT OUTER JOIN dblp ON arxiv.title = dblp.title;

postgres=# SELECT * FROM arxivdblp;
```