# Homework 1

Jacob Taylor Cassady

CECS 535: Introduction to Databases

The following database schema is given:

`RESEARCHER(rid,name,institution,city,country)`
`PAPER(title,journal,volume,number,year)`
`AUTHOR(resid,title)`

where `rid` is the identifier (primary key) of `RESEARCHER`, `name` is the researcher's name, `institution` is where the researcher works, and `city` and `country` the city and country where the institution is located; `title` is the paper identifier (primary key) of `PAPER`, `journal` is the journal where it was published (in volume `volume` and number `number`), and `year` is the year it appeared. Finally, in `AUTHOR` `resid` is a foreign key for `RESEARCHER` and `title` is a foreign key for `PAPER`. A researcher may write several papers, and papers may be jointly written by several researchers; this means that the key of `Author` is `(resid,title)`.

JSCHEMA <- RESEARCHER **JOIN[**rid=resid**]** AUTHOR **JOIN**[title=ptitle] **RENAME**[ptitle <- title] PAPER

# 1 List the names of authors of any paper published in journal "Databases" in 2019 in volume 12.

RESULT <- **PROJECT[**name**]SELECT**[journal = 'Databases' and volume = 12 and year = 2019]JSCHEMA

# 2 List the titles of any paper where at least one of the authors is from an institution in Boston, USA.

RESULT <- **PROJECT[**title**]SELECT**[city="Boston" and country="USA"]JSCHEMA

# 3 List the names of authors who have published a paper in either "Nature" or "Science" (journals).

NATURE <- **PROJECT[**name**]SELECT**[journal="Nature"]JSCHEMA

SCIENCE <- **PROJECT[**name**]SELECT**[journal="Science"]JSCHEMA

RESULT <- NATURE **UNION** SCIENCE

# 4 List the names of authors who have published a paper in both "Nature" and in "Science" (journals).

NATURE <- **PROJECT[**name**]SELECT**[journal="Nature"]JSCHEMA

SCIENCE <- **PROJECT[**name**]SELECT**[journal=”Science”]JSCHEMA

RESULT <- NATURE **INTERSECT** SCIENCE

# 5 List the names of authors who have never published a paper in “Nature” (journal).

NATURE <- **PROJECT[**name**]SELECT**[journal=”Nature”]JSCHEMA

RESULT <- **PROJECT**[name]JSCHEMA **-** NATURE

# 6 List the names of authors who have published two or more papers in “Nature” (journal).

TWO_OR_MORE <- AUTHOR **JOIN**[resid=resid’ and title != title’] **RENAME**[resid’<-resid, title’<-title] AUTHOR

RESULT <- **PROJECT**[name]**SELECT**[journal=’Nature’](RESEARCHER **JOIN**[resid=resid’] **RENAME**[resid’<-resid] TWO_OR_MORE)

# 7 List the names of authors who have published a paper in “Nature” but never in “Science” (journals).

NATURE <- **PROJECT[**rid**]SELECT**[journal=”Nature”]JSCHEMA

SCIENCE <- **PROJECT[**rid**]SELECT**[journal=”Science”]JSCHEMA

RESULT = **PROJECT**[name] RESERCHER **JOIN**[rid=rid’] **RENAME[**rid’<-rid**]**(NATURE – SCIENCE)

# 8 List the names of authors who have published a paper in “Nature” (journal) alone (i.e. without co-authors).

TWO_OR_MORE_AUTHORS <- AUTHOR **JOIN**[resid!=resid’ and title = title’] **RENAME**[resid’<-resid, title’<-title] AUTHOR

TWO_OR_MORE_SCHEMA<- RESEARCHER **JOIN[**rid=resid**]** TWO_OR_MORE_AUTHORS **JOIN**[title=ptitle] **RENAME**[ptitle <- title] PAPER

RELEVANT_SCHEMA <- JSCHEMA – TWO_OR_MORE_SCHEMA

RESULT <- **PROJECT[**name**]SELECT**[journal=”Nature”]RELEVANT_SCHEMA

# 9 List the title of papers where all authors are from MIT (institution). Note: this includes sole authors too.

AUTHORED_PAPERS <- **PROJECT[**title**]**AUTHOR

NON_MIT_AUTHORED_PAPERS <- **PROJECT[**title**]SELECT**[institution != “MIT”]RESEARCHER **JOIN[**rid=resid**]** AUTHOR

RESULT <- AUTHORED_PAPERS – NON_MIT_AUTHORED_PAPERS

# 10 List the pairs of authors who have co-authored a paper. Note: if a paper has 3 authors (say, Larry, Curly and Moe), then (Larry,Curly), (Curly,Moe) and (Larry,Moe) are co-authors. It’s ok to list both (Larry,Curly) and (Curly,Larry) in the answer.

AUTHOR_ONE <- **PROJECT**[title, rid, name] **(**RESEARCH **JOIN**[rid=resid] AUTHOR)

AUTHOR_TWO <- **PROJECT**[title, rid, name] **(**RESEARCH **JOIN**[rid=resid] AUTHOR)

RESULT <- **PROJECT**[AUTHOR_ONE.name, AUTHOR_TWO.name]**SELECT**[AUTHOR_ONE.rid != AUTHOR_TWO.rid and AUTHOR_ONE.title = AUTHOR_TWO.title](AUTHOR_ONE **CARTESIAN PRODUCT** AUTHOR_TWO)