On the Naming of Database Objects in the SQL Databases of Some Existing Software

Erki Eessaar

Department of Software Science,

Tallinn University of Technology,

Estonia

erki.eessaar@taltech.ee

Outline

- Background and research questions.
- How we searched the occurrences of naming problems of database objects?
- The results.
  - Statistics.
  - An example.
- Conclusions and future work.

93 characters out of **197** (**47%**) are a part of a name, i.e., identifier

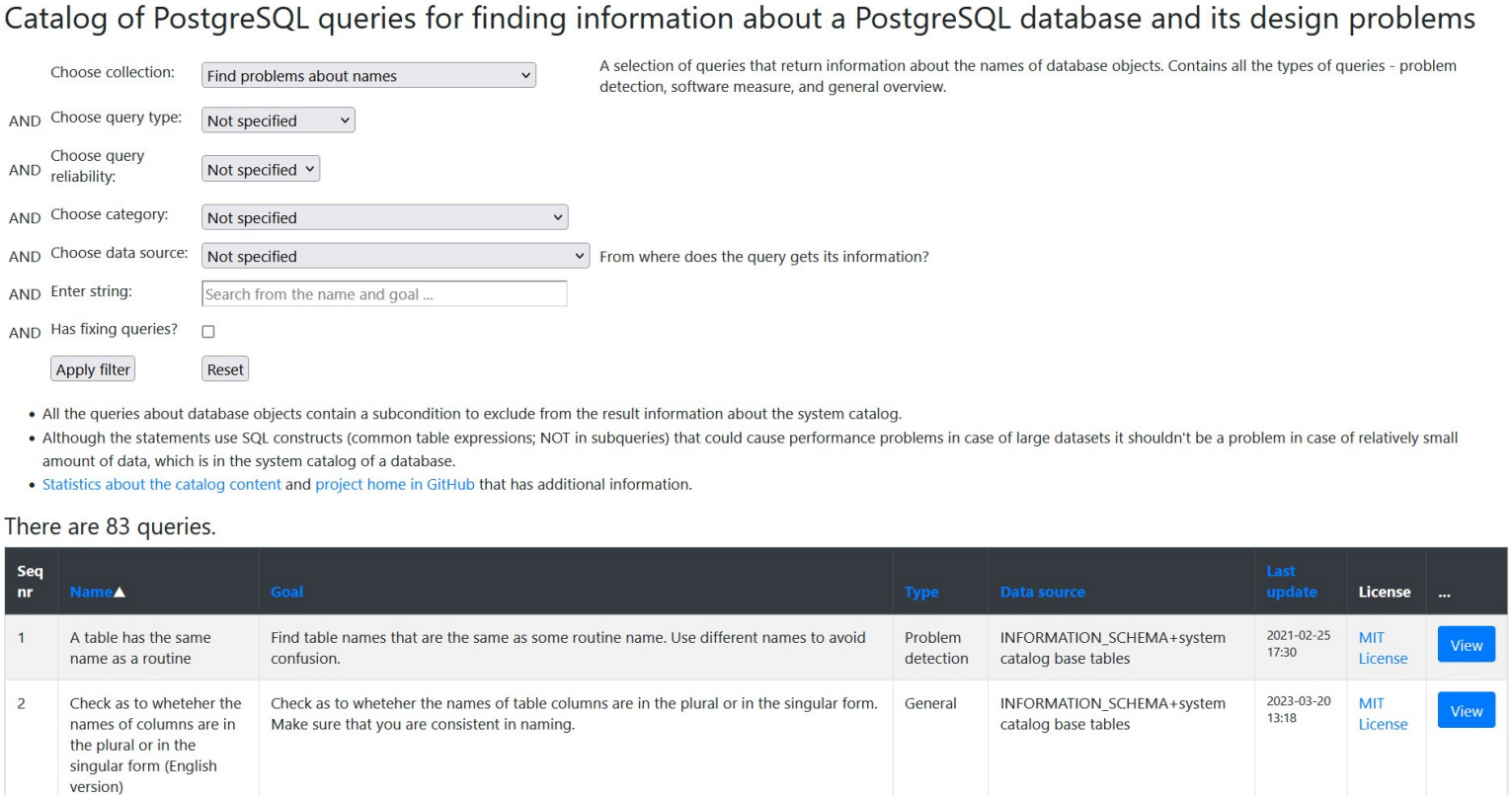
A SQL statement

```
CREATE TABLE Person (person_code
SERIAL,
e_mail VARCHAR(254) NOT NULL,
given_name VARCHAR(50) NOT NULL,
CONSTRAINT pk_person PRIMARY KEY
(person_code),
CONSTRAINT ak_person_e_mail UNIQUE
(e_mail));
```

The research questions

- Is it *possible* and *feasible* to investigate the names of SQL database objects by making **queries** based on the **system catalog** of the database?
- What **problems with naming** exist in the SQL databases of existing programs that have a *long development history*?
- Can the **lexicon bad smells** and **linguistic antipatterns** of software elements also occur in the names of SQL database objects?

The catalog



Catalog of PostgreSQL queries for finding information about a PostgreSQL database and its design problems

Choose collection: Find problems about names

A selection of queries that return information about the names of database objects. Contains all the types of queries - problem detection, software measure, and general overview.

AND Choose query type: Not specified

AND Choose query reliability: Not specified

AND Choose category: Not specified

AND Choose data source: Not specified — From where does the query gets its information?

AND Enter string: Search from the name and goal ...

AND Has fixing queries? ☐

Apply filter Reset

- All the queries about database objects contain a subcondition to exclude from the result information about the system catalog.
- Although the statements use SQL constructs (common table expressions; NOT in subqueries) that could cause performance problems in case of large datasets it shouldn't be a problem in case of relatively small amount of data, which is in the system catalog of a database.
- Statistics about the catalog content and project home in GitHub that has additional information.

There are 83 queries.

Seq nr	Name▲	Goal	Type	Data source	Last update	License	...
1	A table has the same name as a routine	Find table names that are the same as some routine name. Use different names to avoid confusion.	Problem detection	INFORMATION_SCHEMA+system catalog base tables	2021-02-25 17:30	MIT License	View
2	Check as to whetеher the names of columns are in the plural or in the singular form (English version)	Check as to whetеher the names of table columns are in the plural or in the singular form. Make sure that you are consistent in naming.	General	INFORMATION_SCHEMA+system catalog base tables	2023-03-20 13:18	MIT License	View

https://github.com/erki77/database-design-queries

We have developed a large set of PostgreSQL system catalog-based queries for searching database design problems of PostgreSQL databases.

The catalog (2)

- Many of the queries directly point to problem occurrences.
  - Mistakes.
  - Design smells.
    - Will cause later maintenance problems.
- Each such query documents a design problem.
  - The absolute majority of these could appear in the databases of any SQL DBMS.

- A long development history, still actively used
- Use a PostgreSQL database

The analysis – databases

- FusionForge
  - An open source development management and team collaboration software.
  - Development started in **2001**.
  - **2612** named objects. **1428** different names (**54%**).
- LedgerSMB
  - An open source enterprise resource planning software.
  - Development started in **2006**.
  - **4072** named objects. **2198** different names (**53%**).

The analysis – databases (2)

- OTRS Community Edition
  - An open source ticketing software, which can be used to track and manage issues that need resolving.
  - Development started in **2001**.
  - **1827** named objects. **1015** different names (**55%**).
- Stansoft
  - A Linux financial accounting software.
  - **2398** named objects. **2251** different names (**93%**).

Resulting catalog of naming problems

- In total, we identified **30** problems in the analyzed databases.
- Many have *more than one* **sign**, i.e., subproblems.
  - The collection of the used queries for this research contains **54** *problem detection queries*.
- **36.7% (11)** of the problems were present in all the databases.

Resulting catalog of problems (2)

- We did not find any literature reference to **76.7%** (**23**) of these problems.
- We searched the occurrences of a larger set of problems (**132** problem detection queries).
  - Present only those problems that had at least occurrence in at least one of the databases.
  - We created **21** problem detection queries based on the *lexicon bad smells* and *linguistic antipatterns*.

A classification of the problems (problem area)

- General
  - **11** problems
    - **6** not in the literature
- Candidate keys
  - **3** problems
    - **2** not in the literature
- Foreign keys/ relationships
  - **3** problems
    - **None** in the literature
- Tables, views, columns
  - **10** problems
    - **9** not in the literature
- Routines, parameters
  - **3** problems
    - **None** in the literature

A classification of the problems (problem reason)

- Imprecision
  - **17** problems. For instance, extreme contraction, the same name is used in multiple contexts, too generic candidate and foreign key column names.
- Inconsistencies
  - **9** problems. For instance, in the use of
    - writing style,
    - prefixes and suffixes,
    - singular and plural.

7 foreign key columns in FusionForge

An example

- Candidate key and foreign key column names are very similar, but not identical, e.g., *user_id* and *userid* (difference by one symbol).
  - SELECT … FROM Users INNER JOIN Forum_attachment **ON Users.user_id= Forum_attachment.userid**;
  - SELECT … FROM Users INNER JOIN Forum_attachment **USING (user_id)**;
- **Other problems**: inconsistent use of snake_case and plural/singular, too generic foreign key column name.

Conclusions

- All the databases had a lot of different problems with naming of database objects.
- We presented a lot of problems that have not been published before.
- Using queries to search naming problems is possible and feasible – executing all the queries of the collection based on a database takes about **30 seconds**.

Future work

- Investigating other databases in terms of the same problems.
  - Perhaps development practices of **commercial systems** lead to *different outcomes*.
- Investigating the patterns of name changes over time.
- Investigating the impact to the usage and to the maintainability of databases.

Thank you for your attention!

Questions?

Erki Eessaar – erki.eessaar@taltech.ee

Reference to the catalog:
https://github.com/erki77/database-design-queries

- Collections "Find problems about names" and "Lexicon bad smells and linguistic antipatterns".