SQL, a query language for relational databases

Silvio Peroni
silvio.peroni@unibo.it – https://orcid.org/0000-0003-0530-4305 – @essepuntato

Data Science (A.Y. 2021/2022)
Second Cycle Degree in Digital Humanities and Digital Knowledge
Alma Mater Studiorum - Università di Bologna

DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA

Summary of the previous lectures (1/2)

A **datum** is a declarative statement subject-predicate-object that, through the predicate, either **attributes** a **literal** (i.e. a value such as a string, a number, etc.) to a subject **entity** *or* it **relates** such a subject **entity** with another **entity**

Each entity, being used either as subject or object of a statement, is characterised by **a unique identifier**

The **same entity** can be used as subject or object in one or more data, while a literal **cannot be used** as subject in any datum

An attribute is intrinsically **part of** the entity to which it is associated – modifying the value of an attribute affect **only** the entity to which it refers to

A **data model** is an abstract, simplified and formal representation of some data related to a system or a real domain, and enables us to describe what a data collection is about and to check data correctness

A data model permit one to specify **classes** of entities, their **attributes** and **relations**

Summary of the previous lectures (2/2)

Depending on the structure in which data are stored (or exposed), you need to approach the queries to datasets from a different angle

- With **tabular data**, often you have to combine tables between them to obtain bigger tables which contain the query requirements and the related answer
- With **graph data**, you explore the graph starting from fixed points (i.e. known entities, values, predicates) to find a pattern that is compliant with the query

A **database** as a **collection of data** which organised, stored and accessed electronically, which can be created through a database management system (DBMS)

A **transaction** is a unit of work performed (compliant with **ACID properties**) within a DBMS against a database and usually represents any change in a database

Any question about the previous lecture?

SQL

The Structured Query Language (SQL) is a query language used and designed for managing data in a relational database management system

It is a standard that is implemented in all relational database management systems, and allows one to create tables, to populate them, and finally to query them using a particular syntax

Be aware: even if it is a standard, porting SQL code from a database management system to another may require some changes

A good tutorial is available at: https://www.sqlitetutorial.net/

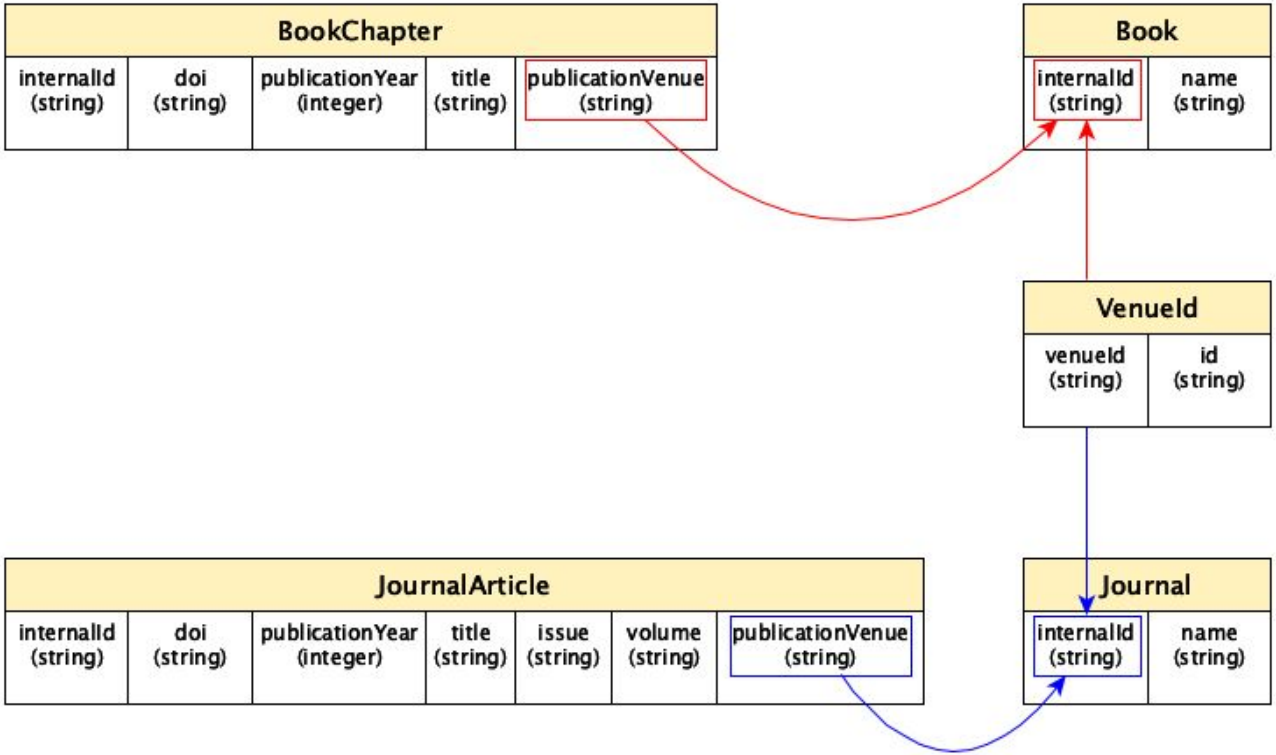
Basic SQL syntax for queries

`SELECT <columns>`	A comma-separated list of columns to have as result of the query (“*” means *all the columns*)
`FROM <tables>`	The table(s) from which to retrieve the data
`WHERE <conditions>`	[Optional clause] A condition that must hold to include the row of a table in the result of the query
`;`	

Questions to remind for designing a query:

- Which tables do you need?
- Which columns do you select?

Exemplar tables for queries



Data in the tables

VenueId

venueId	id
venue-0	1531-6912
venue-1	1367-5931
venue-2	9780470291092
venue-3	1027-3662

Book

internalId	name
venue-2	Proceedings of the 5th Annual Conference on Composites and Advanced Ceramic Materials: Ceramic Engineering and Science Proceedings

Journal

internalId	name
venue-0	Comparative and Functional Genomics
venue-1	Current Opinion in Chemical Biology
venue-3	Journal of Theoretical Medicine

BookChapter

internalId	doi	publicationYear	title	publicationVenue
publication-2	10.1002/9780470291092.ch20	1981	Mechanisms of Toughening in Ceramic Matrix Composites	venue-2

JournalArticle

internalId	doi	publicationYear	title	issue	volume	publicationVenue
publication-0	10.1002/cfg.304	2003	Development of Computational Tools for the Inference of Protein Interaction Specificity Rules and Functional Annotation Using Structural Information	4	4	venue-0
publication-1	10.1016/s1367-5931(02)00332-0	2002	In vitro selection as a powerful tool for the applied evolution of proteins and peptides	3	6	venue-1

Queries

Retrieve complete information about all journal articles

Retrieve the titles of all journal articles

Retrieve the title of the journal article with DOI “10.1016/s1367-5931(02)00332-0”

Retrieve the title of the publication with DOI “10.1016/s1367-5931(02)00332-0”

Return the name of the journal of the article with DOI “10.1016/s1367-5931(02)00332-0”

Return the id and name of the journal of the article with DOI “10.1016/s1367-5931(02)00332-0”

Query 1

Retrieve complete information about all journal articles

- Which tables do you need? JournalArticle
- Which columns do you select? * *(all columns)*

```
SELECT *
FROM JournalArticle;
```

JournalArticle

internalId	doi	publicationYear	title	issue	volume	publicationVenue
publication-0	10.1002/cfg.304	2003	Development of Computational Tools for the Inference of Protein Interaction Specificity Rules and Functional Annotation Using Structural Information	4	4	venue-0
publication-1	10.1016/s1367-5931(02)00332-0	2002	In vitro selection as a powerful tool for the applied evolution of proteins and peptides	3	6	venue-1

Query 2

Retrieve the titles of all journal articles

- Which tables do you need? JournalArticle
- Which columns do you select? title

```
SELECT title
FROM JournalArticle;
```

JournalArticle

internalId	doi	publicationYear	title	issue	volume	publicationVenue
publication-0	10.1002/cfg.304	2003	Development of Computational Tools for the Inference of Protein Interaction Specificity Rules and Functional Annotation Using Structural Information	4	4	venue-0
publication-1	10.1016/s1367-5931(02)00332-0	2002	In vitro selection as a powerful tool for the applied evolution of proteins and peptides	3	6	venue-1

Query 3

Retrieve the title of the journal article with DOI “10.1016/s1367-5931(02)00332-0”

- Which tables do you need? JournalArticle
- Which columns do you select? title

```
SELECT title
FROM JournalArticle
WHERE
doi='10.1016/s1367-5931(02)00332-0';
```

JournalArticle

internalId	doi	publicationYear	title	issue	volume	publicationVenue
publication-0	10.1002/cfg.304	2003	Development of Computational Tools for the Inference of Protein Interaction Specificity Rules and Functional Annotation Using Structural Information	4	4	venue-0
publication-1	10.1016/s1367-5931(02)00332-0	2002	In vitro selection as a powerful tool for the applied evolution of proteins and peptides	3	6	venue-1

Query 4

Retrieve the title of the publication with DOI “10.1016/s1367-5931(02)00332-0”

- Which tables do you need? BookChapter, JournalArticle
- Which columns do you select? title

```
SELECT title
FROM BookChapter
WHERE
doi='10.1016/s1367-5931(02)00332-0'
UNION
SELECT title
FROM JournalArticle
WHERE
doi='10.1016/s1367-5931(02)00332-0';
```

BookChapter

internalId	doi	publicationYear	title	publicationVenue
publication-2	10.1002/9780470291092.ch20	1981	Mechanisms of Toughening in Ceramic Matrix Composites	venue-2

JournalArticle

internalId	doi	publicationYear	title	issue	volume	publicationVenue
publication-0	10.1002/cfg.304	2003	Development of Computational Tools for the Inference of …	4	4	venue-0
publication-1	10.1016/s1367-5931(02)00332-0	2002	In vitro selection as a powerful tool for the applied evolution of …	3	6	venue-1

Query 5

Return the name of the journal of the article with DOI “10.1016/s1367-5931(02)00332-0”

- Which tables do you need? JournalArticle, Journal
- Which columns do you select? name

```
SELECT name
FROM JournalArticle LEFT JOIN Journal ON
     JournalArticle.publicationVenue
     ==
     Journal.internalId
WHERE doi='10.1016/s1367-5931(02)00332-0';
```

Journal

internalId	name
venue-0	Comparative and Functional Genomics
venue-1	Current Opinion in Chemical Biology
venue-3	Journal of Theoretical Medicine

JournalArticle

internalId	doi	publicationYear	title	issue	volume	publicationVenue
publication-0	10.1002/cfg.304	2003	Development of Computational …	4	4	venue-0
publication-1	10.1016/s1367-5931(02)00332-0	2002	In vitro selection as a powerful tool for the applied evolution of …	3	6	venue-1

JournalArticle ⋈ Journal

internalId	doi	publication Year	title	issue	volume	publication Venue	internalId	name
publication-0	10.1002/cfg.304	2003	Dev…	4	4	venue-0	venue-0	Comparative and …
publication-1	10.1016/s1367-5931(02)00332-0	2002	In…	3	6	venue-1	venue-1	Current Opinion in Chemical Biology

Query 6

```
SELECT id , name
FROM JournalArticle LEFT JOIN Journal ON
     JournalArticle.publicationVenue == Journal.internalId
     LEFT JOIN VenueId ON Journal.internalId == VenueId.venueId
WHERE doi='10.1016/s1367-5931(02)00332-0';
```

Return the id and name of the journal of the article with DOI “10.1016/s1367-5931(02)00332-0”

- Which tables do you need? JournalArticle Journal VenueId
- Which columns do you select? id name

Journal

internalId	name
venue-0	Comparative and Functional Genomics
venue-1	Current Opinion in Chemical Biology
venue-3	Journal of Theoretical Medicine

JournalArticle

internalId	doi	publicationYear	title	issue	volume	publicationVenue
publication-0	10.1002/cfg.304	2003	Dev…	4	4	venue-0
publication-1	10.1016/s1367-5931(02)00332-0	2002	In…	3	6	venue-1

VenueId

venueId	id
venue-0	1531-6912
venue-1	1367-5931
venue-2	9780470291092
venue-3	1027-3662

JournalArticle ⋈ Journal ⋈ VenueId

internalId	doi	pub…	title	issue	volume	pub…	internalId	name	venueId	id
publication-0	10.1002/cfg.304	2003	De…	4	4	venue-0	venue-0	Comp…	venue-0	1531-6912
publication-1	10.1016/s1367-5931(02)00332-0	2002	In…	3	6	venue-1	venue-1	Curr…	venue-1	1367-5931

Do you want to try them with real data?

Install SQLite on your computer – see https://www.sqlite.org/download.html, and also this guide (https://www.sqlitetutorial.net/download-install-sqlite/) for Windows users (for Apple users, it is enough to install the DBMS using the installer)

Run the the tool specifying the database file contained in the same directory (documentation at https://sqlite.org/cli.html)
`sqlite3 publications.db` (it is `sqlite3.exe` in Windows)

Set the output mode to markdown to improve readability
`sqlite>` **`.mode markdown`**

Execute the SQL query
`sqlite>` **`SELECT * FROM Journal;`** (remember the ";" at the end of the query)

Close SQLite when you finished
`sqlite>` **`.exit`**

End

SQL, a query language for relational databases

Silvio Peroni
silvio.peroni@unibo.it – https://orcid.org/0000-0003-0530-4305 – @essepuntato

Data Science (A.Y. 2021/2022)
Second Cycle Degree in Digital Humanities and Digital Knowledge
Alma Mater Studiorum - Università di Bologna

DIPARTIMENTO DI FILOLOGIA CLASSICA E ITALIANISTICA