

Gutenberg-Project

*Made by: Murched Kayed, Hallur vid Neyst & Zaeem Shafiq*



May 30, 2019

COPENHAGEN BUSINESS academy

Nørgaardsvej 30, DK-2800 Kgs. Lyngby

# Introduction

This rapport is about the solution we made for the Gutenberg project, so we will speak about which two databases and technologies we used to solve it. This rapport will also show how we imported the data from the files into the databases.

We agreed as a group that we are going to use Java to solve this project, since we all 3 had good skills with Java. We agreed also to use MangoDB and SQL as databases, since we are going to work with documents so it will be an advantage to use MangoDB which is a document database, we agreed to use SQL because all of us had good skills using it.

* How data is modeled in the database.
* How data is modeled in your application.
* How the data is imported.

We created a Java program to import the data into the databases.

What the program does:

1. It generate a cities.csv file with the object we agreed we need to solve the queries, it is a simplify version of the city text file that we downloaded from this link: <http://download.geonames.org/export/dump/cities15000.zip>.
2. The program reads the RDF file for each book, then the program checks after the books (that we downloaded them from Gutenberg webpage into to a cloud machine at digitalocean, we used vagrant for that), to create a list of the cities mentions.
3. After the program reads and generated the need files, it imports the files into the databases (each of the databases has a docker container).

* Behavior of query test set. Including a discussion on how much of the query runtime is influenced by the DB and what is influenced by the application frontend.
* Your recommendation, for which database to use in such a project for production.