

# Towards More Effective Querying of Medical Literature using Alexandria3K

Author:

Bas Verlooy

s.j.verlooy@student.tudelft.nl

## How can Alexandria3K be useful in performing literature surveys?

### 1 Background

#### PubMed

- Medical specific data
- Comparing how papers evolved over time
- Can set precedent for further research
- 325,000 items tagged as literature review in PubMed [1]

### 2 Methodology

- Implementing PubMed dataset in Alexandria3K
- Reevaluating literature survey concerning pathogens researched in articles
- Statistical analysis regarding software tools in articles

### 3 Dataset

- 1.7 billion rows
  - 156 GB
  - 36 million publications
- Compared to Crossref
- Crossref 1.45x bigger
  - Crossref 3.77x more publications

### 6 Conclusions

- Comparable results to the original
- Manual confirmation still needed  
Future work
- More analysis could be done to confirm on CrossRef
- Extend Alexandria3K with text extraction for example

### 4 Literature reevaluation

- 12 different pathogens searched using SQL queries
- 3 different regions searched: globally, Gulf Region, Bahrain
- Bahrain results manually compared

#### Differences for Bahrain

- 2x pathogen listed was different
- 2x no mention of Bahrain in available data
- 1x not indexed in PubMed
- 2x article not included in original found with Alexandria3K

### 5 Statistical analysis

- 12 software tools analyzed over 3 different years
- 16% not found
- Manual confirmation still needed
- Queries possible within second using FTS

