

# Research on Voice to Query & Databases

Bachelorprojekt "Ask Your Repository"

Luise Benkert

## Process

1. Input from voice recognition:
  - JSON files including data type, author, date and other relevant keywords
2. [Optional] Preprocessing of input data
  - E.g. case shift, synonyms, matching
3. Automatic Query Generation
4. Search database and return relevant data

## Challenges

- Unstructured data (images, repositories, diagrams, log files, code, plain text)
- Continuously growing data
- High variance in search terms (synonyms, spelling, matching)
- Interrelationships between data (commits, authors, files)

## Requirements

- Open Source
- Modular and portable / integrable
- Scalable
- Secure
- Fast
- Robust / reliable

## Databases

1. Key-Value Database → Hash table of Key-Value pairs
  - most simple and versatile data structure
2. Document-based Database → Collection of documents with internal structure
  - particularly good for document-style data sets
3. Column-based Database → Relational database that stores data by column
  - simple but performant alternative to row-based relational database
4. Graph-based Database → Network of data nodes
  - particularly good for highly interrelated data and join-heavy/nested queries
5. Search Engine → Data Storage in schema-free JSON documents
  - particularly good for text-based data (offers full-text search including thesaurus, suggestions, matching etc.)

	Key-Value	Document-based	Column-based	Graph-based	Search Engine
Scalability	- -	+ +	+ + +	+ + +	+ + +
Portability	+ + +	+ +	+ +	+ + +	+ +
Performance	+	+ +	+ +	+ + +	+ +
Simplicity	+ + +	+	+ + +	-	+

## Further Research

<https://www.3pillarglobal.com/insights/exploring-the-different-types-of-nosql-databases>

<http://www.dataversity.net/review-pros-cons-different-databases-relational-versus-non-relational/>

<https://jaxenter.com/evaluating-nosql-performance-which-database-is-right-for-your-data-107481.html>

[http://lucene.apache.org/solr/guide/7\\_5/index.html](http://lucene.apache.org/solr/guide/7_5/index.html)

<http://cassandra.apache.org/doc/latest/>