

Database Structure

- **Collection**
 - Key: timestamp of commit
 - format: "YYYY-MM-DD 12:00:00"
 - Records: each file in the repo at the time of the commit
 - Requires storing all artifacts of the commit regardless of change
 - Additional Record: artifact metrics (for analysis view)
- **Record (artifact metrics)** - one record for a repo
 - Number of documents
 - Key: "num_doc"
 - [num_req, num_src, diff_st, diff_ts]
 - All of the values in the list will be numbers
 - 'diff_st' means the value of the difference from source files (requirement files) to target files (source code files)
 - 'diff_ts' means the value of the difference from target files (source code files) to source files (requirement files)
 - Vocabulary Size
 - Key: "vocab_size"
 - [vocab_req, vocab_src, diff_st, diff_ts]
 - Average Number of Tokens Used
 - Key: "avg_tokens"
 - [token_req, token_src, diff_st, diff_ts]
 - Requirement Vocabulary
 - Key: "rec_vocab"
 - {'token1' : [count, freq], 'token2' : [count, freq], 'token3' : [count, freq]}
 - Note: this does not need to be formatted as this matched the direct output from DS4SE's Vocab method, which returns the three most common tokens and their corresponding counts/frequencies
 - Source Code Vocabulary
 - Key: "src_vocab"
 - {'token1' : [count, freq], 'token2' : [count, freq], 'token3' : [count, freq]}
 - Shared Vocabulary
 - Key: "shared_vocab"
 - {'token1' : [count, freq], 'token2' : [count, freq], 'token3' : [count, freq]}
- **Record (individual artifacts)** - a record for every artifact in repo
 - Dictionary of values:
 - Artifact name

- Key: "name"
- Artifact type
 - Key: "type"
 - two types:
 - requirement: "req"
 - source code: "src"
- Artifact content
 - Key: "content"
 - Full file content
- Traceability links
 - Key: "links"
 - List of Tuples: [(target1, [(tech1, val)...(tech7, val)]), ... (targetN, [(tech1, val)...(tech7, val)])]
- Orphan - there are no traceability links to other artifacts
 - Key: "orphan"
 - True/False
- Security-related
 - Key: "security"
 - True/False/Not a requirements file