



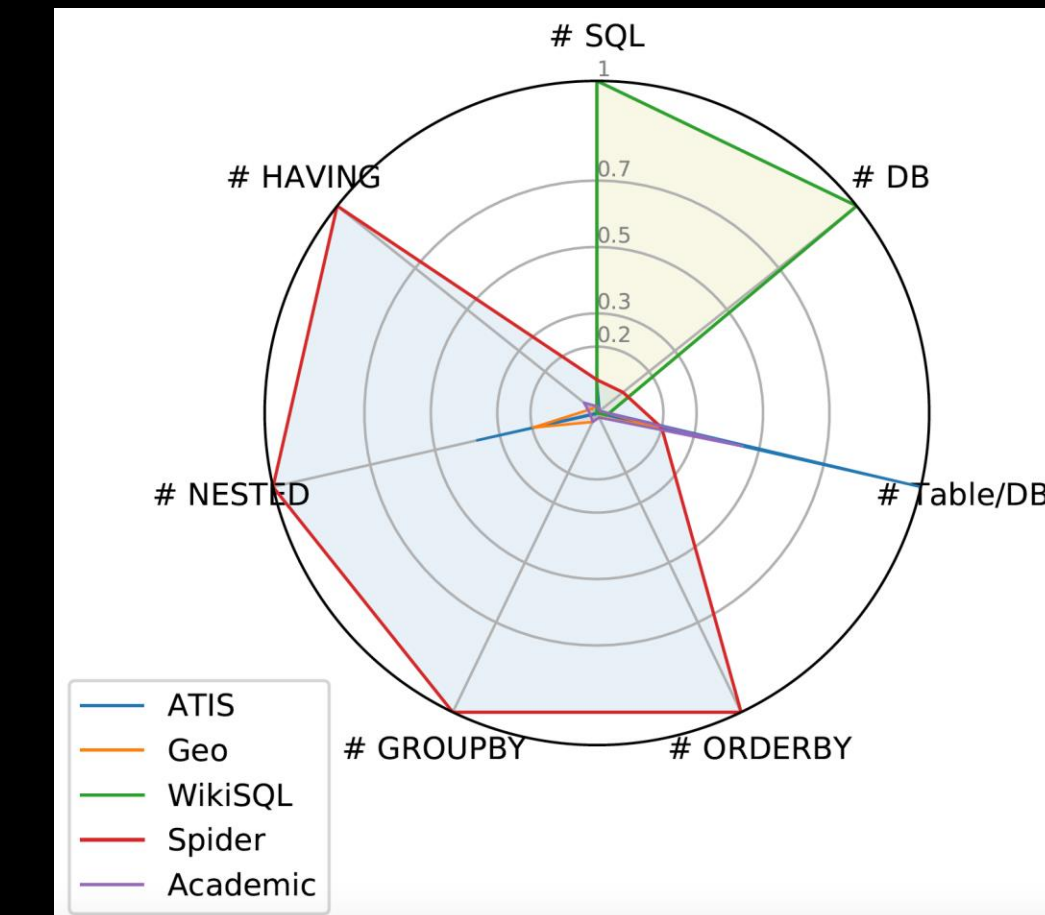
Seth Cram



Khoi Nguyen

AI/ML BASED NATURAL LANGUAGE INTERFACES TO DATABASES

Authors: Seth Cram, Khoi Nguyen
Client: Dr. Jamil (University of Idaho)



OBJECTIVE

To allow non-technical users to access information from electronic database systems.

VALUE PROPOSITION

- Electronic database systems exist in many fields
- Database access is gated by knowledge of query languages like SQL
 - Query languages have increased search granularity compared to typical search engines
- Solution: Natural Language to SQL (NL2SQL) system
- Our implementation: a user interface hosted on the Web
 - Inputs: question in natural language (English), a database file
 - Output: appropriate SQL query, data queried

BACKGROUND

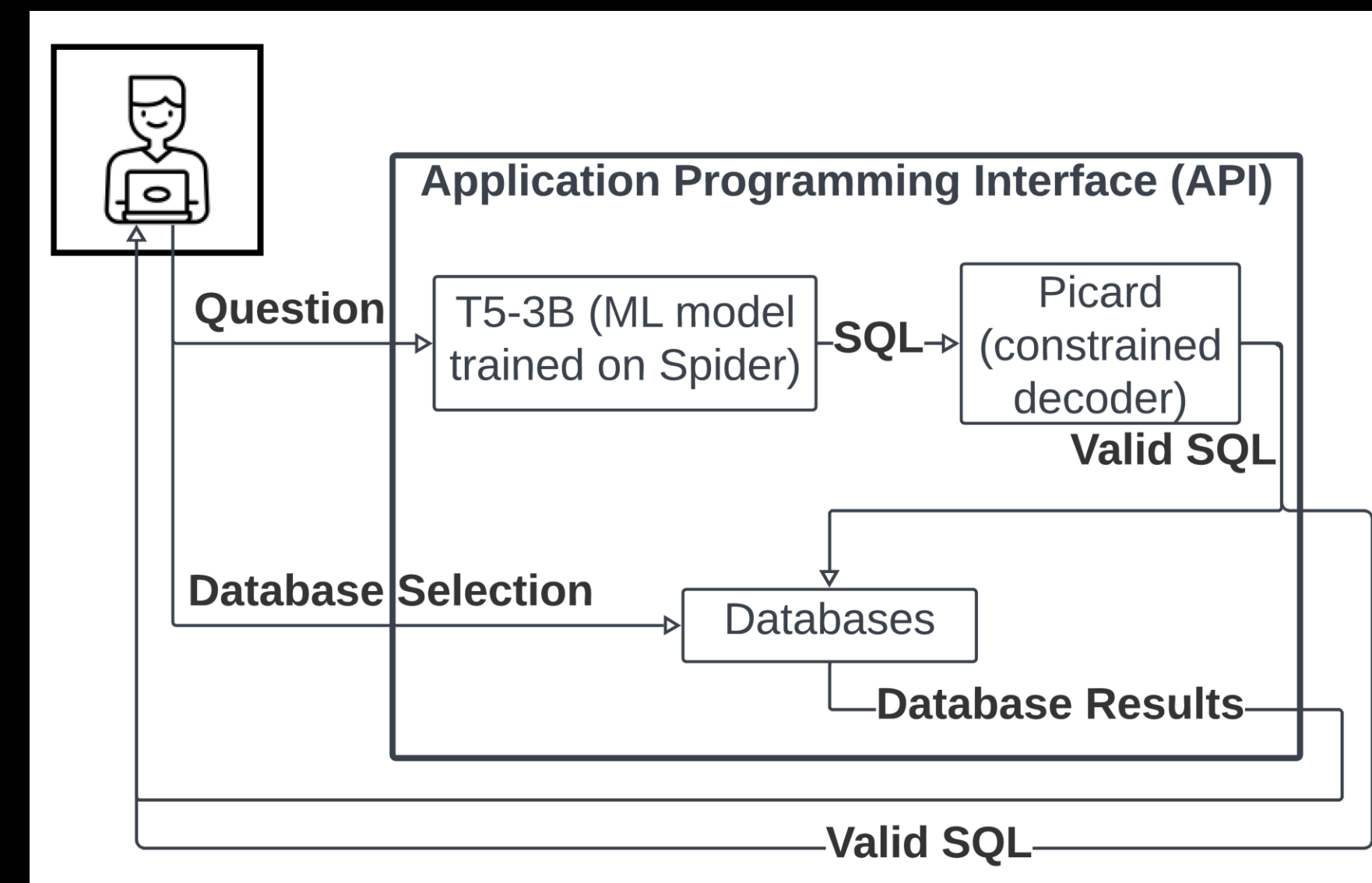
- Campaigned several NL2SQL systems against one another to find the optimal solution
- Deciphered top performing system's development process

KEY REQUIREMENTS

- Should maintain reasonable efficacy on a "general" database (e.g. one from outside the training set)
- Can generate SQL queries joining multiple tables, performing aggregation, and nested queries
- Allows for separate input of a natural language question and database schema
- System can run on any machine with internet access

CONCEPT DEVELOPMENT

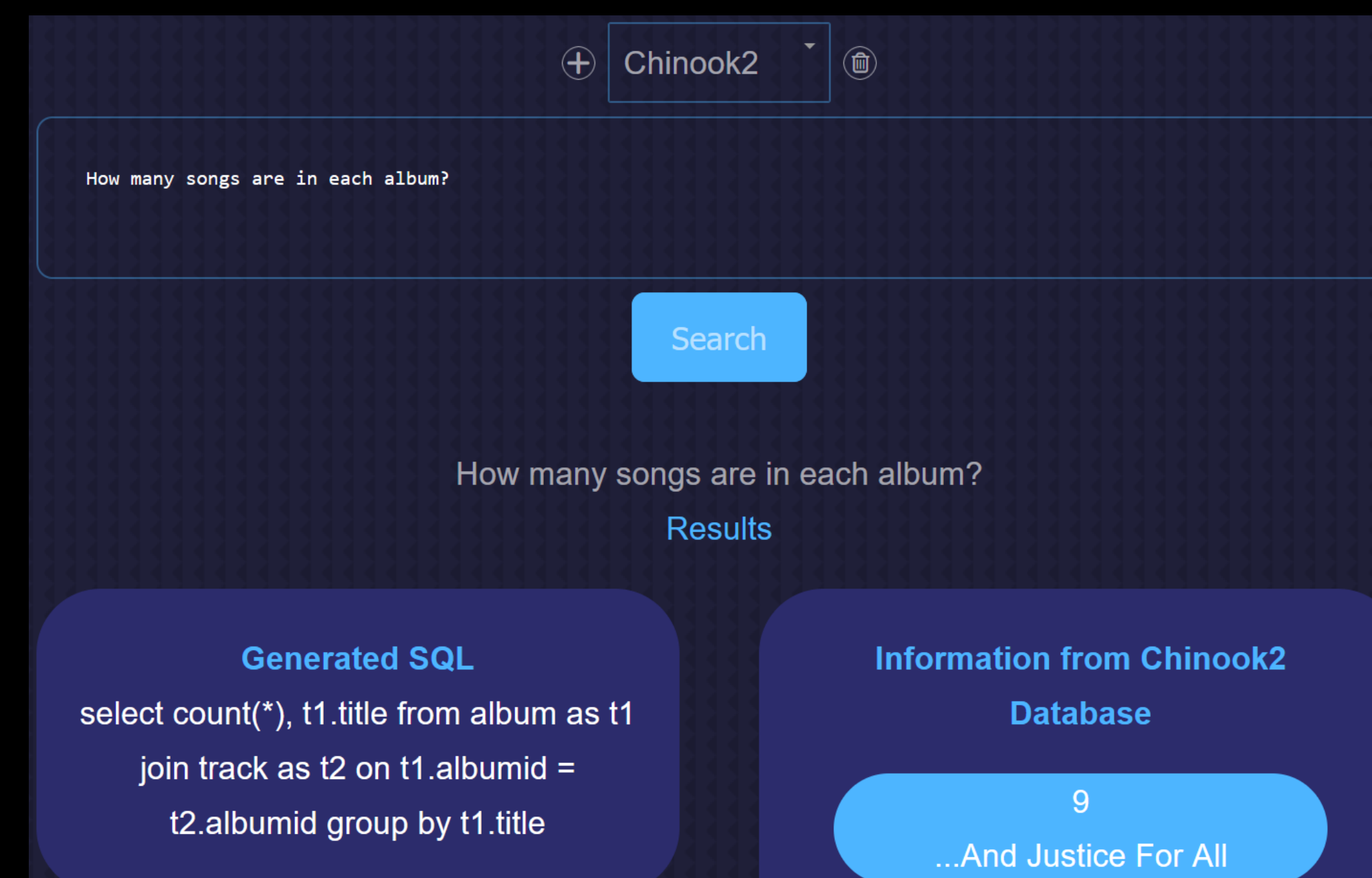
User's Interaction with the System



- Technologies:
 - Python
 - FastAPI
 - React.js
 - Node.js
 - HTML
 - CSS

- Several API methods added to improve accessibility:
 - Store uploaded databases or SQLite files as databases
 - Allow retrieval of uploaded database(s)

FINAL DESIGN



Responsive and flexible User Interface

- Capabilities:
 - Single-clause questions (can be complex)
 - Generates SQLite queries, takes SQLite files and/or databases

VALIDATION

- Initial model: trained on Spider dataset
 - Best performing open source system
 - 71.9% exact set match and 75.1% value execution accuracy
 - More accurate results when databases have contents
- Client assignment
 - Individual clauses have reasonable accuracy
 - Multi-clause questions often miss later clauses

CONCLUSION

- For future development, each selectable database should be visualizable via an ER diagram
- Model does well in generalization but struggles with complexity
- NL2SQL-specific ML models still have difficulties with more complex natural language inputs, but other technologies like ChatGPT are rapidly revolutionizing the field

ACKNOWLEDGEMENTS

- Mentor: Sebastian Garcia
- Lead Instructor: Dr. Chakhchoukh