

Dr. Exel



chatbot

Members:

Tom Amon

Aaron Campbell

Daniel Fitzick

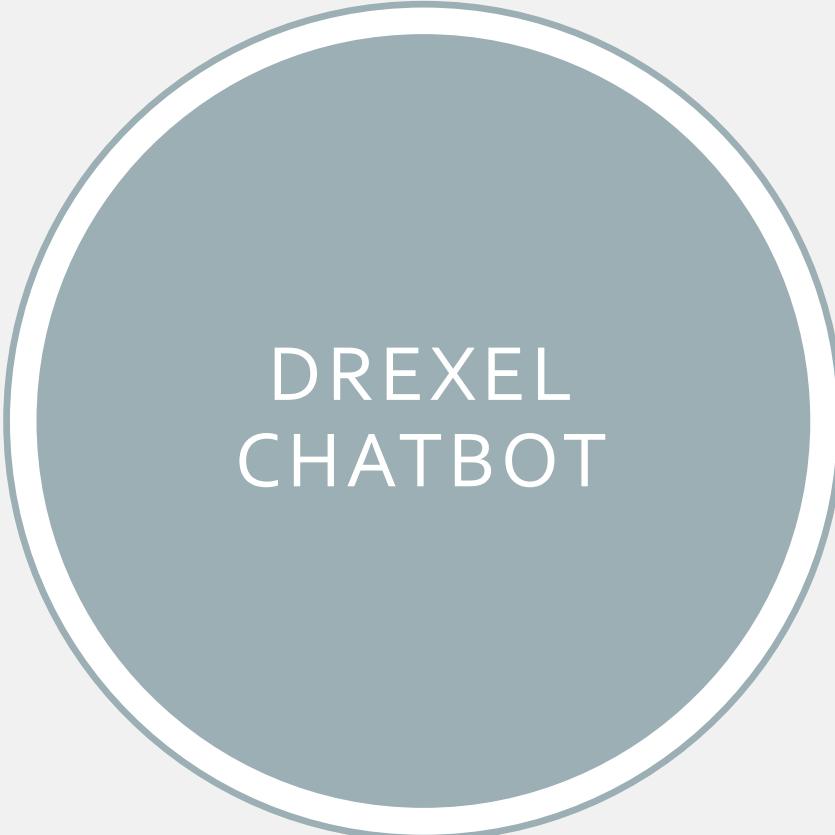
Shishir Kharel

Hoa Vu

Nanxi Zhang

Stakeholder:

Marcello Balduccini



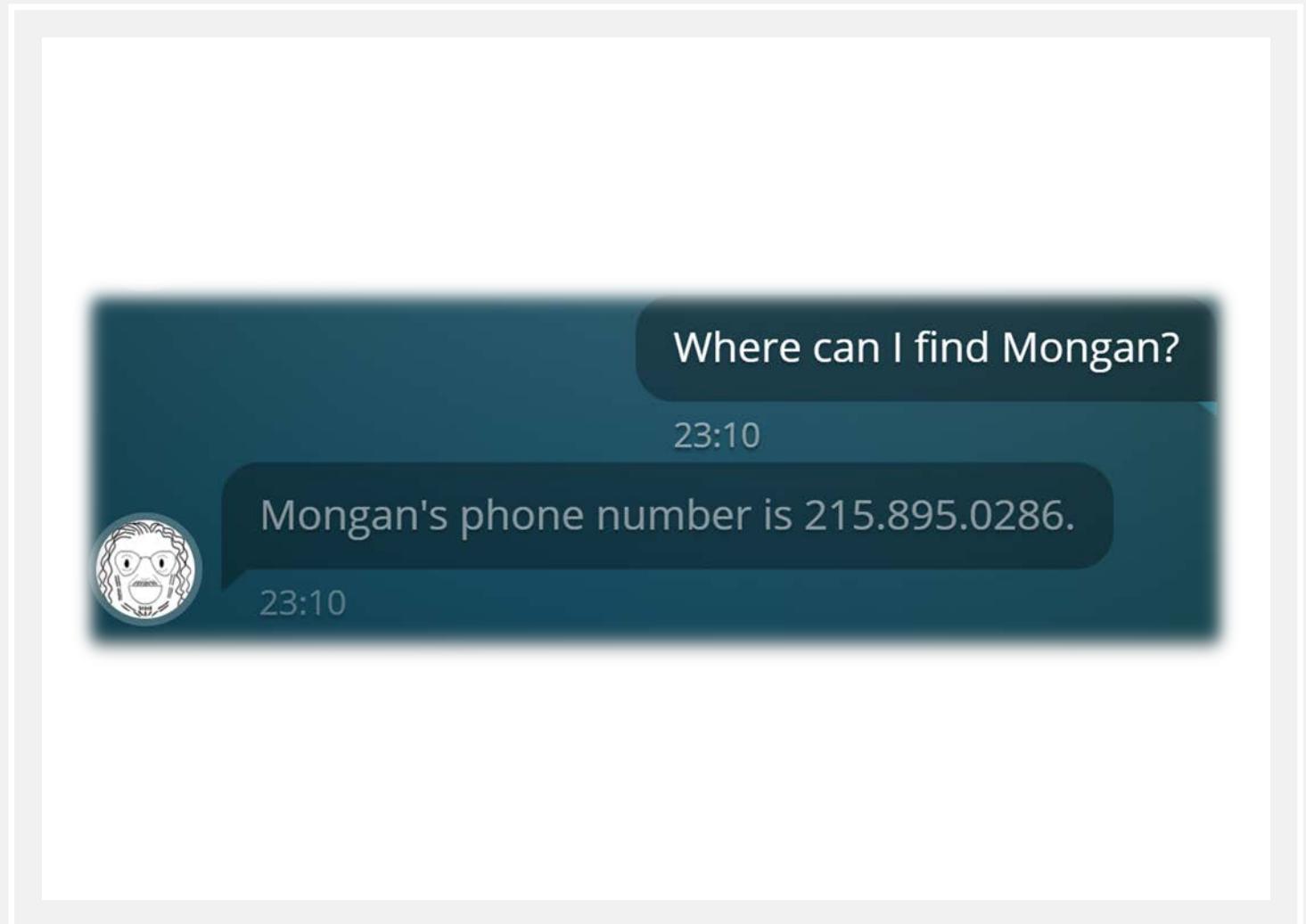
DREXEL CHATBOT

- Introduction & Features
- Overview of Algorithm
- System Architecture & Design
- Management
- Demo
- Limitations & Technical Challenges

INTRODUCTION & FEATURES

WHAT IS DREXEL CHATBOT

- Question answering system for Drexel
- RESTful API
- Demo applications (Web, Mobile and SMS)



WHY DREXEL CHATBOT

Compared to Google:

- More efficient
- More user friendly
- Gets results faster

Compared to Google Assistant/Siri:

- Optimized for Drexel
- More information about Drexel

WHY DREXEL CHATBOT

Why API

- Expandable to other applications
- Friendly to other developers
- Research project
- Focus on the process

FEATURES

- Input and output in natural language
- API
 - Input by URL parameters
 - Output formatted in JSON
 - Concurrent users
- Demo applications
 - Simple interfaces
 - Enter and send questions; get responses

SUPPORTED QUESTIONS

Faculty

- Email
- Website
- Title
- Office
- Phone number
- Picture
- Department
- Education
- Publications
- Research interests

Buildings

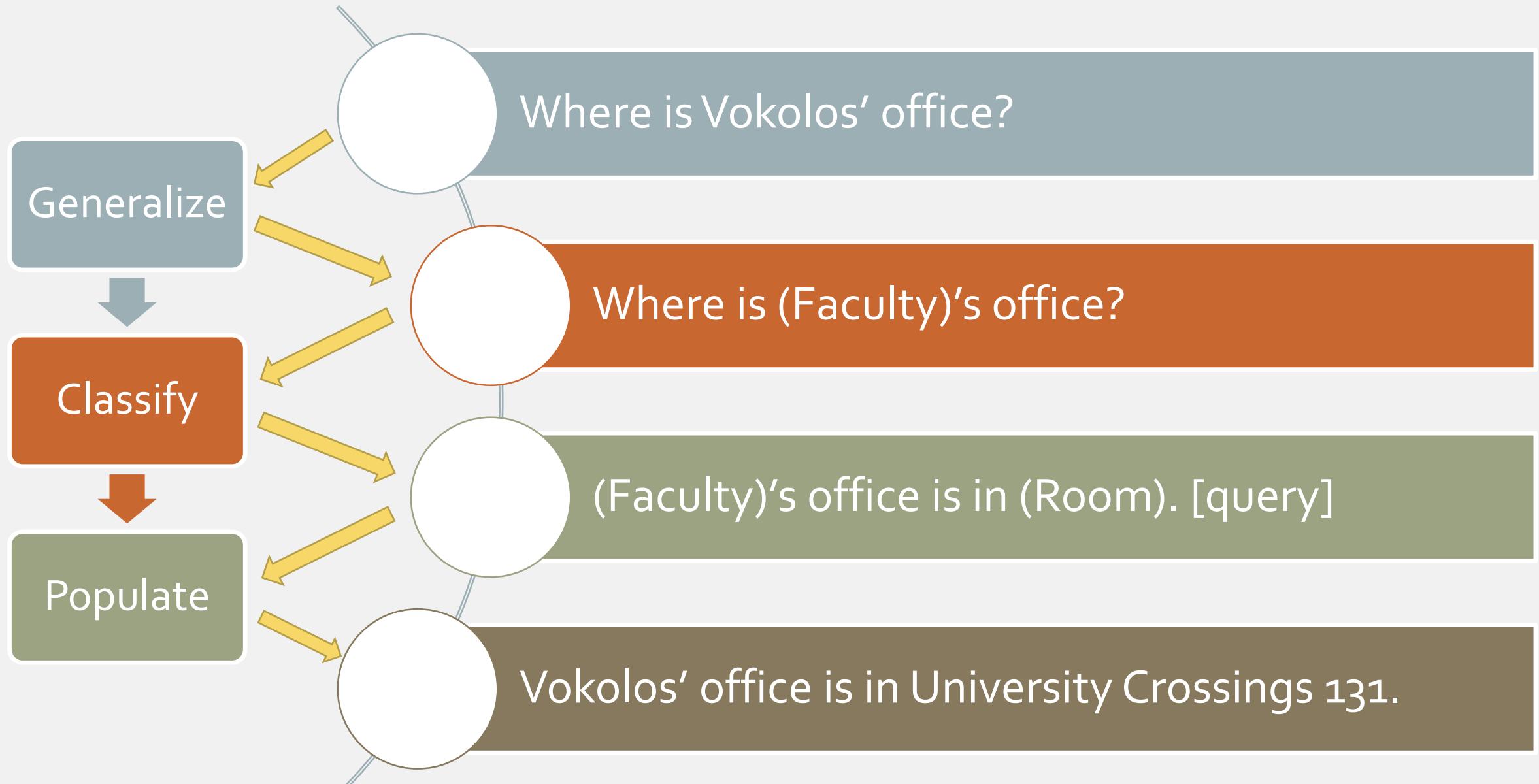
- Address
- Schedule
- Picture
- Website

OVERVIEW OF ALGORITHM

INSPIRATION

Google's "A neural conversational model" (2015)

- I am seeing an error related to VPN
 - What is the error message that you are getting when connecting to VPN using NetworkConnect ?
- Connection refused
 - May I know the version of NetworkConnect you are using?
- **Wheres Marcello Balduccini office?**
 - **Marcello Balduccini office is Rush 233C.**



Generalize

Where is Vokolos' office?

Classify

Where is (faculty)'s office?

Populate

(faculty)'s office is in (room). [query]

Vokolos' office is in University Crossings 131.

Generalize



Classify



Populate

Where is Vokolos' office?

Where is (faculty)'s office?

(faculty)'s office is in (room). [query]

Vokolos' office is in University Crossings 131.

Generalize



Classify



Populate

Where is Vokolos' office?

Where is (faculty)'s office?

(faculty)'s office is in (room). [query]

Vokolos' office is in University Crossings 131.

ALGORITHM DESIGN RATIONALE

- Less training data
- Modular
- Adaptable

SYSTEM ARCHITECTURE & DESIGN

FRONT END
APPLICATION

DREXEL CHATBOT
API SERVICE

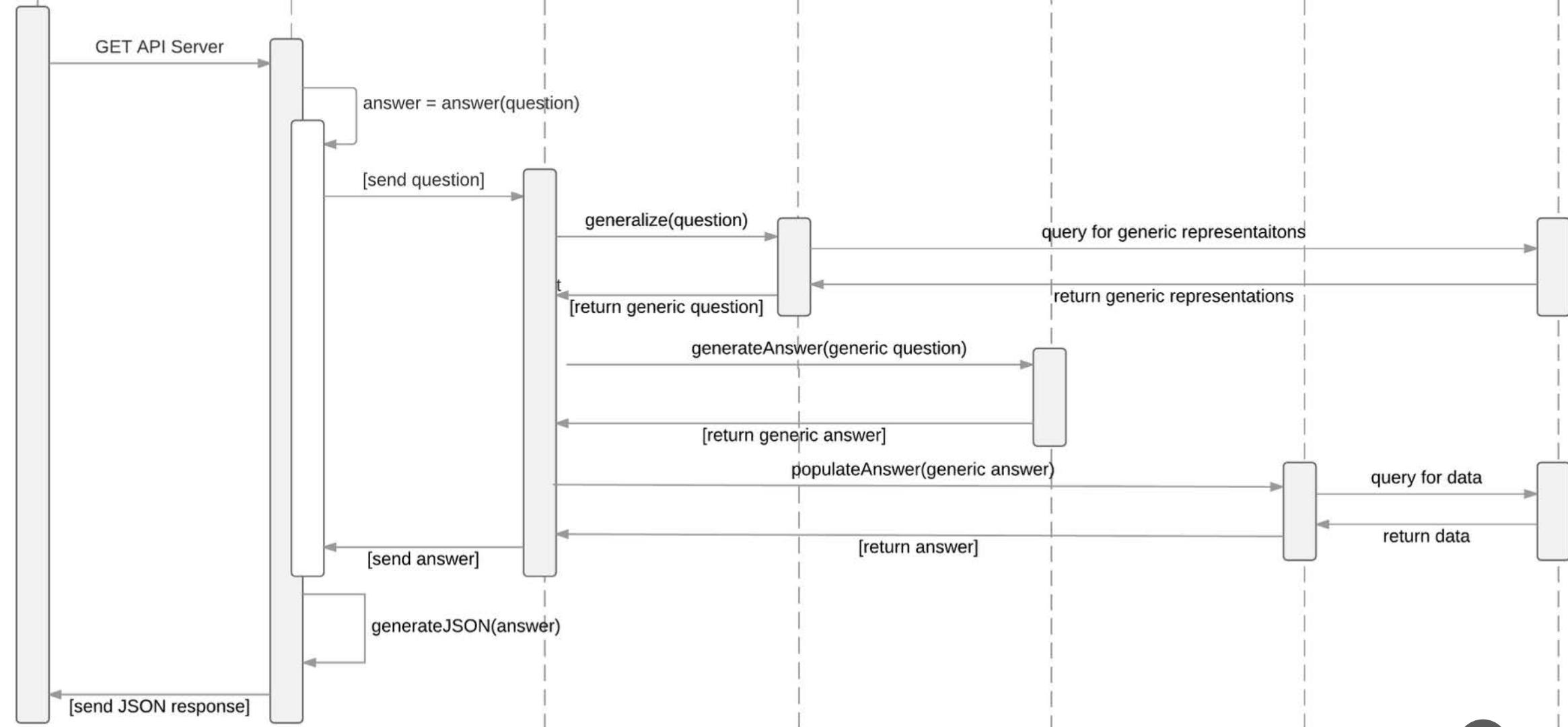
DREXEL
CHATBOT MAIN

GENERIC
QUESTION
CONSTRUCTION

GENERIC
ANSWER
CONSTRUCTION

GENERIC ANSWER
POPULATION

DATABASE



MAIN LANGUAGES

- Frontend
 - Java
 - Java Script
 - HTML & CSS
- API: Java
- Backend: Python



API



- Basic Model-View-Controller
- Calls backend on each GET request
- Each requests gets separate process

GENERIC QUESTION CONSTRUCTION

NLTK

- “Generalize” step
- Natural Language Processing
- Identify keywords
 - Part of speech tagging
 - Database lookup
- Replace keywords with generic form

GENERIC ANSWER CONSTRUCTION



- “Classify” step
- Neural network

GENERIC ANSWER CONSTRUCTION

Where is (Faculty)'s office?	1
Where is Professor (Faculty)'s office?	1
Can I get the address of (Faculty)'s office?	1
Can I get the address of Professor (Faculty)'s office?	1
What is (Faculty)'s office?	1

- 18 answer categories
- ~200 training questions
- ~7000 training question after adding noise (misspelling)

GENERIC ANSWER POPULATION

- Simplest component
- Query the database
- Replace generic form

DATABASE



- RDF Database
- No tables or keys
- Extensibility
- One-to-many

INFORMATION EXTRACTION

- Populates the database
- 20 scripts
- Creates TTL(turtle) files

<#Denise-Agosto>

```
cb:name "Denise Agosto" ;  
cb:room "Rush 214B" ;  
cb:email "dea22@drexel.edu" ;
```

ERROR HANDLER

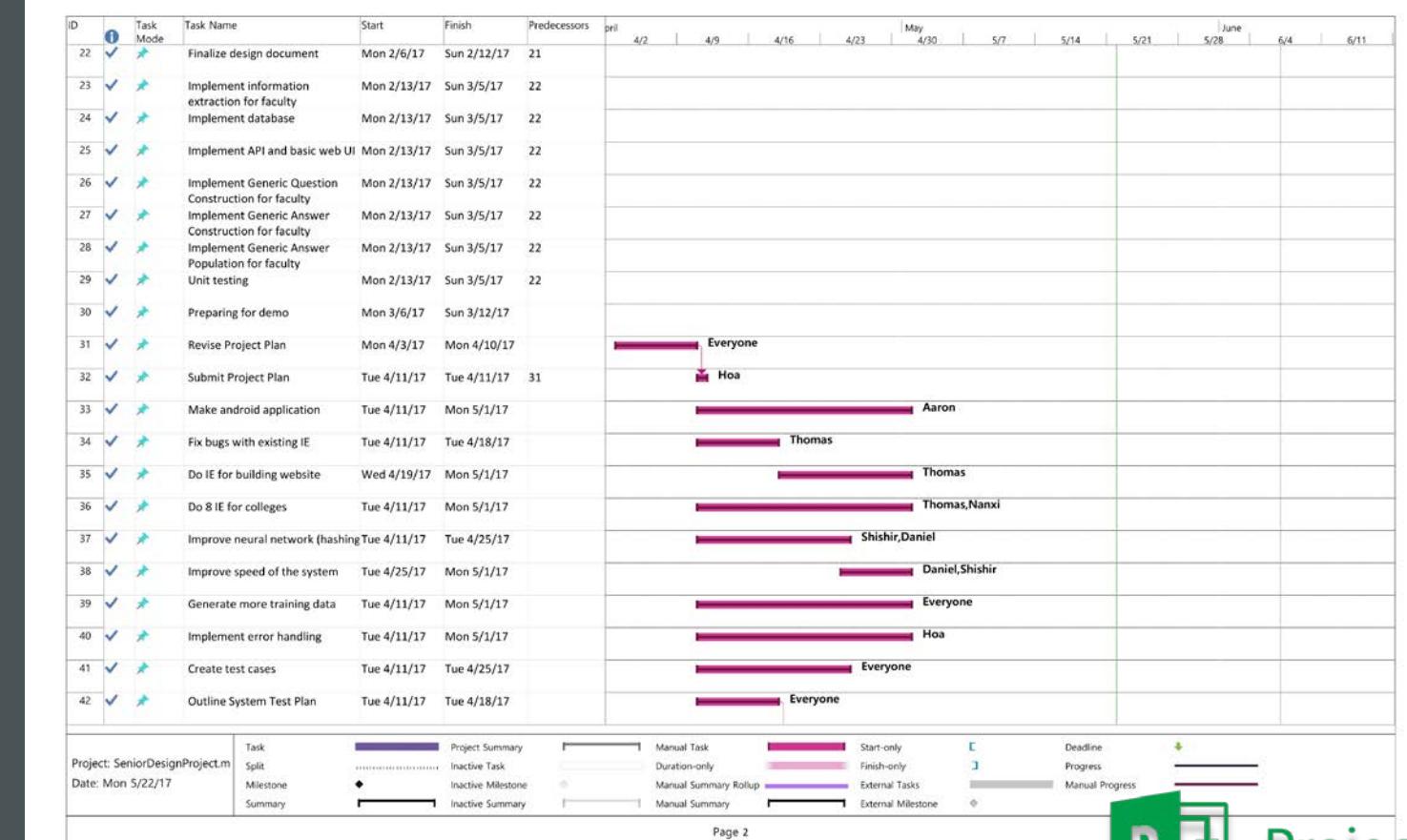
- Allows failing more gracefully
- 2 types of errors
 - “I don’t understand the question”
 - “Sorry I don’t have an answer for that. But I found this website: [http:...](http://)”

MANAGEMENT

COMMUNICATION



PLANNING



REPOSITORY & BUG TRACKING



TESTING

- Unit testing
- System testing
- User testing

DEMO

LIMITATIONS & TECHNICAL CHALLENGES

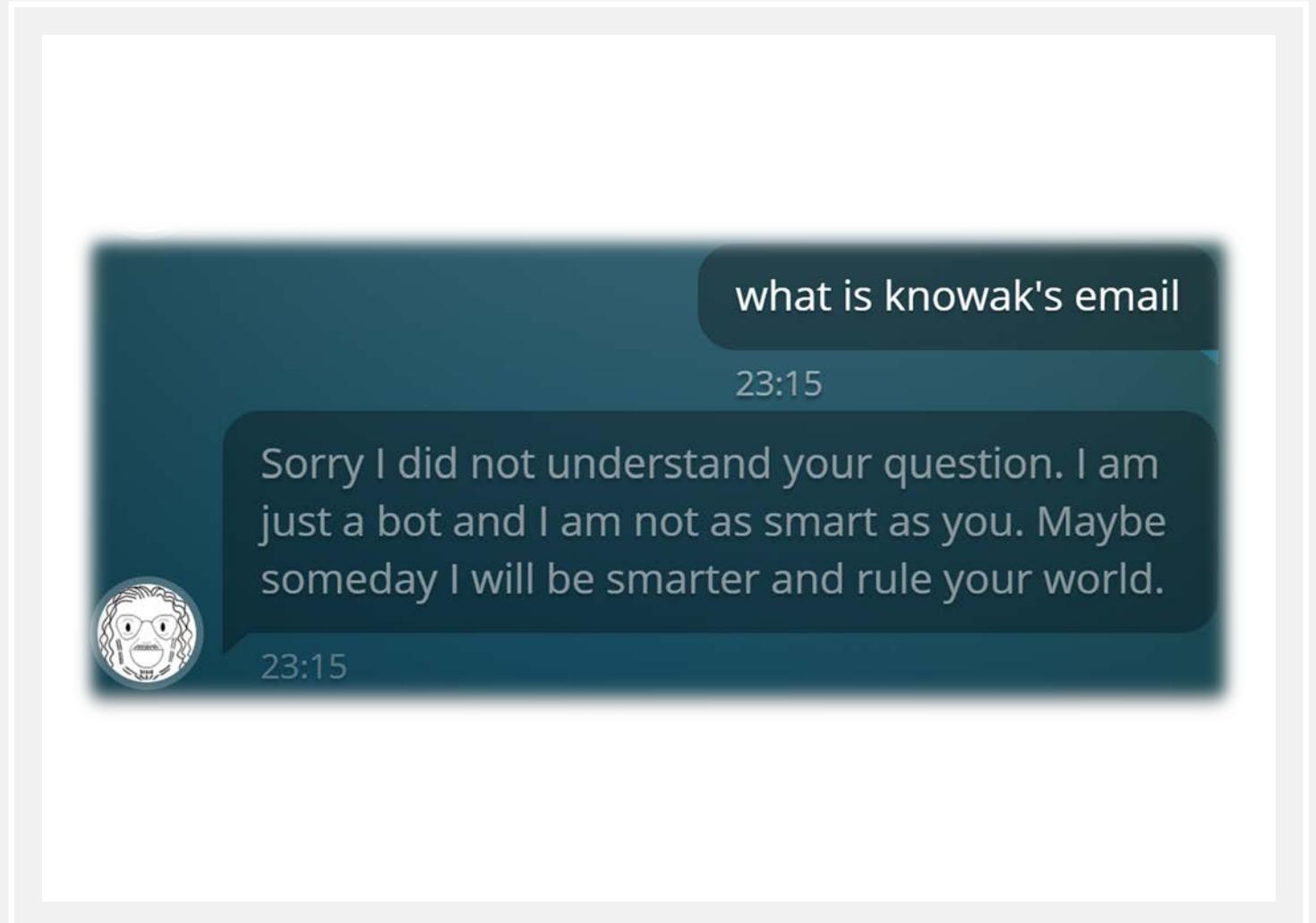
LIMITATIONS

Input question

- No compound
- Context-unaware
- Keyword-dependent

NLTK

- Misidentification



TECHNICAL CHALLENGES

- Database speed
- Neural network speed
- Lack of data

CONCLUSION

QUESTIONS

THANK YOU

Aaron, Daniel, Hoa, Nanxi,
Shishir, Tom, and Bun

