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 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
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?
- Where is Marcello Balduccini's office?
  - Marcello Balduccini's office is Rush 233C.
Where is Vokolos’ office?

Vokolos’ office is in University Crossings 131.

(Faculty)’s office is in (Room). [query]

Where is (Faculty)’s office?

Classify

Populate

Generalize
Where is Vokolos’ office?

Where is (faculty)’s office?

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

Vokolos’ office is in University Crossings 131.
Where is Vokolos’ office?

Where is (faculty)’s office?

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

Vokolos’ office is in University Crossings 131.
Vokolos’ office is in University Crossings 131.
Algorithm Design Rationale

- Less training data
- Modular
- Adaptable
SYSTEM ARCHITECTURE & DESIGN
Front End Application \rightarrow Drexel Chatbot API Service \rightarrow Drexel Chatbot Main \rightarrow Generic Question Construction \rightarrow Generic Answer Construction \rightarrow Generic Answer Population \rightarrow Database

GET API Server

answer = answer(question)

[send question]

generalize(question)

query for generic representations

return generic representations

[return generic question]

generateAnswer(generic question)

[return generic answer]

populateAnswer(generic answer)

query for data

return data

[send answer]

generateJSON(answer)

[send JSON response]

CCI Class of 2017 Senior Design Project - Drexel Chatbot
MAIN LANGUAGES

• Frontend
  • Java
  • Java Script
  • HTML & CSS
• API: Java
• Backend: Python
- Basic Model-View-Controller
- Calls backend on each GET request
- Each request gets separate process
GENERIC QUESTION CONSTRUCTION

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

NLTK
“Classify” step
Neural network
GENERIC ANSWER CONSTRUCTION

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

```xml
<#Denise-Agosto>
    cb:name "Denise Agosto" ;
    cb:room "Rush 214B" ;
    cb:email "dea22@drexel.edu" ;
</#Denise-Agosto>
```
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:...”
MANAGEMENT
<table>
<thead>
<tr>
<th>ID</th>
<th>Task Name</th>
<th>Start</th>
<th>Finish</th>
<th>Predecessors</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>Finalize design document</td>
<td>Mon 2/13/17</td>
<td>Sun 3/12/17</td>
<td>25</td>
</tr>
<tr>
<td>23</td>
<td>Implement information extraction for faculty</td>
<td>Mon 2/13/17</td>
<td>Sun 3/12/17</td>
<td>22</td>
</tr>
<tr>
<td>24</td>
<td>Implement database</td>
<td>Mon 2/13/17</td>
<td>Sun 3/12/17</td>
<td>22</td>
</tr>
<tr>
<td>25</td>
<td>Implement Art and basic webs site</td>
<td>Mon 2/13/17</td>
<td>Sun 3/12/17</td>
<td>22</td>
</tr>
<tr>
<td>26</td>
<td>Implement Generic Question Construction for Faculty</td>
<td>Mon 2/13/17</td>
<td>Sun 3/12/17</td>
<td>22</td>
</tr>
<tr>
<td>27</td>
<td>Implement Generic Answer Construction for Faculty</td>
<td>Mon 2/13/17</td>
<td>Sun 3/12/17</td>
<td>22</td>
</tr>
<tr>
<td>28</td>
<td>Implement Generic Answer Population for Faculty</td>
<td>Mon 2/13/17</td>
<td>Sun 3/12/17</td>
<td>22</td>
</tr>
<tr>
<td>29</td>
<td>Unit testing</td>
<td>Mon 2/13/17</td>
<td>Sun 3/12/17</td>
<td>22</td>
</tr>
<tr>
<td>30</td>
<td>Preparing for demo</td>
<td>Mon 3/10/17</td>
<td>Sun 3/12/17</td>
<td>22</td>
</tr>
<tr>
<td>31</td>
<td>Review Project Plan</td>
<td>Mon 4/10/17</td>
<td>Mon 4/10/17</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Submit Project Plan</td>
<td>Tue 4/11/17</td>
<td>Tue 4/11/17</td>
<td>31</td>
</tr>
<tr>
<td>33</td>
<td>Make android application</td>
<td>Tue 4/11/17</td>
<td>Mon 4/17/17</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Fix bugs with existing if</td>
<td>Tue 4/11/17</td>
<td>Tue 4/18/17</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Do IT for tracking website</td>
<td>Wed 4/12/17</td>
<td>Mon 4/17/17</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Do IT for crawlers</td>
<td>Tue 4/11/17</td>
<td>Mon 4/17/17</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Improve neural network (debugging)</td>
<td>Tue 4/11/17</td>
<td>Tue 4/15/17</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Improve speed of the system</td>
<td>Mon 4/17/17</td>
<td>Mon 4/17/17</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>Generate more training data</td>
<td>Tue 4/11/17</td>
<td>Mon 4/17/17</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>Implement error handling</td>
<td>Tue 4/11/17</td>
<td>Mon 4/17/17</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>Create test cases</td>
<td>Tue 4/11/17</td>
<td>Tue 4/12/17</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>Outline System Test Plan</td>
<td>Tue 4/11/17</td>
<td>Tue 4/13/17</td>
<td></td>
</tr>
</tbody>
</table>
TESTING

- Unit testing
- System testing
- User testing
LIMITATIONS & TECHNICAL CHALLENGES
LIMITATIONS

Input question
• No compound
• Context-unaware
• Keyword-dependent

NLTK
• Misidentification

what is knowak's email

23:15

Sorry I did not understand your question. I am just a bot and I am not as smart as you. Maybe someday I will be smarter and rule your world.
TECHNICAL CHALLENGES

• Database speed
• Neural network speed
• Lack of data
CONCLUSION
THANK YOU

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