

Building a Natural Language Interface to Database system for Medicine

Ashutosh Rai
Earlham College
801 National Rd W
Richmond, IN, 47374
arai13@earlham.edu

1. INTRODUCTION

A natural language interface to a database is a system that allows the user to access information stored in a database by typing requests expressed in some natural language. Modern Intelligent Personal Assistants(IPA) like Siri, Cortana and Google Now could be considered some popular instances of such a system. However, these are rather general purpose programs that complete tasks and answer questions related to a range of fields rather than focusing on a particular field. My senior capstone will present a natural language interface to database system with medicine as the domain of knowledge. Several aspects of the system such as the design, architecture, tools, background and future ideas will be discussed in the paper.

2. BACKGROUND

Late sixties and early seventies saw Natural Language Interface to Database systems appear on the scene for the first time. However, the systems were designed with a very specific databases in mind so the systems could not be easily adapted to a different database. Today, all the different aspects of such systems have progressed a lot. Natural Language Processing has improved a lot to produce a more realistic human-like conversation, systems are more adaptable and portable, and such systems have been integrated into other convenient areas like Facebook messenger, Google and so on. However, it is still an actively researched field with much more to be done.

Among the natural language interface to database systems present, a system that presents medicine and medical health as its knowledge domain is hard to find. The domain could contain different topics related to medical health related to diseases, drugs, treatments and so on. For my senior capstone, I aim to build such a system and will discuss the building process and other related aspects of the system in the paper. The building process will be presented from the point where I started exploring the related field of Artificial Intelligence, Intelligent Personal Assistants to be more

specific, and decided on focusing on a natural language interface to database system with the knowledge domain of medicine. I will then go on to describe the different designs I studied, the different tools I explored, researched I did and why I chose a specific set of these architectures and tools to work with while building such a system. Finally, I will also discuss about how I plan to improve this system by working on maintaining the context of the conversation for a more realistic interaction between the user and the system.

3. PRELIMINARY RESULTS

Much of the preliminary work related to my research has been related to studying different natural language interface to database systems. This would include their background, design, tools used, uses and so on. I also learned about various architectures that are used in the field of natural language processing to understand human-like interactions in terms of computational work. Currently I am looking at a keyword search based design through Artificial Intelligence Markup Language(AIML) and building and integrating a database.

4. PROPOSED SOLUTION

I see three main aspects to this project:

- Building a chatbot that handles the conversation and understands the requests/questions from the user. This would be the Natural Language Processing part of the system. If possible, I would also like to work on maintaining the context.
- Creating/compiling a database that has reliable information related to the domain of knowledge, medicine for this case.
- Integrating the database in to the chatbot and setting it up so that the information can be accessed in an organized manner.

As of now, I still am looking at different choices of collection of tools to build the natural language interface to database system and have not made my final decision. However, I am leaning on using AIML and keyword and pattern based design to process the interaction. AIML(Artificial Intelligence Markup Language) is an XML dialect that was developed by Richard Wallace to create natural language software. The chatbot will be designed to match different keywords and

patterns to understand the user's statement and respond accordingly, accessing and presenting the database whenever needed.

5. TIMELINE

The tentative timeline for my research project stands as following:

- By October 24: Think more about the design and start exploring AIML more
- By October 31: Design a basic chatbot, create a simple database and connect them. Start planning the first draft.
- By November 5: Start putting the first draft together. Continue working on the chatbot and the database, and improving the design.
- By November 16: Finish and submit the first draft. Start building the natural language interface to database system in full swing. Start working on the second draft.
- By December 12: Finish and submit the second draft. Continue working on the system.
- By December 16: Finish and submit the third draft. Continue working on the system.

6. REFERENCES

- [1] I. Androutsopoulos, G. D. Ritchie, and P. Thanisch. Natural language interfaces to databases—an introduction. *Natural language engineering*, 1(01):29–81, 1995.
- [2] A. Komninos and S. Stamou. Healthpal: an intelligent personal medical assistant for supporting the self-monitoring of healthcare in the ageing society. In *Proceedings of UbiHealth*. Citeseer, 2006.
- [3] K. Myers, P. Berry, J. Blythe, K. Conley, M. Gervasio, D. L. McGuinness, D. Morley, A. Pfeffer, M. Pollack, and M. Tambe. An intelligent personal assistant for task and time management. *AI Magazine*, 28(2):47, 2007.
- [4] D. Q. Nguyen, D. Q. Nguyen, and S. B. Pham. A vietnamese text-based conversational agent. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, pages 699–708. Springer, 2012.
- [5] W. Wobcke, A. Nguyen, V. H. Ho, and A. Krzywicki. The smart personal assistant: An overview. In *AAAI Spring Symposium: Interaction Challenges for Intelligent Assistants*, pages 135–136, 2007.