Awesomebot thesis

1. introduction

(introduces the doc in an interesting manner to the reader, what expect of this doc)

* 1. Background

(clarify the importance and the necessity of the paper, explain concepts)

* What is a chatbot
* With all data
* New technologies make it possible

A chatbot, also called conversational agent, is a program that chats with a user.

Research on person-machine interface is influenced by the competition on Turing’s test in 1950: giving the is illusion that a program thinks a meaningful dialogue.

The Turing test is a proposed artificial intelligence test based on the ability of a machine to mimic human conversation. Described by Alan Turing in 1950 in his publication Computing Machinery and Intelligence, this test consists of putting a human in a blind verbal confrontation with a computer and another human.

If the person initiating the conversations is not able to tell which of his or her interlocutors is a computer, the computer software can be considered to have passed the test.

Translated from French “Test de Turing" wikipedia’s page <https://fr.wikipedia.org/wiki/Test_de_Turing>

A chatbot is generally composed of the same components a human would use to communicate: an input (eyes and/or ears of the bot), an algorithm (brain of the bot) and an output (mouth of the bot)

For the first step, a user formulates his request in natural language by writing through a command line interface or using his voice, in the case of voice the data has to be captured via a microphone and transformed to text using speech to text recognition (not covered in this thesis). The program then uses his “brain” to create a response, this part is composed of a database of words and sentences linked with multiple algorithms, to create his memories so it can search through his memories to respond in the best way possible. After finding the best response, the bot outputs directly on the screen the response, it can also output the response through speaker using text to speech transformation (not covered in this thesis).

On the one hand with all the information available on the internet, it is relatively easy to obtain information on a subject, but on the other hand this massive and scattered availability can make it difficult to obtain quality information, because anyone can write on the internet pretending to be a professional.

It is therefore becoming necessary to create platforms that bring together information on important topics in an organized way, while promoting the quality and integrity of the information.

Thanks to recent technological advances it is now possible to create conversational agents capable of mastering a subject and conversing about it with a human.

* 1. Personal Motivation

(why I choose this subject)

* Love and work in AI
* Always wanted since I started prog to make a chatbot, like Jarvis in Marvel movies
* Passionate about human-like AIs, understanding of consciousness and so on
* TV shows like Westworld, 100% Real Humans

I personally wanted to invest myself in this project because since childhood I have always been intrigued by robots, I grew up with series and movies about artificial intelligence: Matrix, 100% Real Humans, West World, Jarvis in Marvel movies and so on.

When I started programming about 4 years ago, I wanted to understand how it was possible to create a robot to talk with. With the arrival of chatbot such as Siri on Apple, Alexa by Amazon and others, the idea seemed more and more within reach.

So, I jumped on the opportunity when I saw that a chatbot topic was available as a thesis topic for Kent's master's degree in cyber security.

* 1. Research Aims and Objectives

(what do we want to achieve) #medicine

/\*

The objective of this thesis is therefore to create an intelligent conversational agent capable of learning from a subject by giving it data, that can engage with humans and provide them with convincing, accurate and detailed help.

Although the possibility of topics can be infinite, we will direct the chatbot towards medicine. The goal is not to try to make him understand subjects that are too complex and multifaceted, but to bring together information on diseases and medications in an accessible way for everyone.

\*/

// aims != objectives

Chabots are a complex and extensive technology. When considering the use of this technology to create a learning and informative environment, the challenge of making it sound like a human with a lot of knowledge about different subjects is very complex.

To address this problem, this research aims to:

* + Create an intelligent conversational agent that can engage with humans and provide them with accurate and detailed help.
  + Consider that the agent must sound like a human to make its answers convincing and to make the user want to exchange more.
  + Evaluate the suitability and the versatility of the agent.

Specifically, the objectives are to:

* + Develop a conversational agent which implements an algorithm to learn from any given data and another one to find accurate data in its database to respond to the user.
  + Conduct an evaluation of the agent proposed using a web testing platform.
  1. Thesis Outline

(what will be discussed in the following chapters)

1. literature review

* medkiosk: an embodied conversational intelligence via deep learning
  + intro: describes an interactive kiosk, defined as a self-service computer terminal integrated with specialized hardware and software designed to display information and perform a variety range of functions to a public exhibit. This research intends to escalate the productivity and efficiency in medical institutions by offering the capabilities to provide immediate reply as well as initiating a conversation, like conversing with the experienced customer service assistant.
  + Methodology: the proposed agent is build using natural language processing (NLP), machine learning and deep learning, as those technologies can ensure smarter decision making in providing accurate, reliable, and up-to-date information.
  + Result: This research aims to design and develop a framework for revolutionizing medical kiosk which incorporates an intelligent chatbot agent to increase the productivity and efficiency in hospital by providing the capabilities of answering routine and frequently asked questions besides permitting relevant searches as well as initiating a conversation. The medical kiosk would be placed in hospital to be used in real-time environment for data collections.
  + Critic: this paper provides a clear overview of what could be a medicine oriented conversational agent, that could interact in medical centres to help users as a human could.
* health education kiosk for low-literacy patients served by community-based clinics
  + intro: describes a computer-based kiosk for the delivery health care information that targets low-literacy indigent persons.
  + Methodology: determining what types of images and messages were suited for the specific audience: Hispanics in the U.S. doing a survey at a large community event, respondents were surveyed about their use and perception of interactive communication technologies.
  + Result: working prototype in clinic. The goal for the kiosk was to engage low literacy Hispanics with technology capable of teaching health care information. Kiosk use results support the survey results, Hispanics are just as likely to use technology as Non-Hispanics.
  + Critic: does not focus on the content the kiosk / agent can teach to users but more about one population's interest in technology than another population's interest in technology.
* a low-cost community healthcare kiosk
  + intro: This paper presents a low-cost community healthcare kiosk, which is used in public areas and developed to help people on daily healthcare and long-term chronic disease monitoring.
  + Methodology: This kiosk integrates some low-cost biological devices and controls them with a built-in industrial computer. A touchscreen based Graphic User Interface and a remote service interface is offered too. It currently supports monitoring ECG, blood pressure, temperature, blood oxygen saturation and weight. The results will be automatically sent to the remote server with 128bit encrypted https protocol and without any user information except an ID number which is delivered by the trusted third-party who will support the kiosk based medical service. Users can query their records via internet and receive reminders and help in case their records should show abnormality.
  + Result: The current version of the healthcare kiosk serves as a useful prototype for low-cost community healthcare analyses. It offers the basic biological data measurements and supports extensions on both kiosk side and service side based on a client server architecture.
  + Critic: This paper completes the previous paper by adding a biological interface to measure user’s blood pressure, temperature, blood oxygen saturation and weight, as well as an ECG. But this paper does not talk about artificial intelligence features like the MedKiosk, features that would be useful to help diagnose users in real time.
* UE-based optimization for self-service community healthcare kiosk
  + Intro: Self-service community healthcare kiosks seen in previously are electronic, some small factors have an impact on the measuring result, leading to deviation or even mistakes. The paper researches on the impact that physical posture acts on the precision and deduces the functional relations between postures and deviation.
  + Methodology: This paper presents an analytically-solving method of the problem to be the design methodology, considering optimization objective and comfort.
  + Result: The conclusion is that a better model design helps to adjust the posture, raise the comfort-degree of users, and get the more precise result.
  + Critic: This paper really helps in the context of physical self-service healthcare kiosk, by optimizing the environment and user posture.
* Checkpoint based multi-version concurrency control mechanism for remote healthcare system
  + Intro: In this paper, a non-blocking concurrency control mechanism is proposed towards handling concurrency among read and write transactions in a Remote Healthcare System (RHS).
  + Methodology:
  + Result: The proposed method allow a fast and comparatively less time consuming retrieval of data when accessed from a Checkpoint, as remote healthcare system demands the storage of all older data as a history of the patient diseases, this mechanism is capable to save all the consistent version efficiently, non-blocking process synchronization implementation for contentious read/write operations is less hazardous as every concurrent update in patient data creates a new version when necessary without any back-off.
  + Critic: This paper proposes a method that allows to share resources between remote healthcare systems, for both read and write operations, this aspect of data sharing in health environment is crucial to update data between agents.
* NFC enabled intelligent hospital appointment and medication scheduling
  + Intro: Patient Appointment and medication Scheduling is necessary to manage and keep efficient tracking of day to day functionalities in health sector.
  + Methodology: The proposed method is an NFC based intelligent appointment scheduling system: the patient on tapping NFC card at the appointment kiosk is provided with timing slot to make appointment which takes into consideration time at which card was tapped.
  + Result: The proposed method allows to schedule appointment easily in healthcare environment.
  + Critic: This paper raises the idea of hospital appointment scheduling on a kiosk in healthcare environment, which is a very useful feature to add to a medical kiosk.

3 design and implementation

- explain what are design and implementation

Software design is the process of accomplishing goals using a set of primitive components and subjects to constraints. It is composed of all the activity involved in conceptualizing, framing, implementing, and modifying complex systems to follow requirements specification before programming the software itself.

Software implementation describes concretely how the program is made, the technologies used and the process of achieving the result. The creation of a working software is a combination of coding, verification, testing and debugging.

3.1 design and method

* Recall goal of the project
* First step: creation of organised database easy to use (parsing NHS, etc…)
* Second step: creation of functions that parse database documents to make precise requests
* Third step: infinite loop with CLI to create interface
* Third .1 step: sanitize user’s input
* Fourth step: the bot needs to know what subject we are talking about to select good document
* Fifth step: algorithm that evaluate most adequate sentence in document compared to user’s request
* Sixth step: create a more human (think Turing) bot by adding chatty lines, asking to repeat question differently, hello, thanks, goodbye, etc…

3.2 implementation

* Describes each previous step explaining the code

Figure 1

The Data Extraction Program is a Python3 script that creates the database used by Awesome-Bot, it is composed of a web scraper, a document parser, and a database.

The web scraper retrieves data from a website, in the case of the NHS’s website, it uses the main medicine page (or conditions) to loop through every medicine available, then it save the content with its title (e.g. Coronavirus).

The web scraper uses the beautifulsoup4 library to download https content.

The document parser parses every document saved by the web scraper; it is adapted for each desired website. Its goal is to create a clean corpus, composed of blocks of texts linked to titles.

After this process, the corpuses are saved in the database, in the desired section.

Figure 2

Because Awesome-Bot could be used for subjects others than health, it has to be specified at the start of the program which subject it should load.

This can be done using the database loader: e.g. database.load(“Medicines”) and/or database.load(“Conditions”)

Figure 3

Awesome-Bot is coded in Python3, its process flow is described as follows:

The user and Awesome-Bot interact through a Command Line Interface (CLI), when the user sends a message to Awesome-Bot, the message is sanitized (transform to lowercase letters, delete useless punctuation, delete useless whitespaces) the Subject Identifier searches if the message contains a key word of a subject it has loaded previously, if yes it uses his Search Engine to find the subject that matches the most with user’s key words, in order to send back the best response using the Sentence Evaluation module, it not it uses its Chatty Response module to keep the discussion going.

3.2

\section{Implementation}

As described in Figure 1, a program that parses NHS's website must be used to create the database of Awesome-Bot, it'll save data from every listed condition and medicine on the NHS's website.

First we get the content of those pages with the function get\\_all\\_content(string) with the first parameter being a string that specifies what to extract, 'medicines' or 'conditions'.

\begin{center}

\includegraphics[scale=1.0]{images/get\_all\_content.JPG}

\textbf{Figure 4: get\\_all\\_content(string) function call}

\end{center}

The get\\_all\\_content(string) function itself will retrieve a list of every topic with the function get\\_list(string) by specifying the subject i.e 'medicines'. Using the BeautifulSoup module in Python 3, we can retrieve every topic listed on the subject's page.

\begin{center}

\includegraphics[scale=0.8]{images/get\_list.JPG}

\textbf{Figure 5: get\\_list(string) function}

\end{center}

Then for each topic in the topic list, we can create the desired url to retrieve data. The subject is 'medicines' or 'conditions' and the topic is the medicine or condition itself, as described in figure 6.

\begin{center}

\includegraphics[scale=0.8]{images/get\_all\_content\_fct.JPG}

\textbf{Figure 6: get\\_all\\_content(string) function}

\end{center}

The program now needs to parse the page of each topic of every subject with the url.