

# A Graph Based Chatbot for Cancer Patients

Belfin R V

Department of Computer Science  
and Engineerig  
*Karunya Institute of Technology  
and Sciences*  
Coimbatore, India  
researchbelfin@gmail.com

Shobana A J

Department of Computer Science  
and Engineering  
*Karunya Institute of Technology  
and Sciences*  
Coimbatore, India

Megha Manilal

Department of Computer Science  
and Engineering  
*Karunya Institute of Technology  
and Sciences*  
Coimbatore, India

Ashly Ann Mathew

Department of Computer Science  
and Engineering  
*Karunya Institute of Technology  
and Sciences*  
Coimbatore, India

Blessy Babu

Department of Computer Science  
and Engineering  
*Karunya Institute of Technology  
and Sciences*  
Coimbatore, India

**Abstract—** World Health Organization (WHO) reports that, the second major cause of death is cancer. Life of people who have cancer is daunting. Their heart is open to all negative emotions like anger, fear, depression, guilt, helplessness, etc. People around them are also finding it difficult to cope with it. They require constant support to discuss their problems with and to provide them with factual information. This paper introduces a potential solution to provide them with what they are seeking for- a chatbot. The proposed chatbot is a cancer chatbot designed only for people dealing with cancer. People can ask about anything and everything about cancer-symptoms, treatments, survival and so on. The bot is trained by information collected from various cancer forums which have a wide range of information about cancer. Sentiment analysis is used to identify the mood of the users so that the bot can give a human-like behavior and comfort them.

**Keywords—** Chatbot, Healthcare, Natural Language Processing, Web Scraping, Neo4j

## I. INTRODUCTION

Cancer is triggered by irrepressible growth and spread of anomalous cells. Most of the people spot cancer at the very last stage. In 2018, 18 million cases of cancer are reported around the world. Out of these 8.5 million are in women and 9.6 million are in men. [1] American Institute of cancer research reports that Lung cancer and breast cancer are the two common cancers found so far. Cancer patients are entirely despaired to expect for a longer and healthier lives. According to the World Health Organization (WHO) survey, depression has a significant influence on health once it combines with any long-term disease. Depression is rapidly becoming one of the most challenging phases in the health sector. Communication helps a lot to improve one's mental health. The problem gets partially solved if a patient can open up to someone what they are going through, but sometimes none will be available at the right time. This is why cancer chatbot is introduced.

People can communicate with chatbot which uses Natural Language Processing (NLP) whenever they require. NLP is a component of artificial intelligence (AI) which makes the computers closer to a human-level understanding of language. NLP interpret the text and analyze it systematically. AI enables the bot to analyze the conversation and learn to handle the situations on its own. The vast amount of information related to cancer is collected

from the web and stored its database allows these bots to provide accurate and efficient information based on the user's requirements. From the messages of the user, chatbot collects enough keywords and starts leading the conversation by asking questions to the user and trying to shortlist a few diseases that the user may be suffering. Once it gets ample amount of data, it finds the most likely disease and can respond to their concerns with information about treatments, symptoms and so on, even by suggesting remedies.

## II. HISTORY

The evolution of chatbot began in the year 1950 from a paper titled "Computing Machinery and Intelligence" which was published by Alan Turing. In this paper, he presented the idea of the machine being intelligent and having the ability to think. This led to the development of Artificial Intelligence and the first chatbot –Eliza (1966) by Joseph Weizenbaum for the people who have psychological issues. He followed the key matching technique. It identified the keywords and the pattern which correctly matches with the set of pre-programmed conditions. After identifying the match, it generated the appropriate responses. Though it made users to feel like as if they were communication with human, it failed to pass the Turing test- test to check the intelligence of a computer.

In 1972, PARRY was developed by Kenneth Colbo to diagnosis paranoid schizophrenia. Few psychiatrists were assigned to test the system. Only 48% of psychiatrists were able to find the contrast between the real person and machine. Even A.L.I.C.E which was developed by Richard in 1995 also failed to pass the Turing test.

Chatbots gained popularity in all the domains like banking, e-commerce, healthcare, education within a few years. Apple's Siri (2011), Microsoft's Cortana (2014) and Tay(2016), Amazon's Alexa(2014), Google Allo (2016) are some of the famous chatbots. Many medical chatbots exist in the medical field, and healthcare field like Sensely is virtual nurse app, Ada, ABBI, Babylon, SafeDrugBot to provide information about the drugs used during breastfeeding. However, none of them are focusing on cancer alone.

### III. CHATBOT DESIGN

The first component in the architecture of the chatbot is the one who is interacting with the bot. Here the user is a cancer patient or anyone who needs to know about cancer.

Messages by the user get forwarded to the bot through the channels like Facebook, WhatsApp, WeChat etc. Then the message is sent to a parser, and it resolves the sentence into its components and identifies the keywords from it. When a message is passed to parser it runs its natural language engine and performs searching. It reads the entire message and identifies the intent and the entity of the message. E.g., if the user asks "What is the symptom of lung cancer?" and the bot identifies the intent as "symptoms" and the entity as "lung cancer."

Subsequently, intent and entities fetch the appropriate data from the cancer database. This database acts as the knowledge base of the bot. Responder takes the fetched data and produces human-like output. Finally, the output from the intelligent bot communicated to the user as a message. The flow diagram of the Cancer Chatbot has been visualized in the Fig.1.

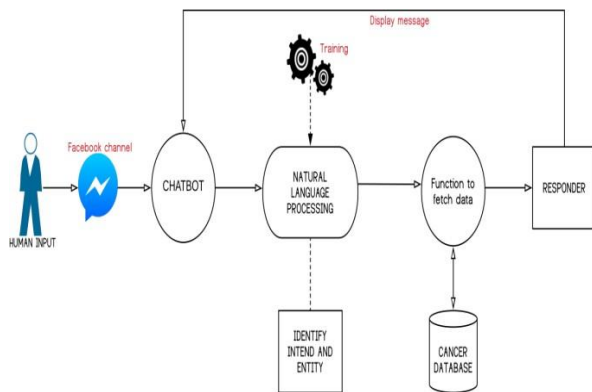


Fig. 1. Architecture Diagram of Chatbot

### IV. KNOWLEDGE BASE OF CHATBOT

#### A. Data collection (web scraping)

Internet consist of a large amount of data. The chatbot does not take data directly from the web during the conversation. It is web scrapped from cancer forums and stored in the local database with the help of python and BeautifulSoup. BeautifulSoup is a Python library for extracting data from the website. Scrapped data set is then stored in CSV (comma separated values) format. The dataset includes all the types of cancer and its related information.

#### B. Data preprocessing

Real world data is often incomplete, inconsistent or lacking in individual behavior or trends. In order to convert the raw data into an understandable format, it is pre-processed. The whole pre-processing is done using natural language tool kit (NLTK) which is available in python. The process of preprocessing is visualized in the Fig.2.

a) Tokenization: The First step is to tokenize the raw data by separating text into words, phrases, symbols, etc.

For example,

There are many symptoms for breast cancer in women. [2]

‘There’, ‘are’, ‘many’, ‘symptoms’, ‘for’, ‘breast’, ‘cancer’, ‘in’, ‘women’, ‘.’

b) Converting to lowercase: Since the program is case sensitive, all the generated token is converted to similar case. Otherwise, it will recognize the same words with different cases as different words.

c) Punctuation Removal: The punctuation is removed from the sentence as it does not make any sense while developing the model.

d) Stop Words Removal: Stop words are the unwanted words which could not contribute deeper meaning of the text. Stop word removal could save memory in the database and the processing time. This process is accomplished by importing corpus.

From the above example, all the stop words are removed.

‘many’, ‘symptoms’, ‘breast’, ‘cancer’, ‘women’

e) Stemming: Stemming helps to bring down a word to its word stem or Root word that affixes to suffixes. The Root of the word is known as a lemma. The sample stemming process is shown in the Table 1.

TABLE I. STEMMING WORDS

Form	Suffix	Stem
Symptoms	-s	Symptom
Itching	-ng	Itchy
Nipples	-s	Nipple

f) Lemmatization: Lemmatization is a method of combining the varied forms of a word. This process identifies the root form of a word. The sample lemmatization process is shown in the Table 2.

TABLE II. LEMMATIZING WORDS

Form	Morphological information	Lemma
Puckering	Gerund of verb pucker	Pucker
Leaking	Gerund of the verb leak	Leak

#### C. Modelling

This data is then converted to a graph model using a graph database- Neo4j. This stores data as edges and nodes. This process helps to efficiently process highly connected data and identify the relationship between different data easily.

The Fig.3 describes a simple graph model of the symptoms of a few cancers. The red nodes denote the name of the cancers, and blue node denotes their corresponding symptoms.

#### D. Detecting Cancer

The engine will shortlist cancer by considering the data provided by the user. Then it suggests remedies and related information to the user.

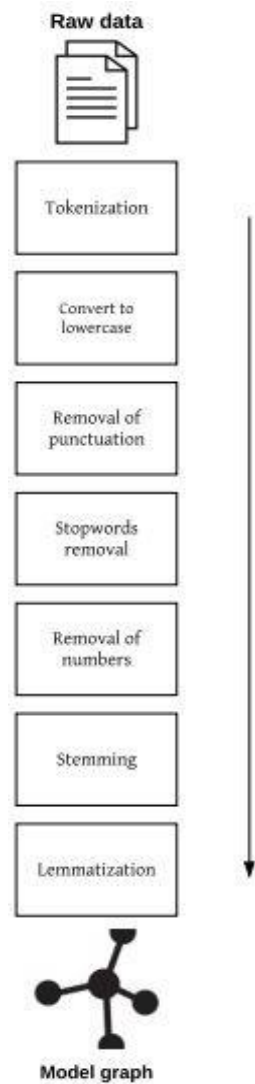


Fig. 2. Data Pre-processing steps

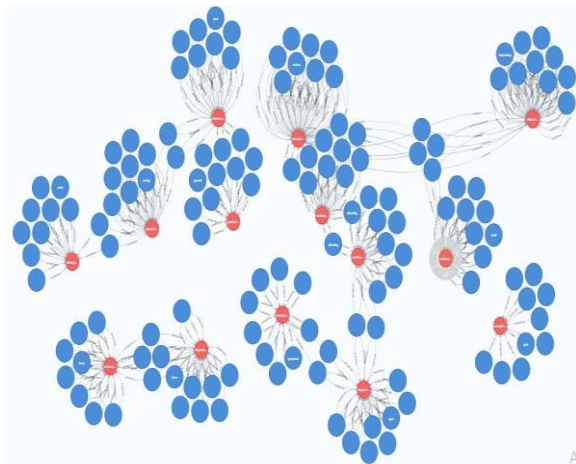


Fig. 3. Graph Data Model

### E. Detecting Cancer

The engine will shortlist cancer by considering the data provided by the user. Then it suggests remedies and related information to the user.

If the user asks about symptoms of any particular cancer, the Cancer bot lists all its symptoms and explain to them if

the user needs to know. The same can be done with other details like stages of cancer and treatments.

## V. RESULT AND DISCUSSION

Cancer patients sometimes hesitate to share their problems with others. Cancer bot will let them share their problems regarding cancer and their helpless situation. The chatbot is increasingly being used to handle tasks, free humans and make them focus on their health. Chatbot lets the users interact without any barriers; it impersonates the human conversation either through text or audio. The constant communication with the bot can bring a positive attitude to the patient. It can learn from previous conversations and get some valuable experience, and become more active as time passes by. A User can chat with the bot like how he can chat with a human. In order to build an intelligent chatbot, there are a lot of program inputs out there that are accessible in the open and make.

When developing a chatbot, the developer must know about the exact purpose and requirements to accomplish them. There are many new concepts for a developer to learn, such as intents and entities. However, there are plentiful resources, such as video tutorials that can be utilized to build a chatbot. When developing a chatbot, the development and testing phases are much more intertwined. The developer needs to continually update the knowledge base of the bot and update them according to the current trend.

A Chatbot named 'C Bot' is created and trained it by feeding a few questions and their corresponding answers from the cancer forum. It responds according to the trained data. The communication between the bot and the human is given in the Fig.4 and Fig.5.

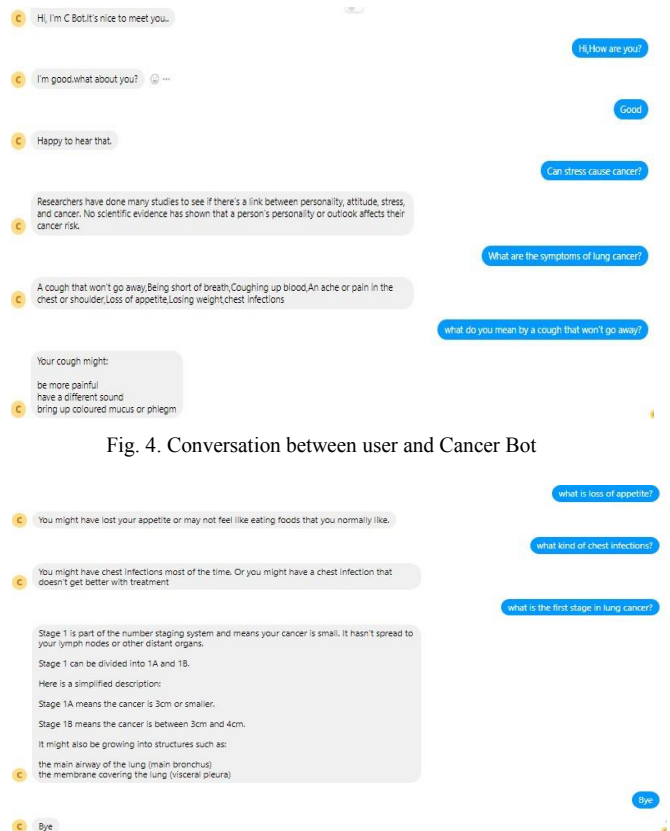


Fig. 4. Conversation between user and Cancer Bot

Fig. 5. Conversation between user and Cancer Bot

## VI. CONCLUSION

This paper explains the healthcare chatbots and how they get developed. It determined that the modern chatbots perform at a very high standard to provide a reliable response to users compared to the traditional chatbots. The cancer chatbot would have a more significant influence on the life of cancer patients and would help them to clarify their doubt regarding cancer. Unlike existing chatbots which focused on various domains of healthcare apart from cancer, this chatbot focused only on cancer. Through an interactive session with this cancer bot, it is possible for the patients to receive relevant information regarding cancer. This is the best solution for people who are busy with their job schedules. They do not need to wait in the queue for hours to get an appointment with a doctor every time instead they can chat with the bot. Also, it would be a great relief for those who want to discuss their cancer-related issues. Our chatbot contains a lot of cancer-related data collected from various cancer forums which would make our chatbot more dynamic and specific.

## VII. FUTURE WORK

Cancer bots will explore a new flexible way for the people dealing with cancer. In our bot, we implemented only limited features like it supports only text-based chatting. Several improvements can be made by adding new features. The following are some of the features that can be added to improve the chatbot functionalities:

- Qualities of the extracted data can be improved by expanding the database by adding more data.
- Images can also be stored in the database that the bot can send to the user as a reference.
- A bot could be trained to recognize the images sent by the user.
- Additional languages can be trained to the bot.
- chat templates can be generalized by clustering related topics and grouping related replies and improve cohesiveness among the consecutive chat by understanding the styles of replies and improve the understanding of the bot.
- Voice recognition features
- Cancer patients have many doubts in a single query. This changes the formulation of the intent classification into a multi-label classification. Different ways could be included to solve these issues.
- The user can send feedback about the bot services so that its services can be improved according to the user's needs.
- Sentiment analysis can be done to know the sentiment of the user.
- The Bot can also suggest specialized hospitals for cancer and cancer-related sites from the web directly.
- Bot can be trained to access data directly from the web.

## REFERENCES

- [1] <https://www.wcrf.org/dietandcancer/cancer-trends/worldwide-cancer-data>
- [2] <https://www.cancerresearchuk.org/about-cancer/breast-cancer/symptoms>
- [3] (Parker et al., 2001)Parker, P. A., Baile, W. F., De Moor, C., Lenzi, R., Kudelka, A. P., & Cohen, L. (2001). Breaking bad news about cancer: Patients' preferences for communication. *Journal of Clinical Oncology*, 19(7), 2049–2056. <https://doi.org/10.1200/JCO.2001.19.7.2049>
- [4] (Rarhi, Bhattacharya, Mishra, & Mandal, 2017)Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., Neill, S. O., ... McTear, M. (2018). Best Practices for Designing Chatbots in Mental Healthcare – A Case Study on iHelp. *Proceedings of British HCI 2018*, 1–5.
- [5] Rarhi, K., Bhattacharya, A., Mishra, A., & Mandal, K. (2017). Automated Medical Chatbot. *SSRN Electronic Journal*, 0–2. <https://doi.org/10.2139/ssrn.3090881>
- [6] (Cameron et al., 2018)Cameron, G., Cameron, D., Megaw, G., Bond, R., Mulvenna, M., Neill, S. O., ... McTear, M. (2018). Best Practices for Designing Chatbots in Mental Healthcare – A Case Study on iHelp. *Proceedings of British HCI 2018*, 1–5.
- [7] (Loder & Nicholas, 2018)Loder, J., & Nicholas, L. (2018). Confronting Dr. Robot Creating a
- [8] people- powered future for AI in health, (May), 37. Retrieved from [http://www.nesta.org.uk/sites/default/files/confronting\\_dr\\_robot.pdf](http://www.nesta.org.uk/sites/default/files/confronting_dr_robot.pdf)
- [9] (Ashiwal, Tripathi, & Miri, 2016)Ashiwal, P., Tripathi, P., & Miri, R. (2016). Web Information Retrieval Using Python and BeautifulSoup, 4(Vi), 335–339.
- [10] (Uma K, 2017)Uma K, M. H. (2017). Data Collection Methods and Data Pre- processing Techniques for Healthcare Data Using Data Mining. *International Journal of Scientific & Engineering Research*, 8(6), 1131–1136. Retrieved from <http://www.ijser.org>
- [11] (S & R, 2016)S, V., & R, J. (2016). Text Mining: open Source Tokenization Tools – An Analysis. *Advanced Computational Intelligence: An International Journal (ACII)*, 3(1), 37–47. <https://doi.org/10.5121/acii.2016.3104>
- [12] (Dootio & Wagan, 2018)Dootio, M. A., & Wagan, A. I. (2018). Automatic stemming and lemmatization process for sindhi text, (December 2017). (Strong et al., 2008)Strong, V., Waters, R., Hibberd, C., Murray, G., Wall, L., Walker, J., ... Sharpe, M. (2008). Management of depression for people with cancer (SMaRT oncology 1): a randomised trial. *The Lancet*, 372(9632), 40–48. [https://doi.org/10.1016/S0140-6736\(08\)60991-5](https://doi.org/10.1016/S0140-6736(08)60991-5)
- [13] (Strong et al., 2008)Strong, V., Waters, R., Hibberd, C., Murray, G., Wall, L., Walker, J., ... Sharpe, M. (2008). Management of depression for people with cancer (SMaRT oncology 1): a randomised trial. *The Lancet*, 372(9632), 40–48. [https://doi.org/10.1016/S0140-6736\(08\)60991-5](https://doi.org/10.1016/S0140-6736(08)60991-5)
- [14] (Kiran Chintalapudi, Bharadwaj, Kolagatla, & Nikhil Sampara Harsha, 2018)Kiran Chintalapudi, S., Bharadwaj, A., Kolagatla, S., & Nikhil Sampara Harsha, R. (2018). Building an artificially intelligent, interactive and online chatbot. *International Journal of Advance Research Ideas and Innovations in Technology*, 4(2), 1202–1207. Retrieved from <https://www.ijariit.com/manuscripts/v4i2/V4I2-1567.pdf>
- [15] (Huang, Zhou, & Yang, 2007)Huang, J., Zhou, M., & Yang, D. (2007). Extracting chatbot knowledge from online discussion forums. *IJCAI International Joint Conference on Artificial Intelligence*, 423–428. <https://doi.org/10.1016/j.tins.2003.11.005>
- [16] (Shawar & Atwell, 2003)Shawar, B., & Atwell, E. (2003). Using dialogue corpora to train a chatbot. *Proceedings of the Corpus Linguistics 2003*, (January), 681–690. <https://doi.org/10.13140/2.1.1455.7122>
- [17] (Hill, Randolph Ford, & Farreras, 2015)Hill, J., Randolph Ford, W., & Farreras, I. G. (2015). Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Computers in Human Behavior*, 49, 245–250. <https://doi.org/10.1016/j.chb.2015.02.026>
- [18] (Am, 2013)Am, S. (2013). Sitara™ AM335x ARM® Cortex™-A8 Microprocessors (MPUs). *October, 12*(April), 2493–2537. <https://doi.org/10.1.1.231.4614>

- [19] (Issn, 2015)Issn, U. (2015). Literature review about Neo4j graph database as a feasible alternative for replacing RDBMS, *18*(2), 135–139.
- [20] (P.F, K.R, & Varghese, 2016)P.F, M. F., K.R, R., & Varghese, S. M. (2016). Outcome Analysis Using Neo4j Graph Database. *International Journal on Cybernetics & Informatics*, 5(2), 229–236. <https://doi.org/10.5121/ijci.2016.5225>