# C:\Users\esu\Desktop\jimma.jpg

**Sub thematic Title 3:** **Medical Information Assistant using Chatbot System for Afaan Oromoo Language**

**By**

**Gemechu Brihanu Adeba (PI)**

**Tafary Kababa Tolassa (Co-PI)**

**This Proposal is Submitted to Research and Publication Director of Jimma Institute of Technology, Jimma University**

Jimma

November, 2019

# **Summary of the Project**

The main aim of this research project is to develop a Medical information assistant for Afaan Oromoo users using a chatbot system. Chatbot system now a day is used for different purposes. For example, they are used for flight booking, language learning, online shopping, Computer system diagnosis and soon. The chatbot system gives users a natural way of conversation when they use it. They can use Natural language question like “What is the best pesticide?”. Then The system understands the question of the user and answers it from its knowledge bases. The conversation will go on until user ends the conversation.

Chatbot system can be rule-based or corpus-based system. The corpus-based approach of chatbot system usually need large data set of user’s real conversation possible conversation. Since we don’t have large corpus for Afaan Oromoo, in our case, we are going to use a rule-based approach, in which user’s question answered using predefined language rules. So, for answer generation, we are going to collect data and develop rules for question and answer patterns of Health-related Information in Afaan Oromoo.

Contents

[1](#_Toc24715249)

[**Summary of the Project** 2](#_Toc24715250)

[**1** **Introduction** 4](#_Toc24715251)

[Background 4](#_Toc24715252)

[Chatbot Components 5](#_Toc24715253)

[**2** **Statement of the Problem** 6](#_Toc24715254)

[**3** **Objectives of the Project** 6](#_Toc24715255)

[3.1 General Objectives of the Project 6](#_Toc24715256)

[3.2 Specific Objectives of the project 6](#_Toc24715257)

[**4** **Methodology** 7](#_Toc24715258)

[4.1 Data Collection Methods 8](#_Toc24715259)

[4.2 Rule development 8](#_Toc24715260)

[4.3 Performance Evaluation Techniques 8](#_Toc24715261)

[4.4 Tool Development 9](#_Toc24715262)

[**5** **Expected outcome** 9](#_Toc24715263)

[**6** **Significance of this Project** 10](#_Toc24715264)

[**7** **Budget** 11](#_Toc24715265)

[**8** **References** 12](#_Toc24715266)

# **Introduction**

## Background

Digital Information Assistant also known as a Chatbots are systems that can carry on extended conversations with the goal of mimicking the unstructured conversational or ‘chats’ characteristic of human-human interaction [1]. When it is First started people use Chatbot system for Entertainment purposes. For example, Microsoft’s XiaoIce system [2], which chats with people on text messaging platforms, is just for entertaining purposes for lonely people. Yet starting from the very first system, ELIZA [3], chatbots have also been used for practical purposes, such as testing theories of psychological counseling. Now a days it is widely used in businesses like flight booking, language learning, Information Retrieving (IR), and mostly in assisting customer in online business [4]. As Information retrieval tool Chatbot system gives user flexibility of asking a question in a natural language. Unlike the traditional IR system, the response from such system is not the document which contains the information rather it is the Information it self. For example, if user ask the system by saying “What is the best coffee in Ethiopia?” the system will answer by displaying answer text like “The best coffee in Ethiopia is Jimma Coffee.”. Conversation can goon if the user also asks for the price of Jimma Coffee. This is what differs it from simple question answering.

When we come to health care, chatbots or health bots are intended to provide personalized health assistance and therapy information to patients, provide relevant products and services to patients, as well as suggest diagnoses and recommend treatments based on patient symptoms. Chatbots in health care may have the potential to provide patients with access to immediate medical information, recommend diagnoses at the first sign of illness, or connect patients with suitable health care centers in their vicinity [9,10]. Theoretically, in some instances, chatbots may be better suited to help patient needs than a human physician because they have no biological gender, age, or race and elicit no bias toward patient demographics. Chatbots do not get tired, fatigued, or sick, and they do not need to sleep; they are cost-effective to operate and can run 24 hours a day, which is especially useful for patients who may have medical concerns outside of their doctor’s operating hours.

## Chatbot Components

Even if the detail implementation or algorithm used in Chatbot system may differ from system to system, most of the chatbot system have the following model [5].

1. **Question Analysis: -** This stage is primarily focused on extracting meaning from user question in natural language or determining the function of the text/sentence in user’s question (e.g. is this a question, suggestion, offer, or command). To extract meaning from user’s question text, the system will convert unstructured text written into a chatbot to structured grammatical data objects, which will be further processed by the system.
2. **Response Generation: -** Response generation is arguably the most central component of the chatbot architecture. As input, the Response Generator (RG) receives a structured representation of the user question. In most architecture of chatbot system, the response selector has access to three key components it will use to make its decision about what to respond to user:
   1. a knowledge database (Rules) / data corpus, which will differ in content based on implementation.
   2. a dialogue history corpus, which will only exist in more complex models, and
   3. an external data source, which provides the bot with intelligence (e.g. a dog is an animal).
3. **Knowledge Base Creation**

In order for chatbot system to work as an intelligent information assistant there should be a knowledge base for it. They way data stored is depend on the type of method used for response generation. For example, if it is rule-based system, there will be collection of rules for the bot system to refer when it answers user question. If it is corpus based there will be data set from where the answer is generated either using machine learning technique or using Infor0mation Retrieval model depending on the type of the bot.

1. **Dialogue manager**

The unique feature of dialogue system or Chatbot is that they respond to user in very human like conversation. For example, if the user asks for “What is the date of today?” the don’t just respond by saying “Nov, 10, 2019”. Instead the answer it like human by saying “the date of today is Nov, 10, 2019”. This kind of feature is handled by dialogue manager of the chatbots. Chatbot system uses language tricks to generate human like responses.

# **Statement of the Problem**

Digital Information Assistant systems are become very common these days. Peoples use computer application to buy and sell goods online without human intervention. Also, people can get different information on how to use different tools from this digital assistant system. Unfortunately for under resourced language like Afaan Oromoo it is difficult to get such services. Afaan Oromo is among the most widely spoken and used Afro-Asiatic languages [6]. In Ethiopia, it is an official and mass communication language of Oromia regional state, which is the largest region in Ethiopia. Besides being an official working language of Oromia regional state, Afaan Oromo is the instructional medium for primary and secondary schools throughout the region. It is used by Oromo people, who are the largest ethnic group in Ethiopia, which amounts to 34.4% of the total population [7]. As a result, there are large portion of our population who want to get information in this language. Most of the population who speak this language lives in ruler area, where it is difficult to get medical consultation or health guide. So, in this work we are going to develop the Medical Information Assistant in local languages, specifically in Afaan Oromoo Language.

# **Objectives of the Project**

The following are general and specific objectives of the project;

## General Objectives of the Project

The main objective of this project is to develop Intelligent Medical Information Assistant for Afaan Oromo using chatbot system.

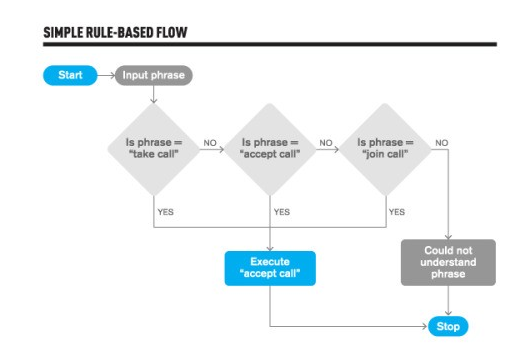
## Specific Objectives of the project

In order to achieve the general objective of the project, the following specific objectives are identified.

* Literature search and review of existing technologies and techniques pertaining to the development of Chatbot System
* Design question and answer pattern rule for Afaan Oromoo Text
* Apply natural language processing (like stemming) to extract meaning (key word extraction) of user question.
* Design Pattern matching algorithm to generate response for user question.
* Design architecture for implementation of Medical Information Assistant for Afaan Oromoo.
* Evaluate the system and release for public use.
* Publish the experimentation process for others to use

# **Methodology**

Generally, Chatbot system can use two approaches [1]; the first one is rule-based system and the second one is corpus based. Corpus based approach needs a large amount of conversational data so that it can retrieve answers from it or trained on it to generate appropriate answer for user’s question. Afaan Oromoo is under resourced language, that means we don’t have rich data set in this language. Due to this we are going to follow rule-based system. Rule based scheme looks like the following during the conversation. User’s question is going to be matched with the rule and if it succeeds the answer will be generated if not the system may prompt that it did not understand the question. The main challenge with this approach is that it needs thousands of rules to create real conversational agents.



*Fig 1: Simple rule-based conversational system.*

## Data Collection Methods

In order to create rule-based dialogue agent we have to gather documents in the domain of our interest, that is in Medical sector. So, document related to health will be collected from different areas like from Oromia Health Bureau, and Oromia Broadcasting Network Health related archive.

## Rule development

After Document is generated, we are going to develop question-answer pattern rules. Rules are collection of question pattern and response patterns in a natural language. So, the rules for Afaan Oromoo question and Response pattern will be developed manually. Many dialogue systems are rule-based, and one of the most popular mechanisms of representing rules is AIML (Artificial Intelligence Markup Language). In our work we are going to store those rules in AIML format.

## Performance Evaluation Techniques

Different evaluation criteria can be applied, depending on the goals a given chatbot has been created to fulfil or tasks to perform. In other words, the evaluation criteria depend on the metrics applied at conceptual, operational, and qualitative levels. As for quality, here cohesion, cooperation, likeability, engagement, trust, reduction of frustration or ability to comment and provide feedback play a role. Cognitive linguistic quality criteria include, broadly speaking, conversational flow, understanding, and accuracy [8].

## Tool Development

The following tool will be used for completion of our project.

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Material** | **Description** | **Remark** |
|  | Laptop computer | More convenient to use than desktop because of mobility and enables efficient use of time | Already available |
|  | Java SDK 8 | The Chatbot system will be developed using Java programming language | Freely available online |
|  | Java Runtime Environment (JRE) | The final Chatbot system requires JRE to run | Freely available online |
|  | AIML | To store rules as knowledge base of the bot system | Freely available online |
|  | Word Processing software | Used for report writing | Already available |

# **Expected outcome**

When the project is completed, the final outcomes will be the Medical Information Assistant Bot as a free tool, that can be used as conversational agent, which answers health related question in Afaan Oromo text and a published article that describes knowledge acquired from the research project.

# **Significance of this Project**

Developing automatic conversational agent or chatbot involves in creating rules for question and answers patterns in Afaan Oromoo language. These developed rules can be used by different researchers to further enrich this language with digital tools. On the other hand, the knowledge and experience acquired during development of this system will be published in reputable journal. The published article can help other researchers on local languages to have clear understanding of what is required when working on chatbot system and share from our experience.

The main beneficiary of our system will be our people who lives in the area, where it is difficult to get health information. This system will provide information they want at any time by just in their natural language. So, this will increase their productivity and health, hence it will have positive impact on our nation’s economy.

# **Budget**

In order to feasible the work of this project economically, the cost breakdown of this project will be designed as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Budget Breakdown** | | | | | |
| **No** | **Duties** | **#of people** | **Wage/day** | **#days** | **Total Cost in Birr** |
| 1 | Document Collection | 5 | 206 | 16 | 16480 |
| 2 | Data Cleaning | 2 | 206 | 29 | 11948 |
| 3 | Preparation of question and answer pattern rules | 7 | 206 | 15 | 21630 |
| Sub Total | | | | | 50058 |
| Contingencies 10% | | | | | 5005.8 |
| Grand Total | | | | | **55063.80** |

Total budget required will be **fifty-five thousand sixty-three birr and eighty cents** (**55,063.80 ETB)**.

# **References**

1. Daniel Djurafsky, James H. Martin (2006). Speech and Language Processing, Prentice-Hall, Inc
2. Microsoft (2014). http://www.msxiaoice.com..
3. Weizenbaum, J. (1966). ELIZA – A computer program for the study of natural language communication between man and machine. Communications of the ACM, 9(1), 36–45
4. Shawar, Bayan Abu. (2007) “Chatbots: Are They Really Useful.” Ldv Forum.
5. Jack Cahn (2017) “CHATBOT: Architecture, Design, & Development”, Senior Thesis (EAS499) University of Pennsylvania
6. Anteneh, Getachew and Ado, Derib. (2008). Language Policy in Ethiopia: History and Current Trends. Ethiopia Journal of Education and Science, 2(1), 37-62.
7. Alemayehu, W. (2005). Rule Based Syntactic Disambiguation Parser For Amharic Sentence. Master’s thesis, Addis Ababa University. Master’s thesis, Addis Ababa University, Addis Ababa, Ethiopia.
8. Hiroshi Yamaguchi, Maxim Mozgovoy, Anna Danielewicz-Betz (2018), “A Chatbot Based on AIML Rules Extracted from Twitter Dialogues”, Communication Papers of the Federated Conference on Computer Science and Information Systems pp. 37–42, ISSN 2300-5963 ACSIS, Vol. 17
9. Amato F, Marrone S, Moscato V, Piantadosi G, Picariello A, (2017) Sansone C. “Chatbots Meet eHealth: Automatizing Healthcare” , CEUR Workshop Proceedings..
10. Morris RR, Kouddous K, Kshirsagar R, Schueller SM. Towards an artificially empathic conversational agent for mental health applications: system design and user perceptions. J Med Internet Res. 2018 Jun 26;20(6)