NGHIÊN CỨU TRỢ LÝ ẢO ỨNG DỤNG TRÍ TUỆ NHÂN TẠO

RESEARCH ASSISTANCE OF ARTIFICIAL INTELLIGENCE APPLICATION

Nguyễn Tiến Dũng, Phạm Trung Thiên*, Lê Ngọc Dũng, Đỗ Văn Tỉnh, Vũ Xuân Tú, Nguyễn Văn Hiệp, Nguyễn Ngọc Thể

Khoa Cơ khi, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp

Đến Tòa soạn ngày 12/02/2022, chấp nhận đăng ngày 11/05/2022

Tóm tắt: Thế giới bùng nổ cuộc Cách mạng công nghiệp 4.0 với các ứng dụng robot và trí tuệ nhân tạo đi sâu vào cuộc sống cũng như sinh hoạt hàng ngày. Các robot cũng như thiết bị tự động đang trở nên thông minh hơn theo cách của chúng để tương tác với cả con người cũng như giao tiếp giữa các thiết bị với nhau. Lĩnh vực xử lý ngôn ngữ tự nhiên trong trí tuệ nhân tạo là công nghệ nhận biết ngôn ngữ tự nhiên của con người ngày càng phát triển và được ứng dụng rộng rãi. Những nghiên cứu này sẽ hỗ trợ tương tác tự nhiên giữa người và máy, trong đó máy sẽ học cách hiểu ngôn ngữ của con người, điều chỉnh và tương tác chủ động. Bài báo này nghiên cứu trợ lý ảo có thể chủ động giao tiếp, tương tác với người sử dụng bằng công nghệ trí thông minh nhân tạo. Mô hình huấn luyện nhận diện giọng nói và phân tích giọng nói phát triển trên thư viện Speech recognition, phân tích âm thanh qua thư viện Pyaudio và playsound, nguồn dữ liệu tìm kiếm truy vấn trên cơ sở dữ liệu trực tuyến. Công nghệ xây dựng trợ lý ảo hoàn toàn sử dụng thư viện mã nguồn mở, không bị hạn chế bởi đám mây lưu chữ thông tin và dữ liệu huấn luyện đầu vào. Trợ lý ảo có thể ứng dụng rộng rãi trong giao tiếp, giáo dục, và các ngành dịch vụ đạt độ chính xác hơn 90% và nhóm đang tiếp tục cải thiện tối ưu hóa hệ thống.

Từ khóa: Trợ lý ảo, ngôn ngữ tự nhiên, trí tuệ nhân tạo.

Abstract: The Fourth Industrial Revolution is bursting with robotics and artificial intelligence which go deeply into our daily life. Robots and automatic devices are becoming more intelligent in their way to interact with humans and other devices. Natural language processing is the technology that is widely developed and applied. These studies will support the interaction between humans and machines; machines will study human language and adjust it properly. In particular, this virtual assistant can communicate with users who use artificial-intelligence technology. The training model for speech recognition and speech analysis is developed on the Speech recognition library, audio analysis through the Pyaudio and playsound libraries, and a search engine query data source on an online database. The technology building virtual assistants uses open source, which is not limited to store information and training data. Virtual assistants will be widely used in communication, education, and service.

Keywords: Assistance, natural language, artificial intelligence

1. GIỚI THIỆU

Ngày nay sự phát triển của các hệ thống trí tuệ nhân tạo (AI) có khả năng tổ chức tương tác giữa người và máy một cách tự nhiên (thông qua giọng nói, giao tiếp, cử chỉ, nét mặt...) đang ngày càng phổ biến. Một trong những hướng được nghiên cứu và phổ biến nhất là hướng tương tác, dựa trên sự hiểu biết của

máy móc về ngôn ngữ tự nhiên của con người. Nó không còn là con người học cách giao tiếp với máy móc, mà máy móc học cách giao tiếp với con người, khám phá hành động, thói quen, hành vi của họ và cố gắng trở thành trợ lý được cá nhân hóa của họ. Công việc tạo ra và cải tiến các trợ lý được cá nhân hóa như vậy đã diễn ra trong một thời gian dài. Các hệ thống này không ngừng cải tiến và cải tiến, vượt ra ngoài máy tính cá nhân và đã tạo dựng vững chắc cho mình trong các thiết bị di động và tiện ích khác nhau. Trợ lý ảo (có thể được gọi là trợ lý kỹ thuật số, trợ lý giọng nói hay là trợ lý AI) là một ứng dụng lập trình hướng nhiệm vụ, nhận dạng giọng nói của con người và thực hiện các lệnh được phát âm bởi người dùng. Nền tảng của nó là AI và năng suất của nó dựa vào việc lưu trữ hàng triệu từ và hàng triệu cụm từ. Không giống như các thiết bị nhận dạng giọng nói đầu tiên mà các nhà khoa học đang nghiên cứu vào những năm 40-50 của thế kỷ XX, các trợ lý kỹ thuật số hiện đại không bị hạn chế bởi một mẫu ngôn ngữ hoặc từ vựng nhất định.

Vào những năm 1960, Bộ Quốc phòng Hoa Kỳ đã quan tâm đến loại công việc này và bắt đầu đào tạo máy tính để bắt chước lý luận cơ bản của con người. Công việc này đã mở đường cho tự động hóa và lý luận chính thức mà chúng ta thấy trong các máy tính ngày nay. Năm 1966 Báo cáo của Ủy ban Tư vấn xử lý ngôn ngữ tự động (ALPAC) của chính phủ Hoa Kỳ nêu chi tiết về sự thiếu tiến bộ trong nghiên cứu dịch máy, một sáng kiến lớn của chiến tranh lạnh với lời hứa dịch tự động tiếng Nga. Năm 1970 các nhà nghiên cứu tại Đại học Carnegie Mellon ở Pittsburgh, Pennsylvania cùng với sự hỗ trợ của Bộ Quốc phòng Hoa Kỳ và Cơ quan Dự án Nghiên cứu Quốc phòng Tiên tiến (DARPA) - đã tạo ra chiếc máy Harpy. Nó có thể hiểu gần 1.000 từ, gần bằng từ vựng của một đứa trẻ ba tuổi. Vào tháng 4 năm 1997, Dragon NataturalSpeaking là phần mềm chỉnh sửa chính tả đầu tiên có thể hiểu khoảng 100 từ và biến nó thành nội dung có thể đọc được. Năm 1982 Bộ Thương mại Quốc tế và Công nghiệp Nhật Bản khởi động dự án Hệ thống máy tính thế hệ thứ năm đầy tham vọng. Mục tiêu của FGCS là phát triển hiệu năng giống như siêu máy tính và một nền tảng để phát triển trí tuệ nhân tạo AI. 2005 STANLEY, một chiếc xe tự lái, chiến thắng DARPA Grand Challenge. Quân đội Hoa Kỳ bắt đầu đầu tư vào các robot tự hành như "Big Dog" của Boston Dynamic và "PackBot" của iRobot và trợ lý ảo. 2008 Google tạo ra những bước đột phá trong nhận dạng giọng nói và giới thiệu tính năng này trong ứng dụng iPhone đưa trợ lý ảo giọng nói phổ biến thương mại ra thị trường. Một trong những trợ lý giọng nói phổ biến nhất là Siri của Apple, Amazon Echo, ứng với tên Alex từ Amazon, Cortana từ Microsoft, Google Assistant từ Google.

Xử lý ngôn ngữ tự nhiên (natural language processing - NLP) là một nhánh của trí tuệ nhân tạo tập trung vào các ứng dụng trên ngôn ngữ của con người. Trong trí tuệ nhân tạo thì xử lý ngôn ngữ tự nhiên là một trong những phần khó nhất vì nó liên quan đến việc phải hiểu ý nghĩa ngôn ngữ - công cụ hoàn hảo nhất của tư duy và giao tiếp. Trợ lý ảo (có thể được gọi là trợ lý kỹ thuật số, trợ lý giọng nói hay là trợ lý AI) này là một ứng dụng nhận dạng giọng nói của con người và thực hiện các lệnh được phát âm bởi người dùng) nhằm phát triển một trợ lý cá nhân được điều khiển bằng giọng nói đang thực hiện nhiều việc như

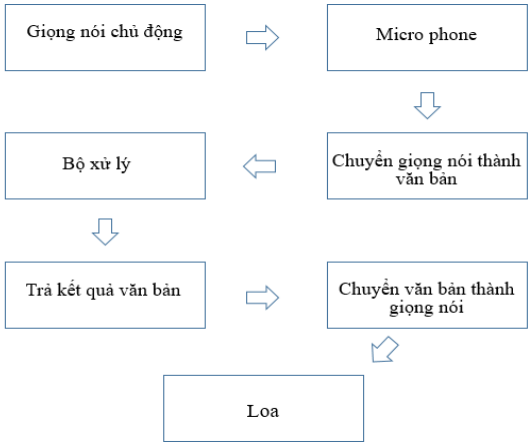
tự chủ nghe và trả lời các câu hỏi với tư duy logic trí tuệ nhân tạo, tìm kiếm thông tin trên internet khi được hỏi và trả lời các thông tin tìm kiếm được, chủ động các tác vụ hành động tương tác với thiết bị ngoại vi như mở ứng dụng, điều khiển thiết bị ngoại vi. Trợ lý ảo có thể mở và khởi chạy các ứng dụng web và bộ nhớ cục bộ của máy tính người dùng.

2. NGUYÊN LÝ HOẠT ĐỘNG

2.1. Nguyên lý chung

Nguyên lý chung của các hệ thống trợ lý ảo đều dựa trên phương pháp học máy và sử dụng một lượng lớn dữ liệu được thu thập từ nhiều nguồn khác nhau, sau đó được đào tạo về chúng, nguồn của dữ liệu này đóng vai trò quan trọng, có thể là hệ thống tìm kiếm, các nguồn thông tin khác nhau hoặc mạng xã hội. Số lượng thông tin từ các nguồn khác nhau xác định bản chất của trợ lý, do đó có thể là kết quả chính xác hoặc không chính xác tùy thuộc nguồn dữ liệu. Xây dựng hệ thống có thể tự huấn luyện từ nguồn dữ liệu nhờ nghiên cứu xây dựng hoặc sử dụng các module và thư viện mã nguồn mở. Mỗi hệ thống trợ lý ảo có các phương pháp tiếp cận để học tập, các thuật toán và kỹ thuật khác nhau, nhưng nguyên tắc xây dựng các hệ thống như vậy vẫn xấp xỉ giống nhau. Các công nghệ được sử dụng để tạo ra các hệ thống tương tác thông minh với con người bằng ngôn ngữ tự nhiên như giọng nói kích hoạt, nhận dạng giọng nói tự động, dạy sang giọng nói ((Teach-To-Speech), sinh trắc học giọng nói (Voice biometrics), trình quản lý hộp thoại (Dialog manager), hiểu ngôn ngữ tự nhiên (Natural language understanding) và công nhận thực thể được đặt tên (Named entity recognition).

2.2. Nguyên lý hệ thống

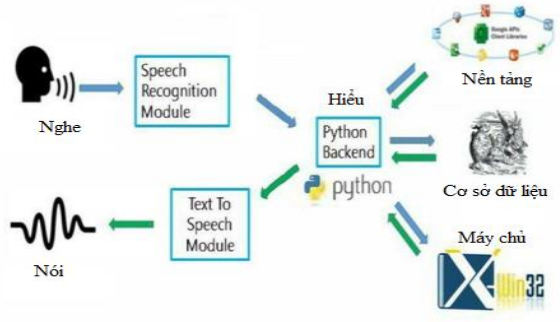


Hình 1. Sơ đồ nguyên lý hệ thống

Giọng nói chủ động được nói vào microphone, microphone thu nhận tín hiệu âm thanh để nhận dạng giọng nói và chuyển đổi đầu vào giọng nói thành văn bản chữ. Giọng nói được nhận diện dựa trên mô hình đã được huấn luyện để nhận biết âm thanh đó là gì và chuyển sang văn bản đúng nghĩa. Văn bản này sau đó được đưa đến bộ xử lý trung tâm để xác định bản chất của lệnh và gọi tập lệnh liên quan cho chấp hành. Sau khi bộ xử lý nhận dạng và hiểu được lệnh yêu cầu, chạy thuật toán tìm kiếm và xử lý thông tin nhận được. Khi có kết quả sau thuật toán xử lý, kết quả sẽ được in ra dạng text văn bản. Kết quả văn bản lại được chuyển đổi thành dạng âm thanh và phát ra loa để kết thúc chu trình một câu lệnh giao tiếp.

3. XÂY DỰNG HỆ THỐNG

3.1. Cấu trúc hệ thống



Hình 2. Cấu trúc hoạt động hệ thống

Hình 2 mô tả cấu trúc hệ thống trợ lý ảo giọng nói có ba module chính là nghe, hiểu và nói ngoài ra còn có các nền tảng và cơ sở dữ liệu cung cấp tri thức. Toàn bộ hệ thống chạy trên ngôn ngữ lập trình Python với các thư viện python.

A. Nghe: Nhận dạng giọng nói hệ thống sử dụng hệ thống nhận dạng giọng nói trực tuyến sử dụng thư viện Speech_recognition để chuyển đổi đầu vào bằng giọng nói thành văn bản. Người dùng có thể lấy văn bản từ kho tài liệu đặc biệt được tổ chức trên máy chủ mạng máy tính tại trung tâm thông tin từ micrô là được lưu trữ tạm thời trong hệ thống, sau đó được gửi đến đám mây của Google để nhận dạng giọng nói. Văn bản tương đương sau đó được nhận và đưa đến bộ xử lý trung tâm.

B. Hiểu: Đây là não bộ của trợ lý ảo sử dụng robot_brain. Phần hỗ trợ Python nhận đầu ra từ môđun nhận dạng giọng nói và sau đó xác định xem lệnh hay giọng nói đầu ra là một lệnh gọi API, trích xuất ngữ cảnh và lệnh gọi hệ thống. Đầu ra sau đó được gửi trở lại chương trình phụ trợ python để đưa ra yêu cầu xuất cho người dùng.

C. Trích xuất ngữ cảnh: Trích xuất ngữ cảnh (CE) là nhiệm vụ trích xuất tự động thông tin có cấu trúc từ không có cấu trúc và / hoặc bán cấu trúc tài liệu có thể đọc được bằng máy. Trong hầu hết các trường hợp, hoạt động này liên quan đến việc xử lý các văn bản ngôn ngữ của con người bằng các phương thức tự nhiên xử lý ngôn ngữ (NLP). Các hoạt động gần đây trong xử lý tài liệu đa phương tiện như chú thích tự động và trích xuất nội dung ra khỏi hình ảnh / âm thanh / video có thể được coi là kết quả kiểm tra trích xuất ngữ cảnh.

E. Cuộc gọi hệ thống: Trong máy tính, lệnh gọi hệ thống là cách lập trình trong đó chương trình máy tính yêu cầu một dịch vụ từ hệ điều hành mà nó được thực thi. Điều này có thể bao gồm các dịch vụ liên quan đến phần cứng (ví dụ: truy cập ổ đĩa cứng), tạo và thực thi các quy trình mới và giao tiếp với các dịch vụ nhân tích hợp chẳng hạn như lập lịch quy trình. Hệ thống cung cấp cuộc gọi một giao diện thiết yếu giữa quy trình và hệ điều hành.

F. Nói: Chuyển văn bản thành giọng nói Text-to-Speech (TTS) đề cập đến khả năng máy tính đọc to văn bản. Công cụ TTS chuyển đổi văn bản viết thành một phiên âm biểu diễn, sau đó chuyển đổi biểu diễn âm vị thành dạng sóng có thể phát ra dưới dạng âm thanh. Các công cụ TTS với các ngôn ngữ, phương ngữ và từ vựng chuyên ngành có sẵn thông qua các nhà xuất bản bên thứ ba

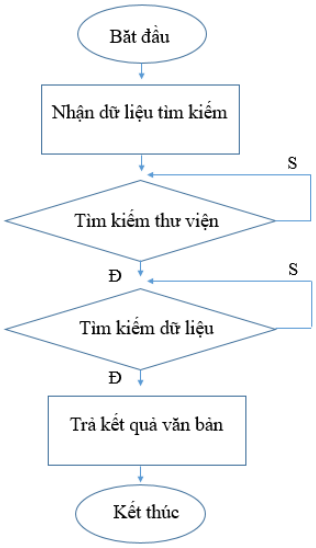
3.2. Thiết kế hệ thống

```
import os
import playsound
import speech_recognition as sr
import time
import sys
import ctypes
import wikipedia
import datetime
import json
import re
import webbrowser
import smtplib
import requests
import urllib
import urllib.request as urllib2
from selenium import webdriver
from selenium.webdriver.common.keys import Keys
from webdriver_manager.chrome import ChromeDriverManager
from time import strftime
from gtts import gTTS
from youtube_search import YoutubeSearch
```

Hình 3. Thư viện hỗ trợ

Hệ thống trợ lý ảo giọng nói khởi chạy trên nền tảng window với ngôn ngữ python. Để khởi chạy hệ thống nghe, hiểu, nói nghiên cứu sử dụng các thư viên hỗ trợ như hình 3. Điển hình là thư viện speech_recognition để nhận dạng âm thanh giọng nói và thư viên Playsound để phát âm thanh ra loa. Ngoài ra hệ thống còn sử dụng các thư viện thời gian, thời tiết…

Xử lý ngôn ngữ tự nhiên là một trong những mảng nghiên cứu khó nhất vì nó liên quan đến việc phải hiểu ý nghĩa ngôn ngữ tư duy và giao tiếp chưa kể ở dạng tiếng việt. Nghiên cứu có thể xử lý ngôn ngữ tiếng việt ở nhiều vùng miền, cả nam lẫn nữ, tương đối phức tạp. Nghiên cứu này nhóm sử dụng thư viện tiếng việt tích hợp hỗ trợ trên hệ thống "language = 'vi'". Tuy chất lượng xử lý tiếng việt chưa thực sự tuyệt vời với tất cả các giọng vùng miền nhưng với giọng điệu ngôn ngữ phổ thông xử lý hoàn toàn tốt.



Hình 4. Thuật toán tìm kiếm dữ liệu

Nghe với thư viện speech_recognition (sr) có chức năng là nhận dạng giọng nói để chuyển âm thanh thành văn bản. Âm thanh được đọc vào microphone của máy tính sau đó được xử lý qua hàm listen của sr.Recognition rồi lưu dữ liệu âm thanh vào biến audio. Dữ liệu âm thanh audio thu được sẽ được nhận dạng ở ngôn ngữ tiếng việt trong hàm r.recognize_google để chuyển thành dạng văn bản rồi lưu dữ liệu vào biến text. Nếu dữ liệu âm thanh audio không lỗi tức là hàm r.recognize_google có thể nhận dạng được audio để chuyên thành text thì hàm get_audio() sẽ được trả về giá trị là text còn nếu dữ liệu audio bị lỗi mà hàm r.recognition_google không nhận dạng được thì hàm get_audio() sẽ được trả về giá trị là 0.

Hiểu: Robot_Brain sau khi nhận tín hiệu nghe được sẽ truy suất vào các thư viện được hỗ trợ trong chương trình như thư viện ngày, giờ, duyệt web, truy suất hệ thống, khởi chạy ứng dụng... Ưu điểm của sử dụng thư viện tích hợp sẽ giúp hệ thống bớt cồng kềnh về việc lưu trữ dữ liệu đám mây, không cần cung cấp dữ liệu dạy học cho hệ thống mà sử dụng nền tảng có sẵn. Hình 4 là thuật toán truy suất dữ liệu trên hệ thống qua các thư viện hỗ trợ và kho dữ liệu trực tuyến của hệ thống.

Nói: Kết quả dữ liệu tìm kiếm trên hệ thống của thư viện đã được dạy học và kiểm chứng được trích xuất kết quả dạng văn bản. Văn bản kết quả được chuyển thành dữ liệu định dạng âm thanh và phát ra loa. Hình 5 là cấu trúc tập lệnh chuyển đổi.

```
def get_audio():
    print("\nHuman Robot: \tĐang nghe \t --__-- \n")
    r = sr.Recognizer()
    with sr.Microphone() as source:
        print("Tôi: ", end='')
        audio = r.listen(source, phrase_time_limit=8)
        try:
            text = r.recognize_google(audio, language="vi-VN")
            print(text)
            return text.lower()
        except:
            print("...")
            return 0
```

Hình 5. Tập lệnh nói của trợ lý ảo

3. KẾT QUẢ THẢO LUẬN

Tiến hành thử nghiệm giao tiếp ngẫu nghiên với Trợ lý ảo mỗi lần 50 câu lấy kết quả. Bảng kết quả đánh giá:

Thử nghiệm	Số câu đúng	Số câu sai	Độ chính xác
1	28/50	22/50	56%
2	31/50	19/50	62%
3	24/50	26/50	48%
4	39/50	11/50	78%
5	43/50	7/50	90%

Đánh giá kết quả thử nghiệm

Dựa trên bảng kết quả có thể thấy độ chính xác sẽ tăng dần. Có thể giải thích vì:

Lần 1: Giao tiếp với trợ lý ảo thì có những câu trợ lý ảo chưa được huấn luyện nên sẽ dẫn tới nó không hiểu và trả lời sai. Vì thế độ chính xác sẽ thấp.

Lần 2: Những câu trả lời mới sẽ được huấn luyện lại cho trợ lý ảo hiểu, nên lần sau gặp câu đó nó sẽ trả lời đúng ý của người dùng. Vì thế độ chính xác sẽ tăng thêm.

Lần 3: Giao tiếp ở những nội dung khác nhau, do huấn luyện chưa có nội dung đó nên trả lời sai vì thế độ chính xác cũng thấp.

Lần 4, 5: Khi được huấn luyện tiếp, độ chính xác sẽ tăng và người dùng nói đúng nội dung trợ lý ảo được huấn luyện

Trong bài báo này, Nhóm nghiên cứu đưa ra thiết kế bằng mã nguồn mở môđun phần mềm với sự hỗ trợ của thư viện python. Hướng nghiên cứu này giúp xây dựng hệ thống đơn giản hơn, dễ dàng thêm các tính năng bổ sung mà không làm ảnh hưởng đến các chức năng hiện tại của hệ thống. Nó không chỉ hoạt động theo lệnh của con người mà còn đưa ra phản hồi cho người dùng trên cơ sở truy vấn được hỏi hoặc các từ được nói bởi người dùng chẳng hạn như mở các tác vụ và hoạt động.

Với kết quả thực nghiệm mô hình hệ thống qua 50 câu hỏi bất kỳ các nội dung có kết quả thống kê: Thời gian phản hồi 2 giây; độ chính xác thông tin trả về 90%; dữ liệu truy vấn theo thời gian thực chính xác 99%; Phạm vi thông tin và ưu tiên kết quả trả lời phụ thuộc thuật toán google.

Tuy nhiên, trong quá trình nhận dạng giọng nói sẽ gặp phải sự phức tạp do nhiễu. Có rất nhiều yếu tố khác có thể đóng một vai trò gây nhiễu và làm ảnh hưởng tới kết quả nhận dạng cũng như chạy thuật toán gây sai số sấp xỉ 10%. Tiếng ồn xung quanh có thể dễ dàng khiến thiết bị nhận dạng giọng nói đi chệch hướng. Hay giọng đặc trưng vùng miền mà thư viện với dữ liệu chưa đươc huấn luyện kỹ.

5. KẾT LUẬN

Nghiên cứu này đã trình bày một nghiên cứu thiết kế trợ lý ảo giọng nói ứng dụng xử lý ngôn ngữ tự nhiên của trí tuệ nhân tạo. Kết quả nghiên cứu có thể áp dụng trong thực tế cuộc sống như giảng dạy làm ví dụ cho sinh viên về khả năng xử lý ngôn ngữ tự nhiên trong trí tuệ nhân tạo hay làm trợ lý trả lời các câu hỏi của sinh viên trong lĩnh vực trợ lý đã được huấn luyện dữ liệu. Hướng nghiên cứu phát triển tiếp theo sẽ tự xây dựng cơ sở dữ liệu và thư viện để huấn luyện cho trợ lý những mảng kiến thức dữ liệu mới đặc thù chưa có sẵn trên thư viện hỗ trợ.

TÀI LIỆU THAM KHẢO

[1] G. Bohouta, V. Z. Këpuska, "*Comparing Speech Recognition Systems (Microsoft API Google API And CMU Sphinx)*", Int. Journal of Engineering Research and Application 2017, (2017).

[2] Hill, J., Ford, W.R. and Farreras, I.G., "*Real conversations with artificial intelligence: A comparison between human– human online conversations and human–chatbot conversations*". Computers in Human Behavior, 49, pp.245-250, (2015)

[3] M. Bapat, H. Gune, and P. Bhattacharyya, "*A paradigm-based finite state morphological analyzer for marathi*," in Proceedings of the 1st Workshop on South and Southeast Asian Natural Language Processing (WSSANLP), pp. 26–34, (2010).

[4] G. Muhammad, Y. Alotaibi, M. N. Huda, " *Pronunciation variation for asr: A survey of the "Automatic speech recognition for bangla digits*," literature," Speech Communication, vol. 29, no. in Computers and Information Technology, 2009. 2, pp. 225–246, (1999).

[5] S.R. Eddy, "Hidden Markov models", *Current opinion in structural biology*", vol. 6, no. 3, pp. 361–365, (1996).

[6] Srivastava and S. Prakash, "*An Analysis of Various IoT Security Techniques: A Review*," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 355- 362, doi: 10.1109/ICRITO48877.2020.9198027, (2020).

[7] Saijshree Srivastava, Surya Vikram Singh, Rudrendra Bahadur Singh, Himanshu Kumar Shukla," *Digital Transformation of Healthcare: A blockchain study*" International Journal of Innovative Science, Engineering & Technology, Vol. 8 Issue 5, May (2021).

Thông tin liên hệ: **Phạm Trung Thiên**

Điện thoại: 0963284444 - Email: ptthien.ck@uneti.edu.vn

Khoa Cơ khí, Trường Đại học Kinh tế - Kỹ thuật Công nghiệp.