**TRƯỜNG ĐẠI HỌC CẦN THƠ**

**KHOA CÔNG NGHỆ THÔNG TIN & TRUYỀN THÔNG**



**BÁO CÁO NGHIÊN CỨU**

**NIÊN LUẬN KHOA HỌC MÁY TÍNH**

## **XÂY DỰNG ỨNG DỤNG PHÁT HIỆN**

## **TỪ KHÓA TRONG CÂU NÓI**

**Sinh viên thực hiện:**

Họ tên: Vũ Thái Hà

MSSV: B2016960

**Giảng viên hướng dẫn:**

TS. Phạm Nguyên Hoàng

**Cần Thơ, 8/2023**

## Ứng dụng

* [Kích hoạt trợ lý ảo: Mô hình KWS được sử dụng để kích hoạt các trợ lý ảo như “Hey Siri” cho trợ lý Siri của Apple hoặc “OK Google” cho trợ lý Google Assistant1](https://bkhost.vn/blog/phan-mem-ung-dung/" \t "https://www.bing.com/_blank).
* [Điều khiển bằng giọng nói: Mô hình KWS cũng được sử dụng trong các ứng dụng điều khiển bằng giọng nói, cho phép người dùng điều khiển các thiết bị thông minh bằng giọng nói1](https://bkhost.vn/blog/phan-mem-ung-dung/" \t "https://www.bing.com/_blank).
* [Tìm kiếm bằng giọng nói: Mô hình KWS cũng được sử dụng trong các ứng dụng tìm kiếm bằng giọng nói, cho phép người dùng tìm kiếm thông tin bằng cách nói ra các từ khóa](https://bkhost.vn/blog/phan-mem-ung-dung/" \t "https://www.bing.com/_blank)

## Blog

* [keyword-spotting · GitHub Topics · GitHub](https://github.com/topics/keyword-spotting)
* [TinyML Made Easy: KeyWord Spotting (KWS)](https://www.hackster.io/mjrobot/tinyml-made-easy-keyword-spotting-kws-5fa6e7)
* [Keyword Spotting (KWS) in Voice Recognition](https://picovoice.ai/blog/keyword-spotting-voice-recognition/)
* [“Simple audio recognition: Recognizing keywords”](https://www.tensorflow.org/tutorials/audio/simple_audio" \t "https://www.bing.com/_blank)
* [“Launching the Speech Commands Dataset”](https://ai.googleblog.com/2017/08/launching-speech-commands-dataset.html" \t "https://www.bing.com/_blank)

## Bài báo và các nghiên cứu

* “Spoken keyword detection using autoassociative neural networks” by S. [Jothilakshmi](https://link.springer.com/article/10.1007/s10772-013-9208-2" \t "https://www.bing.com/_blank)
* [“Low-Power Audio Keyword Spotting using Tsetlin Machines” by cair/TsetlinMachine](https://paperswithcode.com/task/keyword-spotting" \t "https://www.bing.com/_blank)
* [“Keyword extraction: Issues and methods” by Nazanin Firoozeh, Adeline Nazarenko, Fabrice Alizon, and Béatrice Daille](https://www.cambridge.org/core/journals/natural-language-engineering/article/abs/keyword-extraction-issues-and-methods/84BFD5221E2CA86326E5430D03299711" \t "https://www.bing.com/_blank)
* [“The Main Issues on Keyword Retrieval from Audio Streams” by Zhang Weitao, Mijit Ablimit, and Askar Hamdulla](https://ieeexplore.ieee.org/document/9526663" \t "https://www.bing.com/_blank)
* [“Deep Spoken Keyword Spotting: An Overview”](https://arxiv.org/abs/2111.10592" \t "https://www.bing.com/_blank)
* [1811.07684.pdf (arxiv.org)](https://arxiv.org/pdf/1811.07684.pdf)

1. **Kết luận**

* [KWS (Keyword Spotting) là một kỹ thuật nhận dạng giọng nói tự động, được sử dụng để phát hiện các từ khóa trong một luồng âm thanh1](https://blogdaytinhoc.com/tong-hop-tai-lieu-tu-hoc-ke-toan-thuc-hanh-tu-co-ban-toi-nang-cao-204" \t "https://www.bing.com/_blank). [KWS có thể được sử dụng trong nhiều ứng dụng khác nhau, chẳng hạn như kích hoạt trợ lý ảo, điều khiển bằng giọng nói, và tìm kiếm bằng giọng nói1](https://blogdaytinhoc.com/tong-hop-tai-lieu-tu-hoc-ke-toan-thuc-hanh-tu-co-ban-toi-nang-cao-204" \t "https://www.bing.com/_blank).
* [KWS có thể được phân loại thành hai loại chính: KWS độc lập và KWS phụ thuộc1](https://blogdaytinhoc.com/tong-hop-tai-lieu-tu-hoc-ke-toan-thuc-hanh-tu-co-ban-toi-nang-cao-204" \t "https://www.bing.com/_blank). [KWS độc lập là loại KWS có thể hoạt động mà không cần kết nối với mạng1](https://blogdaytinhoc.com/tong-hop-tai-lieu-tu-hoc-ke-toan-thuc-hanh-tu-co-ban-toi-nang-cao-204" \t "https://www.bing.com/_blank). [Nó thường được sử dụng trong các thiết bị di động và các thiết bị IoT](https://www.bing.com/search?pglt=43&q=document+of+keyword+detection+in+audio&cvid=0ab3f6403c324a668efd75c4f7b8d4a4&aqs=edge..69i57j69i64j0l3.16096j0j1&FORM=ANAB01&PC=U531)[1](https://blogdaytinhoc.com/tong-hop-tai-lieu-tu-hoc-ke-toan-thuc-hanh-tu-co-ban-toi-nang-cao-204" \t "https://www.bing.com/_blank). [KWS phụ thuộc là loại KWS cần kết nối với mạng để hoạt động1](https://blogdaytinhoc.com/tong-hop-tai-lieu-tu-hoc-ke-toan-thuc-hanh-tu-co-ban-toi-nang-cao-204" \t "https://www.bing.com/_blank). [Nó thường được sử dụng trong các ứng dụng trên máy tính và điện thoại thông minh1](https://blogdaytinhoc.com/tong-hop-tai-lieu-tu-hoc-ke-toan-thuc-hanh-tu-co-ban-toi-nang-cao-204" \t "https://www.bing.com/_blank).
* [Có nhiều phương pháp khác nhau để thực hiện KWS, bao gồm cả các phương pháp truyền thống và các phương pháp học sâu1](https://blogdaytinhoc.com/tong-hop-tai-lieu-tu-hoc-ke-toan-thuc-hanh-tu-co-ban-toi-nang-cao-204" \t "https://www.bing.com/_blank). [Các phương pháp truyền thống bao gồm các phương pháp dựa trên mô hình Markov ẩn (HMM) và các phương pháp dựa trên mạng nơ-ron1](https://blogdaytinhoc.com/tong-hop-tai-lieu-tu-hoc-ke-toan-thuc-hanh-tu-co-ban-toi-nang-cao-204" \t "https://www.bing.com/_blank). [Các phương pháp học sâu bao gồm các mô hình như mạng nơ-ron tích chập (CNN) và mạng nơ-ron hồi quy (RNN)1](https://blogdaytinhoc.com/tong-hop-tai-lieu-tu-hoc-ke-toan-thuc-hanh-tu-co-ban-toi-nang-cao-204" \t "https://www.bing.com/_blank).
* [Mạng nơ-ron tích chập (CNN), mạng nơ-ron hồi quy (RNN), và mô hình Transformer là ba loại mạng nơ-ron phổ biến được sử dụng trong các bài toán học sâu](https://www.bing.com/search?pglt=43&q=document+of+keyword+detection+in+audio&cvid=0ab3f6403c324a668efd75c4f7b8d4a4&aqs=edge..69i57j69i64j0l3.16096j0j1&FORM=ANAB01&PC=U531)[1](https://ai.stackexchange.com/questions/20075/why-does-the-transformer-do-better-than-rnn-and-lstm-in-long-range-context-depen" \t "https://www.bing.com/_blank).
* Mỗi loại mạng có những ưu và nhược điểm riêng khi được áp dụng trong việc phát hiện từ khóa trong âm thanh.
* [CNN: Mạng nơ-ron tích chập thường được sử dụng trong các bài toán xử lý ảnh và video](https://www.bing.com/search?pglt=43&q=document+of+keyword+detection+in+audio&cvid=0ab3f6403c324a668efd75c4f7b8d4a4&aqs=edge..69i57j69i64j0l3.16096j0j1&FORM=ANAB01&PC=U531)[1](https://ai.stackexchange.com/questions/20075/why-does-the-transformer-do-better-than-rnn-and-lstm-in-long-range-context-depen" \t "https://www.bing.com/_blank). [Ưu điểm của CNN là khả năng trích xuất đặc trưng tự động từ dữ liệu đầu vào, giúp giảm thiểu công đoạn tiền xử lý dữ liệu](https://www.bing.com/search?pglt=43&q=document+of+keyword+detection+in+audio&cvid=0ab3f6403c324a668efd75c4f7b8d4a4&aqs=edge..69i57j69i64j0l3.16096j0j1&FORM=ANAB01&PC=U531)[1](https://ai.stackexchange.com/questions/20075/why-does-the-transformer-do-better-than-rnn-and-lstm-in-long-range-context-depen" \t "https://www.bing.com/_blank). [Tuy nhiên, nhược điểm của CNN là khó khăn trong việc xử lý các dữ liệu có cấu trúc tuần tự dài](https://www.bing.com/search?pglt=43&q=document+of+keyword+detection+in+audio&cvid=0ab3f6403c324a668efd75c4f7b8d4a4&aqs=edge..69i57j69i64j0l3.16096j0j1&FORM=ANAB01&PC=U531)[1](https://ai.stackexchange.com/questions/20075/why-does-the-transformer-do-better-than-rnn-and-lstm-in-long-range-context-depen" \t "https://www.bing.com/_blank).
* [RNN: Mạng nơ-ron hồi quy thường được sử dụng trong các bài toán xử lý ngôn ngữ tự nhiên và dự đoán chuỗi thời gian](https://www.bing.com/search?pglt=43&q=document+of+keyword+detection+in+audio&cvid=0ab3f6403c324a668efd75c4f7b8d4a4&aqs=edge..69i57j69i64j0l3.16096j0j1&FORM=ANAB01&PC=U531)[1](https://ai.stackexchange.com/questions/20075/why-does-the-transformer-do-better-than-rnn-and-lstm-in-long-range-context-depen" \t "https://www.bing.com/_blank). [Ưu điểm của RNN là khả năng xử lý các dữ liệu có cấu trúc tuần tự](https://www.bing.com/search?pglt=43&q=document+of+keyword+detection+in+audio&cvid=0ab3f6403c324a668efd75c4f7b8d4a4&aqs=edge..69i57j69i64j0l3.16096j0j1&FORM=ANAB01&PC=U531)[1](https://ai.stackexchange.com/questions/20075/why-does-the-transformer-do-better-than-rnn-and-lstm-in-long-range-context-depen" \t "https://www.bing.com/_blank). [Tuy nhiên, nhược điểm của RNN là khó khăn trong việc học các phụ thuộc dài hạn](https://www.bing.com/search?pglt=43&q=document+of+keyword+detection+in+audio&cvid=0ab3f6403c324a668efd75c4f7b8d4a4&aqs=edge..69i57j69i64j0l3.16096j0j1&FORM=ANAB01&PC=U531)[1](https://ai.stackexchange.com/questions/20075/why-does-the-transformer-do-better-than-rnn-and-lstm-in-long-range-context-depen" \t "https://www.bing.com/_blank).
* [Transformer: Mô hình Transformer được giới thiệu để giải quyết các vấn đề về phụ thuộc dài hạn và tính toán tuần tự trong RNN](https://www.bing.com/search?pglt=43&q=document+of+keyword+detection+in+audio&cvid=0ab3f6403c324a668efd75c4f7b8d4a4&aqs=edge..69i57j69i64j0l3.16096j0j1&FORM=ANAB01&PC=U531)[1](https://ai.stackexchange.com/questions/20075/why-does-the-transformer-do-better-than-rnn-and-lstm-in-long-range-context-depen" \t "https://www.bing.com/_blank). [Ưu điểm của Transformer là khả năng xử lý các dữ liệu có cấu trúc tuần tự một cách hiệu quả và cho phép tính toán song song1](https://ai.stackexchange.com/questions/20075/why-does-the-transformer-do-better-than-rnn-and-lstm-in-long-range-context-depen" \t "https://www.bing.com/_blank). [Tuy nhiên, nhược điểm của Transformer là chi phí tính toán cao khi xử lý các chuỗi dài1](https://ai.stackexchange.com/questions/20075/why-does-the-transformer-do-better-than-rnn-and-lstm-in-long-range-context-depen" \t "https://www.bing.com/_blank).
* There are many models that are suitable for small-footprint keyword spotting (KWS) with small datasets and limited computational resources. [Some of the popular models include convolutional neural networks (CNNs), recurrent neural networks (RNNs), and Transformers1](https://arxiv.org/abs/1912.05124" \t "https://www.bing.com/_blank). Each of these models has its own strengths and weaknesses, and the best model for a particular KWS task may depend on various factors such as the size of the dataset, the computational resources available, and the desired trade-off between accuracy and efficiency.
* [One approach that has been shown to be effective for small-footprint KWS is the use of attention-based end-to-end neural models2](https://www.isca-speech.org/archive_v0/Interspeech_2018/pdfs/1204.pdf" \t "https://www.bing.com/_blank). [These models aim to simplify the pipelines of building a production-quality KWS system by using an encoder and an attention mechanism to automatically extract relevant features from the input audio signal2](https://www.isca-speech.org/archive_v0/Interspeech_2018/pdfs/1204.pdf" \t "https://www.bing.com/_blank). [Another approach is to use graph convolutional networks (GCNs) to encode non-local relations in the convolutional feature maps1](https://arxiv.org/abs/1912.05124" \t "https://www.bing.com/_blank).
* [1204.pdf (isca-speech.org)](https://www.isca-speech.org/archive_v0/Interspeech_2018/pdfs/1204.pdf)

Chốt phương pháp thực hiện, phạm vi, keyword

Liệt kê tất cả keyword ít nhất 10 keyword