

BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM

---



**PHẠM THỊ LAN HƯƠNG**

**Xây dựng hệ thống truy vấn video nông nghiệp  
hướng ngữ nghĩa có sử dụng Ontology**

**LUẬN VĂN THẠC SĨ**

Chuyên ngành: Công nghệ thông tin

Mã số ngành: 60480201

TP. HỒ CHÍ MINH, tháng 01 năm 2015  
BỘ GIÁO DỤC VÀ ĐÀO TẠO

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM**

---



**PHẠM THỊ LAN HƯƠNG**

**Xây dựng hệ thống truy vấn video nông nghiệp  
hướng ngữ nghĩa có sử dụng Ontology**

**LUẬN VĂN THẠC SĨ**

Chuyên ngành: Công nghệ thông tin

Mã số ngành: 60480201

**CÁN BỘ HƯỚNG DẪN KHOA HỌC: PGS.TS. VŨ HẢI QUÂN**

TP. HỒ CHÍ MINH, tháng 01 năm 2015

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH TẠI**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ TP. HCM**

Cán bộ hướng dẫn khoa học: PGS.TS. VŨ HẢI QUÂN

Luận văn Thạc sĩ được bảo vệ tại Trường Đại học Công nghệ TP. HCM  
ngày 06 tháng 02 năm 2015

Thành phần Hội đồng đánh giá Luận văn Thạc sĩ gồm:

<b>TT</b>	<b>Họ và tên</b>	<b>Chức danh Hội đồng</b>
1	PGS.TS. Lê Hoài Bắc	Chủ tịch
2	PGS.TS. Quản Thành Thơ	Phản biện 1
3	TS. Lư Nhật Vinh	Phản biện 2
4	TS. Cao Tùng Anh	Ủy viên
5	TS. Võ Đình Bảy	Ủy viên, Thư ký

Xác nhận của Chủ tịch Hội đồng đánh giá Luận sau khi Luận văn đã được  
sửa chữa (nếu có).

Chủ tịch Hội đồng đánh giá LV

*TP. HCM, ngày tháng năm 2015*

## **NHIỆM VỤ LUẬN VĂN THẠC SĨ**

Họ tên học viên: Phạm Thị Lan Hương

Giới tính: Nữ

Ngày, tháng, năm sinh: 08/12/1982

Nơi sinh: Ninh Bình

Chuyên ngành: Công nghệ thông tin

MSHV: 1241860008

### **I- Tên đề tài:**

Xây dựng hệ thống truy vấn video nông nghiệp hướng ngữ nghĩa có sử dụng Ontology.

### **II- Nhiệm vụ và nội dung:**

Xây dựng bộ nhận dạng tiếng nói để chuyển dữ liệu audio từ các video chăn nuôi sang dạng văn bản trích, từ đó lập chỉ mục cho hệ thống tìm kiếm.

Xây dựng bộ Ontology cho nhánh chăn nuôi trong nông nghiệp Việt Nam để mở rộng câu truy vấn từ đó mở rộng và nâng cao chất lượng tìm kiếm.

Xây dựng hệ thống truy vấn video hướng ngữ nghĩa có sử dụng Ontology.

**III- Ngày giao nhiệm vụ:** 02/4/2014

**IV- Ngày hoàn thành nhiệm vụ:** 19/01/2015

**V- Cán bộ hướng dẫn:** PGS.TS. VŨ HẢI QUÂN

**CÁN BỘ HƯỚNG DẪN**

**KHOA QUẢN LÝ CHUYÊN NGÀNH**

PGS.TS. Vũ Hải Quân



## **LỜI CAM ĐOAN**

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi. Các số liệu, kết quả nêu trong Luận văn là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Tôi xin cam đoan rằng mọi sự giúp đỡ cho việc thực hiện Luận văn này đã được cảm ơn và các thông tin trích dẫn trong Luận văn đã được chỉ rõ nguồn gốc.

**Học viên thực hiện Luận văn**

**Phạm Thị Lan Hương**

## LỜI CẢM ƠN

Trong quá trình thực hiện luận văn thạc sĩ, tôi đã gặp không ít khó khăn, để đạt được kết quả này ngoài những nỗ lực của bản thân, tôi còn nhận được rất nhiều sự giúp đỡ từ phía quý Thầy, gia đình và bạn bè. Giờ đây sau khi đã hoàn thành luận văn này, tôi xin ghi vào đây những lời biết ơn đầy trân trọng.

Đầu tiên, tôi xin tỏ lòng kính trọng và biết ơn chân thành, sâu sắc nhất đến **PGS.TS Vũ Hải Quân**, người thầy đã định hướng đề tài nghiên cứu, tận tình hướng dẫn và tạo mọi điều kiện thuận lợi cho tôi hoàn thành luận văn này.

Kế đến, xin cảm ơn quý Thầy khoa Công nghệ thông tin, Trường Đại học Công nghệ TP. HCM đã tận tình hướng dẫn và cung cấp đầy đủ các thông tin giúp tôi hoàn thành luận văn.

Xin cảm ơn quý Thầy đã nhiệt tình giảng dạy và truyền đạt những kiến thức, kinh nghiệm quý báu cho tôi trong suốt khóa học.

Xin cảm ơn bạn Phạm Minh Nhật và các bạn phòng thí nghiệm AILAB thuộc Trường Đại học Khoa Học Tự Nhiên TP. HCM đã giúp đỡ tôi rất nhiều trong quá trình làm luận văn.

Xin cảm ơn gia đình đã bên tôi trong suốt chặng đường khó khăn này.

Cuối cùng xin cảm ơn các anh chị đồng nghiệp, các bạn sinh viên đã hỗ trợ tôi hoàn thành luận văn.

Mặc dù đã cố gắng hết sức để hoàn thành luận văn, song không thể tránh khỏi sai sót. Kính mong nhận được nhận xét và sự đóng góp của quý Thầy Cô và bạn bè.

**Học viên thực hiện**

**Phạm Thị Lan Hương**

## TÓM TẮT

Nội dung nghiên cứu chính của luận văn là xây dựng một hệ thống truy vấn video hướng ngữ nghĩa cho thông tin chăn nuôi trong nông nghiệp với sự hỗ trợ của hệ thống mở rộng câu truy vấn tự động từ một Ontology.

Việc truy vấn video được xây dựng mang tính hướng ngữ nghĩa dựa trên công nghệ nhận dạng tiếng nói, nghĩa là các video ban đầu sẽ được rút trích kênh audio sau đó đưa vào bộ nhận dạng tiếng nói để chuyển thể lời nói trong audio sang dạng văn bản trích. Từ văn bản trích, bộ lập chỉ mục sẽ đánh chỉ mục cho hệ thống tìm kiếm.

Ứng dụng của phương thức truy vấn video hướng ngữ nghĩa nhằm mang lại kết quả truy vấn gần với mong muốn tìm kiếm hơn, đỡ tốn chi phí cho việc chú thích bằng tay thông tin của video. Mô hình nhận dạng tiếng nói được xây dựng theo học mẫu, thống kê trên dữ liệu tiếng nói với tổng thời lượng là 4 giờ 34 phút 47 giây. Kết quả thực nghiệm đạt 85,23% độ chính xác nhận dạng.

Việc xây dựng bộ Ontology áp dụng vào hệ thống truy vấn thông tin video nhằm mở rộng câu truy vấn của người dùng từ đó mở rộng và nâng cao kết quả tìm kiếm.



## **ABSTRACT**

The main research content of the thesis is to build a system user query semantic video information for livestock in agriculture with the support of the system to expand queries automatically from an Ontology.

The query is constructed nature video oriented semantics based on speech recognition technology, meaning that the original video channel audio will be extracted and then put into the speech recognition to speech in the audio adaptation to textual criticism. From the extracted text, the index will index the search system.

Application of the method of query semantics video to yield results close to the desired query looking for more, less expensive for manual annotation of video information. Speech recognition model is built to study samples, statistical data on the total amount of time the voice was 4 hours 34 minutes 47 seconds. The experimental results achieved 85.23% recognition accuracy.

The construction of the Ontology applied to information retrieval system video to expand the user's query from which to expand and improve search results.

## MỤC LỤC

LỜI CAM ĐOAN .....	i
LỜI CẢM ƠN .....	ii
TÓM TẮT .....	iii
ABSTRACT .....	iv
MỤC LỤC.....	v
DANH MỤC CÁC TỪ VIẾT TẮT .....	ix
DANH MỤC CÁC HÌNH ẢNH .....	x
DANH MỤC CÁC BẢNG.....	xii
CHƯƠNG 1. TỔNG QUAN.....	1
1.1 Tổng quan và bối cảnh đề tài. ....	1
1.2 Một số hướng tiếp cận cho bài toán truy vấn video hướng ngữ nghĩa.....	2
1.2.1 Hướng tiếp cận dựa vào đặc trưng hình ảnh.....	2
1.2.2 Hướng tiếp cận dựa vào đặc trưng âm thanh.....	3
1.2.3 Hướng tiếp cận tổng hợp .....	3
1.3 Hướng tiếp cận của đề tài.....	3
1.4 Các công trình nghiên cứu liên quan.....	4
1.5 Mục tiêu của đề tài .....	7
1.6 Ý nghĩa khoa học và thực tiễn của đề tài .....	8
1.6.1 Ý nghĩa khoa học.....	8
1.6.2 Ý nghĩa thực tiễn .....	8
1.7 Bố cục của luận văn .....	8
CHƯƠNG 2. CƠ SỞ LÝ THUYẾT .....	10
2.1 Hệ thống tìm kiếm thông tin .....	10
2.1.1 Giới thiệu chung về hệ thống tìm kiếm thông tin.....	10
2.1.1.1 Định nghĩa về hệ thống tìm kiếm thông tin .....	10

2.1.1.2	Mục tiêu và chức năng của một hệ thống tìm kiếm thông tin.....	10
2.1.1.3	Kiến trúc chung của một hệ thống tìm kiếm thông tin .....	11
2.1.1.4	Phân loại hệ thống tìm kiếm thông tin .....	12
2.1.1.5	Tiêu chí để đánh giá một hệ thống tìm kiếm thông tin .....	12
2.1.2	Hệ thống tìm kiếm dựa trên từ khóa .....	13
2.1.2.1	Bộ thu thập thông tin - Robot .....	14
2.1.2.2	Bộ lập chỉ mục - Index.....	14
2.1.2.3	Bộ truy vấn (bộ tìm kiếm).....	14
2.1.3	Hệ thống tìm kiếm dựa trên khái niệm (hướng ngữ nghĩa).....	15
2.1.3.1	Bộ thu thập thông tin.....	16
2.1.3.2	Bộ lập chỉ mục khái niệm .....	17
2.1.3.3	Bộ truy vấn.....	18
2.2	Nhận dạng tiếng nói.....	19
2.2.1	Dẫn nhập .....	19
2.2.2	Trích chọn đặc trưng.....	21
2.2.3	Mô hình ngữ âm (Acoustic Model) .....	22
2.2.4	Mô hình ngôn ngữ (LM).....	25
2.2.5	Phép tìm kiếm .....	27
2.2.5.1	Đánh giá kết quả nhận dạng.....	28
2.2.5.2	Đánh giá mô hình ngôn ngữ.....	29
2.2.6	Tổ chức đồ thị tìm kiếm kết quả nhận dạng .....	30
2.2.6.1	Tổ chức đồ thị tìm kiếm .....	30
2.2.6.2	Đồ thị tìm kiếm với uni-gram và bi-gram .....	31
2.2.6.3	Đồ thị tìm kiếm dựa trên mô hình uni-gram .....	32
2.2.6.4	Đồ thị tìm kiếm dựa trên mô hình bi-gram .....	32

2.2.7 Đồ thị tìm kiếm với ngữ cảnh một âm vị (monophone), ngữ cảnh có xét âm vị trái phải chỉ trong từ (triphone within-word) và ngữ cảnh có xét âm vị trái phải trong câu nói (triphones cross-word).....	33
2.3 Ontology .....	35
2.3.1 Khái niệm.....	35
2.3.2 Mục đích xây dựng Ontology .....	35
2.3.3 Yêu cầu khi xây dựng Ontology.....	35
2.3.4 Các thành phần của Ontology.....	36
2.3.5 Phương pháp xây dựng Ontology .....	37
CHƯƠNG 3. HỆ THỐNG TRUY VẤN VIDEO NÔNG NGHIỆP.....	38
HƯỚNG NGỮ NGHĨA CÓ SỬ DỤNG ONTOLOGY.....	38
3.1 Kiến trúc tổng quan của hệ thống.....	38
3.2 Các thành phần của hệ thống.....	39
3.2.1 Bộ nhận dạng tiếng nói tiếng Việt .....	39
3.2.1.1 Công đoạn huấn luyện.....	39
3.2.1.2 Công đoạn nhận dạng.....	40
3.2.2 Ứng dụng truy vấn video .....	40
3.2.2.1 Lập chỉ mục kho dữ liệu .....	41
3.2.2.2 Xây dựng ứng dụng truy vấn video .....	48
3.2.2.3 Xây dựng Ontology cho hệ thống .....	54
3.3 Đặc tả phần mềm .....	55
3.3.1 Môi trường phát triển.....	55
3.3.2 Thiết kế kiến trúc.....	56
3.3.2.1 Sơ đồ lớp: .....	56
3.3.2.2 Diễn giải các lớp xử lý chính .....	56
CHƯƠNG 4. THỰC NGHIỆM .....	62
4.1 Cấu hình máy thử nghiệm .....	62

4.2	Thực nghiệm.....	62
4.2.1	Thực nghiệm đánh giá độ chính xác của mô hình nhận dạng tiếng nói .....	62
4.2.1.1	Tập dữ liệu huấn luyện mô hình ngữ âm .....	62
4.2.1.2	Tập dữ liệu huấn luyện mô hình ngôn ngữ .....	63
4.2.1.3	Tập dữ liệu test.....	63
4.2.1.4	Độ đo đánh giá .....	63
4.2.1.5	Kết quả thực nghiệm .....	64
4.2.2	Kết quả xây dựng Ontology chăn nuôi, phương pháp đánh giá và thực nghiệm truy vấn của hệ thống.....	65
4.2.2.1	Ontology chăn nuôi.....	65
4.2.2.2	Phương pháp đánh giá Ontology.....	76
4.2.2.3	Kết quả thực nghiệm .....	77
4.2.3	Demo ứng dụng .....	80
	CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN.....	84
5.1	Kết luận .....	84
5.1.1	Tổng kết.....	84
5.1.2	Những đóng góp của đề tài .....	84
5.1.3	Hạn chế của đề tài.....	84
5.2	Hướng phát triển.....	85
	TÀI LIỆU THAM KHẢO.....	86

**DANH MỤC CÁC TỪ VIẾT TẮT**

ASR	Automatic Speech Recognition
AQE	Automatic Query Expansion
OOV	Out-Of- Vocabulary
RTMP	Real Time Messaging Protocol
WER	Word Error Rate
WAR	Word Accuracy Rate
HMM	Hidden Markov Model
IR	Information Retrieval
LM	Language Model
AM	Acoustic Model

## DANH MỤC CÁC HÌNH ẢNH

Hình 2.1. Kiến trúc chung của hệ thống tìm kiếm thông tin.....	11
Hình 2.2. Hệ thống tìm kiếm dựa trên từ khóa .....	13
Hình 2.3. Hệ thống tìm kiếm dựa trên khái niệm .....	16
Hình 2.4. Quá trình nhận dạng tiếng nói.....	19
Hình 2.5. Các thành phần cơ bản của hệ thống nhận dạng tiếng nói.....	21
Hình 2.6. Mô hình HMM với 4 trạng thái.....	23
Hình 2.7. Minh họa tiến trình xây dựng mô hình từ dựa trên các phone.....	25
Hình 2.8. Tiến trình xây dựng đồ thị kết ghép từ LM, AC và từ điển phát âm .....	28
Hình 2.9. Đồ thị âm vị của từ không.....	31
Hình 2.10. Đồ thị "từ" xây dựng trên mô hình uni-gram.....	32
Hình 2.11. Đồ thị "từ" xây dựng trên mô hình uni-gram có sử dụng đỉnh NULL .....	32
Hình 2.12. Đồ thị "từ" xây dựng trên mô hình bi-gram.....	32
Hình 2.13. Đồ thị đơn âm vị.....	33
Hình 2.14. Đồ thị tri-gram within-word.....	34
Hình 2.15. Đồ thị tri-phone cross-word .....	34
Hình 3.1. Kiến trúc của hệ thống truy vấn video hướng ngữ nghĩa có sử dụng Ontology .....	39
Hình 3.2. Cách lưu trữ của mô hình chỉ mục ngược [17] .....	43
Hình 3.3. Mô hình tổng quát xây dựng chỉ mục ngược [17].....	44
Hình 3.4. Dãy các token đã được chỉnh sửa [17].....	45
Hình 3.5. Sắp xếp token theo từ chỉ mục và docID [17].....	46
Hình 3.6. Thông tin Dictionary và Postings của chỉ mục [17] .....	47
Hình 3.7. Hai danh sách Posting của “nông” và “nghiệp”.....	49
Hình 3.8. Kết quả của thuật toán trộn 2 danh sách posting.....	50
Hình 3.9. Minh họa Con trỏ nhảy .....	50

Hình 3.10. Tính cân bằng của việc đặt con trỏ nhảy.....	52
Hình 3.11. Truy vấn với dữ liệu chỉ mục theo từ riêng biệt.....	52
Hình 3.12. Minh họa lập chỉ mục từ theo vị trí.....	53
Hình 3.13. Dữ liệu chỉ mục theo nhóm từ và truy vấn .....	54
Hình 3.14. Sơ đồ các lớp xử lý của phần mềm .....	56
Hình 4.1. Kết quả nhận dạng của mô hình bi-gram .....	64
Hình 4.2. Mô hình tổng thể các khái niệm của Ontology chăn nuôi .....	67
Hình 4.3. Giao diện trang chủ của ứng dụng demo .....	80
Hình 4.4. Giao diện trang tìm kiếm thuật ngữ của Ontology trong chăn nuôi .....	81
Hình 4.5. Giao diện khái niệm thuật ngữ của Ontology trong chăn nuôi .....	81
Hình 4.6. Giao diện khái niệm thuật ngữ của Ontology trong chăn nuôi .....	81
Hình 4.7. Giao diện trang tìm kiếm thông tin về chăn nuôi khi không sử dụng Ontology (ví dụ: Kỹ thuật chăn nuôi dê) .....	82
Hình 4.8. Giao diện trang tìm kiếm thông tin về chăn nuôi khi có sử dụng Ontology..	82
Hình 4.9. Giao diện trang hiển thị nội dung chi tiết của video .....	83



## DANH MỤC CÁC BẢNG

Bảng 3.1. Chỉ mục của các tài liệu tương ứng với các thuật ngữ .....	42
Bảng 4.1. Cấu hình máy thử nghiệm.....	62
Bảng 4.2. Bảng thống kê dữ liệu huấn luyện mô hình ngữ âm.....	62
Bảng 4.3. Bảng thống kê dữ liệu huấn luyện mô hình ngôn ngữ.....	63
Bảng 4.4. Bảng thống kê dữ liệu test .....	63
Bảng 4.5. Phân loại các khái niệm của Ontology chăn nuôi.....	66
Bảng 4.6. Danh sách các thuật ngữ của Ontology .....	68
Bảng 4.7. Độ phủ của các lượt truy vấn.....	78
Bảng 4.8. Độ chính xác của các lượt truy vấn .....	79

# CHƯƠNG 1. TỔNG QUAN

## 1.1 Tổng quan và bối cảnh đề tài.

Internet đã đóng góp một vai trò vô cùng quan trọng trong đời sống và được xem như là một trong những phát minh vĩ đại tạo ra một bước ngoặt lớn trong lịch sử văn minh của nhân loại. Cùng với sự bùng nổ và phát triển nhanh chóng của World Wide Web, Internet trở thành nơi lưu trữ và chia sẻ tài nguyên trên toàn cầu.

Vì vậy, thông tin được lưu trữ trên Internet ngày càng nhiều và phong phú. Internet trở thành kho tri thức khổng lồ của nhân loại. Nhu cầu tìm kiếm thông tin trong vô số các thông tin được lưu trữ trên Internet là một yêu cầu hết sức cần thiết. Các hệ thống tìm kiếm hay còn được gọi là các hệ thống truy vấn thông tin đã ra đời nhằm mục đích đáp ứng nhu cầu này.

Truy vấn thông tin (Information Retrieval - IR) là việc tìm kiếm thông tin (thường là các tài liệu) ở một dạng không có cấu trúc (thông thường là văn bản) để thỏa mãn nhu cầu thông tin từ những nguồn thông tin khổng lồ được lưu trữ trên các máy tính. Truy vấn thông tin là lĩnh vực nghiên cứu nhằm tìm ra giải pháp để có được các thông tin cần thiết trong một khối lượng lớn dữ liệu. Một hệ thống tìm kiếm thông tin có hai chức năng chính: lập chỉ mục (indexing) và truy vấn (interrogation). Trong đó, lập chỉ mục là giai đoạn phân tích tài liệu (document) để xác định các chỉ mục (term/index term) biểu diễn nội dung của tài liệu, truy vấn là giai đoạn tìm các tài liệu phù hợp với từ khóa (keyword) đã được đưa vào trước đó.

Đề tài nghiên cứu của luận văn này chủ yếu tập trung vào bài toán truy vấn thông tin video – một dạng thức của hệ thống truy vấn thông tin.

Về cách thức, truy vấn thông tin video khác hẳn những hệ thống truy vấn thông tin trên tài liệu văn bản. Đối với các hệ thống truy vấn thông tin dựa trên tài liệu văn bản, các tài liệu sẽ trực tiếp đưa vào để xây dựng bộ chỉ mục. Đối với hệ thống truy

vấn thông tin video thì đánh chỉ mục dựa trên lời thoại của video và vị trí xuất hiện của từ khóa trong đó.

Những hệ thống truy vấn video phổ biến thường hỗ trợ tìm kiếm theo từ khóa mô tả được đưa vào hệ thống một cách thủ công, ví dụ như youtube [23]. Phương pháp tìm kiếm video dựa vào từ khóa này thường chưa hỗ trợ đầy đủ mong muốn tìm kiếm vì người dùng nhiều khi không thể mô tả chính xác đoạn video mà mình muốn tìm giống như từ khóa mô tả đoạn video đó đã được đưa vào hệ thống.

Hướng tiếp cận khác tốt hơn là xây dựng hệ thống hỗ trợ truy vấn video dựa vào nội dung hay còn gọi là hệ thống truy vấn thông tin video hướng ngữ nghĩa.

## **1.2 Một số hướng tiếp cận cho bài toán truy vấn video hướng ngữ nghĩa**

Một hệ thống truy vấn video lý tưởng phải kết hợp cả nội dung hình ảnh lẫn thông tin thoại của lời nói. Tuy nhiên, do những hạn chế trong lĩnh vực xử lý ảnh và thị giác máy tính, việc xây dựng hoàn chỉnh một hệ thống truy vấn lý tưởng vẫn còn gặp nhiều khó khăn. Do đó, các nghiên cứu về truy vấn video hướng ngữ nghĩa thường tập trung theo ba hướng chính:

- Truy vấn video dựa trên các đặc trưng về hình ảnh
- Truy vấn video dựa trên các đặc trưng về âm thanh.
- Các phương pháp tổng hợp kết quả truy vấn trên hình ảnh và âm thanh.

### **1.2.1 Hướng tiếp cận dựa vào đặc trưng hình ảnh**

Hệ thống truy vấn video dựa vào đặc trưng hình ảnh sẽ hỗ trợ truy vấn các sự vật, sự việc dựa trên các đặc trưng cấp thấp và đặc trưng cấp cao. Đặc trưng cấp thấp bao gồm đặc trưng về màu sắc, đặc trưng về hình dáng. Đặc trưng về màu sắc bao gồm lược đồ màu, vector liên kết màu, đặc trưng tương quan màu. Đặc trưng về hình dáng

bao gồm lược đồ hệ số góc, vector liên kết hệ số góc. Đặc trưng cấp cao bao gồm đặc trưng màu, đặc trưng vân, đặc trưng vị trí. Ưu điểm của hướng tiếp cận dựa vào đặc trưng hình ảnh giúp cho hệ thống truy vấn có ngữ nghĩa hơn. Khuyết điểm của hướng tiếp cận này là phải tổ chức cấu trúc dữ liệu phức tạp để lưu trữ cho nhiều loại đặc trưng khác nhau.

### 1.2.2 Hướng tiếp cận dựa vào đặc trưng âm thanh

Hệ thống truy vấn video dựa vào đặc trưng âm thanh sẽ hỗ trợ truy vấn video dựa trên lời thoại của các nhân vật xuất hiện trong video. Cụ thể, lời thoại trong video sẽ được chuyển sang văn bản thông qua bộ nhận dạng tiếng nói. Các văn bản này sẽ được xây dựng chỉ mục và việc truy vấn sẽ được thực hiện trên văn bản như trong các hệ thống truy vấn thông tin dựa trên tài liệu văn bản khác.

Ưu điểm của hướng tiếp cận này cũng là dễ dàng xác định chính xác từ khóa nằm ở vị trí nào trong đoạn video để đoạn video đó sẽ được phát ngay vị trí xuất hiện từ khóa. Cũng như hướng tiếp cận dựa vào đặc trưng hình ảnh, hướng tiếp cận này cũng có khuyết điểm là hiệu năng của nó còn phụ thuộc vào độ chính xác của bộ nhận dạng tiếng nói.

### 1.2.3 Hướng tiếp cận tổng hợp

Đó là sự kết hợp của hai hướng tiếp cận ở trên – hướng tiếp cận dựa vào đặc trưng hình ảnh và hướng tiếp cận dựa vào đặc trưng âm thanh. Sự kết hợp này sẽ cho ra kết quả tìm kiếm chính xác hơn, giảm được sự chênh lệch ngữ nghĩa giữa từ khóa tìm kiếm và kết quả trả về của hệ thống.

## 1.3 Hướng tiếp cận của đề tài

Tuy bài toán có ba hướng tiếp cận, và hướng tiếp cận *tổng hợp kết quả truy vấn trên hình ảnh, âm thanh* là tốt nhất, nhưng do phạm vi quá rộng của nó, nên hướng tiếp cận dựa vào đặc trưng âm thanh được chọn làm nhánh để nghiên cứu trong đề tài này.

Một cách khái quát, đề tài này hướng đến mục tiêu xây dựng hệ thống truy vấn thông tin video hướng ngữ nghĩa dựa vào đặc trưng âm thanh cụ thể là lời thoại trong video. Đề tài ứng dụng công nghệ nhận dạng tiếng nói vào việc chuyển nội dung âm thanh của video thành văn bản trích để lập dữ liệu chỉ mục cho hệ thống tìm kiếm. Cụ thể, các video ban đầu sẽ được rút trích các kênh âm thanh (audio). Bộ nhận dạng tiếng nói sẽ được xây dựng nhằm mục đích chuyển thể (hay nhận dạng) kênh audio sang dạng văn bản trích (transcription). Các transcription nhận dạng được sẽ đưa vào bộ lập chỉ mục của hệ thống truy vấn thông tin để xây dựng chỉ mục. Dữ liệu chỉ mục kết quả sẽ là cơ sở dữ liệu cho hệ thống truy vấn thông tin video.

Nhưng vì độ chính xác của nhận dạng tiếng nói chưa cao nên đề tài còn áp dụng Ontology để có thể mở rộng câu truy vấn từ đó làm tăng số lượng kết quả liên quan bù lại cho việc giảm độ chính xác.

Mỗi thể loại video có một đặc trưng riêng, ở đây, thể loại video chăn nuôi trong nông nghiệp được chọn để làm thực nghiệm vì tính ứng dụng thực tiễn của nó.

## **1.4 Các công trình nghiên cứu liên quan**

Đối với những hệ thống truy vấn sử dụng nhận dạng tiếng nói tự động (ASR), các công trình trước đây bao gồm rất nhiều hướng tiếp cận và lĩnh vực khác nhau. Sau đây sẽ trình bày một số công trình đáng chú ý, cùng với hướng tiếp cận và kết quả đạt được. Nói chung thì mọi hệ thống truy vấn dựa trên ASR đều gặp phải vấn đề sai sót trong nhận dạng, dẫn đến sai sót trong quá trình tìm kiếm.

- Ebru Arisoy và các cộng sự [1] đề xuất một hệ thống tự động nhận dạng và truy tìm video cho tin tức Thổ Nhĩ Kỳ. Hệ thống này giải quyết vấn đề có nhiều từ nằm ngoài từ điển (OOV) bằng cách sử dụng đơn vị nhận dạng phụ từ.
- Ciprian Chelba và các cộng sự [2] đưa ra một hệ thống nhằm truy vấn những bài

giảng bằng audio. Họ kết luận rằng khi sử dụng tiếng nói, dấu rằng sai số từ có cao thì vẫn làm tăng hiệu suất truy vấn.

- SpeechBot [3] là một hệ thống đánh chỉ mục và truy vấn audio đầy đủ trên nền Web. Phiên bản hiện nay có khả năng tìm kiếm trên số lượng lớn, đưa ra kết quả chấp nhận được mặc dù có sai số nhận dạng khá cao.

Một hướng tiếp cận khác cho hệ thống truy vấn sử dụng nhận dạng tiếng nói là mở rộng câu truy vấn tự động (AQE). Mở rộng câu truy vấn tự động là một quá trình đưa vào những từ đồng nghĩa hay gần với câu truy vấn, làm tăng số lượng kết quả liên quan bù lại cho việc giảm độ chính xác.

AQE có một lịch sử khá lâu đời trong truy vấn thông tin, bắt đầu từ những năm 1960. Trong những năm gần đây, nhiều phương pháp đã được đề xuất sử dụng các kỹ thuật khác nhau để tìm những từ liên quan.

- Carpineto và Giovanni [4] chỉ ra rằng kết quả tốt nhất đạt được trong AQE là từ phân tích xác suất. Những kỹ thuật này chủ yếu là dựa trên bộ ngữ liệu và phát hiện ra những mối tương quan giữa các từ bằng cách tính xác suất đồng hiện.
- Một số kỹ thuật phổ biến là từ-khái niệm (concept term) [5], gom nhóm từ [6], và từ điển tương đồng [7]. Mặc dù vậy, phân tích xác suất dựa trên bộ ngữ liệu có thể làm lệch khi những từ mở rộng không thường xuất hiện cùng với từ truy vấn. Một phương pháp khác là sử dụng Ontology tổng quát hoặc của một chủ đề cần quan tâm. Câu truy vấn được biến đổi thành những khái niệm trong hệ thống, sau đó được mở rộng dựa trên những mối quan hệ giữa các khái niệm trong Ontology. AQE sử dụng Ontology cực kỳ phụ thuộc vào chất lượng và tính xác đáng của Ontology được sử dụng.
- Giannis Varelas và các đồng nghiệp [8] cố gắng sử dụng WordNet, một từ điển đồng nghĩa được định nghĩa thủ công, để tìm sự tương đồng ngữ nghĩa giữa các

tài liệu. Kết quả của họ cho thấy AQE với từ điển đồng nghĩa có thể đạt được hiệu suất cao hơn so với mô hình không gian vector. Tuy nhiên, từ điển tự định nghĩa có thể không cung cấp một vùng phủ sóng tốt về các bộ dữ liệu và có xu hướng bị sai lệch và không thống nhất.

- Đối với Ontology cho một chủ đề, ta có nhiều nghiên cứu, tập trung vào mảng sinh học và y dược. Textpresso [9], một hệ thống truy vấn dựa trên Ontology cho sinh học, sử dụng kiến thức chuyên ngành được nhúng trong một Ontology để cải thiện hiệu suất truy vấn. Trong chủ đề y dược, MELISA [10] cho chức năng tương tự sử dụng một Ontology bao gồm 1800 khái niệm.
- AGROVOC [11] là một bộ từ điển có cấu trúc được phát triển bởi Tổ chức Lương thực và Nông nghiệp Thế giới (FAO). Nó bao gồm tất cả những vấn đề như lương thực, nông nghiệp, lâm nghiệp, nuôi trồng thủy sản, v.v... AGROVOC có hơn 32,000 khái niệm và được dịch ra 21 thứ tiếng (không có tiếng Việt). Hiện nay AGROVOC đang được sử dụng để đánh chỉ mục và tìm kiếm thông tin nông nghiệp.
- Từ điển NAL [12] là một từ điển nông nghiệp được xây dựng bởi Thư viện Nông nghiệp Quốc gia của Mỹ (USDA). Nó có trên 98,000 từ, cả tiếng Anh và tiếng Tây Ban Nha.
- Ontology for Vietnamese Language (OVL) - Open version [13] là một Ontology tổng quát (Universal Ontology) được thực hiện bởi Nguyễn Tuấn Đăng và cộng sự thuộc trường Đại học Công nghệ Thông tin. Mục tiêu tác giả xây dựng Ontology này là để đóng góp cho những nghiên cứu về xử lý ngôn ngữ tiếng Việt, xây dựng tri thức phổ quát trong nhiều lĩnh vực bằng tiếng Việt. Dữ liệu của Ontology là dữ liệu tổng quát về các lĩnh vực gồm 10 lĩnh vực chính như: Khoa học, Pháp luật, Chính trị, Kinh doanh, Thể thao, Văn hóa du lịch, Xã hội,

Vi tính, Viễn thông, Ô tô xe máy. Tuy nhiên, dữ liệu của Ontology mang tính phổ quát, không tập trung vào một lĩnh vực (domain) cụ thể.

Từ việc khảo sát các công trình nghiên cứu liên quan, tác giả tin rằng khi thu hẹp chủ đề tìm kiếm, phương pháp dựa trên Ontology sẽ có kết quả tốt vì phương pháp tính xác suất có thể làm lệch kết quả khi những từ mở rộng không thường xuất hiện cùng với từ truy vấn.

Vấn đề chính của những Ontology sẵn có là chúng không có tiếng Việt, nên chúng không thể ứng dụng trực tiếp vào ngữ cảnh Việt Nam. Hoặc Ontology tiếng Việt nhưng không tập trung vào một lĩnh vực cụ thể ví dụ như lĩnh vực chăn nuôi.

## 1.5 Mục tiêu của đề tài

Mục tiêu nghiên cứu chính của luận văn là xây dựng một hệ thống truy vấn video hướng ngữ nghĩa cho thông tin chăn nuôi trong nông nghiệp với sự hỗ trợ của hệ thống mở rộng câu truy vấn tự động từ một Ontology.

Khác biệt của nghiên cứu này so với các nghiên cứu hiện nay là chủ đề tìm kiếm chăn nuôi trong nông nghiệp và làm việc trên tiếng Việt, không giống như đại đa số các nghiên cứu trên tiếng Anh.

Để thực hiện mục tiêu đó, luận văn cần thực hiện những công việc cụ thể sau:

- Xây dựng bộ nhận dạng tiếng nói cho lĩnh vực chăn nuôi. Luận văn thu thập dữ liệu rồi huấn luyện ra một bộ nhận dạng. Với cơ sở lý thuyết và công cụ được kế thừa từ phòng AILAB thuộc trường Đại học Khoa học Tự nhiên TP. HCM.
- Xây dựng một Ontology nhánh chăn nuôi trong nông nghiệp cho tiếng Việt.
- Tích hợp một hệ thống truy vấn video nông nghiệp hướng ngữ nghĩa có sử dụng Ontology.



- Thực nghiệm đánh giá độ chính xác của bộ nhận dạng tiếng nói.
- Thực nghiệm để đánh giá hệ thống truy vấn và Ontology đã xây dựng thông qua việc so sánh kết quả truy vấn của hệ thống khi không sử dụng Ontology và khi có sử dụng Ontology.
- Chương trình demo.

## **1.6 Ý nghĩa khoa học và thực tiễn của đề tài**

### **1.6.1 Ý nghĩa khoa học**

- Thử nghiệm phương pháp ASR trong truy vấn video nông nghiệp hướng ngữ nghĩa.
- Góp phần xây dựng một Ontology nhánh chăn nuôi có thể duy trì và mở rộng.
- Đóng góp bộ dữ liệu video chăn nuôi cho nghiên cứu khoa học về sau.

### **1.6.2 Ý nghĩa thực tiễn**

- Hệ thống truy vấn thông tin video nông nghiệp được triển khai sẽ giúp ích cho người dùng trong việc tìm kiếm thông tin hữu ích liên quan đến việc chăm sóc vật nuôi.
- Bên cạnh đó khi hệ thống được triển khai sẽ có nhiều ứng dụng cho các lĩnh vực khác như: dịch vụ truy vấn dữ liệu video của đài truyền hình, search engine cho các công ty kinh doanh về nông nghiệp, v.v...

## **1.7 Bố cục của luận văn**

Luận văn được chia thành 5 chương, chương 1 đã được trình bày ở phần trước, các chương còn lại được mô tả như sau:

- Chương 2 nói về các cơ sở lý thuyết được sử dụng trong đề tài

- Chương 3 mô tả chi tiết các thành phần của hệ thống truy vấn video hướng ngữ nghĩa có sử dụng Ontology và đặc tả phần mềm demo của hệ thống.
- Chương 4 là thực nghiệm đánh giá độ chính xác của mô hình nhận dạng tiếng nói. Trình bày kết quả xây dựng Ontology chăn nuôi, phương pháp đánh giá và chương trình demo.
- Chương 5 đưa ra kết luận và hướng phát triển của đề tài.

## **CHƯƠNG 2. CƠ SỞ LÝ THUYẾT**

Chương này trình bày cơ sở lý thuyết được ứng dụng trong đề tài, bao gồm lý thuyết về hệ thống tìm kiếm thông tin trong đó có hệ thống tìm kiếm thông tin dựa trên từ khóa và hệ thống tìm kiếm thông tin dựa trên khái niệm (ngữ nghĩa), về nhận dạng tiếng nói và cuối cùng là Ontology.

### **2.1 Hệ thống tìm kiếm thông tin**

#### **2.1.1 Giới thiệu chung về hệ thống tìm kiếm thông tin**

##### **2.1.1.1 Định nghĩa về hệ thống tìm kiếm thông tin**

Kowalski định nghĩa như sau: “Hệ thống tìm kiếm thông tin là một hệ thống có khả năng lưu trữ, truy tìm và duy trì thông tin. Thông tin trong những trường hợp này có thể bao gồm văn bản, hình ảnh, âm thanh, video và những đối tượng đa phương tiện khác”. [21]

Định nghĩa khác: “Một hệ thống tìm kiếm thông tin là hệ thống có nhiệm vụ tìm kiếm trong kho dữ liệu các tài liệu liên quan đến nhu cầu người dùng (thể hiện dưới dạng câu truy vấn)”. [19]

Thuật ngữ “liên quan” ở đây được đánh giá theo 2 cách: liên quan theo người dùng và liên quan theo hệ thống. Khi hai đánh giá liên quan này trùng nhau thì chúng ta có được một hệ thống tìm kiếm lý tưởng. Hai tiêu chí cơ bản để đánh giá một hệ thống tìm kiếm thông tin là độ bao phủ và độ chính xác được trình bày trong mục 2.1.1.5.

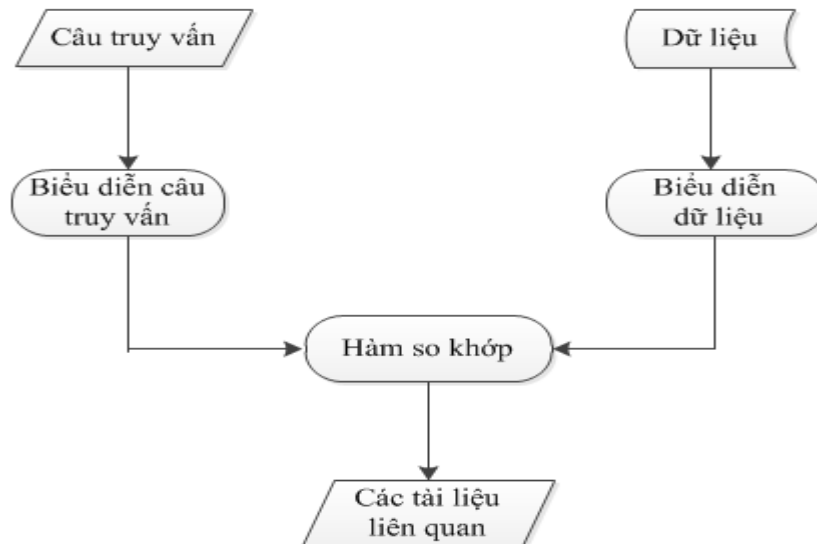
##### **2.1.1.2 Mục tiêu và chức năng của một hệ thống tìm kiếm thông tin**

Mục tiêu của một hệ thống tìm kiếm thông tin là tìm kiếm và trả về các tài liệu có liên quan (relevant) đến nhu cầu của người dùng.

Một hệ thống tìm kiếm thông tin có hai chức năng chính, đó là lập chỉ mục (indexing) và tra cứu hay tìm kiếm (interrogation). Lập chỉ mục là giai đoạn phân tích tài liệu để rút trích các đơn vị thông tin từ tài liệu và biểu diễn lại tài liệu bởi các đơn vị thông tin đó. Đơn vị thông tin có thể là từ (word), hoặc phức tạp hơn là cụm từ (phrase), khái niệm (concept). Tra cứu là giai đoạn tìm kiếm trong cơ sở dữ liệu những tài liệu phù hợp với nội dung câu truy vấn.

### 2.1.1.3 Kiến trúc chung của một hệ thống tìm kiếm thông tin

Trong giai đoạn tra cứu, nhu cầu thông tin của người sử dụng được đưa vào hệ thống dưới dạng một câu truy vấn (query) bằng ngôn ngữ tự nhiên hay một dạng thức qui ước nào đó. Câu truy vấn và tập dữ liệu sẽ được phân tích và biểu diễn thành một dạng biểu diễn bên trong. Hệ thống sẽ sử dụng một hàm so khớp (matching function) để so khớp biểu diễn của câu hỏi với các biểu diễn của các tài liệu để tìm và trả về các tài liệu có liên quan (relevance). Một hệ thống tìm kiếm thông tin có thể được biểu diễn như trong hình 2.1 [18]



**Hình 2.1. Kiến trúc chung của hệ thống tìm kiếm thông tin**

#### 2.1.1.4 Phân loại hệ thống tìm kiếm thông tin

Phân loại theo cách xây dựng từ chỉ mục: có hai cách:

- Cách thứ nhất: là tập chỉ mục được xây dựng từ tập từ hay cụm từ được rút trích từ chính nội dung của tài liệu, cách lập chỉ mục này gọi là lập chỉ mục free-text.
- Cách thứ hai: là dựa vào một cấu trúc phân lớp có sẵn, phân loại tài liệu theo một danh mục tiêu đề đề mục có sẵn. Tập chỉ mục trong cách làm này là tồn tại trước và độc lập với tài liệu, cách lập chỉ mục này gọi là lập chỉ mục controlled vocabulary.

Phân loại theo đơn vị thông tin: có hai cách:

- Hệ thống tìm kiếm thông tin dựa trên từ khóa: sử dụng từ khóa biểu diễn tài liệu và câu truy vấn. Đây là cách làm phổ biến của các hệ thống tìm kiếm trước đây.
- Hệ thống tìm kiếm thông tin dựa trên khái niệm: sử dụng khái niệm biểu diễn tài liệu và câu truy vấn.

#### 2.1.1.5 Tiêu chí để đánh giá một hệ thống tìm kiếm thông tin

Để đánh giá một hệ thống tìm kiếm thông tin, người ta sử dụng đến hai độ đo cơ bản là độ chính xác (precision) và độ bao phủ (recall). Những độ đo này đo sự thỏa mãn của người dùng với các tài liệu mà hệ thống tìm thấy. Cho  $S$  là tập các tài liệu được tìm thấy (liên quan theo hệ thống). Cho  $U$  là tập các tài liệu liên quan theo đánh giá của người dùng. Khi đó, độ chính xác và độ bao phủ sẽ được định nghĩa như sau:

*Độ chính xác:* là sự tương ứng giữa số tài liệu mà hệ thống tìm thấy có liên quan đến câu truy vấn theo người dùng trên tổng số các tài liệu tìm thấy của hệ thống [19].

$$\text{Độ chính xác} = \frac{|S \cap U|}{|S|}$$

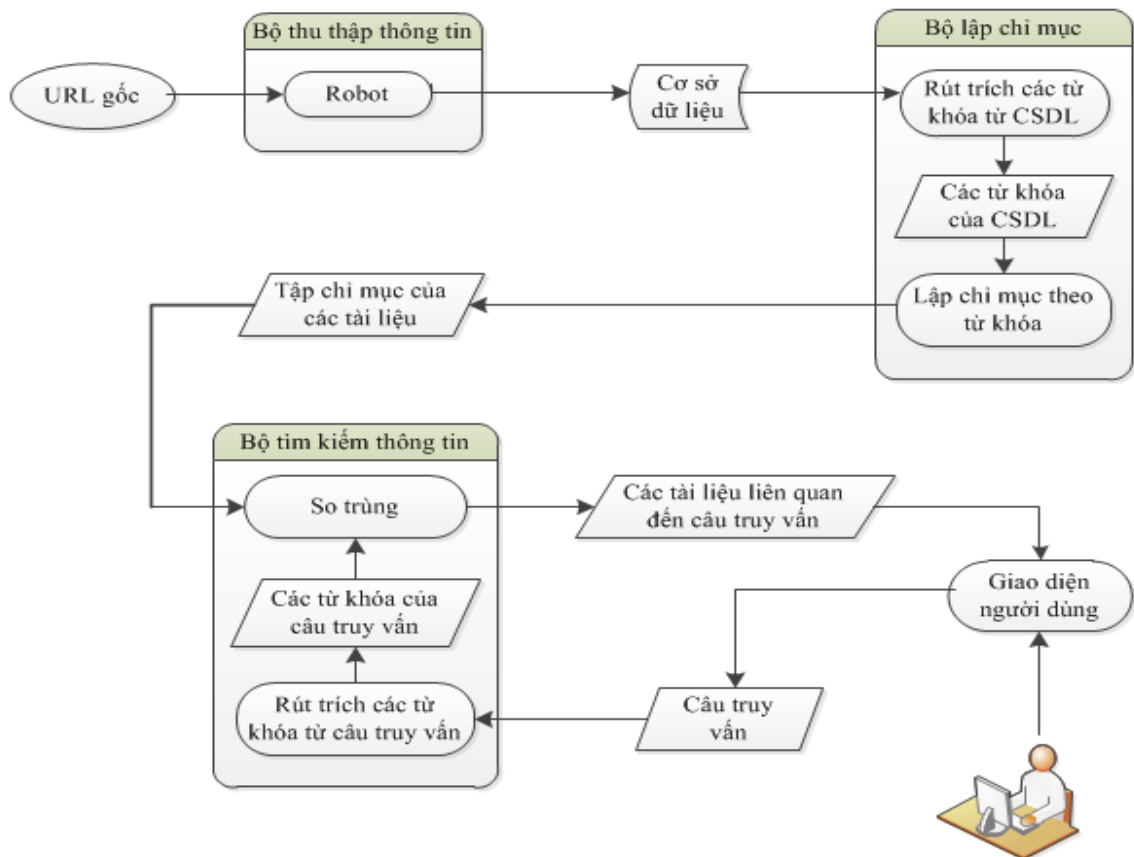
Độ chính xác 100% nghĩa là tất cả các tài liệu mà hệ thống tìm thấy đều liên quan đến câu truy vấn theo người dùng.

*Độ bao phủ:* là sự tương quan giữa số tài liệu hệ thống tìm thấy được đánh giá là liên quan theo người dùng trên tổng số các tài liệu có liên quan theo người dùng [19].

$$\text{Độ bao phủ} = \frac{|S \cap U|}{|U|}$$

Độ bao phủ là 100% có nghĩa là hệ thống tìm thấy tất cả các tài liệu liên quan. Thông thường, khó đáp ứng được cả hai độ đo này cùng một lúc.

### 2.1.2 Hệ thống tìm kiếm dựa trên từ khóa



**Hình 2.2. Hệ thống tìm kiếm dựa trên từ khóa**

Một hệ thống tìm kiếm trên Web có 3 thành phần chính: bộ thu thập thông tin, bộ lập chỉ mục, và bộ truy vấn [19]. Tính năng và chi tiết hoạt động của từng thành phần được giới thiệu trong các tiêu mục dưới đây.

#### **2.1.2.1 Bộ thu thập thông tin - Robot**

Robot là một chương trình tự động duyệt qua các cấu trúc siêu liên kết để thu thập tài liệu và nó nhận về tất cả tài liệu có liên kết với tài liệu này. Robot được biết đến dưới nhiều tên gọi khác nhau : spider, Web wanderer hoặc Web worm,... Những tên gọi này đôi khi gây nhầm lẫn, như từ ‘spider’, ‘wanderer’ làm người ta nghĩ rằng robot tự nó di chuyển và từ ‘worm’ làm người ta liên tưởng đến virus. Về bản chất robot chỉ là một chương trình duyệt và thu thập thông tin từ các site theo đúng giao thức Web. Những trình duyệt thông thường không được xem là robot do thiếu tính chủ động, chúng chỉ duyệt web khi có sự tác động của con người.

#### **2.1.2.2 Bộ lập chỉ mục - Index**

Hệ thống lập chỉ mục hay còn gọi là hệ thống phân tích và xử lý dữ liệu, thực hiện việc phân tích, trích chọn những thông tin cần thiết (thường là các từ đơn, từ ghép, cụm từ quan trọng) từ những dữ liệu mà robot thu thập được và tổ chức thành cơ sở dữ liệu riêng để có thể tìm kiếm trên đó một cách nhanh chóng, hiệu quả. Hệ thống chỉ mục là danh sách các từ khoá, chỉ rõ các từ khoá nào xuất hiện ở trang nào, địa chỉ nào.

#### **2.1.2.3 Bộ truy vấn (bộ tìm kiếm)**

Bộ phận tìm kiếm có nhiệm vụ so khớp câu truy vấn của người dùng với tập chỉ mục đã lập của các tài liệu để đánh giá độ liên quan của các tài liệu với câu truy vấn và trả về các tài liệu liên quan, được sắp xếp theo độ liên quan của nó với câu truy vấn. Động cơ tìm kiếm có thể tương tác với người dùng (user) thông qua giao diện Web, để có thể hiệu chỉnh dần kết quả trả về cho phù hợp với nhu cầu thông tin của người dùng.

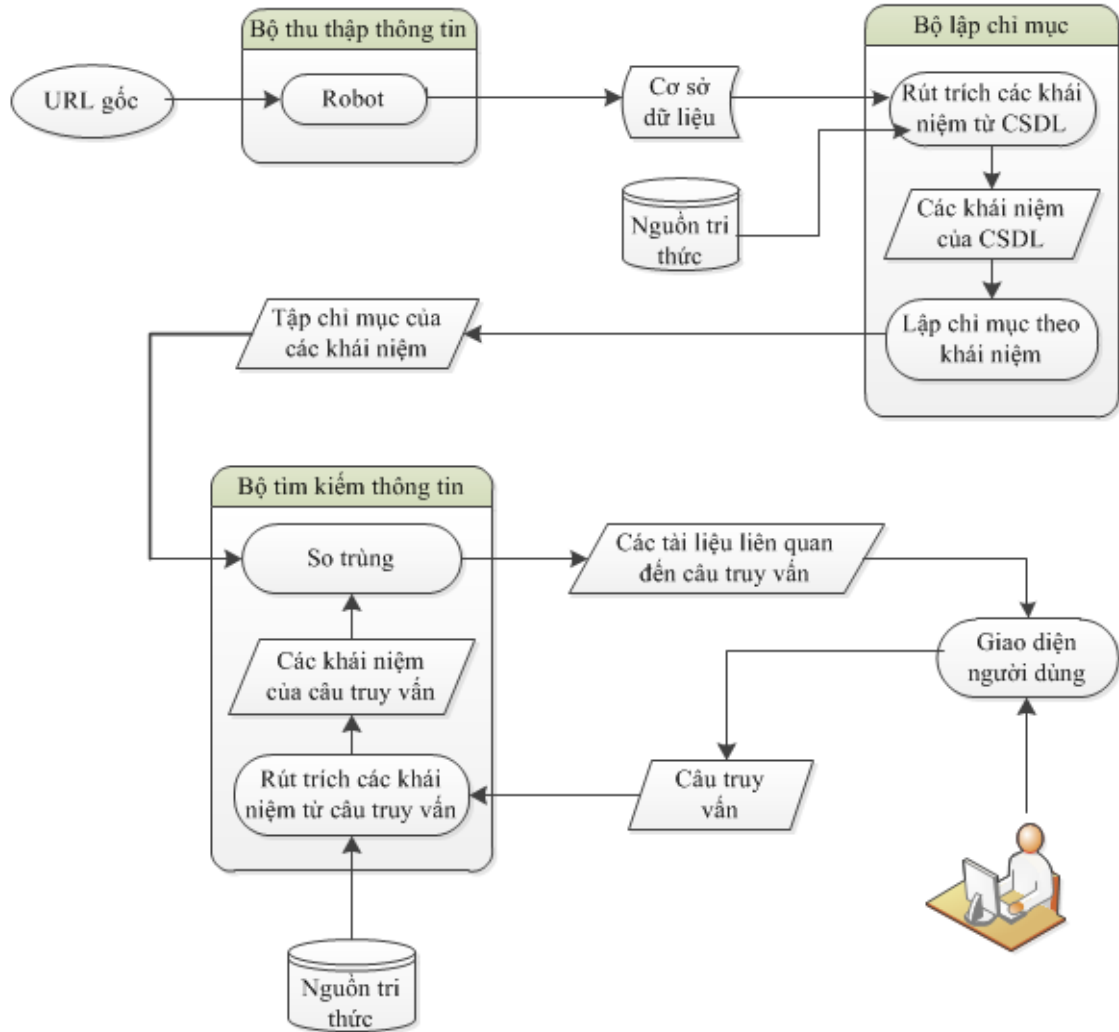
Đối với những động cơ tìm kiếm tìm theo từ khóa, tìm kiếm từ là tìm kiếm các trang mà những từ trong câu truy vấn (query) xuất hiện nhiều nhất, ngoại trừ *stopword* (các từ quá thông dụng như mạo từ a, an, the, v.v...). Một từ càng xuất hiện nhiều trong một trang thì trang đó càng được chọn để trả về cho người dùng. Và một trang chứa tất cả các từ trong câu truy vấn thì tốt hơn là một trang không chứa một hoặc chứa một số từ. Ngày nay, hầu hết các động cơ tìm kiếm đều hỗ trợ chức năng tìm cơ bản và nâng cao, tìm từ đơn, từ ghép, cụm từ, danh từ riêng, hay giới hạn phạm vi tìm kiếm như trên đề mục, tiêu đề, đoạn văn bản giới thiệu về trang Web, v.v...

### **2.1.3 Hệ thống tìm kiếm dựa trên khái niệm (hướng ngữ nghĩa).**

Trong mô hình tìm kiếm thông tin dựa trên khái niệm, nội dung của một đối tượng thông tin được mô tả bởi một tập các khái niệm [20]. Hệ thống tìm kiếm dựa trên khái niệm cũng có chức năng, nguyên lý hoạt động, và các bộ phận cấu thành như một hệ thống tìm kiếm tổng quát. Tuy nhiên, khác biệt lớn nhất giữa hệ tìm kiếm dựa trên khái niệm và hệ tìm kiếm dựa trên từ khóa ở hai điểm sau:

- Hệ tìm kiếm dựa trên từ khóa sẽ sử dụng từ khóa để lập chỉ mục, trong khi hệ tìm kiếm dựa trên khái niệm sử dụng khái niệm để lập chỉ mục.
- Để rút trích khái niệm, hệ tìm kiếm dựa trên khái niệm cần sử dụng đến nguồn tri thức về lĩnh vực nhất định nào đó.





**Hình 2.3. Hệ thống tìm kiếm dựa trên khái niệm**

Kiến trúc chung hệ thống tìm kiếm dựa trên khái niệm như hình 2.3. Nó được cấu thành từ 3 bộ phận chính, đó là bộ thu thập thông tin, bộ lập chỉ mục khái niệm, và bộ truy vấn. Chi tiết của các bộ phận này như sau:

### 2.1.3.1 Bộ thu thập thông tin

Giống bộ thu thập thông tin trong một hệ thống tìm kiếm dựa trên từ khóa. Nó có chức năng thu thập các trang web trên Internet và lưu trữ lại trong cơ sở dữ liệu. Chức

năng này được thực hiện lặp đi lặp lại thường xuyên để cập nhật những trang Web mới vào trong bộ cơ sở dữ liệu.

### 2.1.3.2 Bộ lập chỉ mục khái niệm

Điều khác biệt cơ bản nhất giữa một động cơ tìm kiếm theo khái niệm và động cơ tìm kiếm theo từ khóa nằm ở bộ phận lập chỉ mục. Đây cũng là bộ phận quan trọng nhất trong toàn bộ hệ thống. Với những động cơ tìm kiếm dựa trên từ khóa, hệ thống sẽ lập chỉ mục theo từ khóa, với những động cơ tìm kiếm dựa trên khái niệm, hệ thống sẽ lập chỉ mục theo khái niệm.

Để có bộ khái niệm, hệ thống cần thực hiện công việc rút trích toàn bộ các khái niệm trong cơ sở dữ liệu để phục vụ cho quá trình lập chỉ mục. Như vậy, trong bộ lập chỉ mục sẽ có 2 nhiệm vụ rất quan trọng là *rút trích các khái niệm* từ tập cơ sở dữ liệu và *lập chỉ mục* cho các tài liệu dựa trên các khái niệm đó.

#### Quy trình chung của rút trích khái niệm

Rút trích khái niệm là nhiệm vụ khó khăn nhất của một hệ thống tìm kiếm dựa trên khái niệm. Quá trình này gồm hai giai đoạn chính là: rút trích các từ chỉ mục trong tài liệu và so khớp các cụm từ này với nguồn tri thức.

Giai đoạn rút trích các cụm từ trong tài liệu:

- Đầu tiên, một tài liệu sẽ được đưa vào để tách thành các thành phần khác nhau như danh từ, cụm danh từ, động từ, cụm động từ, tính từ, cụm tính từ, v.v...
- Tiếp theo, hệ thống bắt đầu tạo ra các biến thể từ các thành phần đó. Giai đoạn so khớp các cụm từ này với nguồn tri thức.
- Sau khi đã có tập các biến thể, hệ thống sẽ xem xét xem những biến thể nào có trong cơ sở tri thức chứa các khái niệm thì sẽ đưa vào thành tập

ứng viên.

- Sau đó, tập ứng viên này sẽ được đánh giá và cho điểm theo những tiêu chí nhất định nào đó và sắp xếp lại theo điểm số.
- Cuối cùng là việc chọn lựa các ứng viên để đưa vào tập khái niệm. Hệ thống sẽ tìm ra những ứng viên phù hợp nhất để tạo thành khái niệm, gọi là các khái niệm được rút trích từ tài liệu.

### **2.1.3.3 Bộ truy vấn**

Cũng giống như bộ truy vấn của hệ tìm kiếm dựa trên từ khóa. Bộ truy vấn của hệ thống dựa trên khái niệm có chức năng lấy nội dung câu truy vấn do người dùng nhập vào, sau đó so trùng với tập chỉ mục đã được lập của các tài liệu để tìm ra các tài liệu liên quan đến câu truy vấn.

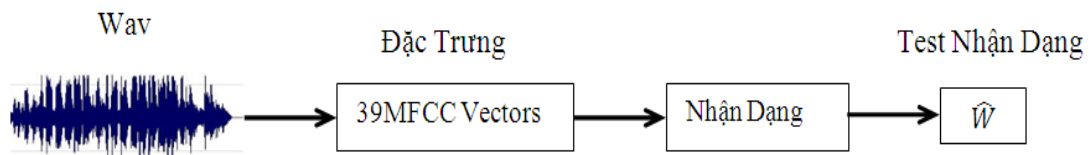
Để so trùng với tập chỉ mục đã được lập của các tài liệu, trước tiên hệ thống cần phải rút trích khái niệm từ câu truy vấn. Việc rút trích các khái niệm từ câu truy vấn tương tự như quá trình rút trích khái niệm của các tài liệu.

Tùy thuộc vào cách lập chỉ mục cho tập khái niệm như thế nào mà sẽ có những cách so trùng câu truy vấn với tập chỉ mục của tài liệu khác nhau. Nếu như bộ lập chỉ mục sử dụng các mô hình truyền thống, cách bộ truy vấn thông tin so trùng các khái niệm cũng giống như trong hệ thống tìm kiếm dựa trên từ khóa truyền thống. Nếu một cấu trúc khái niệm biểu diễn tập khái niệm của các tài liệu đã được xây dựng trong quá trình lập chỉ mục, thì cần xây dựng thêm một cấu trúc khái niệm để biểu diễn tập khái niệm của câu truy vấn. Sau đó, việc tìm kiếm mới có thể được thực hiện dựa trên việc so trùng hai cấu trúc khái niệm.

## 2.2 Nhận dạng tiếng nói

### 2.2.1 Dẫn nhập

Chương trình nhận dạng tiếng nói thực chất là một thiết bị tự động chuyển tín hiệu tiếng nói sang thành dạng văn bản mà đơn vị cơ sở của nó là từ. Các từ nhận dạng được này có thể là kết quả cuối cùng cho các ứng dụng như điều khiển hệ thống bằng giọng nói, trả lời điện thoại tự động. Chúng cũng có thể là những kết quả trung gian, đóng vai trò là đầu vào cho những chương trình xử lý ngôn ngữ tiếp theo, như dịch máy, tìm kiếm thông tin, v.v...



*Hình 2.4. Quá trình nhận dạng tiếng nói*

Trong thực tế, chương trình nhận dạng thường dựa trên một tập từ vựng hữu hạn, nhằm giới hạn các từ kết quả. Để trình bày bài toán nhận dạng tiếng nói, chúng ta cần phát biểu nó một cách hình thức thông qua những công thức toán học.

Gọi  $X$  là dữ liệu ngữ âm mà dựa vào đó trình nhận dạng sẽ quyết định chuỗi từ đã được phát âm. Để không mất tính tổng quát, chúng ta có thể giả thiết rằng  $X$  bao gồm một chuỗi  $T$  kí hiệu, lấy từ một bảng kí hiệu  $\mathbf{X}$  nào đó.

$$X = x_1, x_2, x_3, \dots, x_T, x_i \in \mathbf{X} \quad (1)$$

Kí hiệu  $x_i$  có thể được ràng buộc theo thời gian. Cụ thể, ta có thể xem như nó được sinh ra ở thời điểm  $i$  (bằng cách gán cho nó chỉ số  $i$ ).

Gọi  $W$  là một chuỗi  $N$  từ, nằm trong bộ từ vựng hữu hạn,  $\mathbf{W}$ .

$$W = w_1, w_2, w_3, \dots, w_N, w_i \in \mathbf{W} \quad (2)$$

Nếu  $P(W/X)$  định nghĩa xác suất điều kiện mà chuỗi từ bất kì  $W$  được phát âm khi tín hiệu ngữ âm quan sát được là  $X$  thì trình nhận dạng sẽ quyết định chuỗi từ nhận dạng được theo công thức 3 sau:

$$\tilde{W} = \arg \max_W P(W|X) \quad (3)$$

Điều này có nghĩa là trình nhận dạng sẽ chọn chuỗi gần giống nhất (hay nói cách khác, chuỗi từ có xác suất cao nhất), dựa trên tín hiệu quan sát được là  $X$ . Qui tắc *Bayes* cho phép ta viết lại vế phải của phương trình 3 như sau:

$$P(W|X) = \frac{P(W) \cdot P(X|W)}{P(X)} \quad (4)$$

Trong đó  $P(W)$  là xác suất mà chuỗi từ  $W$  sẽ được phát âm,  $P(X|W)$  là xác suất điều kiện mà chuỗi từ  $W$  phát âm khi tín hiệu ngữ âm quan sát được là  $X$  và  $P(X)$  là xác suất thu được tín hiệu  $X$ . Cụ thể,  $P(X)$  được biểu diễn theo phương trình 5 sau:

$$P(X) = \sum_{W'} P(W') \cdot P(X|W') \quad (5)$$

Tiến trình cực đại xác suất trong phương trình 3 độc lập với biến  $X$  nên từ các phương trình 3, 4 ta suy hàm mục tiêu của trình nhận dạng để tìm chuỗi từ là:

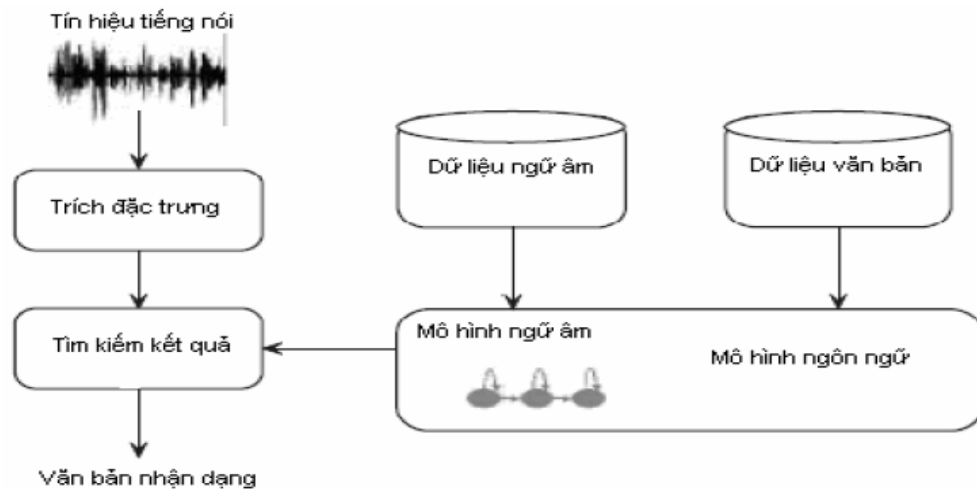
$$\tilde{W} = \arg \max_W P(W) \cdot P(X|W) \quad (6)$$

Một vấn đề đối với phương trình 6 là hai thừa số trong phép nhân có không gian rất khác nhau. Nếu chỉ thuần túy thực hiện phép nhân như trong phương trình 6 thì kết quả nhận dạng sẽ bị chi phối hoàn toàn bởi các giá trị ngữ âm hay nói cách khác, mô hình ngôn ngữ hầu như không có tác động gì đến kết quả nhận dạng. Để cân bằng cho tích xác suất này, ta sẽ thêm vào phương trình 6 một hệ số  $\gamma$ , gọi là hệ số mô hình ngôn ngữ. Khi đó, phương trình 6 có thể được viết lại như phương trình 7 sau:

$$\tilde{W} = \arg \max_W P(W)^{\gamma} \cdot P(X|W) \quad (7)$$

Hình 2.5 minh hoạ các thành phần cơ bản của một hệ thống nhận dạng tiếng nói, trong đó:

- *Mô hình ngữ âm* (Acoustic Models – AM): liên quan đến việc biểu diễn tri thức liên quan đến tín hiệu ngữ âm, âm vị, ngữ điệu, v.v...
- *Mô hình ngôn ngữ* (Language Models - LM): liên quan đến việc biểu diễn tri thức của các từ, chuỗi từ, hình thành nên câu.



**Hình 2.5. Các thành phần cơ bản của hệ thống nhận dạng tiếng nói**

Phần sau sẽ tóm tắt nội dung của từng thành phần.

### 2.2.2 Trích chọn đặc trưng

Mục tiêu của trình trích chọn đặc trưng là tham số hoá tín hiệu tiếng nói thành chuỗi vector đặc trưng, chứa thông tin liên quan đến âm thanh của câu nói. Đối với bất kì hệ thống nhận dạng tiếng nói nào, các đặc trưng ngữ âm cũng phải có các tính chất sau:

- Có khả năng phân biệt tốt để phân biệt các tiếng nói có phát âm giống nhau.
- Cho phép xây dựng các mô hình thống kê mà không đòi hỏi quá nhiều dữ liệu huấn luyện.
- Có những tính chất thống kê mang tính bất biến đối với các giọng nói khác nhau, cũng như môi trường thu âm.

Dĩ nhiên không có một đặc trưng đơn lẻ nào thoả mãn tất cả các yêu cầu trên. Do đó sử dụng một tập hợp các đặc trưng là một giải pháp. Các đặc trưng này thông thường là kết quả của việc nghiên cứu phân tích tiếng nói, mã hoá tiếng nói và ngữ âm.

Trong các hệ thống nhận dạng tiếng nói, tín hiệu tiếng nói thường được lấy mẫu ở tốc độ từ 6kHz đến 20kHz sau đó qua bước xử lý để sinh ra chuỗi vector chứa các giá trị, được gọi một cách tổng quát là tham số. Mỗi vector lưu khoảng từ 10 đến 40 tham số và được tính sau mỗi 10 tới 20 mili giây. Giá trị của các tham số trong chuỗi vector này dùng để ước lượng xác suất cho các phần tín hiệu tiếng nói hay âm vị. Hiện nay, các trình trích chọn đặc trưng trong các hệ thống nhận dạng tiếng nói thường dùng:

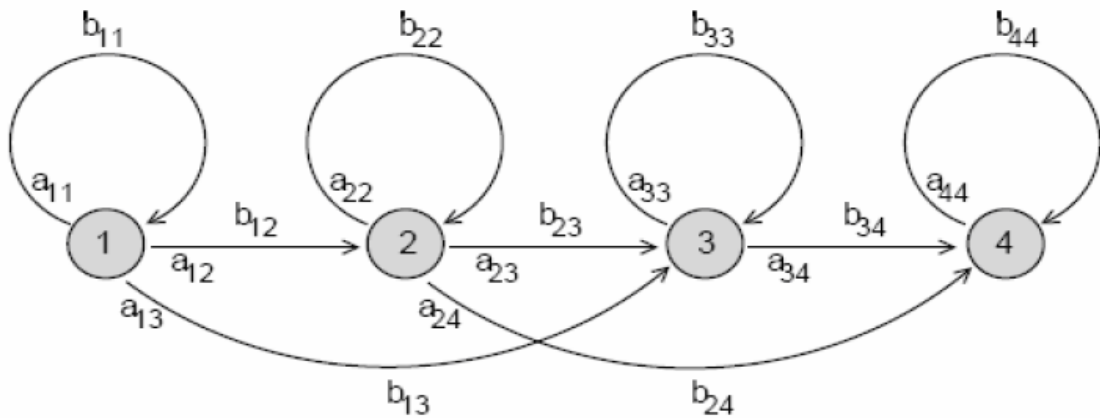
- *Đặc trưng phổ ngắn* (Short-Time Spectral Features) dựa trên phép biến đổi Fourier rời rạc hoặc mã hoá dự báo tuyến tính, áp dụng trên những đoạn tiếng nói nhỏ để rút ra các đặc trưng phổ.
- *Đặc trưng phổ có tần số biến đổi* (Frequency-Warped Spectral Features) dựa trên tỉ lệ các tần số biến đổi không đồng nhất để trích chọn ra các tần số Mel (MFCC).

Chi tiết về phép trích chọn đặc trưng được trình bày trong [15].

### 2.2.3 Mô hình ngữ âm (Acoustic Model)

Như đã đề cập đến trong phương trình 6, trình nhận dạng phải xác định giá trị xác suất  $P(X/W)$  khi phát âm chuỗi từ  $W$  với vector tham số tương ứng  $X$ . Do vậy để tính  $P(X/W)$  chúng ta cần một mô hình thống kê. Mô hình thống kê ấy trong trường

hợp này là Hidden Markov Model (HMM). HMM được cấu thành từ hai tiến trình mang tính chất thống kê ngẫu nhiên. Một chuỗi Markov ẩn nhằm giải thích các biến đổi về thời gian và một trình quan sát, giải thích các biến đổi về phổ tiếng nói ở thời điểm hiện tại. Phép kết hợp này đã được chứng minh là có đủ khả năng để giải quyết hầu hết các trường hợp nhập nhằng trong phát âm cũng như đủ mềm dẻo để áp dụng cho những hệ thống lớn với bộ từ vựng khoảng vài chục đến hàng trăm nghìn từ.



**Hình 2.6. Mô hình HMM với 4 trạng thái**

Hình 2.6 minh họa về một HMM có 4 trạng thái. Một cách hình thức, ta có thể tóm tắt các định nghĩa chính về HMM như sau. Đặt:

- $x$  là biến biểu diễn các giá trị quan sát.
- $s$  là các biến biểu diễn trạng thái của mô hình.

Khi đó HMM có thể được biểu diễn bởi các tham số sau đây:

$$A \equiv \{a_{ij}/s_i, s_j \in S\}: \text{các xác suất dịch chuyển} \quad (8)$$

$$B \equiv \{b_{ij}/s_i, s_j \in S\}: \text{các xác suất kết} \quad (9)$$



$$\pi \equiv \{\pi_{ij}\}: \text{các xác suất khởi tạo} \quad (10)$$

Trong đó:

$$a_{ij} \equiv p(s_t = s_j / s_{t-1} = s_i) \quad (11)$$

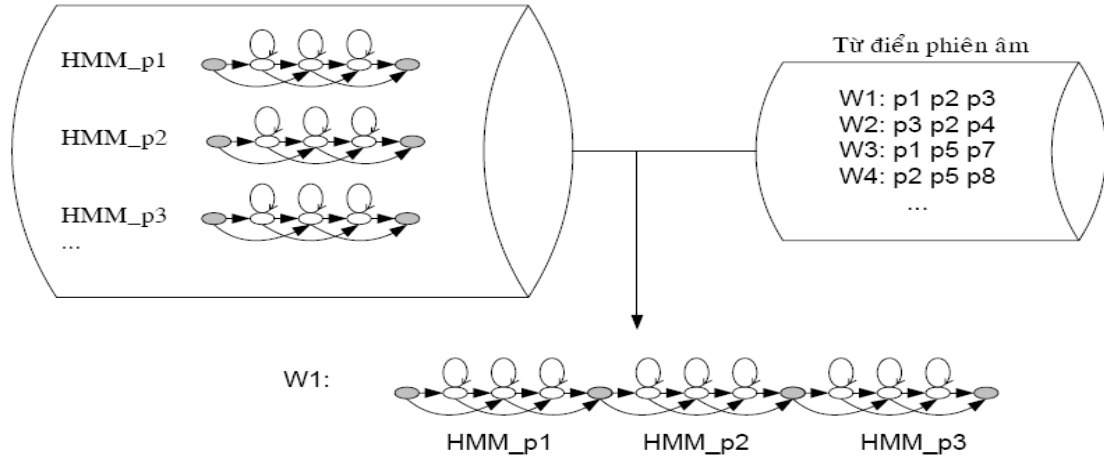
$$b_{ij}(o) \equiv p(o_t = o / s_{t-1} = s_i, s_t = s_j) \quad (12)$$

$$\pi_{ij} \equiv p(s_o = s_i) \quad (13)$$

Chi tiết về HMM được trình bày trong [14].

Thông thường đối với các hệ thống nhận dạng tiếng nói với bộ từ vựng lớn, người ta thường dựa vào một tập các đơn vị nhận dạng cơ sở - đơn vị ngữ âm. Các đơn vị này nhỏ hơn mức từ, thường được gọi là âm vị hoặc *phone*, hay *semi-syllable* hoặc *syllable*, tùy thuộc vào phép chọn. Khi đó mô hình từ sẽ được xây dựng dựa trên việc kết nối các mô hình nhỏ hơn này, dựa trên từ điển phát âm và cấu trúc tương ứng của từ. Trong hầu hết các hệ thống nhận dạng tiếng nói, các đơn vị ngữ âm cơ sở được mô hình bởi HMM.

Thông thường việc chọn lựa cấu trúc cho HMM, các hàm phân bố xác suất do người phát triển hệ thống quyết định. Tuy nhiên toàn bộ các giá trị tham số sau đó sẽ được ước lượng một cách tự động. Phép huấn luyện (training) trong nhận dạng tiếng nói bản chất là phép cực đại xác suất theo tham số mô hình. Trong trường hợp tổng quát, các tham số của mô hình thống kê được ước lượng theo thuật toán lặp sao cho sau mỗi bước, xác suất trên tập dữ liệu huấn luyện luôn tăng cho đến khi vượt ngưỡng cho trước thì dừng.



**Hình 2.7. Minh họa tiến trình xây dựng mô hình từ dựa trên các phone.**

### 2.2.4 Mô hình ngôn ngữ (LM)

Phương trình 6 đòi hỏi việc tính xác suất  $P(W)$  cho mỗi chuỗi từ  $W$ . Trong lĩnh vực nhận dạng tiếng nói,  $P(W)$  được gọi là mô hình ngôn ngữ. Đúng như tên gọi của nó, LM cung cấp thông tin về cú pháp, ngữ nghĩa và phỏng đoán qua giá trị xác suất  $P(W)$  đối với chuỗi từ  $W$ . Xác suất  $P(W)$  có thể được biểu diễn bởi:

$$P(W) = \prod_{i=1}^N P(w_i | w_1, \dots, w_{i-1}) \quad (14)$$

$$= \prod_{i=1}^N P(w_i | h_i) \quad (15)$$

$$\cong \prod_{i=1}^N P(w_i | w_{i-n+1}^{i-1}) \quad (16)$$

Trong đó  $h_i = w_1, \dots, w_{i-1}$  là chuỗi từ đứng ngay trước  $w_i$  hay còn gọi là ngữ cảnh của  $w_i$ . Mức độ khó trong việc ước lượng các xác suất  $P(w_i | h_i)$  tỉ lệ với chiều dài của chuỗi từ  $h_i$ . Do đó người ta thường giả thiết rằng chuỗi từ này tuân theo tiến trình Markov bậc  $(n - 1)$  như trong phương trình 16. Các LM tương ứng trong trường hợp

này được gọi là n-gram LM. Hiện nay,  $n = 2$ ,  $n = 3$  tương ứng với *bigram* LM và *trigram* LM được sử dụng trong hầu hết các hệ thống nhận dạng tiếng nói. Phần sau trình bày một số chi tiết về trigram LM. Trong trường hợp  $n = 2$ , xác suất của chuỗi từ trong phương trình 14 sẽ có dạng:

$$P(W) \approx \prod_{i=1}^N P(w_i | w_{i-2}, w_{i-1}) \quad (17)$$

Để ước lượng các xác suất trigram, ta sử dụng một tập lớn dữ liệu văn bản huấn luyện (training corpus) để ước lượng các tần số trigram:

$$f_3(w_3 | w_1, w_2) = \frac{c_{123}}{c_{12}} \quad (18)$$

Trong đó  $c_{123}$ ,  $c_{12}$  là số lần xuất hiện của cụm từ  $\{w_1, w_2, w_3\}$  và  $\{w_1, w_2\}$  trong tập dữ liệu huấn luyện. Đối với tập từ vựng có kích thước  $V$ , sẽ có tất cả  $V^3$  các trigram khác nhau. Do đó sẽ có rất nhiều những trigram trong số này không xuất hiện trong tập dữ liệu huấn luyện; cũng có nghĩa là chúng sẽ có giá trị xác suất bằng không. Giá trị xác suất bằng không sẽ làm cho tích xác suất trong phương trình 6 bằng không, dẫn đến những kết quả nhận dạng không mong muốn. Để giải quyết vấn đề này, ta cần một phương pháp ước lượng “tron” cho xác suất của những cụm từ không xuất hiện. Một trong số những phương pháp đó là phép nội suy tuyến tính cho các tần số của unigram, bigram, trigram và một hàm phân bố đồng nhất trên tập từ vựng, như được mô tả trong phương trình 19 sau:

$$P(w_3 | w_1, w_2) = \lambda_3 f_3(w_3 | w_1, w_2) + \lambda_2 f_2(w_2 | w_1) + \lambda_1 f_1(w_1) + \lambda_0 \frac{1}{V} \quad (19)$$

Trong đó  $f_1(\cdot)$ ,  $f_2(\cdot)$  là giá trị tần số của unigram và bigram tương ứng và các hệ số  $\lambda_i$  được xác định bằng thực nghiệm. Ngoài phép nội suy nói trên, còn một số phương pháp khác cũng thường được sử dụng như *back-off* LM, *Good-Turing Discounting* LM.

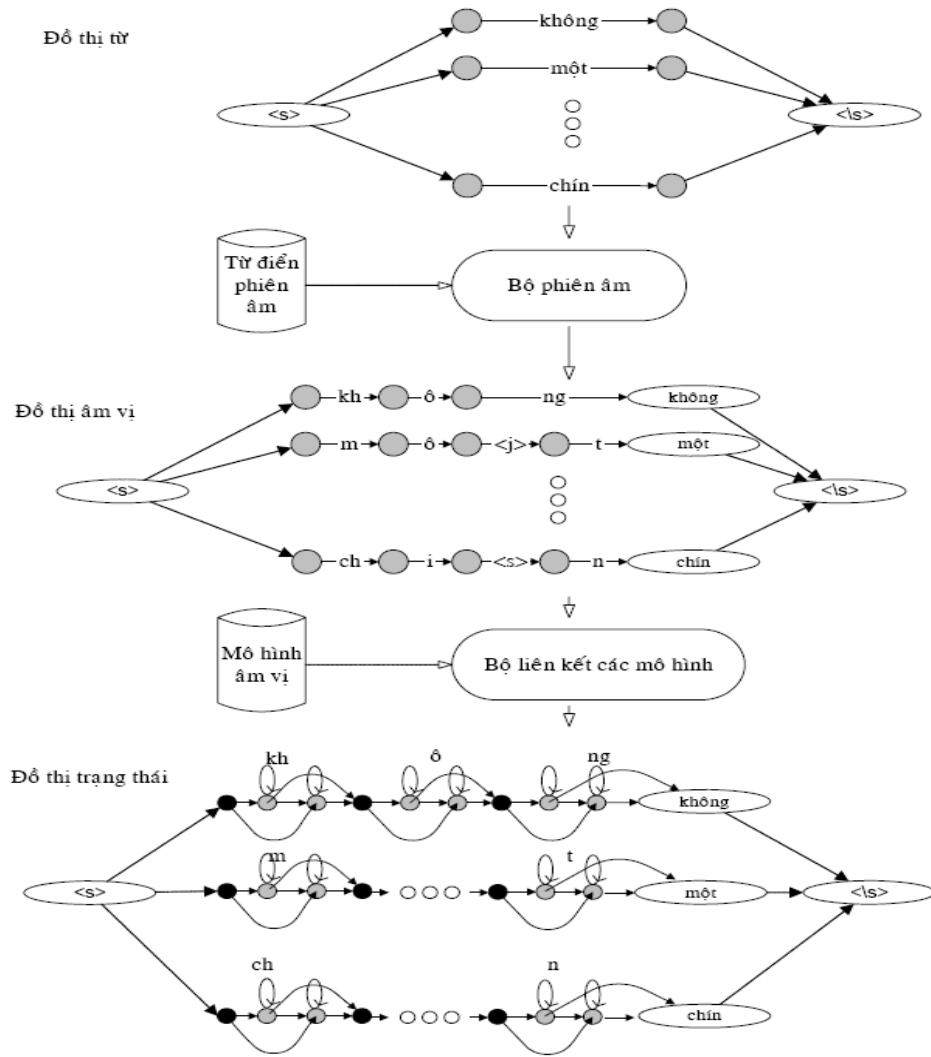
### 2.2.5 Phép tìm kiếm

Việc quyết định chọn lựa chuỗi từ ứng với tín hiệu ngữ âm về mặt bản chất là việc tìm kiếm tối ưu trên đồ thị được xây dựng bằng cách kết-ghép các mô hình ngôn ngữ, mô hình ngữ âm và từ điển phát âm. Ở mức chi tiết nhất, đồ thị là phép kết ghép của những HMM như minh hoạ trong hình 2.8. Từ hình ta thấy, việc xây dựng đồ thị kết-ghép cho tiến trình nhận dạng bao gồm ba bước:

- **Bước 1:** LM được biểu diễn theo đồ thị trong đó các cạnh sẽ được gán nhãn là các từ tương ứng. Chúng cũng có thể được gán nhãn là NULL, với một giá trị xác suất nào đó.
- **Bước 2:** mỗi cạnh trong đồ thị LM sẽ được thay thế bằng một chuỗi hoặc thậm chí một đồ thị trong đó các cạnh của đồ thị mới này sẽ được gán nhãn là các phone hay syllable, tương ứng với tập các qui tắc trong từ điển phát âm.
- **Bước 3:** là bước cuối cùng trong đó mỗi cạnh được gán nhãn phone từ bước 2 sẽ được thay thế bằng tập các HMM tương ứng.

Kết quả cuối cùng chính là đồ thị kết ghép như minh hoạ trong hình 2.8. Với kiến trúc đồ thị này, tất cả các tri thức về ngữ âm, về ngôn ngữ, về thống kê đều được tích hợp trong đó. Bài toán nhận dạng tiếng nói trở thành bài toán tổ chức và tìm kiếm trên đồ thị. Các thách thức ở đây bao gồm:

- *Kích thước:* đồ thị cần được tổ chức và tối ưu sao cho có kích thước càng nhỏ (thông qua số đỉnh và số cạnh) càng tốt. Tuy nhiên vẫn phải đảm bảo tính đầy đủ cho toàn bộ thông tin tri thức được tích hợp trong đó.
- *Độ chính xác:* phép tìm kiếm trên đồ thị tuân theo ràng buộc cơ bản, đó là phải đảm bảo sao cho kết quả tìm kiếm là gần giống với chuỗi từ đã được phát âm nhất.



**Hình 2.8. Tiến trình xây dựng đồ thị kết ghép từ LM, AC và từ điển phát âm**

**2.2.5.1 Đánh giá kết quả nhận dạng**

Chuẩn đánh giá được sử dụng trong hầu hết các hệ thống nhận dạng tiếng nói là tỉ lệ lỗi từ (*Word Error Rate – WER*). *WER* dựa trên sự khác biệt giữa chuỗi từ sinh ra từ chương trình nhận dạng so với chuỗi từ gốc ban đầu. Việc tính *WER* được thực hiện qua hai bước. Bước thứ nhất sẽ tính khoảng cách khác biệt nhỏ nhất (*minimum edit*

*distance*) giữa chuỗi từ nhận dạng được và chuỗi từ gốc. Kết quả của phép tính này trả về giá trị nhỏ nhất của số từ thay thế (*substitutions*), số từ thêm (*insertions*) và số từ xoá (*deletions*) cần thiết để biến đổi chuỗi từ nhận dạng được về chuỗi từ ban đầu. Sau đó, *WER* được định nghĩa như sau:

$$WER = 100. \frac{\text{Insertions} + \text{Substitutions} + \text{Deletions}}{N_c} \quad (20)$$

Trong đó  $N_c$  là tổng số từ trong chuỗi từ ban đầu.

Ngược lại với tỉ lệ lỗi từ là tỉ lệ từ chính xác (*Word Accuracy Rate*). Giá trị *WAR* thường được sử dụng trong các hệ thống nhận dạng tiếng Trung Quốc. Gần giống với tỷ lệ từ chính xác là tỷ lệ từ đúng (*Word Correct Rate*) nhưng *WCR* không quan tâm đến từ thêm vào (*insertion*) trong khi tính khoảng cách khác biệt nhỏ nhất. Kết quả nhận dạng trong luận văn này dựa trên *WAR*.

### 2.2.5.2 Đánh giá mô hình ngôn ngữ

Việc so sánh hiệu quả giữa các mô hình ngôn ngữ khác nhau là cần thiết trong quá trình xây dựng hệ thống nhận dạng tiếng nói. Cách đơn giản nhất là áp dụng chúng trên cùng một trình nhận dạng. Mô hình ngôn ngữ tốt là mô hình ngôn ngữ cho kết quả nhận dạng với độ chính xác cao. Đây được xem như cách tốt nhất để đánh giá tính hiệu quả của một LM. Tuy nhiên chi phí cho phép đánh giá này là quá lớn. Một cách tiếp cận khác trong việc đánh giá tính hiệu quả của LM là dựa vào lí thuyết thông tin mà cụ thể là độ đo entropy. Ý tưởng chính của phép đánh giá này là dựa trên giá trị trung bình của logarit của xác suất trên mỗi từ, nằm trong đoạn văn bản mới, không tham gia vào quá trình huấn luyện.

Kí hiệu  $p$  là phân bố xác suất chính xác của đoạn văn bản  $x$  chứa  $k$  từ. Khi đó độ đo *entropy* trên mỗi từ được định nghĩa là:

$$H = \lim_{n \rightarrow \infty} -\frac{1}{k} \sum_x p(x) \cdot \log_2 p(x) \quad (21)$$

Để xác định xác suất của đoạn văn bản này, ta sẽ sử dụng LM của mình, kí hiệu  $p \sim$ , khác với phân bố  $p$ . Trước tiên, ta có thể tính giá trị logarit trung bình của xác suất cho mỗi từ như sau:

$$lp_k = \lim_{n \rightarrow \infty} -\frac{1}{k} \sum_{i=1}^k p(x) \cdot \log_2 p(x) \quad (22)$$

Ta có thể chứng minh  $\lim_{k \rightarrow \infty} lp_k = lp \geq H$  rằng, nghĩa là giá trị logarit trung bình của xác suất không nhỏ hơn độ đo entropy của dữ liệu test. Hiển nhiên, mục tiêu của ta là tìm LM sao cho giá trị logarit trung bình của xác suất càng gần với độ đo entropy của dữ liệu văn bản càng tốt.

Một phép đo liên quan đến giá trị logarit trung bình của xác suất, gọi là độ rối *perplexity*, định nghĩa bởi:  $2^{lp}$  rất thường được dùng để đánh giá tính hiệu quả của các LM. Theo định nghĩa ta có thể thấy: giá trị *perplexity* càng nhỏ thì LM càng có khả năng dự báo chính xác từ sẽ xuất hiện. Các LM cho tiếng Việt trong luận văn này cũng được đánh giá bằng độ đo *perplexity*.

## 2.2.6 Tổ chức đồ thị tìm kiếm kết quả nhận dạng

Mục này trình bày sự hình thành các đồ thị tìm kiếm dựa trên các mô hình ngữ âm và ngôn ngữ và các thuật toán tìm kiếm kết quả trên đó.

### 2.2.6.1 Tổ chức đồ thị tìm kiếm

Xem xét lại hình 2.8, gần giống như đồ thị hai phía, đồ thị từ đơn giản hơn. Đồ thị gồm  $2 \cdot NW + 2$  đỉnh với  $NW$  là số từ vựng. Trong đó, có 1 đỉnh phát  $\langle s \rangle$ , 1 đỉnh thu  $\langle /s \rangle$ , đỉnh phát nối tới  $NW$  đỉnh có chỉ số từ 1 tới  $NW$ , mỗi đỉnh  $i$  (từ 1 tới  $NW$ ) sẽ có cung nối tới đỉnh  $(i + NW)$  và cung  $(i, i + NW)$  lưu thông tin từ thứ  $i$ . Các đỉnh  $(NW + 1)$

tới (NW + NW) sẽ có cung nối tới đỉnh thu </s>. Việc đặt từ trên cạnh là dựa trên tính chất có chiều thời gian trong từ. Điều này sẽ thuận tiện cho việc giãn đồ thị ra (chi tiết hóa đồ thị).

Khi chuyển sang đồ thị âm thì âm vị là quan trọng, từ trở thành là điểm dừng. Vì vậy, đồ thị âm sẽ được giãn ra từ đồ thị từ bằng việc thay đồ thị âm vị của từng từ vào cung của đồ thị đó hình 2.8 thể hiện đồ thị âm vị của từ “không”.



**Hình 2.9. Đồ thị âm vị của từ không**

Đồ thị âm vị của một từ gồm  $N_{av} + 1$  đỉnh. Với  $N_{av}$  là số âm vị cấu thành từ đó và một đỉnh cuối là đỉnh thể hiện đã tích lũy được từ đó. Chiều thời gian được chuyển qua cho các âm vị, "từ" trở thành đỉnh kết thúc một giai đoạn. Các đỉnh được nối lại thành một chuỗi và âm vị được gán lên cạnh như Hình 2.9. Và được ghép vào đồ thị âm vị như hình 2.8.

Tương tự như giãn đồ thị "từ" sang đồ thị âm vị, đồ thị trạng thái được xây dựng bằng cách thay đồ thị trạng thái của từng âm vị vào cạnh âm vị tương ứng. Đồ thị âm vị của "từ" phụ thuộc vào mô hình ngữ âm, nghĩa là dựa trên cách phiên âm của từ đó. Đồ thị trạng thái của âm vị sẽ dựa trên mô hình markov ẩn. Với hình 2.8, mô hình Markov được chọn là mô hình trái-phải và có 4 trạng thái.

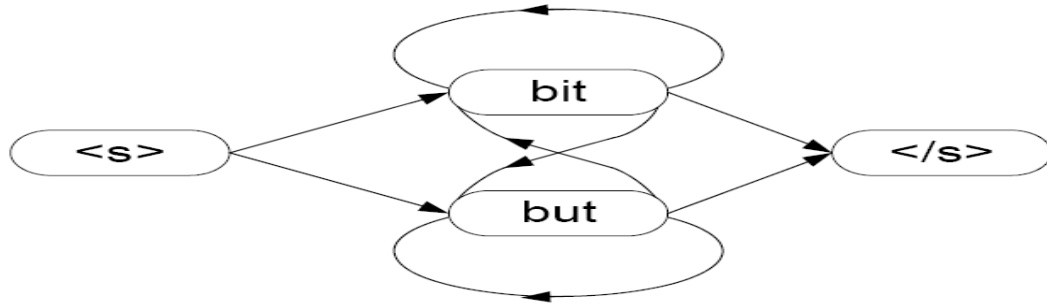
### 2.2.6.2 Đồ thị tìm kiếm với uni-gram và bi-gram

Việc chọn mô hình ngôn ngữ cho hệ thống sẽ quyết định đồ thị “từ” và có nghĩa là quyết định khung của đồ thị tìm kiếm kết quả nhận dạng.



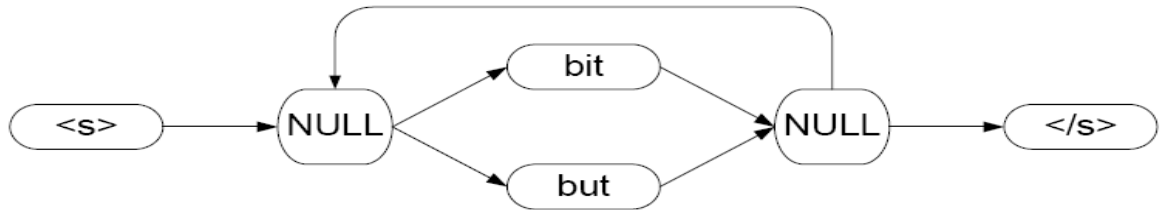
### 2.2.6.3 Đồ thị tìm kiếm dựa trên mô hình uni-gram

Ý tưởng của uni-gram là một câu gồm có nhiều từ đơn liên kết lại.



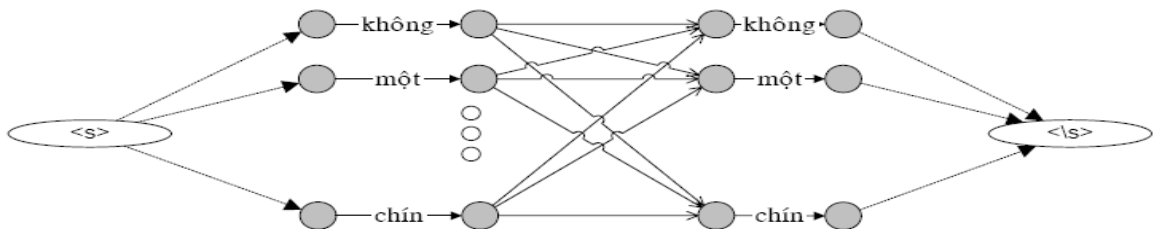
*Hình 2.10. Đồ thị "từ" xây dựng trên mô hình uni-gram*

Để giảm độ phức tạp cung nối cho đồ thị, 2 đỉnh NULL được thêm vào làm trung gian để quay trở lại đỉnh khác trước khi đi tới đỉnh kết thúc. Hình 2.11 thể hiện điều này.



*Hình 2.11. Đồ thị "từ" xây dựng trên mô hình uni-gram có sử dụng đỉnh NULL*

### 2.2.6.4 Đồ thị tìm kiếm dựa trên mô hình bi-gram

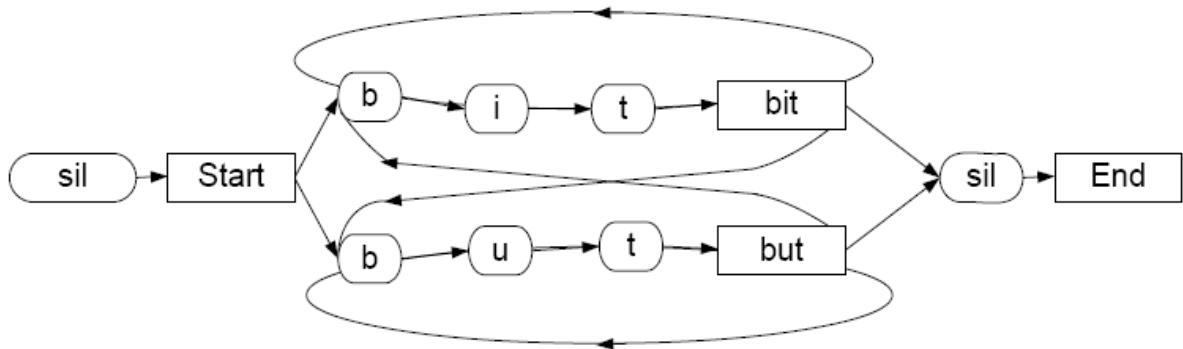


*Hình 2.12. Đồ thị "từ" xây dựng trên mô hình bi-gram*

Hình 2.12 thể hiện ý tưởng của mô hình bi-gram. Mô hình xây dựng dựa trên thông tin ngữ cảnh có xét "từ" đứng trước đó là xác suất  $P(W2 | W1)$ , xác suất  $W2$  biết trước  $W1$  đã xuất hiện trước đó. Tuy nhiên, không phải lúc nào thông tin  $P(W2|W1)$  cũng tốt. Chính vì vậy, nếu  $P(W2 | W1)$  quá nhỏ thì nên dùng lại  $P(W2)$ .

### 2.2.7 Đồ thị tìm kiếm với ngữ cảnh một âm vị (monophone), ngữ cảnh có xét âm vị trái phải chỉ trong từ (triphone within-word) và ngữ cảnh có xét âm vị trái phải trong câu nói (triphones cross-word)

Quay lại bài toán giãn đồ thị, xét bước chuyển từ đồ thị "từ" sang đồ thị "âm vị", nếu chỉ thay đồ thị âm vị của "từ" một cách thuần túy, đồ thị âm vị sinh ra là đồ thị *đơn âm vị* (monophone). Hình 2.13 là đồ thị âm vị được chuyển từ đồ thị "từ" trong hình 2.10.

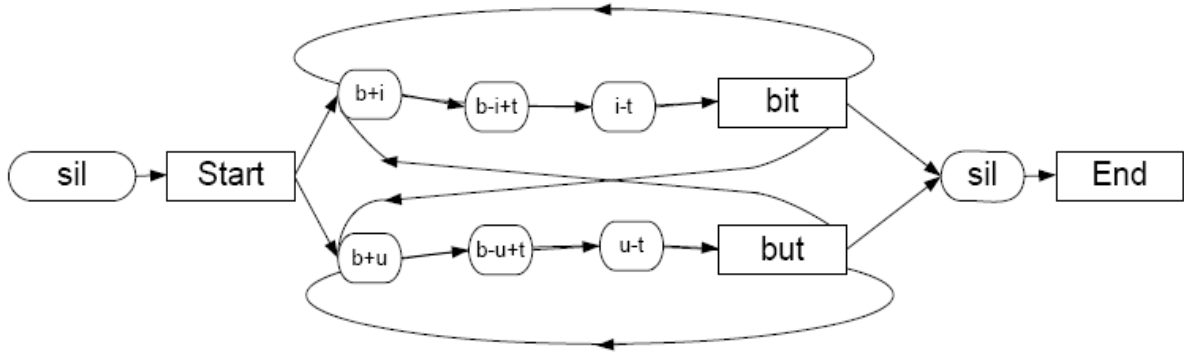


**Hình 2.13. Đồ thị đơn âm vị**

Khi làm việc trên dữ liệu tiếng nói liên tục, âm vị của các từ đều có thay đổi vì ảnh hưởng bởi các âm lân cận. Đặc biệt là tiếng Việt, một ngôn ngữ hữu thanh. Các âm vị có thanh đi chung sẽ thay đổi rất nhiều. Chính vì điều này, các mô hình âm vị có xét đến ngữ cảnh được đề xuất.

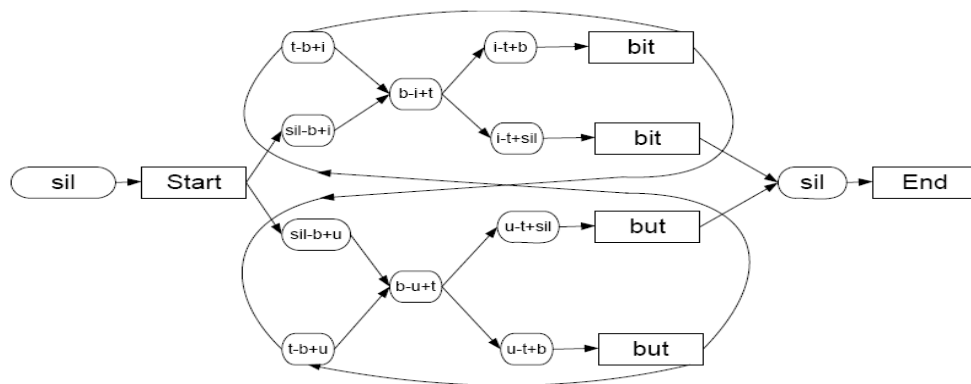
Có hai hướng tiếp cận khi giải quyết vấn đề này. Hướng thứ nhất là phiên âm trực tiếp các âm vị có liên quan đến ngữ cảnh vào trong từ điển. Hướng tiếp cận thứ hai

là dùng mô hình ngữ âm đơn giản nhưng khi chuyển qua đồ thị âm vị, thì tạo ra âm có xét ngữ cảnh. Luận văn đi theo hướng tiếp cận thứ hai. Đồ thị trong hình 2.14 diễn tả đồ thị trong hình 2.13 nhưng có xét đến ngữ cảnh.



**Hình 2.14. Đồ thị tri-gram within-word**

Những trường hợp đầu từ và cuối từ sẽ thiếu đi một ngữ cảnh trái hoặc phải. Trường hợp này là trường hợp biphones. Còn các trường hợp đủ ngữ cảnh trái phải thì được gọi là triphones. Tuy nhiên, trong dữ liệu tiếng nói liên tục, giữa 2 từ không đảm bảo luôn có khoảng lặng, hoặc khoảng lặng này rất ngắn và việc âm vị cuối của từ trước ảnh hưởng lên âm vị đầu của từ sau là hoàn toàn có. Trường hợp đồ thị trong hình 2.14 là trường hợp tri-phones within-word. Hình 2.15 sẽ diễn tả trường hợp đồ thị tri-phones crossword.



**Hình 2.15. Đồ thị tri-phone cross-word**

Rõ ràng đồ thị tri-phone cross-word thể hiện mô hình ngữ âm chính xác hơn. Tuy nhiên, đồ thị khá phức tạp và tốc độ nhận dạng lại chậm hơn các mô hình khác.

## 2.3 Ontology

### 2.3.1 Khái niệm

Ontology là tập từ vựng để mô hình hóa thế giới bên ngoài, nó đưa ra các khái niệm cơ bản và định nghĩa quan hệ giữa các khái niệm đó trong một miền lĩnh vực. Đồng thời Ontology còn cung cấp các ràng buộc, là các giả định cơ sở về ý nghĩa mong muốn của bộ từ vựng.

### 2.3.2 Mục đích xây dựng Ontology

Ontology được xây dựng nhằm các mục đích sau:

- Để chia sẻ kiến thức chung giữa con người hoặc những tác tử phần mềm với nhau.
- Cho phép tái sử dụng kiến thức về một lĩnh vực.
- Làm rõ ràng những giả định thuộc chuyên ngành.
- Có thể phân tích và suy luận kiến thức chuyên ngành vì những thuật ngữ, khái niệm cũng như các mối quan hệ giữa chúng đều được khai báo, đặc tả trong Ontology với cấu trúc có thể suy luận được theo ngữ nghĩa. Cụ thể là do các khái niệm được lưu dưới cấu trúc cây phân cấp, tên của khái niệm và quan hệ là những từ và cụm từ có nghĩa biểu diễn cho những phát biểu có nghĩa.

### 2.3.3 Yêu cầu khi xây dựng Ontology

Ngôn ngữ Ontology cho phép người sử dụng viết rõ ràng, các khái niệm hình thức của mô hình miền. Các yêu cầu chính:

- *Cấu trúc rõ ràng*: đây là điều kiện cần cho máy có thể xử lý thông tin.

- *Ngữ nghĩa hình thức miêu tả ý nghĩa tri thức một cách chính xác:* Ý nghĩa của ngữ nghĩa hình thức tồn tại trong một thời gian dài trong miền toán logic. Việc sử dụng ngữ nghĩa hình thức cho phép con người suy diễn tri thức. Với tri thức trong Ontology chúng ta có thể suy diễn về:
  - **Thành viên của lớp:** Nếu  $x$  là một thể hiện của lớp  $C$  và  $C$  là lớp con của lớp  $D$  thì chúng ta có  $x$  là thể hiện của lớp  $D$ .
  - **Các lớp tương đương:** Nếu lớp  $A$  tương đương với lớp  $B$  và lớp  $B$  tương đương với lớp  $C$ , thì lớp  $A$  cũng tương đương với lớp  $C$ .
  - **Tính nhất quán:** Giả sử chúng ta khai báo  $x$  là thể hiện của lớp  $A$  và  $A$  là lớp con của  $B \cap C$ ,  $A$  là lớp con của lớp  $D$ , Lớp  $B$  và lớp  $D$  không có quan hệ với nhau (disjoint). Thì chúng ta không nhất quán bởi vì  $A$  nên là rỗng nhưng lại có thể hiện là  $x$ . Đây là một dấu hiệu của một lỗi trong Ontology.
  - **Phân loại** : nếu chúng ta khai báo các cặp thuộc tính giá trị đã biết là điều kiện đủ cho thành viên trong một lớp  $A$ , thì nếu một cá thể  $x$  thỏa mãn các điều kiện, chúng ta có thể kết luận  $x$  phải là một thể hiện của  $A$ .
- *Ngữ nghĩa là điều kiện tiên quyết cho việc hỗ trợ suy diễn:* Hỗ trợ suy diễn rất quan trọng bởi vì nó cho phép kiểm tra tính nhất quán của Ontology và tri thức, kiểm tra các quan hệ thừa giữa các lớp, tự động phân loại các thể hiện trong lớp.

### 2.3.4 Các thành phần của Ontology

- Các lớp (Classes) - Khái niệm
- Các cá thể (Individuals)
- Các thuộc tính (Properties)
- Các mối quan hệ (Relations)

Ontology thường phân biệt các nhóm quan hệ khác nhau. Vd:

- Quan hệ giữa các lớp
- Quan hệ giữa các thực thể
- Quan hệ giữa một thực thể và một lớp
- Quan hệ giữa một đối tượng đơn và một tập hợp
- Quan hệ giữa các tập hợp.

### 2.3.5 Phương pháp xây dựng Ontology

Hiện nay không có phương pháp chuẩn nào cho việc xây dựng một Ontology. Khi xây dựng Ontology chúng ta nên dựa vào nhu cầu của ứng dụng sẽ sử dụng nó để thiết kế cho phù hợp.

Quá trình xây dựng một Ontology là một quá trình lặp, thường bắt đầu bằng một phiên bản thô rồi sao đó xem xét, chỉnh sửa, lọc lại Ontology phiên bản trước và thêm vào các chi tiết.

Những khái niệm trong Ontology là những đối tượng thực tế hoặc logic phản ánh thế giới thực và những quan hệ trong Ontology thường là những động từ trong câu mô tả khái niệm trong lĩnh vực.

Theo tài liệu [16] thì phương pháp xây dựng Ontology gồm các bước :

- Bước 1: Xác định lĩnh vực và phạm vi của Ontology
- Bước 2: Xem xét việc sử dụng lại các Ontology có sẵn
- Bước 3: Liệt kê các thuật ngữ quan trọng
- Bước 4: Xác định các lớp và phân cấp của các lớp
- Bước 5: Xác định các thuộc tính
- Bước 6: Xác định ràng buộc của các thuộc tính
- Bước 7: Tạo các thể hiện / thực thể

## **CHƯƠNG 3. HỆ THỐNG TRUY VẤN VIDEO NÔNG NGHIỆP HƯỚNG NGỮ NGHĨA CÓ SỬ DỤNG ONTOLOGY**

Chương này giới thiệu về hệ thống truy vấn video hướng ngữ nghĩa dựa trên công nghệ nhận dạng tiếng nói có sử dụng Ontology. Tính ngữ nghĩa của hệ thống được thể hiện qua việc sử dụng từ, cụm từ nhận dạng được từ kênh âm thanh để lập chỉ mục tìm kiếm cho các đoạn video. Dựa vào bộ chỉ mục đã lập, khi cần truy vấn video bằng từ khoá, đối tượng Ontology được tích hợp trong máy tìm kiếm sẽ mở rộng ra các câu truy vấn có liên quan đến từ khóa trong câu truy vấn của người dùng. Với mỗi câu truy vấn con này, máy tìm kiếm sẽ tìm và trả về những video clip có nói đến từ khóa đó.

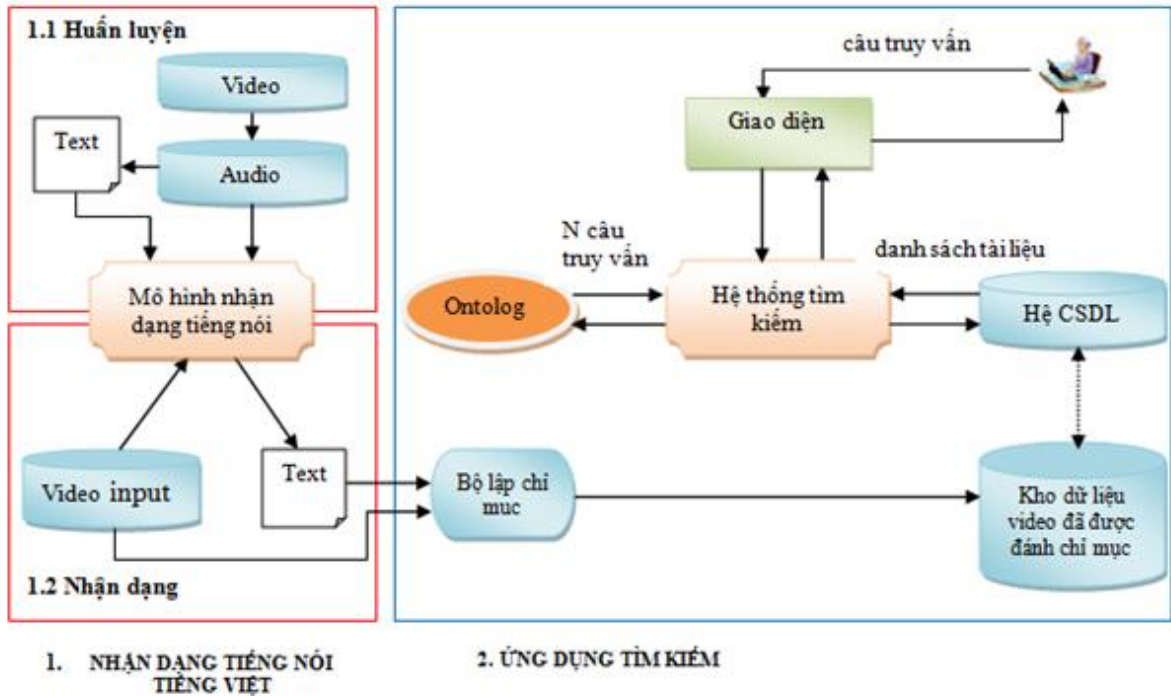
Hệ thống truy vấn thông tin (IR) thường gồm 3 phần chính. Thứ nhất là bộ thu thập dữ liệu (crawler), thứ hai là bộ lập chỉ mục (indexing) và cuối cùng là bộ xử lý truy vấn hay tìm kiếm (searching). Với đề tài này, hệ thống truy vấn thông tin được thực hiện trên dữ liệu video có sẵn nên không cần đến bộ thu thập dữ liệu.

### **3.1 Kiến trúc tổng quan của hệ thống**

Hệ thống gồm hai phần chính: Bộ nhận dạng tiếng nói tiếng Việt và Ứng dụng tìm kiếm.

Dữ liệu audio sẽ được tự động chuyển sang văn bản nhờ bộ nhận dạng giọng nói, sau đó chúng được thu thập và đánh chỉ mục để tạo cơ sở dữ liệu chỉ mục cho máy tìm kiếm. Máy tìm kiếm sẽ nhận câu truy vấn của người dùng dưới dạng văn bản, sau đó đối tượng Ontology sẽ có nhiệm vụ mở rộng câu truy vấn của người dùng thành các câu truy vấn con. Với mỗi câu truy vấn con, máy tìm kiếm sẽ phân tích chúng thành các từ khóa, đối chiếu với cơ sở dữ liệu chỉ mục có sẵn trong hệ thống, máy tìm kiếm sẽ xác định các tài liệu liên quan và trả kết quả về cho người dùng.

Hình 3.1 minh họa hệ thống truy vấn video hướng ngữ nghĩa, có sử dụng Ontology đã được xây dựng.



*Hình 3.1. Kiến trúc của hệ thống truy vấn video hướng ngữ nghĩa có sử dụng Ontology*

## 3.2 Các thành phần của hệ thống

### 3.2.1 Bộ nhận dạng tiếng nói tiếng Việt

Bộ nhận dạng giọng nói đóng vai trò chuyển đổi các tập tin âm thanh thành dạng văn bản tương ứng. Các văn bản này sẽ là đầu vào cho bộ xây dựng dữ liệu chỉ mục cho hệ thống tìm kiếm thông tin video. Bộ nhận dạng này được xây dựng dựa trên cơ sở lý thuyết và công cụ kế thừa từ nhóm nghiên cứu thuộc phòng thí nghiệm AILAB, bao gồm hai công đoạn: công đoạn huấn luyện và công đoạn nhận dạng.

#### 3.2.1.1 Công đoạn huấn luyện

Các video clip về lĩnh vực chăn nuôi ban đầu được trích ra kênh nội dung audio.



Các tập tin audio này sau đó được chia thành các đoạn nhỏ có độ dài tương ứng với một câu nói của phát thanh viên hay người được phỏng vấn trong audio, mỗi đoạn thường có độ dài tối đa không quá 15 giây. Lời của người nói trong các đoạn video được ghi nhận thủ công thành văn bản. Các đoạn audio và văn bản tương ứng là đầu vào cho công đoạn huấn luyện mô hình nhận dạng tiếng nói. Tên tiếng nước ngoài của các loại thuốc dùng trong chăn nuôi, hay những thuật ngữ tiếng nước ngoài sẽ được phiên âm trước khi đưa vào huấn luyện, ví dụ: “Hecta” được chuyển thành “Hét ta”, “Singapore” thành “xin ga bo”. Nguyên nhân của việc chuyển đổi tiếng nước ngoài thành tiếng Việt là vì chúng ta đang cần xây dựng bộ nhận dạng tiếng nói tiếng Việt, các thuật ngữ tiếng nước ngoài thì không nằm trong từ điển tiếng Việt của bộ nhận dạng giọng nói.

### **3.2.1.2 Công đoạn nhận dạng**

Trước khi đưa vào nhận dạng, video clip được chuyển thành audio. Các audio này là đầu vào của mô hình nhận dạng tiếng nói. Đầu tiên, tín hiệu tiếng nói của người nói trong các audio sẽ được rút trích đặc trưng. Đặc trưng này sẽ là đầu vào cho việc tìm kiếm trên đồ thị nhận dạng. Đồ thị nhận dạng được xây dựng bằng cách kết ghép mô hình ngôn ngữ và mô hình ngữ âm. Kết quả nhận dạng là các tập tin văn bản trích tương ứng với lời của người nói trong audio. Văn bản trích này sẽ là đầu vào cho phần xây dựng ứng dụng tìm kiếm bên dưới.

### **3.2.2 Ứng dụng truy vấn video**

Ứng dụng tìm kiếm có vai trò tiếp nhận câu truy vấn từ người dùng, tiếp đến đối tượng Ontology sẽ mở rộng câu truy vấn ra N câu truy vấn có liên quan đến câu truy vấn đó. Với mỗi câu truy vấn con mới này, khi tiến hành tìm kiếm dưới cơ sở dữ liệu sẽ cho ra một tập các kết quả tìm kiếm. Kết quả tìm kiếm cuối cùng là tập hợp của tất cả các kết quả tìm kiếm con. Kết quả tìm kiếm bao gồm thời điểm xuất hiện từ khoá trong video và các video sẽ được play ngay tại thời điểm đó. Phần ứng dụng tìm kiếm gồm 3

công đoạn: lập chỉ mục kho dữ liệu, xây dựng Ontology và xây dựng ứng dụng truy vấn video.

### 3.2.2.1 Lập chỉ mục kho dữ liệu

Sau khi chuyển các đoạn âm thanh sang văn bản, các tập tin văn bản này được đưa vào lập chỉ mục để xây dựng cơ sở dữ liệu tìm kiếm.

Đánh chỉ mục cho các tài liệu là phương pháp thực hiện quét một lần trên các tài liệu và lưu lại danh sách các thuật ngữ (term – từ, cụm từ) có trong tài liệu đó cũng như thông tin đi kèm với mỗi thuật ngữ (vị trí, tần suất, trọng số, v.v...). Các thông tin này sẽ được tổ chức theo một cấu trúc dữ liệu riêng và được gọi là chỉ mục. Lúc này, các thao tác tìm kiếm sẽ được tiến hành dựa trên chỉ mục thay vì thực hiện trực tiếp trên tài liệu.

Một số mô hình đánh chỉ mục phổ biến như: cây hậu tố, chỉ mục ngược, chỉ mục trích dẫn, chỉ mục ngram, ma trận tài liệu – từ chỉ mục. Trong các mô hình đánh chỉ mục đã nêu, chỉ mục ngược là mô hình được sử dụng trong đa số các bộ máy tìm kiếm thông dụng hiện nay. Trong phạm vi cho phép, đề tài này sẽ trình bày mô hình chỉ mục ngược của hệ thống truy vấn thông tin.

Trước tiên, chúng ta sẽ quy ước một số ký hiệu và khái niệm.

- Tài liệu  $d$ : diễn tả những ý tưởng về đề tài nào đó dưới dạng ngôn ngữ tự nhiên, là đơn vị mà một hệ thống IR được xây dựng dựa trên đó.
- Truy vấn  $q$ : diễn tả nhu cầu thông tin cho những tài liệu liên quan đến một đề tài nào đó.
- Từ chỉ mục  $k$ : một đơn vị ngữ nghĩa, một từ, một cụm từ ngắn hay là dạng gốc của một từ. Một tài liệu  $d$  sẽ được biểu diễn bởi một tập các từ chỉ mục  $k_i$ .
- Cơ sở dữ liệu  $DB$ : tập  $n$  tài liệu,  $d_j \in DB, j = 1, 2, \dots, n$

- Bảng từ vựng  $T$ : tập  $t$  từ chỉ mục,  $k_i \in T, i = 1, 2, \dots, t$

Giả sử cho trước một yêu cầu truy vấn sau: “*Những vở kịch nào của Shakespeare xuất hiện Brutus và Caesar nhưng không xuất hiện Calpurnia*”. Với truy vấn đã cho, hệ thống sẽ đọc tất cả các vở kịch của **Shakespeare**, lọc lại những vở kịch nào xuất hiện **Brutus** và **Caesar** và loại bỏ nếu nó có xuất hiện **Calpurnia**.

Khi “trả lời” một truy vấn, hệ thống phải thỏa mãn nhu cầu về tốc độ xử lý. Nghĩa là với tập tài liệu lớn đã cho, hệ thống phải xử lý một cách nhanh nhất có thể. Ngoài ra, hệ thống cũng phải truy tìm dựa trên xếp hạng các kết quả (ranking). Nói cách khác, hệ thống phải có câu trả lời tốt nhất cho một nhu cầu thông tin trong số nhiều tài liệu cùng chứa một số từ nào đó. Một yêu cầu nữa là hệ thống phải thực hiện những thao tác so khớp linh động khi truy vấn.

Chỉ mục của tài liệu (index) tương ứng với tập hợp các thuật ngữ (term) chứa trong nó. Các tài liệu được biểu diễn dưới dạng như bảng 1 bên dưới:

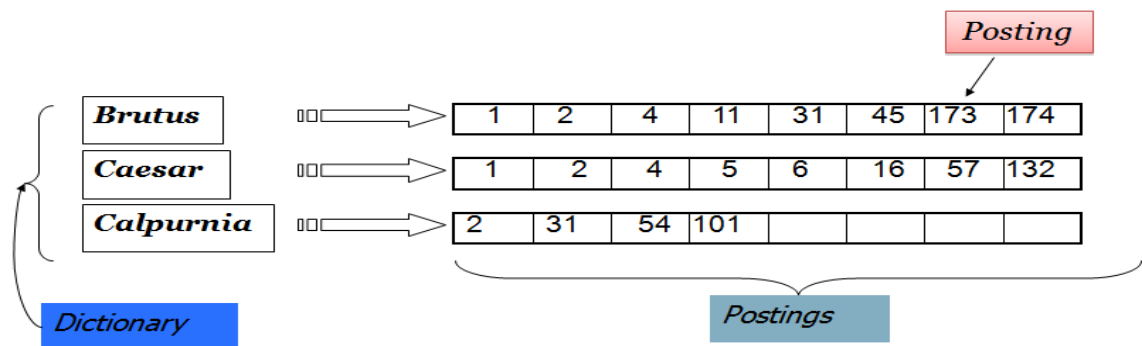
**Bảng 3.1. Chỉ mục của các tài liệu tương ứng với các thuật ngữ**

	$t_1$	$t_2$	$t_3$	$t_4$	...	$t_5$
$d_1$	1	0	1	0	0	1
...	0	1	1	1	0	1
$d_n$	1	1	0	1	1	0

Như bảng trên,  $d_i$  là tài liệu thứ  $i$  trong bộ sưu tập tài liệu (document collection),  $t_j$  là thuật ngữ thứ  $j$  chứa trong tài liệu. Giá trị 1 thể hiện thuật ngữ  $t_j$  có chứa trong tài liệu  $d_i$  và giá trị 0 là ngược lại. Các số 1 trong bảng 1 có thể được thay thế bằng số lần xuất hiện của thuật ngữ trong tài liệu. Mô hình chỉ mục lưu trữ theo bảng trên còn được gọi là mô hình “ma trận tài liệu – từ chỉ mục”, hay có tên gọi vắn tắt là “ma trận từ chỉ mục”.

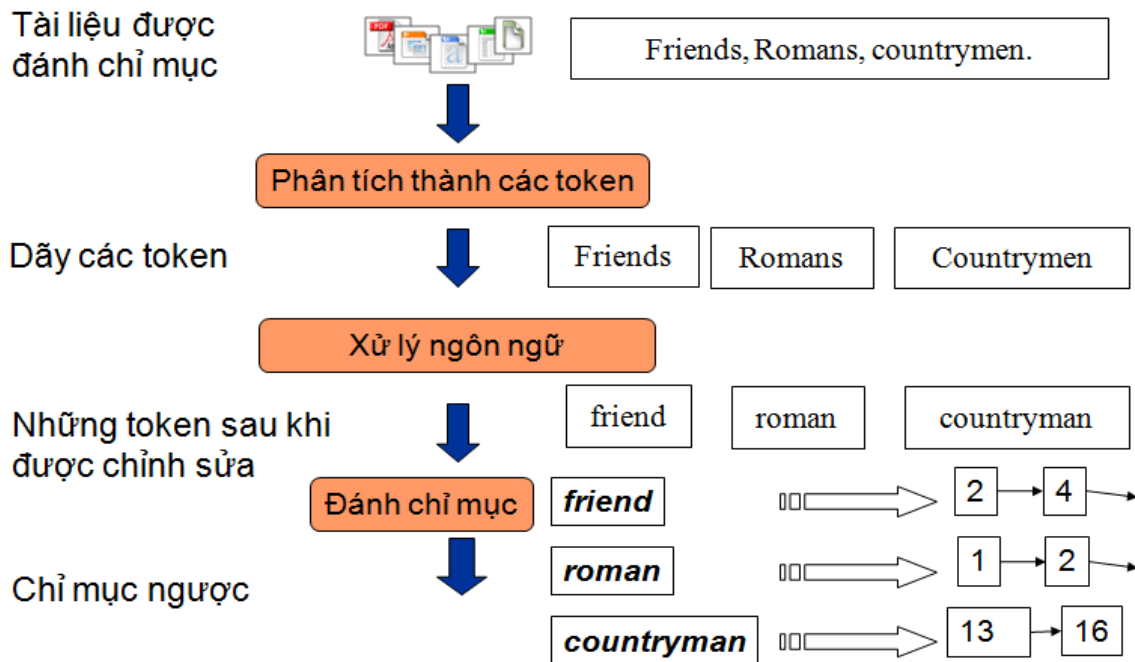
Với cách lưu trữ trên, nếu chúng ta có  $n$  bằng 1 triệu tài liệu và  $t$  bằng 500.000 từ chỉ mục (hay thuật ngữ) thì kích thước của ma trận từ chỉ mục sẽ là  $500.000 \times 1$  triệu. Và chúng ta sẽ có 500 tỉ số 0 và 1 được lưu trữ cho mô hình. Vấn đề về lưu trữ đã phát sinh ở đây. Nếu quan sát kỹ ma trận từ chỉ mục, chúng ta sẽ nhận thấy một điều rất quan trọng là nó có rất ít số thành phần khác 0. Ngoài ra, việc tìm kiếm chỉ quan tâm đến giá trị 1 (nghĩa là hệ thống chỉ tìm những tài liệu nào có chứa thuật ngữ cho trước). Như vậy, chúng ta chỉ cần lưu trữ lại những vị trí xuất hiện 1. Mô hình chỉ mục ngược (Inverted index) giúp chúng ta giải quyết vấn đề này.

Ý tưởng của mô hình chỉ mục ngược là với mỗi từ  $t$ , ta sẽ lưu một danh sách những tài liệu (DocID) có chứa  $t$ . Như trong Hình 3.2, Brutus, Caesar và Calpurnia là các từ hay thuật ngữ. Mỗi từ, sẽ được đi kèm với danh sách các DocID của tài liệu chứa nó, danh sách này được gọi là Postings và mỗi phần tử của nó là một Posting.



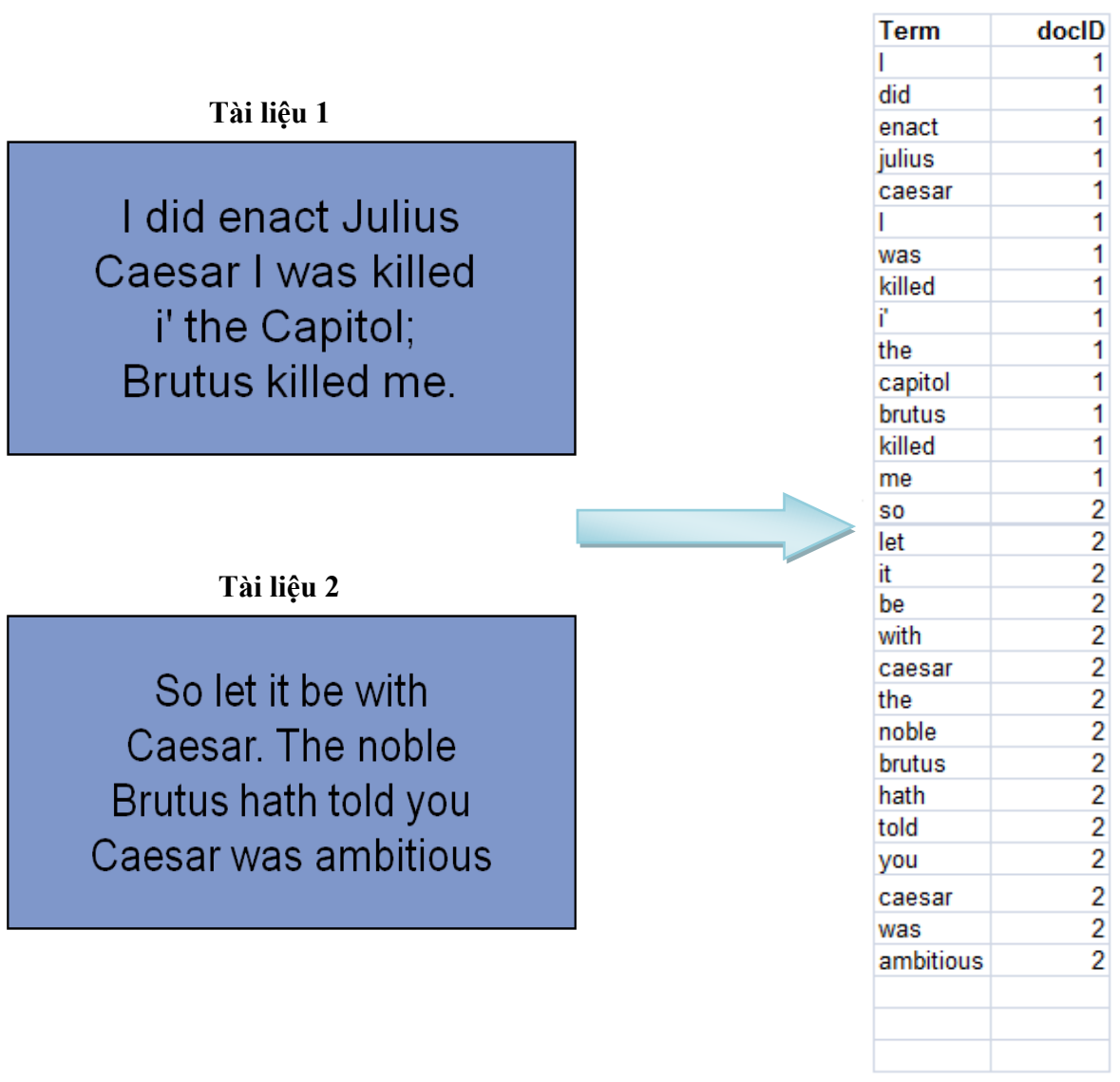
**Hình 3.2. Cách lưu trữ của mô hình chỉ mục ngược [17]**

Mô hình này có tên gọi là chỉ mục ngược do thay vì hệ thống IR sẽ truy vấn bằng cách trả lời câu hỏi “Những từ nào xuất hiện trong tài liệu này” thì nó sẽ trả lời câu hỏi “Những tài liệu nào chứa từ X”. Cấu trúc của mô hình chỉ mục ngược được tối ưu hóa để đưa ra câu trả lời nhanh cho loại câu hỏi này. Phần tiếp theo sẽ trình bày từng bước của việc xây dựng chỉ mục ngược.



**Hình 3.3. Mô hình tổng quát xây dựng chỉ mục ngược [17]**


Hình 3.3 mô tả tổng quát các bước xây dựng chỉ mục ngược. Việc xây dựng bộ chỉ mục gồm ba giai đoạn chính. Trước tiên, các tài liệu cần đánh chỉ mục sẽ được phân tích thành các token. Ở giai đoạn này, các câu trong tài liệu văn bản sẽ được chia nhỏ (split) thành từng âm tiết (syllable). Giai đoạn tiếp theo, các syllable sau đó sẽ qua bước xử lý ngôn ngữ để được chuẩn hóa. Tùy theo mỗi loại ngôn ngữ mà các cách chuẩn hóa sẽ khác nhau. Ví dụ, đối với tiếng anh, friends sẽ được chuẩn hóa thành friend. Hình 3.4 minh họa dãy các token đã được chỉnh sửa,



**Hình 3.4. Dãy các token đã được chỉnh sửa [17]**

Bước tiếp theo sau khi chỉnh sửa các token, chúng ta sẽ sắp xếp chúng lại theo từ chỉ mục và sau đó là đến docID. Việc sắp xếp này sẽ giúp tối ưu hóa trong quá trình xử lý chỉ mục và tìm kiếm. Hình 3.5 minh họa việc sắp xếp các token theo từ chỉ mục và docID.

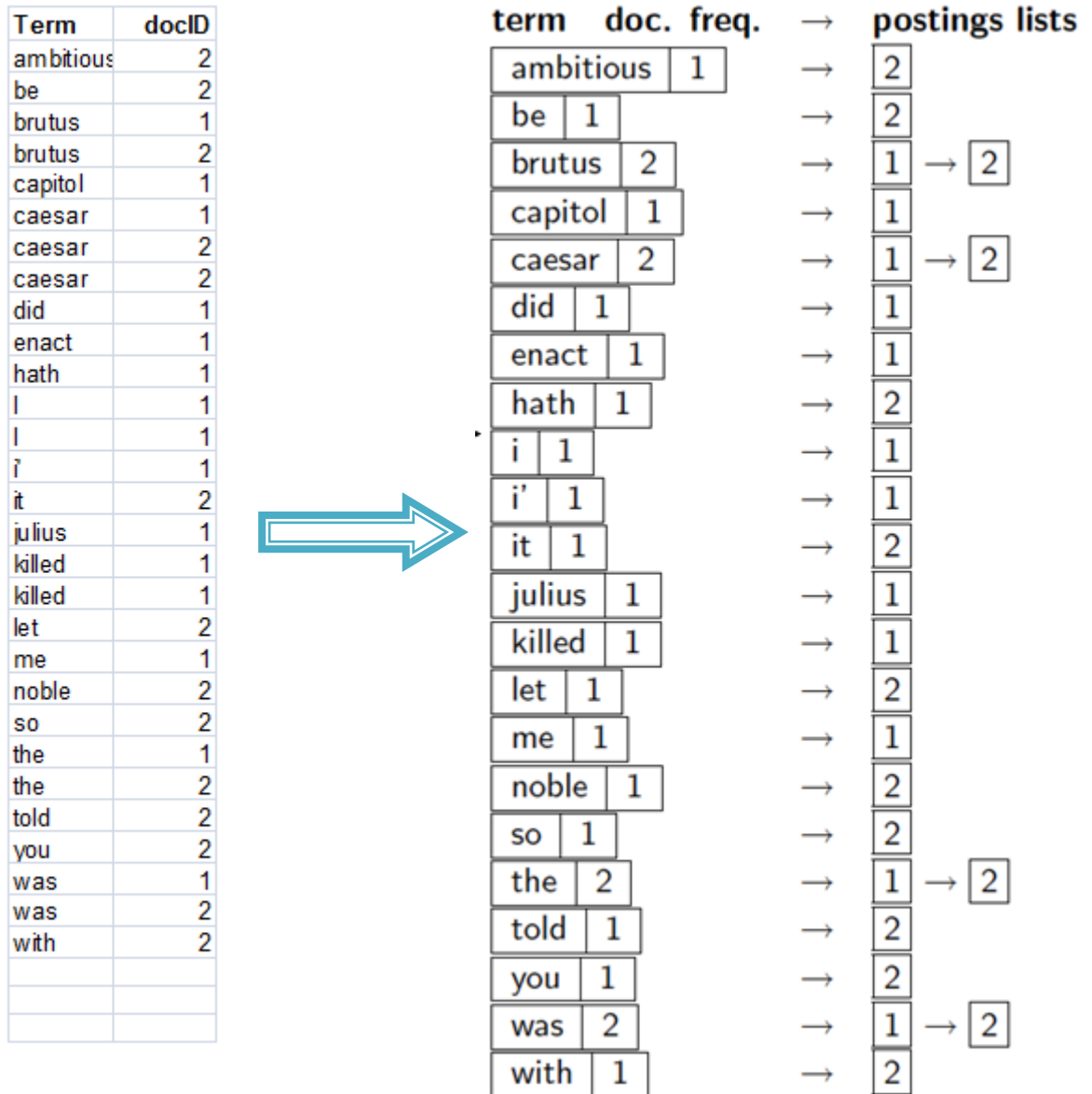
Term	docID
I	1
did	1
enact	1
julius	1
caesar	1
I	1
was	1
killed	1
i'	1
the	1
capitol	1
brutus	1
killed	1
me	1
so	2
let	2
it	2
be	2
with	2
caesar	2
the	2
noble	2
brutus	2
hath	2
told	2
you	2
caesar	2
was	2
ambitious	2



Term	docID
ambitious	2
be	2
brutus	1
brutus	2
capitol	1
caesar	1
caesar	2
caesar	2
did	1
enact	1
hath	1
I	1
I	1
i'	1
it	2
julius	1
killed	1
killed	1
let	2
me	1
noble	2
so	2
the	1
the	2
told	2
you	2
was	1
was	2
with	2

**Hình 3.5. Sắp xếp token theo từ chỉ mục và docID [17]**

Sau bước sắp xếp token theo từ chỉ mục và docID, những thực thể của cùng một từ chỉ mục được nhóm lại. Thông tin chỉ mục được tách thành dictionary và postings kèm theo thông tin tần số xuất hiện tài liệu (như trong Hình 3.6), tần số xuất hiện này cũng chính là kích thước của mỗi posting list được thêm vào.



**Hình 3.6. Thông tin Dictionary và Postings của chỉ mục [17]**

Sau khi chuẩn hóa, mỗi token sẽ được đánh chỉ mục bằng cách lưu trữ danh sách các tài liệu chứa nó theo ý tưởng chỉ mục ngược (inverted index) đã được trình bày ở đầu mục này.



### 3.2.2.2 Xây dựng ứng dụng truy vấn video

Trước tiên từ khoá của người dùng nhập vào sẽ được đối tượng Ontology mở rộng từ khóa tìm kiếm nhằm cung cấp khả năng tìm các từ đồng nghĩa, các nhóm từ xuất hiện trong cùng một ngữ cảnh.

Dựa vào bộ chỉ mục đã lập, khi cần truy vấn video bằng từ khoá, máy tìm kiếm sẽ tìm và trả về những video clip có nói đến từ khoá.

Từ kết quả trả về của máy tìm kiếm là danh sách các tài liệu (các tập tin video) có chứa nội dung tương ứng với từ khóa tìm kiếm, hệ thống tiếp tục xử lý trả về nội dung tập tin video theo thời gian thực nếu người dùng gửi yêu cầu xem tập tin này.

Theo đó, kết quả trả về từ máy tìm kiếm phải bao gồm hai thông tin: tên tập tin video chứa từ khóa tìm kiếm và khoảng thời gian xuất hiện từ khóa tìm kiếm trong tập tin video (khoảng thời gian này đã được tính toán và lưu trữ lại trong quá trình chia đoạn âm thanh, chính là khoảng thời gian bắt đầu của mỗi đoạn âm thanh được chia nhỏ)

Việc xử lý trả kết quả tập tin video theo thời gian thực dựa trên giao thức RTMP (Real Time Messaging Protocol), là giao thức được phát triển bởi Adobe, hỗ trợ việc truyền dữ liệu video trên môi trường mạng theo thời gian thực. Việc xử lý streaming video dựa trên giao thức RTMP gồm hai phần: client và server. Client được tích hợp một trình flash player có khả năng trình diễn dữ liệu video được truyền về từ server và một server hỗ trợ việc truyền dữ liệu theo giao thức RTMP.

Việc truyền nhận dữ liệu theo giao thức RTMP diễn ra như sau: Dữ liệu video gồm hình ảnh (video) và audio sẽ được phân chia thành các đoạn (fragment) trước khi được truyền đi, kích thước của một fragment phụ thuộc vào băng thông đường truyền, thường là 128 bytes cho dữ liệu video và 64 bytes cho dữ liệu audio. Các fragment có thể được truyền đi theo cơ chế ghép kênh và cơ chế kiểm soát lỗi dữ liệu trên đường truyền. Một số kênh truyền được sử dụng trong giao thức RTMP bao gồm: kênh đảm

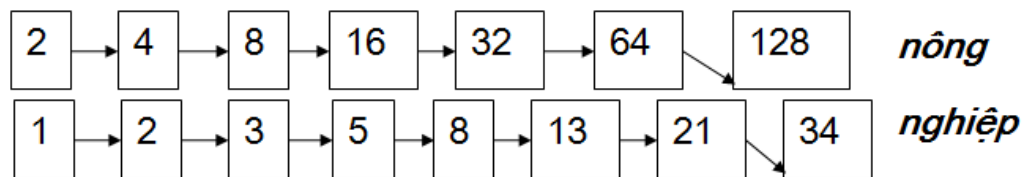
nhiệm RPC (Remote procedure call) sử dụng Active Message Format, kênh đảm nhiệm truyền nhận dữ liệu video, kênh đảm nhiệm truyền nhận dữ liệu audio, kênh đảm nhiệm truyền nhận thông điệp quá tải đường truyền. Dữ liệu đến client sẽ được thực hiện tách kênh và trình diễn thông qua flash player.

- **Xử lý truy vấn thông tin**

Xử lý truy vấn là việc chọn ra một giải thuật để xử lý việc tìm kiếm thông tin một cách hiệu quả nhất. Việc xử lý truy vấn có hai khía cạnh cần bàn đến, đó là xử lý truy vấn theo từ riêng biệt và xử lý truy vấn theo nhóm từ.

- **Truy vấn theo từ riêng biệt**

Ta xét truy vấn đơn giản là **nông** AND **nghiệp**. Trước tiên, hệ thống sẽ xác định Brutus trong dictionary và truy tìm những posting của nó. Tiếp theo, hệ thống sẽ xác định Caesar trong dictionary và truy tìm những posting của nó.



**Hình 3.7. Hai danh sách Posting của “nông” và “nghiệp”**

Sau đó, hai danh sách posting sẽ được trộn lại theo thuật toán trộn (Intersect).

Bên dưới là mã giả của thuật toán trộn

INTERSECT (p1, p2)

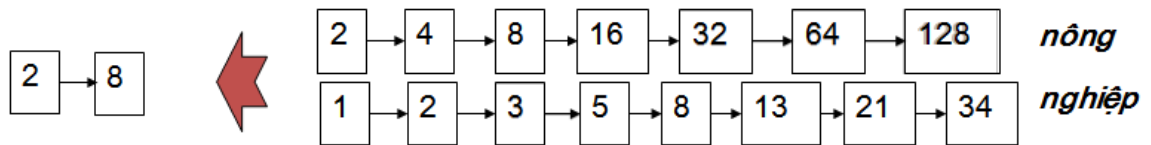
```

1  answer ← < >
2  while p1 ≠ NIL and p2 ≠ NIL
3  do if docID(p1) = docID(p2)
4      then ADD(answer, docID(p1))
5      p1 ← next(p1)
  
```

```

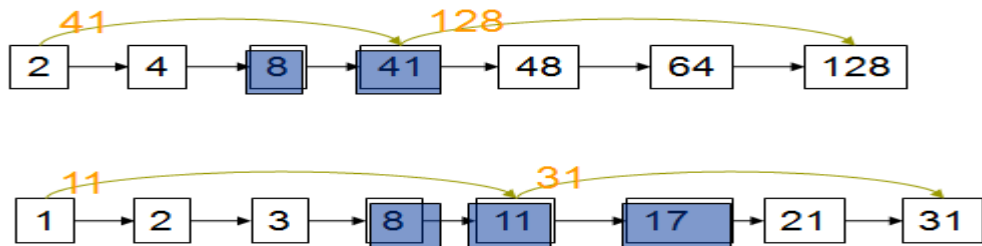
6         p2 ← next(p2)
7     else if docID(p1) < docID(p2)
8         then p1 ← skip(p1)
9         else p2 ← next(p2)
10    return answer

```



**Hình 3.8. Kết quả của thuật toán trộn 2 danh sách posting**

Kết quả của thuật toán trộn hai danh sách posting ở hình 3.7 ta được postID 2 và 8 như trong hình 3.8. Nếu hai danh sách có số phần tử tương ứng là  $m$  và  $n$  thì thuật toán trộn có chi phí là  $O(m+n)$  phép toán (các posting đã được sắp xếp theo docID). Như vậy có thuật toán nào tốt hơn không. Giải thuật con trỏ nhảy được áp dụng trong trường hợp này. Mục đích của giải thuật con trỏ nhảy là không xử lý những phần trong danh sách các posting không tham gia vào kết quả tìm kiếm. Hình bên dưới minh họa con trỏ nhảy



**Hình 3.9. Minh họa Con trỏ nhảy**

Trong hình 3.9, giả sử xét việc xử lý **docID 8** trong mỗi danh sách, chúng ta so khớp nó và tiếp tục, khi đó chúng ta có docID **41** (ở danh sách trên) và **11** (ở danh sách

dưới), **11** thì nhỏ hơn **41** và **docID** tiếp theo nhảy của **11** trong danh sách dưới là **31** và **31** vẫn còn nhỏ hơn **41**. Vì vậy, chúng ta có thể nhảy qua những posting giữa **11** và **31**. Bên dưới là mã giả của thuật toán trộn với con trỏ nhảy.

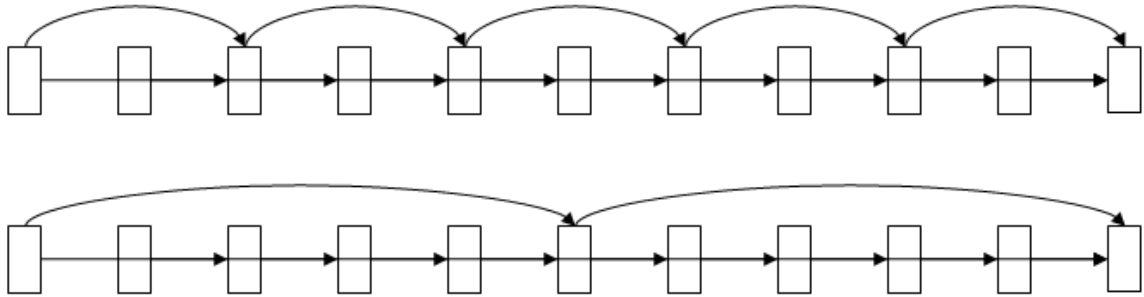
INTERSECTWITHSKIPS(*p1*, *p2*)

```

1  answer ← < >
2  while p1 ≠ NIL and p2 ≠ NIL
3  do if docID(p1) = docID(p2)
4      then ADD(answer, docID(p1))
5          p1 ← next(p1)
6          p2 ← next(p2)
7  else if docID(p1) < docID(p2)
8      then if hasSkip(p1) and (docID(skip(p1)) ≤ docID(p2))
9          then while hasSkip(p1) and (docID(skip(p1)) ≤ docID(p2))
10             do p1 ← skip(p1)
11             else p1 ← next(p1)
12     else if hasSkip(p2) and (docID(skip(p2)) ≤ docID(p1))
13         then while hasSkip(p2) and (docID(skip(p2)) ≤ docID(p1))
14             do p2 ← skip(p2)
15             else p2 ← next(p2)
16  return answer

```

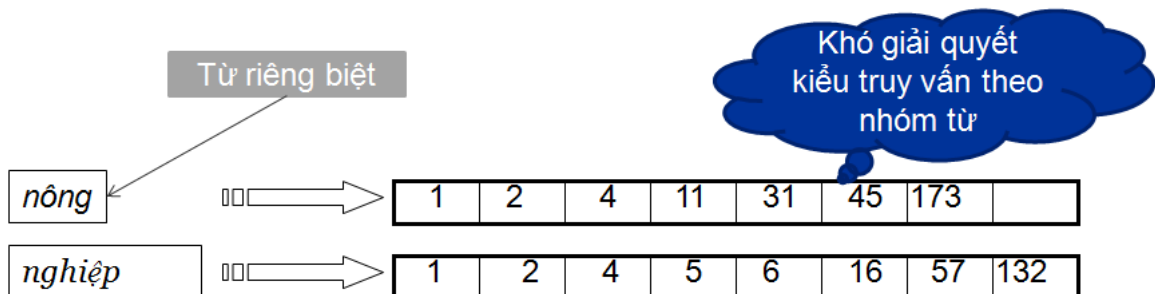
Hiệu năng của giải thuật con trỏ nhảy tùy thuộc vào độ chính xác của việc đặt con trỏ nhảy. Nếu có nhiều con trỏ nhảy trên 1 danh sách thì bước nhảy sẽ ngắn. Tuy nhiên, số lần so sánh con trỏ nhảy sẽ nhiều và tốn không gian lưu trữ con trỏ nhảy. Nếu có ít con trỏ nhảy trên 1 danh sách thì số lần so sánh con trỏ nhảy sẽ ít nhưng bước nhảy sẽ dài hơn và do vậy cũng sẽ có ít những bước nhảy thành công. Hình 14 bên dưới minh họa tính cân bằng của việc đặt con trỏ nhảy. Cách đặt con trỏ nhảy phổ biến là với mỗi danh sách posting chiều dài  $L$ , ta sử dụng  $\sqrt{L}$  con trỏ nhảy chia đều.



**Hình 3.10. Tính cân bằng của việc đặt con trỏ nhảy**

- **Truy vấn theo nhóm từ**

Khi trả lời cho những truy vấn dạng “*nông nghiệp*” – dạng nhóm từ (keyword có nháy kép), thì “*hướng nghiệp ngành nông*” không phải là câu trả lời. Đa số các bộ máy tìm kiếm đều hỗ trợ truy vấn theo nhóm từ. Với cách tổ chức dữ liệu chỉ mục đã trình bày thì khó giải quyết kiểu truy vấn theo nhóm từ.



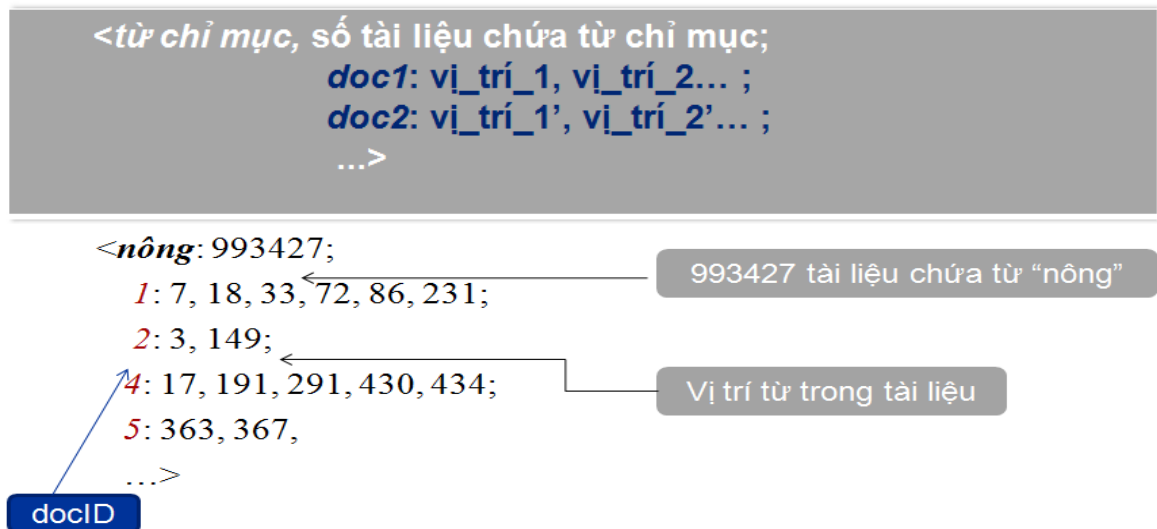
**Hình 3.11. Truy vấn với dữ liệu chỉ mục theo từ riêng biệt**

Để xử lý truy vấn theo nhóm từ, chúng ta có 2 hướng tiếp cận để giải quyết vấn đề trên. Hướng tiếp cận thứ nhất là lập dữ liệu chỉ mục theo cặp từ. Hướng tiếp cận thứ hai là lập dữ liệu chỉ mục theo vị trí.

Với hướng tiếp cận thứ nhất – lập dữ liệu chỉ mục theo cặp từ, chúng ta xem mỗi cặp từ liên tiếp nhau trong tài liệu là một nhóm từ. Các từ ngành, nông, nghiệp sẽ

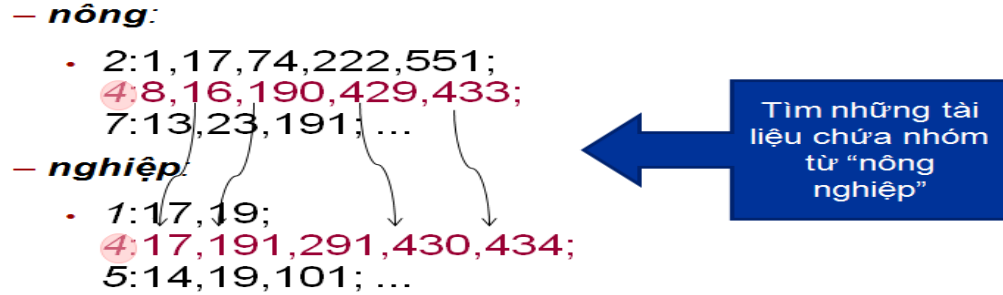
sinh ra những cặp từ “ngành nông” và “nông nghiệp”. Mỗi cặp từ được xem là một từ chỉ mục. Với hướng tiếp cận này thì hệ thống chỉ xử lý những truy vấn theo nhóm từ với hai từ (keyword chỉ chứa 2 từ). Nếu keyword chứa nhiều hơn 2 từ thì hướng tiếp cận này sẽ xử lý bằng cách biểu diễn dưới dạng truy vấn boolean trên các cặp từ. Ví dụ khi truy vấn “nông nghiệp chăn nuôi” tương đương với truy vấn “nông nghiệp AND nghiệp chăn AND chăn nuôi”. Khi xử lý như vậy, chúng ta không thể kiểm chứng việc những tài liệu khớp với truy vấn trên thật sự chứa nhóm từ truy vấn hay không nếu không mở tài liệu ra xem. Số lượng từ chỉ mục lúc này sẽ phình to. Giải pháp chỉ mục theo cặp từ không phải là giải pháp chuẩn cho xử lý theo nhóm từ.

Hướng tiếp cận thứ hai – lập dữ liệu chỉ mục theo vị trí, thì ứng với mỗi từ chỉ mục, ta lưu lại vị trí mà nó xuất hiện theo cách thức sau:



**Hình 3.12. Minh họa lập chỉ mục từ theo vị trí**

Theo hướng tiếp cận này, khi truy vấn nhóm từ, chúng ta trộn tất cả các danh sách <tài liệu : vị trí> để liệt kê tất cả các vị trí chứa nhóm từ cần tìm. Hình 3.13 bên dưới minh họa với nhóm từ “nông nghiệp”.



**Hình 3.13. Dữ liệu chỉ mục theo nhóm từ và truy vấn**

Khi truy vấn nhóm từ “nông nghiệp” ở hình 3.13, trộn hai danh sách posting lại ta được docID 4 có chứa cả hai từ “nông” và “nghiệp”. Chúng ta xét tiếp đến hai danh sách vị trí. Chúng ta thấy vị trí 16 của từ “nông” và vị trí 17 của từ “nghiệp” là hợp lý vì “nông” đứng trước “nghiệp”. Như vậy, chúng ta có được 1 vị trí xuất hiện từ “nông nghiệp” trong docID 4.

### 3.2.2.3 Xây dựng Ontology cho hệ thống

Mục đích của việc xây dựng Ontology trong đề tài này là để mở rộng câu truy vấn hay nói cách khác là mở rộng không gian tìm kiếm.

Lĩnh vực cần xây dựng Ontology ở đây là các thông tin liên quan đến chăn nuôi.

Ontology lưu trữ những thông tin về chăn nuôi bằng tiếng Việt đáp ứng được các nhu cầu của hệ thống như:

- Tìm kiếm thông tin liên quan đến vật nuôi:

Ví dụ: Kỹ thuật nuôi gia súc, dịch bệnh cúm gia cầm, v.v...

Theo khảo sát đã trình bày trong phần các công trình nghiên cứu liên quan, trên thế giới đã có những bộ Ontology được xây dựng cho lĩnh vực nông nghiệp nhưng không có tiếng Việt, chúng không thể ứng dụng trực tiếp vào ngữ cảnh Việt Nam.

Còn ở nước ta hiện này, chưa có Ontology cho lĩnh vực nông nghiệp, cụ thể nhánh chăn nuôi.

Vì vậy Ontology cho đề tài được xây dựng từ đầu dựa vào những nguồn thông tin trên Wikipedia tiếng Việt, các trang web liên quan đến lĩnh vực chăn nuôi Việt Nam.

Bước đầu xây dựng Ontology chăn nuôi với khoảng 76 khái niệm và hơn 300 thuật ngữ dựa trên 2 mối quan hệ: quan hệ liên quan và quan hệ tương đương.

Kết quả xây dựng Ontology sẽ được trình bày ở mục 4.2.2 trong chương 4.

### **3.3 Đặc tả phần mềm**

#### **3.3.1 Môi trường phát triển**

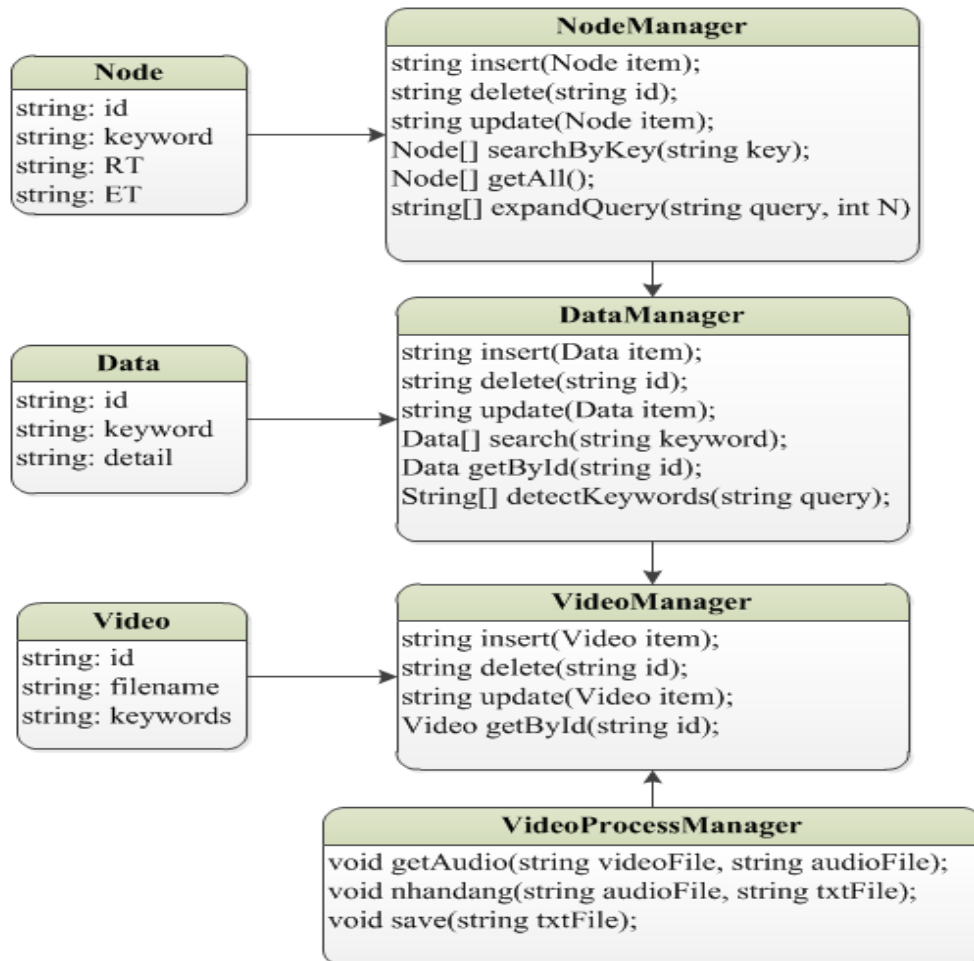
Thông số về môi trường phát triển ứng dụng tìm kiếm gồm:

- Hệ điều hành: Windows.
- Software platform: Microsoft .Net Framework .
- Ngôn ngữ lập trình: C#.Net.
- Cơ sở dữ liệu: SQL Server 2008.



### 3.3.2 Thiết kế kiến trúc

#### 3.3.2.1 Sơ đồ lớp:



*Hình 3.14. Sơ đồ các lớp xử lý của phần mềm*

#### 3.3.2.2 Diễn giải các lớp xử lý chính

- **Phần Ontology**

- \* **Lớp Node** gồm các thuộc tính:

- **id**: id của Node.

- **keyword**: từ khóa đại diện của Node.
- **RT**: danh sách các Node có quan hệ liên quan với Node hiện hành.
- **ET**: danh sách các Node có quan hệ tương đương với Node hiện hành.

\* **Lớp NodeManager** gồm các phương thức:

- **string insert(Node item)**: Thêm một Node vào trong hệ thống Ontology.
  - **Input**: item kiểu Node.
  - **Output**: nếu thành công trả về “done”, nếu có lỗi xảy ra thì trả về thông báo lỗi.
- **string delete(string id)**: Xóa một Node trong hệ thống Ontology dựa vào id của Node.
  - **Input**: id của Node cần xóa.
  - **Output**: nếu thành công trả về “done”, nếu có lỗi xảy ra thì trả về thông báo lỗi.
- **string delete(Node item)**: Cập nhật Node trong Ontology
  - **Input**: item kiểu Node.
  - **Output**: nếu thành công trả về “done”, nếu có lỗi xảy ra thì trả về thông báo lỗi.
- **Node[] searchByKey(string key)**: Tìm kiếm các Node dựa vào keyword
  - **Input**: key kiểu string.

- Output: trả về danh sách các Node có chứa key. Nếu không có sẽ trả về danh sách rỗng.
- **Node[] getAll():** Lấy danh sách tất cả các Node trong hệ thống Ontology.
  - Input: không.
  - Output: trả về danh sách tất cả các Node có trong CSDL.
- **string[] expandQuery(string query, int N):** Mở rộng câu truy vấn.
  - Input: gồm 2 thành phần:
    - câu truy vấn: kiểu chuỗi.
    - cấp tìm kiếm: kiểu số nguyên không âm. Quy định độ sâu tìm kiếm ( mức mở rộng không gian tìm kiếm) trong hệ thống Ontology.
  - Output: trả về danh sách các câu hỏi có liên quan đến câu hỏi nhập vào.
- **Phần quản lý CSDL**
  - \* **Lớp Data** gồm các thuộc tính :
    - **id:** id của data .
    - **keyword:** từ khoá.
    - **detail:** chứa thông tin vị trí xuất hiện của từ khoá trong video có định dạng : [vị trí bắt đầu#vị trí kết thúc#id video] | ...| [vị trí bắt đầu#vị trí kết thúc#id video].
  - \* **Lớp DataManager** gồm các phương thức:
    - **string insert(Data item):** Thêm một data vào trong CSDL.

- Input: đối tượng kiểu Data;
- Output: nếu thành công trả về “done”. Nếu có lỗi xảy ra trả về thông báo lỗi.
- **string delete(string id):** Xoá một Data trong CSDL dựa trên id.
  - Input: id của dòng Data;
  - Output: nếu thành công trả về “done”. Nếu có lỗi xảy ra trả về thông báo lỗi.
- **string update(Data item):** Cập nhật nội dung Data trong CSDL dựa trên id.
  - Input: đối tượng kiểu Data.
  - Output: nếu thành công trả về “done”. Nếu có lỗi xảy ra trả về thông báo lỗi.
- **Data[] search(string keyword):** Tìm kiếm danh sách data dựa trên keyword.
  - Input: từ khoá tìm kiếm.
  - Output: danh sách các Data chứa từ khoá. Nếu không có sẽ trả về danh sách rỗng.
- **Data getById(string id):** Lấy một Data dựa trên id.
  - Input: id của dòng Data;
  - Output: dòng Data có id = id input. Nếu không tồn tại sẽ trả về đối tượng rỗng.
- **String[] detectKeywords(string query):** Nhận diện và trả về danh sách các keywords trong nội dung truyền vào. Việc nhận diện này được thực hiện dựa trên hệ thống Ontology được xây dựng sẵn trước đó.

- Input: câu truy vấn người dùng nhập vào.
- Output: danh sách các keyword trong câu truy vấn.

\* **Lớp Video** gồm các thuộc tính:

- **id**: id của video.
- **filename**: đường dẫn + tên file video.
- **keywords**: danh sách keyword được rút trích trong phần detail dựa trên việc sử dụng Ontology xây dựng trước đó.

\* **Lớp VideoManager** gồm các phương thức:

- **string insert(Video item)**: Thêm một video vào trong CSDL.
  - Input: item kiểu video
  - Output: nếu thành công trả về “done”. Nếu có lỗi xảy ra trả về thông báo lỗi.
- **string delete(string id)**: Xoá một video trong CSDL dựa trên id.
  - Input: id của video.
  - Output: nếu thành công trả về “done”. Nếu có lỗi xảy ra trả về thông báo lỗi.
- **string update(Video item)**: Cập nhật nội dung video trong CSDL dựa trên id.
  - Input: item kiểu Video.
  - Output: nếu thành công trả về “done”. Nếu có lỗi xảy ra trả về thông báo lỗi.
- **Video getById(string id)**: lấy một video dựa trên id.
  - Input: id của video.
  - Output: đối tượng video có id = id input. Nếu không tồn tại

thì trả về rỗng.

- **Phần xử lý Video.**

\* **Lớp VideoProcessManager** có các phương thức:

- **getAudio** (string videoFile, string audioFile): Rút trích file âm thanh từ file video.
  - videoFile: đường dẫn file video.
  - audioFile: đường dẫn file audio.
- **nhandang**(string audioFile, string txtFile): Nhận dạng file âm thanh
  - audioFile: đường dẫn file âm thanh.
  - txtFile: đường dẫn file kết quả nhận dạng.
- **save**(string txtFile): Lưu trữ kết quả nhận dạng vào hệ CSDL
  - txtFile: đường dẫn file chứa kết quả nhận dạng

## CHƯƠNG 4. THỰC NGHIỆM

Chương này trình bày các kết quả thực nghiệm của luận văn, bao gồm: Thực nghiệm đánh giá độ chính xác của mô hình nhận dạng tiếng nói. Trình bày kết quả xây dựng Ontology chăn nuôi, phương pháp đánh giá và chương trình demo.

### 4.1 Cấu hình máy thử nghiệm

Các kết quả thu được với thực nghiệm trên máy tính có cấu hình như bảng sau:

*Bảng 4.1. Cấu hình máy thử nghiệm*

CPU	Intel Core 2 Dual 2.8GHz
RAM	4GB

### 4.2 Thực nghiệm

#### 4.2.1 Thực nghiệm đánh giá độ chính xác của mô hình nhận dạng tiếng nói

##### 4.2.1.1 Tập dữ liệu huấn luyện mô hình ngữ âm

Với tổng thời lượng 4 giờ 34 phút 47 giây tiếng nói được lấy từ các video về lĩnh vực chăn nuôi trên youtube. Dữ liệu tiếng nói này sau đó được cắt thành nhiều đoạn nhỏ có độ dài tối đa không quá 15 giây. Phần văn bản tương ứng với dữ liệu tiếng nói cũng được tạo thủ công (do người gõ) tạo nên từ điển gồm 950 file, 44297 từ.

Trong đó lấy ra tập dữ liệu để huấn luyện mô hình ngữ âm như sau:

*Bảng 4.2. Bảng thống kê dữ liệu huấn luyện mô hình ngữ âm*

Tổng số file	853 file
Tổng thời gian	4,05 giờ
Tổng số từ	39410 từ

#### 4.2.1.2 Tập dữ liệu huấn luyện mô hình ngôn ngữ

Mô hình ngôn ngữ dùng để huấn luyện là mô hình ngôn ngữ bi-gram. Dữ liệu cho việc xây dựng mô hình ngôn ngữ là bộ text được ghi nhận thủ công từ các video chăn nuôi gồm:

**Bảng 4.3. Bảng thống kê dữ liệu huấn luyện mô hình ngôn ngữ**

Tổng số câu	2189 câu
Tổng số từ	39410 từ

#### 4.2.1.3 Tập dữ liệu test

**Bảng 4.4. Bảng thống kê dữ liệu test**

Tổng số file	97 file
Tổng thời gian	30 phút
Tổng số từ	4887 từ

#### 4.2.1.4 Độ đo đánh giá

Hiệu năng của hệ thống nhận dạng tiếng nói thông thường được đánh giá qua sai số từ (WER - Word Error Rate).

$$WER = \frac{s + i + d}{n} \quad (23)$$

Trong đó:

- n là tổng số từ cần nhận dạng
- s là số lỗi sửa (lỗi nhận dạng sai từ)
- i là số lỗi chèn (lỗi chèn thêm từ)
- d là số lỗi xoá (lỗi mất từ)

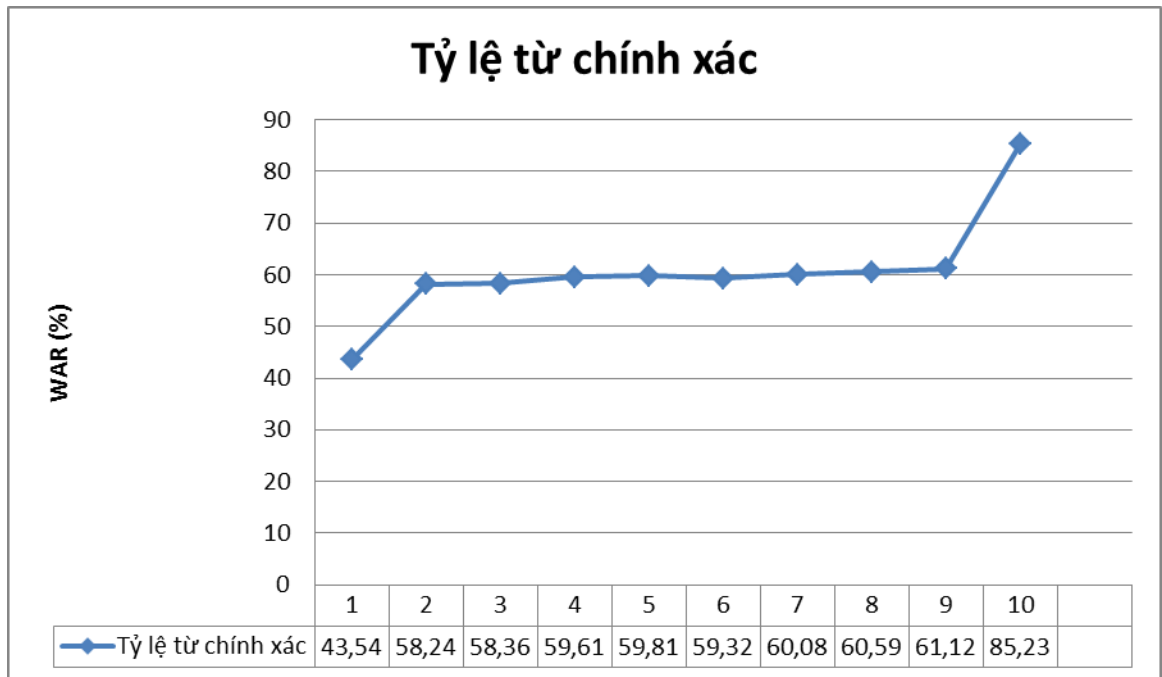


Hệ thống thử nghiệm áp dụng độ chính xác từ (WAR - Word Accuracy Rate) để đánh giá hiệu năng của hệ thống, với:

$$\text{WAR} = 1 - \text{WER} \quad (24)$$

#### 4.2.1.5 Kết quả thực nghiệm

Thực nghiệm được đánh giá dựa trên mô hình ngôn ngữ bi-gram với 10 phương pháp. Để so sánh tỷ lệ từ chính xác từ kết quả nhận dạng với từng phương pháp, ta có biểu đồ ở hình 4.1 như sau:



**Hình 4.1. Kết quả nhận dạng của mô hình bi-gram**

- 1-Monophone.
- 2-Triphone – first pass
- 3-Triphone LDA+MLLT
- 4-Triphone LDA+MLLT with MMI

- 5-Triphone LDA+MLLT with MMI and boosting
- 6-Triphone LDA+MLLT with MMI and MPE (Minimum phone error )
- 7-Triphone LDA+MLLT+SAT
- 8-Triphone LDA+MLLT+SAT+MM
- 9-Triphone LDA+MLLT+SAT and fMMI+MMI
- 10-Triphone SGMM

Ta nhận thấy kết quả nhận dạng của mô hình bi-gram với Triphone SGMM cho hiệu suất cao nhất với độ chính xác khá tốt là 85,23%.

## **4.2.2 Kết quả xây dựng Ontology chăn nuôi, phương pháp đánh giá và thực nghiệm truy vấn của hệ thống.**

### **4.2.2.1 Ontology chăn nuôi**

Ontology chăn nuôi được xây dựng dựa trên sự tham khảo từ các tài liệu trên các website liên quan đến lĩnh vực chăn nuôi. Các khái niệm và quan hệ đã được chính thức hóa và xác nhận trong bảng dữ liệu, bước đầu xây dựng Ontology chăn nuôi có khoảng 76 khái niệm, hơn 300 thuật ngữ.

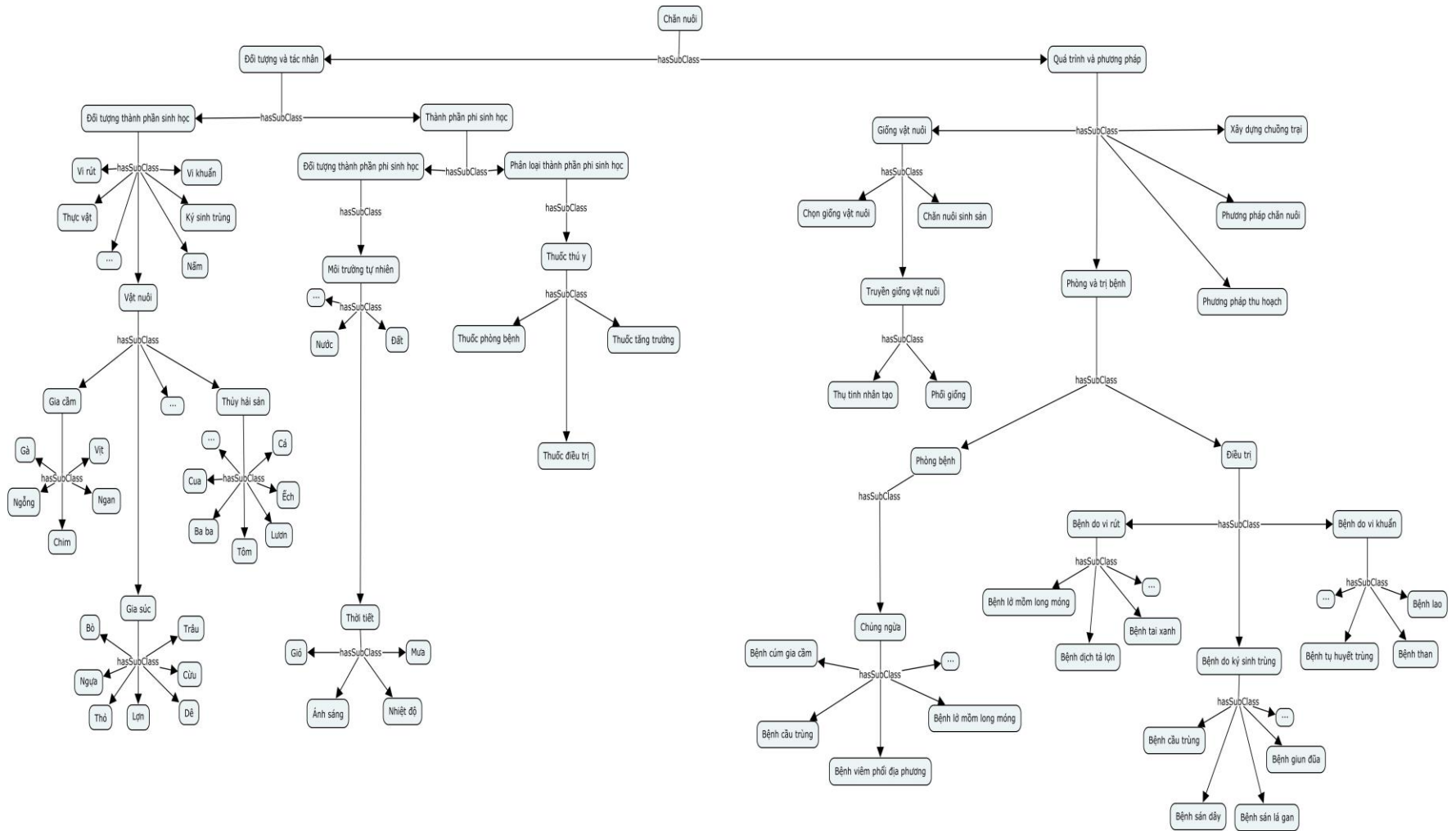
Tuy nhiên, vì điều kiện thời gian không cho phép cũng như giới hạn sự hiểu biết, các khái niệm và thuật ngữ trong lĩnh vực này còn rất hạn chế. Cho nên Ontology còn hạn hẹp với ít mối quan hệ hỗ trợ truy vấn kiến thức. Việc phân lớp dữ liệu và xác định mối quan hệ còn rất hạn chế vì chủ yếu dựa trên các tài liệu tham khảo và phân tích theo ý kiến cá nhân, chưa có sự tham khảo ý kiến từ nhiều chuyên gia trong miền để việc phân lớp và xác định mối quan hệ được chính xác hơn.

Xây dựng bảng phân loại các đối tượng và khái niệm của Ontology chăn nuôi, như bảng 4.5.

**Bảng 4.5. Phân loại các khái niệm của Ontology chăn nuôi**

<b>Đối tượng và tác nhân</b>	<b>Quá trình và phương pháp</b>
Vật nuôi	Chuồng trại
Các bệnh	Giống vật nuôi
Vi khuẩn	Chăn nuôi
Vi rút	Phòng và trị bệnh
Ký sinh trùng	Thu hoạch
Nấm	
Thuốc thú y	
Các yếu tố môi trường	

Dựa vào các khái niệm chính trên, ta có cấu trúc phân cấp các khái niệm của Ontology như hình 4.2 sau:



Hình 4.2. Mô hình tổng thể các khái niệm của Ontology chăn nuôi

### ***Xây dựng bộ thuật ngữ cho Ontology chăn nuôi***

Đề tài xây dựng bộ thuật ngữ Ontology về vật nuôi, bộ thuật ngữ này được xây dựng gồm 2 quan hệ: quan hệ liên quan và quan hệ tương đương (1-1). Các thuật ngữ trong Ontology sẽ được chuẩn hóa và lưu vào CSDL để phục vụ cho việc mở rộng câu truy vấn.

Danh sách các thuật ngữ được thể hiện như ở bảng sau:

***Bảng 4.6. Danh sách các thuật ngữ của Ontology***

<b>STT</b>	<b>Thuật ngữ</b>	<b>Các thuật ngữ liên quan (Quan hệ liên quan)</b>	<b>Các thuật ngữ tương đương (Quan hệ tương đương)</b>
1.	Vật nuôi	Gia súc, Gia cầm, Thủy hải sản	
2.	Gia súc	Lợn, Trâu, Bò, Dê, Thỏ, Cừu, Hươu, Ngựa	Lợn – Heo Trâu – Nghé Bò – Bê Cừu – Trừu
3.	Gia cầm	Gà, Vịt, Ngan, Ngỗng, Chim	Ngan – Vịt xiêm
4.	Chim	Bồ câu, Chim cú	
5.	Thủy hải sản	Tôm, Lươn, Cá, Cua, Ba ba	Ba ba – Cua đĩnh
6.	Lợn	Giống đại bạch, Giống Landrace, Giống lợn bò, Giống Berkshire, Lợn đực giống, Lợn nái sinh sản, Lợn thịt, Phó thương hàn, Dịch tả lợn, Đóng dấu lợn,	Giống đại bạch – Giống Yorkshise large white Giống lợn bò – Giống Landrace

		Tụ huyết trùng, Viêm phổi địa phương, Lở mồm long móng	
7.	bệnh gia súc	bệnh truyền nhiễm, bệnh ở bò, bệnh ở dê, bệnh ở thỏ, bệnh ở cừu, bệnh ở lợn, bệnh ở ngựa, bệnh do vi khuẩn, bệnh do vi rút, bệnh do ký sinh trùng, bệnh do nấm	
8.	bệnh truyền nhiễm	bệnh dịch tả lợn, bệnh giả dại ở lợn, bệnh tụ huyết trùng, bệnh đóng dấu lợn, bệnh phó thương hàn	
9.	Phó thương hàn (Salmonellosis)	phó thương hàn, vi khuẩn, Salmonella, nước uống nhiễm khuẩn, thức ăn, nhiễm khuẩn, sốt cao, ngời bú, bỏ ăn, tiêu chảy, tụ máu, bong vảy, tím ở ngực, tím ở tai, tím ở bụng, viêm phổi, cách ly, tiêm phòng, vắc xin, Samonella A, Samonella B, B Subtiliss, thuốc trợ sức.	phó thương hàn-bệnh phó thương hàn -Salmonellosis ngời bú - lười bú, ít bú, biếng bú bỏ ăn - không ăn, ngưng ăn tiêu chảy-ỉa chảy tụ máu - ứ máu, tụ huyết, ứ huyết cách ly - tách riêng, nhốt riêng sốt cao - thân nhiệt cao, nhiệt độ cao tiêm phòng - chích ngừa,

			<p>phòng ngừa bong vảy-tróc vảy cafein-cà phê in thuốc trợ sức-thuốc bổ vắc xin-vaccine</p>
10.	Dịch tả lợn (Swine fever)	dịch tả lợn, truyền nhiễm, vi rút, ủ bệnh, bỏ ăn, mệt nhọc, giầy dụa, sốt cao, khó thở, nôn mửa, táo bón, tụ huyết, sút nhanh, gầy yếu, nghiêng răng, xuất huyết, kiệt sức, tiêm phòng, tẩy uế, sát trùng, vắc xin, dịch tả, tạo miễn dịch	<p>Dịch tả lợn-bệnh Dịch tả lợn-swine fever truyền nhiễm-lây nhiễm vi rút-Virus ủ bệnh-nung bệnh, mang bệnh bỏ ăn-không ăn, ngưng ăn mệt nhọc-mệt mỏi sốt cao-thân nhiệt cao, nhiệt độ cao nôn mửa-ói mửa tụ huyết-ứ huyết, ứ máu gầy yếu-ôm yếu kiệt sức-mất sức, đuối sức tiêm phòng-tiêm ngừa, chích ngừa vắc xin-vaccine.</p>
11.	Tụ huyết trùng (Pasteurellosis)	tụ huyết trùng, vi khuẩn, serotyp b, typ a, typ b, typ c, typ d, sốt cao, thờ dộc, mệt, nằm bệt, tim rối loạn, nằm li bì, ứ máu, phù thũng, viêm họng, nhiễm	<p>Tụ huyết trùng-bệnh Tụ huyết trùng-pasteurellosis vi khuẩn- pasteurellamultocida vi khuẩn - Serotyp B typ a -tuýp a typ b-tuýp b</p>

		trùng huyết, khó thở, ngò thở, chảy mũi, gầy yếu, sung khớp, táo bón, ỉa chảy, vệ sinh, cách ly, tiêm phòng, vắc xin, kháng sinh, streptomycin, kanamycin, nofloxilin, tiamulin, sulfadimetoxin, gentamixin, sulfamethazin, gentamixin, tylosin, thuốc bổ, cafein, vitamin b1, vitamin c, vitamin b12,	typ c-tuýp c typ d-tuýp d sốt cao-thân nhiệt cao, nhiệt độ cao ứ máu-tụ máu, ứ huyết, tụ huyết gầy yếu-ôm yếu ỉa chảy-tiêu chảy, tách riêng, nhốt riêng tiêm phòng-chích ngừa vắc xin-vaccine thuốc bổ-thuốc trợ sức cafein-cà phê in.
12.	Đóng dấu lợn (Eerysipelas suum)	đóng dấu lợn, vi khuẩn, erysipelothix rhuioopathiae, mang bệnh, sốt, mắt đỏ, bỏ ăn, bỏ uống, nhiễm trùng máu, bại huyết, chết đột ngột, chết giãy dụa, ốm, ủ rũ, lờ đờ, ăn ít, không ăn, da khô, nôn mửa, tiêu chảy, chân run, thở khó, đi táo, ăn kém, uống kém, gầy còm, viêm khớp, hoại tử, phổi thủng, da sung, ỉa chảy, thiếu máu, rụng lông, loét lợi, chăm sóc, sát	đóng dấu lợn-erysipelas suum đóng dấu lợn-swine erysipelas vi khuẩn-Erysipelothix rhuioopathiae mang bệnh-ủ bệnh mang bệnh-nung bệnh sốt-nhiệt độ cao sốt-thân nhiệt cao ôm-bệnh ủ rũ-buồn bã ăn ít-biếng ăn ăn ít-lười ăn tiêm phòng-chích ngừa vắc xin-vaccine



		trùng, tiêm phòng, vắc xin, vô hoạt, tụ máu, nhược độc, huyết thanh đông máu, pen trep, ka ampi, anazin c, bcomplex,	
13.	Lở mồm long móng (LMLM) (Swine foot and muth disease)	lở mồm long móng, lmlm, swine foot and muth disease, vi rút, lây nhiễm, esitero, picorna, entero, cardio, rhino, aphto, typ o, typ vi rút a, typ vi rút c, typ vi rút sat1, typ vi rút sat2, typ vi rút sat3, typ vi rút as doong i, lâm sàn, sốt, nốt loét, mụn loét, không uống, không ăn, long móng, dạ dày loét, tiêm phòng, vắc xin, đa giá, cách ly, thuốc sát trùng, nước vôi, axit fenic, virkin, iodin, kiểm dịch, xét nghiệm, kháng huyết thanh, chăm sóc, trợ sức, sát trùng, methylen, than xoan, lá đào, thuốc đỏ, thuốc tím, formol, acid axetic, acid citric, lá khế, nghệ, tỏi, mỡ kháng sinh,	Lở mồm long móng-bệnh Lở mồm long móng-lmlm Lở mồm long móng- swine foot and muth disease typ vi rút a-typ virus a typ vi rút o-typ virus o typ vi rút c-typ virus c typ vi rút sat1-typ virus sat1 typ vi rút sat2-typ virus sat2 typ vi rút sat3-typ virus sat3 typ vi rút as doong i-typ virus as doong i không ăn-ngưng ăn, không ăn-bỏ ăn không uống-ngưng uống, bỏ uống cách ly-nhốt riêng, tách riêng trợ sức-thuốc bổ.

		dầu lạc, phèn chua, tro bếp, lá ổi, phèn xanh, neomat, manycin, diệt trùng, sulfamid, b-complet, untropin, trị viêm nhiễm, penicilin, hamcoli s, hanoxylin,	
14.	Viêm phổi địa phương (swine enzootie pneumonia)	viêm phổi địa phương, suyễn, vi khuẩn, vi trùng, viêm phế quản, nung bệnh, ngòì thờ, thờ giật, ho dai dẳng, thờ khó, mệt nhọc, ăn ít, viêm kết mạc, có ghèn, gầy còm, vắc xin, respisure, tytan-premix, tm-200, nhót riêng, sát trùng, vimekat, vime c antistress, tiêm, vimespiro fsp, vimefloro fdp, vimetryl, vitamin c, b-complex, nacampho, vimekon	viêm phổi địa phương-bệnh viêm phổi địa phương- swine enzootie pneumonia vi khuẩn- mycoplasma hyopneumoniae nung bệnh-ủ bệnh, mang bệnh mệt nhọc-mệt mỏi ăn ít-biếng ăn, ăn ít-lười ăn gầy còm-gầy nhom vắc xin-vaccine nhót riêng-tách riêng, cách ly tiêm-chích.
15.	Trâu Bò	Trâu, thức ăn thô, cỏ khô, cỏ xanh, rơm, thức ăn tinh, cỏ stylo, cỏ voi, bệnh tiên mao trùng, bệnh sán lá gan, bệnh cầu trùng, bệnh dịch	Trâu – Nghé Bò - Bê

		tả, bệnh tụ huyết trùng, bệnh chướng bụng đầy hơi, hội chứng tiêu chảy, bệnh lở mồm long móng	
16.	Dê	bệnh ỉa chảy, bệnh viêm vú, bệnh giun sán, bệnh đau mắt, bệnh ký sinh trùng, bệnh viêm phổi, hội chứng tiêu chảy	
17.	Gà	cúm gà, tiêu chảy do E. Coli, hen gà, tụ huyết trùng gà, ký sinh trùng máu, thương hàn, cầu trùng, đậu gà, hội chứng giảm đẻ	
18.	Cúm gà	cúm gà, vi rút, nhiễm khuẩn, Orthomyxoviridae, sốt cao, uống nhiều nước, viêm sung, mào tím thâm, khó thở, tiêu chảy, phân xanh vàng, phân lẫn máu, giảm đẻ, mất trứng bằng, sát trùng, tiêm phòng, amylyte, unisol 500, vytrolype, soramin, livercin, zymepro, perfectzyme, kháng sinh,	cúm gà - avian influenza sốt cao - nhiệt độ cao, thân nhiệt cao tiêu chảy - ỉa chảy giảm đẻ - đẻ ít

		moxcolis, amoxy 50, nexymix, tyloguard, doxycycline 150	
19.	Tiêu chảy do E. Coli	vi khuẩn, echerichia coli, bệnh tiêu hóa, bỏ ăn, ủ rũ, mệt mỏi, sốt cao, ỉa chảy, xuất huyết, trứng giảm, trứng không tốt, kiệt sức, chết, sát trùng, ioguard 300, bestagquam s, ultraxide, tiêm phòng, vaccine, amylyte, unisol 500, moxcolis, amoxy 50	bệnh tiêu hóa – bệnh đường ruột trứng giảm – đẻ ít trứng không tốt – trứng xấu
20.	Tôm	trại nuôi, ao nuôi, hồ nuôi, bệnh hoại tử gan tụy cấp tính, bệnh đốm trắng, bệnh đầu vàng, bệnh phân trắng, bệnh hoại tử gan tụy do vi khuẩn, bệnh taura	
21.	Bệnh hoại tử gan tụy cấp tính	vi khuẩn, vibrio parahaemolyticus, gan tụy teo, ruột tôm đứt đoạn, mềm vỏ, tỷ lệ chết cao, chậm lớn, nuôi luân canh, chọn giống tốt	hoại tử gan tụy cấp tính – EMS, AHPND
22.	Bệnh đốm	vi rút, white spot	bệnh đốm trắng – WSD

trắng	syndrome, vi khuẩn, bacterial white spot syndrome, độ cứng nước cao, đốm trắng, bỏ ăn, bơi lờ đờ, dạt vào bờ, đổ thân, chậm lớn, chết rải rác, chọn giống tốt, PCR, lọc nước ao	vi rút – virus white spot syndrome – WSSV bacterial white spot syndrome – BWSS bỏ ăn – ngưng ăn, không ăn chọn giống tốt – chọn giống khỏe
-------	---	--

#### 4.2.2.2 Phương pháp đánh giá Ontology

Phương pháp đánh giá chất lượng của Ontology được thể hiện bằng hai phương pháp đánh giá như sau:

➤ **Đánh giá bởi các chuyên gia:**

Đây là bước đánh giá tính đúng đắn về mặt lý thuyết như các khái niệm, các thuật ngữ, các mối quan hệ liên quan. Bước này cần phải có các chuyên gia trong miền để việc đánh giá được chính xác.

➤ **Đánh giá bởi người dùng:**

Đây là bước đánh giá xem thế nào là một Ontology tốt để đáp ứng được các yêu cầu của người dùng bằng cách trả lời các câu hỏi kiến thức được xác định thông qua giai đoạn đặc tả Ontology. Hiệu quả của việc sử dụng Ontology được đo bằng độ chính xác (Precision) và độ thu hồi (Recall) của các thuật ngữ thông qua các kết quả truy vấn tìm kiếm.

- Độ phủ (recall) của truy vấn cho bởi:

$$\text{Recall} = \frac{\# \text{ retrieved documents}}{\# \text{ relevant documents in the database}} \times 100$$

- Độ chính xác (precision) của truy vấn:

$$\text{Precision} = \frac{\text{\#correctly retrieved documents}}{\text{\# retrieved documents}} \times 100$$

**Trong đó:**

# retrieved documents: số tài liệu truy vấn được.

# relevant documents in the database: số tài liệu liên quan đến câu query thực có trong cơ sở dữ liệu.

# correctly retrieved documents: số tài liệu truy vấn được chính xác

Quá trình xây dựng phát triển Ontology là một quá trình lặp đi lặp lại: một lần xác định mà *Precision* hoặc *Recall* không đáp ứng nhu cầu người sử dụng (ví dụ như khái niệm trong Ontology không đầy đủ), việc phát triển Ontology có thể được cải thiện bằng cách quay lại giai đoạn thu thập kiến thức để bổ sung cập nhật cho Ontology.

Để thấy được hiệu quả của việc sử dụng Ontology. Chúng ta có thể so sánh kết quả truy vấn của hệ thống khi không sử dụng Ontology và kết quả truy vấn của hệ thống khi có sử dụng Ontology thông qua hai độ đo Recall (độ phủ) và Precision (độ chính xác) bằng cách nhập lần lượt 10 câu truy vấn vào hệ thống. Kết quả thực nghiệm được trình bày trong mục 4.2.2.2 như sau.

#### 4.2.2.3 Kết quả thực nghiệm

**Ta quy ước:**

**NoOn:** Khi hệ thống truy vấn không sử dụng Ontology

**On:** Khi hệ thống truy vấn có sử dụng Ontology

▪ **Độ phủ của các lượt truy vấn:**

*Bảng 4.7. Độ phủ của các lượt truy vấn*

TT	Câu truy vấn	# relevant documents in the database	#retrieved documents		% Recall	
			NoOn	On	NoOn	On
1	Kỹ thuật nuôi heo	8	6	8	75	100
2	Dịch bệnh ở gà	16	12	14	75	87,5
3	Phòng bệnh cúm gia cầm	11	8	10	73	91
4	Dịch bệnh ở lợn	20	11	15	55	75
5	Phòng bệnh đốm trắng ở tôm	5	4	4	80	80
6	Kỹ thuật nuôi cá	13	10	10	77	77
7	Kỹ thuật nuôi nhím	3	3	3	100	100
8	Kỹ thuật nuôi dê	9	6	7	66,7	77,8
9	Cách chế biến rơm làm thức ăn cho bò	15	11	13	73	86,7
10	Kỹ thuật nuôi lươn	4	3	3	75	75

▪ **Độ chính xác của các lượt truy vấn:**

*Bảng 4.8. Độ chính xác của các lượt truy vấn*

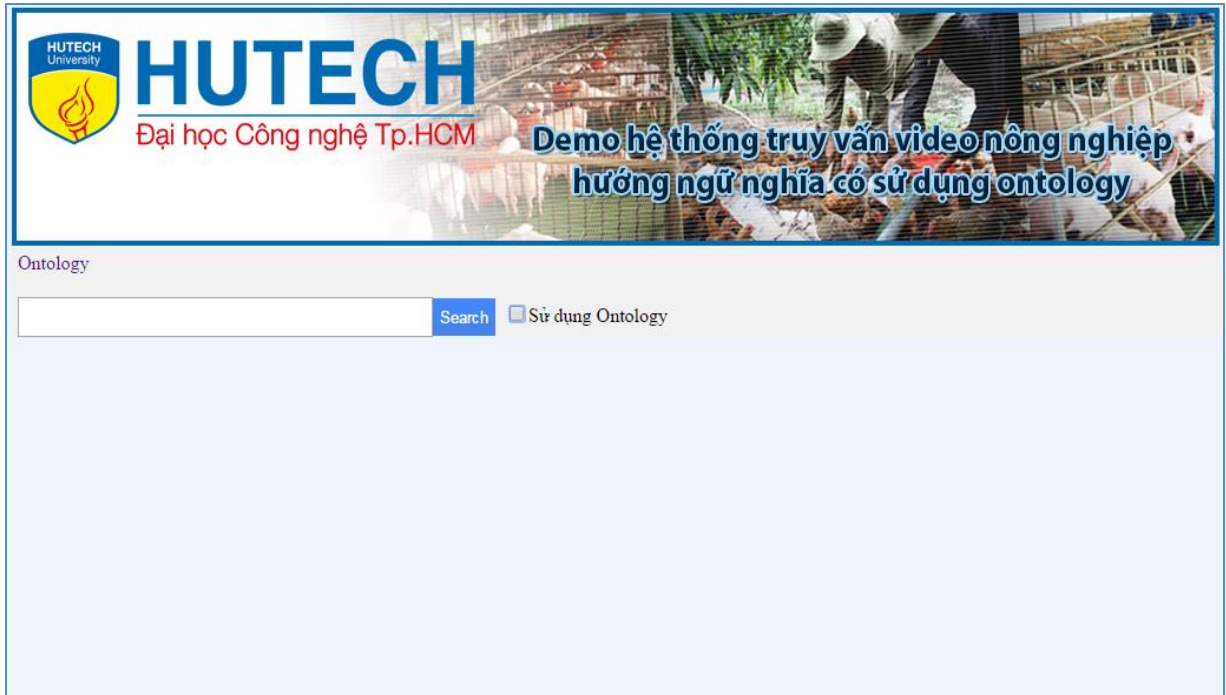
TT	Câu truy vấn	# retrieved documents		# correctly retrieved documents	% Precision	
		NoOn	On		NoOn	On
1	Kỹ thuật nuôi heo	6	8	2	33,5	25
2	Dịch bệnh ở gà	12	14	6	50	43
3	Phòng bệnh cúm gia cầm	8	10	2	25	20
4	Dịch bệnh ở lợn	11	15	2	18	13,5
5	Phòng bệnh đốm trắng ở tôm	4	4	1	25	25
6	Kỹ thuật nuôi cá	10	10	3	30	30
7	Kỹ thuật nuôi nhím	3	3	0	0	0
8	Kỹ thuật nuôi dê	6	7	2	35	29
9	Cách chế biến rom làm thức ăn cho bò	11	13	1	9	7,7
10	Kỹ thuật nuôi lươn	3	3	1	33	33

Qua các kết quả thực nghiệm trên ta nhận thấy hệ thống khi sử dụng Ontology có độ bao phủ lớn hơn so với hệ thống khi không sử dụng Ontology. Vì truy vấn thông tin dựa trên Ontology mở rộng phạm vi tìm kiếm, như những từ đồng nghĩa, từ tiếng Anh, những từ viết tắt, biến thể chính tả, tên thương mại.



### 4.2.3 Demo ứng dụng

Giao diện các trang hiển thị thông tin tìm kiếm như sau:



*Hình 4.3. Giao diện trang chủ của ứng dụng demo*

Trang chủ

1 kết quả

STT	Nội dung
1	Bệnh gia súc

**Hình 4.4. Giao diện trang tìm kiếm thuật ngữ của Ontology trong chăn nuôi**

[Trở về](#)

**NỘI DUNG CHI TIẾT**

- BỆNH GIA SÚC

- Thuật ngữ liên quan : bệnh truyền nhiễm, bệnh ở bò, bệnh ở dê, bệnh ở thỏ, bệnh ở cừu, bệnh ở lợn, bệnh ở ngựa, bệnh do vi khuẩn, bệnh do vi rút, bệnh do ký sinh trùng, bệnh do nấm

**Hình 4.5. Giao diện khái niệm thuật ngữ của Ontology trong chăn nuôi (ví dụ: Bệnh gia súc)**

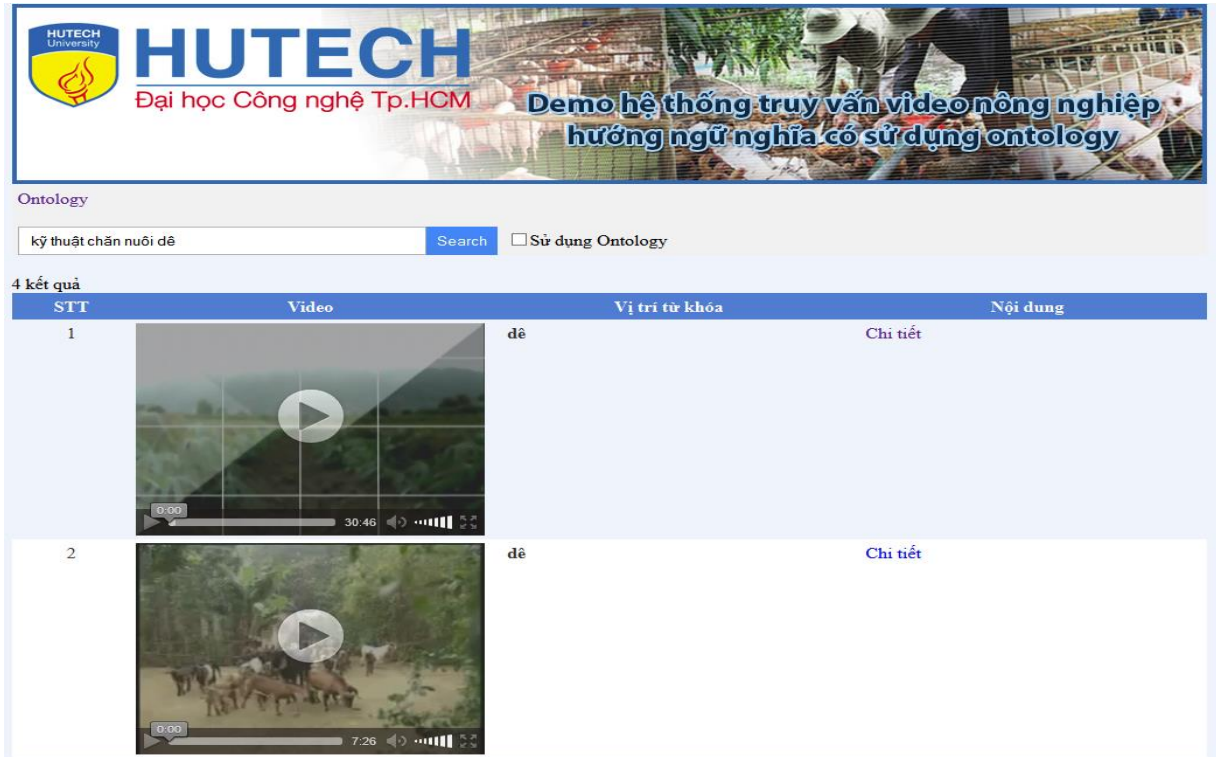
[Trở về](#)

**NỘI DUNG CHI TIẾT**

- BỆNH TRUYỀN NHIỄM

- Thuật ngữ liên quan : bệnh dịch tả lợn, bệnh giả dại ở lợn, bệnh tụ huyết trùng, bệnh đóng dấu lợn, bệnh phó thương hàn

**Hình 4.6. Giao diện khái niệm thuật ngữ của Ontology trong chăn nuôi (ví dụ: Bệnh truyền nhiễm)**





HUTECH University  
Đại học Công nghệ Tp.HCM

Demo hệ thống truy vấn video nông nghiệp hưởng ngữ nghĩa có sử dụng ontology

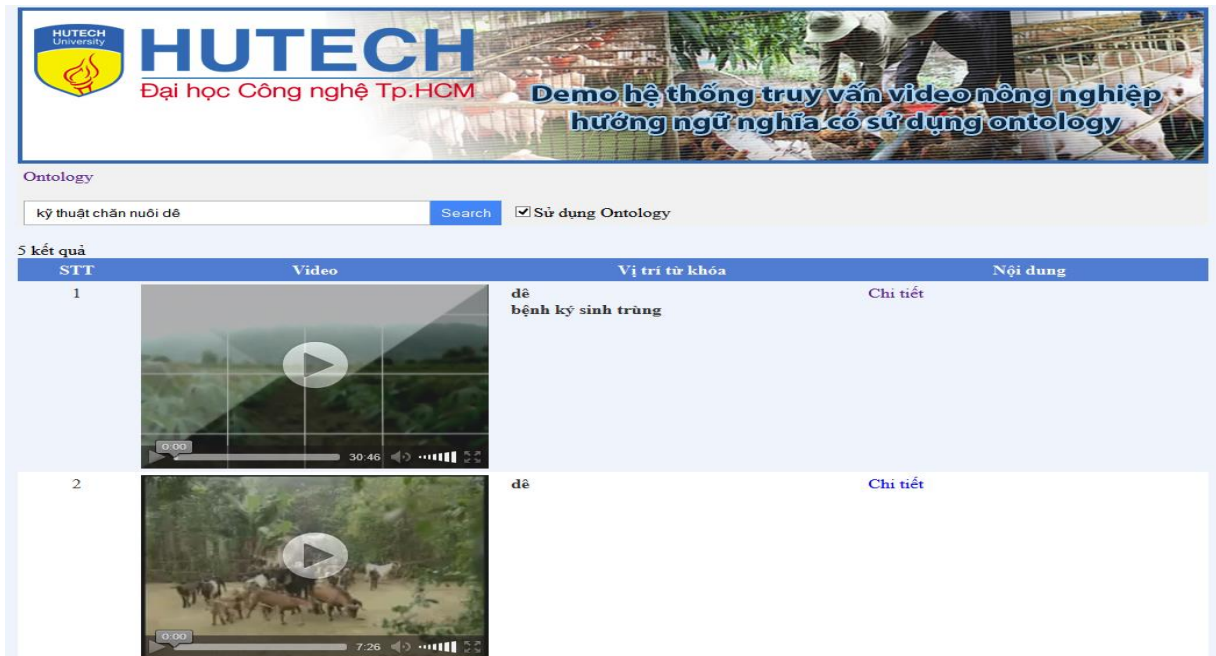
Ontology

kỹ thuật chăn nuôi dê Search  Sử dụng Ontology

4 kết quả

STT	Video	Vị trí từ khóa	Nội dung
1		dê	Chi tiết
2		dê	Chi tiết

**Hình 4.7. Giao diện trang tìm kiếm thông tin về chăn nuôi khi không sử dụng Ontology (ví dụ: Kỹ thuật chăn nuôi dê)**





HUTECH University  
Đại học Công nghệ Tp.HCM

Demo hệ thống truy vấn video nông nghiệp hưởng ngữ nghĩa có sử dụng ontology

Ontology

kỹ thuật chăn nuôi dê Search  Sử dụng Ontology

5 kết quả

STT	Video	Vị trí từ khóa	Nội dung
1		dê bệnh ký sinh trùng	Chi tiết
2		dê	Chi tiết

**Hình 4.8. Giao diện trang tìm kiếm thông tin về chăn nuôi khi có sử dụng Ontology**



[Trở về](#)

#### NỘI DUNG CHI TIẾT

đặc lác có địa hình hướng thấp dần từ đông nam sang tây bắc khí hậu toàn tỉnh được chia thành hai tiểu vùng vùng phía tây bắc có khí hậu nắng nóng khô hanh về mùa khô vùng phía đông nam có khí hậu mát mẻ ôn hòa thời tiết chia làm hai mùa khá rõ rệt là mùa mưa và mùa khô mùa mưa thường bắt đầu từ tháng năm đến tháng mười kèm theo gió tây nam thịnh hành các tháng có lượng mưa lớn nhất là tháng bảy tháng tám tháng chín lượng mưa chiếm tám mươi đến chín mươi phần trăm lượng mưa của năm riêng vùng phía đông do chịu ảnh của đồng trường sơn nên mùa mưa kéo dài hơn tới tháng mười một mùa khô từ tháng mười một đến tháng tư năm sau trong mùa này độ ẩm giảm gió đông bắc thổi mạnh bốc hơi lớn gây khô hạn nghiêm trọng lượng mưa trung bình toàn tỉnh từ một ngàn sáu trăm đến một ngàn tám trăm mi li mét năm trong vùng sinh thái có khí hậu đặc thù của tây nguyên đặc lác có đồng cỏ tự nhiên rộng và đa dạng đây là thế mạnh phát triển chăn nuôi **gia súc** nói chung về đặc điểm sinh lý của con **đê** thì có thể sống ở nhiều vùng khí hậu khác nhau thổ địa khác nhau ví dụ như là vùng sa mạc vùng núi cao vùng đồng bằng đê có thể vẫn sống được riêng ở đặc lác khí hậu tương đối là ôn hòa mùa nóng mùa nóng mùa lạnh chênh lệch nhau không quá cao nuôi đê là tương đối thích hợp và đặc biệt là những điều kiện như đất đai thổ nhưỡng và điều kiện về **thức ăn** ở đặc lác rất phù hợp với cả điều kiện cho con đê sinh trưởng và phát triển một số hộ thì người ta nuôi người ta tận dụng ở cái vùng đất rộng để người ta chăn thả còn một số hộ họ có thể nuôi thâm canh nuôi nhốt thì các điều kiện này đều phù hợp với đặc điểm sinh lý của đê địa hình huyện rộng bắc tương đối bằng phẳng có độ cao trung bình năm trăm mét so với mực nước biển thấp dần từ tây bắc xuống đông nam nhìn chung địa hình có đặc trưng kiểu vùng cao nguyên dãy đồi lượn sóng vùng núi thấp sườn dốc và vùng trũng thấp khí hậu nhiệt đới gió mùa thuận lợi cả về địa hình và thời tiết phù hợp với

**Hình 4.9. Giao diện trang hiển thị nội dung chi tiết của video**

## **CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN**

### **5.1 Kết luận**

#### **5.1.1 Tổng kết**

Luận văn trình bày hệ thống truy vấn thông tin video nông nghiệp hướng ngữ nghĩa. Tính ngữ nghĩa của hệ thống được thể hiện qua việc sử dụng từ, cụm từ nhận dạng được từ kênh âm thanh của video để lập chỉ mục tìm kiếm cho các đoạn video đó.

Áp dụng Ontology vào hệ thống truy vấn thông tin video giúp mở rộng câu truy vấn của người dùng từ đó mở rộng và nâng cao chất lượng tìm kiếm, bên cạnh đó một số lợi ích mà bộ Ontology mang lại như tính dễ mở rộng, tính tái sử dụng.

#### **5.1.2 Những đóng góp của đề tài**

- Phát triển bộ nhận dạng tiếng nói cho lĩnh vực chăn nuôi trong nông nghiệp Việt Nam.

- Xây dựng một Ontology cho lĩnh vực chăn nuôi.

- Tích hợp được một hệ thống truy vấn video về chăn nuôi trong nông nghiệp Việt Nam hướng ngữ nghĩa có sử dụng Ontology.

- Đóng góp bộ dữ liệu huấn luyện cho mô-đun nhận dạng tiếng nói tiếng Việt trong lĩnh vực chăn nuôi với tổng thời gian 4 giờ 34 phút 47 giây gồm 950 file của 44297 từ.

#### **5.1.3 Hạn chế của đề tài**

- Hệ thống truy vấn video trong phạm vi hẹp là lĩnh vực chăn nuôi của Việt Nam.

- Nguồn thu dữ liệu khó khăn, chất lượng video không cao dẫn đến khó khăn cho việc huấn luyện dữ liệu.

- Với mô hình nhận dạng tiếng nói đang dùng, cần phải có bộ dữ liệu huấn luyện khá lớn mới có thể đảm bảo độ chính xác khi nhận dạng.

- Chưa có Ontology nào cho lĩnh vực chăn nuôi của Việt Nam trước đây, vì vậy khi xây dựng Ontology cho đề tài gặp nhiều khó khăn khi không có nền tảng để kế thừa. Bản thân người xây dựng cũng không có chuyên môn trong lĩnh vực chăn nuôi mà tự tìm hiểu tài liệu trên Internet. Vì vậy xây dựng chưa đầy đủ các thuật ngữ, khái niệm trong Ontology, do đó chương trình ứng dụng thể hiện ở mức cơ bản của Ontology trong lĩnh vực chăn nuôi.

## 5.2 Hướng phát triển

Thông qua kết quả đã đạt được cùng với những hạn chế đã nêu, một số vấn đề có thể được xem xét nghiên cứu và phát triển xa hơn nhằm nâng cao hiệu suất nhận dạng và tìm kiếm hiệu quả hơn:

- Mở rộng tập dữ liệu huấn luyện để tăng độ phong phú của dữ liệu huấn luyện giúp cho mô hình nhận dạng đạt kết quả chuẩn xác hơn.

- Mở rộng dữ liệu truy vấn cho các lĩnh vực khác như trồng trọt, y tế, v.v...

- Phát triển bộ nhận dạng giọng nói tốt hơn sao cho bộ nhận dạng không phụ thuộc vào bộ dữ liệu huấn luyện với số lượng lớn.

- Hoàn thiện Ontology chăn nuôi để hỗ trợ quá trình truy vấn thông tin về chăn nuôi hiệu quả hơn, từ đó làm cơ sở để phát triển Ontology sang các lĩnh vực khác.

## TÀI LIỆU THAM KHẢO

- [1] Arisoy, E., Can, D., Parlak, S., Sak, H., & Saraçlar, M (2009), Turkish broadcast news transcription and retrieval. *Audio, Speech, and Language Processing, IEEE Transactions on*, 17(5), 874-883.
- [2] Chelba, C., Hazen, T. J., & Saraçlar, M (2008), Retrieval and browsing of spoken content. *Signal Processing Magazine, IEEE*, 25(3), 39-49.
- [3] Van Thong, J. M., Moreno, P. J., Logan, B., Fidler, B., Maffey, K., & Moores, M (2002), Speechbot: an experimental speech-based search engine for multimedia content on the web. *Multimedia, IEEE Transactions on*, 4(1), 88-96.
- [4] Carpineto, Claudio, Giovanni, Romano (2012), A Survey of Automatic Query Expansion in Information Retrieval. *ACM Comput. Surv. (CSUR)* 44(1).
- [5] Qiu, Y. and Frei, H.-P (1993), Concept based query expansion. In *Proceedings of the 16<sup>th</sup> Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM Press, Pittsburgh, Pennsylvania, USA, 160–169.
- [6] Bast, H., Majumdar, D., and Weber (2007), Efficient interactive query expansion with complete search. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM Press, Amsterdam, The Netherlands, 857–860.
- [7] Curran, J. R., Moens, M (2002). Improvements in Automatic Thesaurus Extraction. In *Proceedings of the ACL-02 Workshop on Unsupervised Lexical Acquisition*, vol. 9, pp. 59--66. Association for Computational Linguistics.



- [8] Varelas, G., Voutsakis, E., Raftopoulou, P., Petrakis, E. G., & Milios, E. E (2005, November). Semantic similarity methods in WordNet and their application to information retrieval on the web. In Proceedings of the 7th annual ACM international workshop on Web information and data management (pp. 10-16). ACM.
- [9] Müller, H. M., Kenny, E. E., Sternberg, P. W. Textpresso (2004), An Ontology-Based Information Retrieval and Extraction System for Biological Literature. PLoS Biology, 2(11), e309.
- [10] Abasolo, J. M., Gomez, M. MELISA (2000), An Ontology-Based Agent for Information Retrieval in Medicine. In Proceedings of the first international workshop on the semantic web (SemWeb2000), pp. 73—82.
- [11] Salton, G. (1989), Automatic Text Processing The transformation Analysis and Retrieval of Information by Computer. Addison-Wesley Longman Publishing: Boston, MA, USA p. 530.
- [12] Kowalski, G.J. and M.T. Maybury (2000), Information Storage and Retrieval Systems: Theory and Implementation. 2nd ed.
- [13] Nguyen Tuan Dang, Nguyen Phi Minh Tri (2010), Building a Universal Ontology for Vietnamese Language, Faculty of Computer Science, University of Information Technolog.
- [14] Ha Nguyen, Quan Vu (Feb 2006), “Selection of Basic Units for Vietnamese Large Vocabulary Continuous Speech Recognition”, The 4th IEEE International Conference on Computer Science - Research, Innovation and Vision of the Future, HoChiMinh City, Vietnam.



- [15] Vu Thang Tat, Nguyen Dung Tien, Luong Mai Chi, Hosom John Paul (2005), Vietnamese large vocabulary continuous speech recognition, Eurospeech, Lisbon, Portugal, p1689-1692.
- [16] Natalya F. Noy and Deborah L. McGuinness (2001), *Ontology Development 101: A Guide to Creating Your First Ontology*, Stanford University, Stanford, CA, 9430.
- [17] Christopher D.Manning, Prabhakar Raghavan, Hinrich Schutze (2008), *An Introduction to Information Retrieval*, Cambridge University Press.
- [18] Vallez, M. and R. Pedraza-Jimenez (2007), *Natural Language Processing in Textual Information Retrieval and Related Topics*, I.S.S.o.t.P.F. University, Editor.
- [19] Salton, G. and M.J. McGill (1986), *Introduction to Modern Information Retrieval*, New York, NY, USA: McGraw-Hill, Inc.
- [20] Haav, H.-M. and T.-L. Lubi (2001), *A Survey of Concept-based Information Retrieval Tools on the Web*, in 5th East-European Conference.
- [21] Kowalski, G.J. and M.T. Maybury (2000: Springer), *Information Storage and Retrieval Systems: Theory and Implementation*. 2nd ed.
- [22] Lê Thúy Ngọc (2008), *Xây dựng hệ tìm kiếm thông tin theo hướng tiếp cận ngữ nghĩa*. Luận văn thạc sĩ, chuyên ngành Khoa học Máy tính, Trường ĐH Khoa học Tự nhiên, TP. Hồ Chí Minh.
- [23] Hệ thống truy vấn video dựa vào tên, chú thích, [www.youtube.com](http://www.youtube.com) (2014)
- [24] Wikipedia (2014)
- [25] <http://vcn.vnn.vn> (2014)
- [26] <http://www.2lua.vn/article/nen-nuoi-giong-heo-nao-nhieu-nac>