

**NHẬP MÔN CÔNG NGHỆ THÔNG TIN**

# ĐỒ ÁN MÔN HỌC

Tài liệu này mô tả đồ án môn học cho môn học theo hướng chuyên ngành Khoa học máy.



Khoa Công nghệ Thông tin  
Đại học Khoa học Tự nhiên TP HCM

# MỤC LỤC

Các nội dung chính .....	1
1 Nội dung đề án .....	2
2 Các công cụ gợi ý .....	4

# ĐỒ ÁN MÔN HỌC

## Các nội dung chính

- Mô tả nội dung đồ án và các kết quả cần đạt được
- Hướng dẫn thực hiện đồ án theo tuần cho giảng viên và sinh viên
- Cách thức đánh giá

# 1

## Nội dung đồ án

### Nội dung chính

Xây dựng một video demo cho một hệ thống tìm kiếm thông tin bằng hình ảnh văn bản. Đồng thời, sinh viên tìm hiểu nội dung kiến thức liên quan.

### Mục tiêu đồ án

Sinh viên thực hiện đồ án sẽ nắm được các kiến thức:

- Kiến thức cơ bản cho một số chủ đề thuộc 3 chuyên ngành KHMT.
- Cách tiến hành thực hiện một dự án theo mô hình linh hoạt.
- Một số công cụ hỗ trợ làm việc cộng tác trên môi trường Internet.
- Xây dựng mô hình sản phẩm theo yêu cầu.

Sinh viên thực hiện đồ án sẽ rèn luyện các kỹ năng:

- Lập kế hoạch, làm việc nhóm, tổ chức và quản lý nhóm
- Tìm hiểu và viết báo cáo
- Đọc hiểu tiếng Anh và các tài liệu chuyên ngành

Xây dựng video demo cho một hệ thống tìm kiếm thông tin bằng hình ảnh văn bản. Đồng thời sinh viên tìm hiểu các nội dung kiến thức liên quan. Biết rằng hệ thống hoạt động theo mô hình sau:

- Đầu vào là một tập tin ảnh của một từ, câu hoặc một đoạn văn bản(text). Đầu ra là các tài liệu có nội dung liên quan đến nội dung của văn bản được tìm thấy từ Internet.
- Hệ thống sẽ chuyển tập tin hình ảnh thành dạng văn bản bằng công cụ OCR. Tiếp theo, văn bản rút được sẽ được sử dụng trực tiếp (đối với trường hợp chỉ là một từ hoặc câu ngắn) hoặc phải qua rút trích từ khóa (đối với trường hợp là một đoạn văn bản) để tìm các tài liệu từ Internet. Việc tìm kiếm các tài liệu này được thực hiện bởi các chương trình tìm kiếm như Google, Yahoo!...

Video demo cần thể hiện được:

- Đầy đủ các bước được đề cập phía trên và theo đúng trình tự.
- Ở mỗi bước, cần thể hiện được kĩ thuật nào được sử dụng, cũng như đầu vào và đầu ra ở mỗi bước.
- Video cần dễ hiểu và trực quan sinh động.

Các nội dung kiến thức sinh viên cần tìm hiểu:

- OCR (optical character recognition)
- Rút trích các từ khóa, là các từ chính thể hiện nội dung văn bản (keyword extractor)
- So sánh 2 công cụ tìm kiếm, ví dụ Google và Yahoo!

### Đánh giá

<b>ĐÁNH GIÁ SẢN PHẨM ĐỀ TÀI A (50% điểm của đề tài)</b>				
	Yếu (0 -49%)	TB(50 – 74%)	Khá (75-89%)	Tốt (90 – 100%)
OCR - Nhận dạng chữ viết (25%)	Biết được các khái niệm cơ bản, ứng dụng	Hiểu được các khái niệm cơ bản, ứng dụng. Liệt kê được phương pháp OCR	Hiểu được các khái niệm, ứng dụng Trình bày các phương pháp OCR	Hiểu các khái niệm, ứng dụng. Trình bày các phương pháp OCR Thử nghiệm các công cụ OCR(có minh chứng)
Rút trích từ khóa (20%)	Biết được các khái niệm cơ bản	Hiểu được các khái niệm cơ bản	Hiểu được các khái niệm. Biết các phương pháp rút trích từ khóa	Hiểu các khái niệm, Biết các phương pháp rút trích từ khóa Thử nghiệm các công cụ (có minh chứng)
So sánh công cụ tìm kiếm (20%)	So sánh được, nhưng số lượng tiêu chí $\leq 3$	So sánh được, nhưng số lượng tiêu chí $\leq 5$	So sánh được, nhưng số lượng tiêu chí $> 5$	So sánh được, nhưng số lượng tiêu chí $> 5$ .

				Có số liệu rõ ràng
Video demo (35%)	Thể hiện đủ các bước của mô hình Hình ảnh nhiều chỗ không phù hợp với nội dung	Thể hiện đủ các bước mô hình Nêu rõ các kĩ thuật sử dụng ở mỗi bước Hình ảnh phù hợp với nội dung Sử dụng hoàn toàn hình ảnh không tự quay	Thể hiện đủ các bước mô hình Nêu rõ các kĩ thuật sử dụng ở mỗi bước. Hình ảnh phù hợp với nội dung <50% hình ảnh sử dụng là tự quay	Thể hiện đủ các bước mô hình Nêu rõ các kĩ thuật sử dụng ở mỗi bước. Hình ảnh phù hợp với nội dung > 50% hình ảnh sử dụng là tự quay

## 2 Các công cụ gợi ý

### A. Các công cụ xử lý hình ảnh

a. IrfanView

b. MS Paint

### B. Các công cụ nhận dạng ảnh

c. VietOCR

A Java/.NET GUI frontend for Tesseract OCR engine. Supports optical character recognition for Vietnamese and other languages supported by Tesseract.

VietOCR is released and distributed under the [Apache License, v2.0](http://vietocr.sourceforge.net/). <http://vietocr.sourceforge.net/>

d. Other OCR tools

các tool OCR tham khảo. Sinh viên được khuyến khích sử dụng và so sánh kết quả của các tool khác nhau. Đa phần các tool này cho phép dùng thử. Sinh viên cần submit thông tin để download bản trial.

[http://en.wikipedia.org/wiki/List\\_of\\_optical\\_character\\_recognition\\_software](http://en.wikipedia.org/wiki/List_of_optical_character_recognition_software)

## C. Các công cụ xử lý phim

### e. Window Movie Maker

Window Live Movie Maker - có sẵn trong máy Windows

### f. Adobe Premiere

Adobe Premiere Pro is a timeline-based video editing software application. It is part of the Adobe Creative Suite, a suite of graphic design, video editing and web

[http://www.adobe.com/cfusion/tdrc/index.cfm?product=premiere\\_elements&loc=us](http://www.adobe.com/cfusion/tdrc/index.cfm?product=premiere_elements&loc=us)

### g. Other video\_editing\_softwares

Các tool tham khảo. Sinh viên được khuyến khích sử dụng và so sánh độ tiện dụng và hiệu quả của các tool khác nhau. Một số tool miễn phí, một số trial. Sinh viên cần submit thông tin để download bản trial.

[http://en.wikipedia.org/wiki/Comparison\\_of\\_video\\_editing\\_software](http://en.wikipedia.org/wiki/Comparison_of_video_editing_software)

### h. Adobe After Effects, tool Roto Brush.

Adobe After Effects is a digital motion graphics and compositing software package published by Adobe Systems, used in the post-production process of filmmaking and television production. Its main uses are the origination of 2D and 2.5D animation, visual effects compositing and finishing (image adjustment, color correction etc.). After Effects can also be used as a basic non-linear editor and a media transcoder.

[http://www.adobe.com/cfusion/tdrc/index.cfm?product=after\\_effects](http://www.adobe.com/cfusion/tdrc/index.cfm?product=after_effects)

### i. Một số tool tham khảo khác.

[http://en.wikipedia.org/wiki/Adobe\\_After\\_Effects#Comparable\\_products](http://en.wikipedia.org/wiki/Adobe_After_Effects#Comparable_products)

## D. Các công cụ xử lý âm thanh



## j. Jet Audio



JetAudio là một trong những trình nghe nhạc vào loại bậc nhất hiện nay, có khả năng chơi hầu hết các định dạng phổ biến và các định dạng khác, JetAudio còn có riêng một bộ Equalizer của riêng mình làm tăng chất lượng âm thanh khi chơi bằng JetAudio...

## k. Audacity



Audacity® là một phần mềm nguồn mở, tự do cho việc ghi âm và chỉnh sửa âm thanh. Rất nhỏ gọn và nhanh.

<http://audacity.sourceforge.net/>

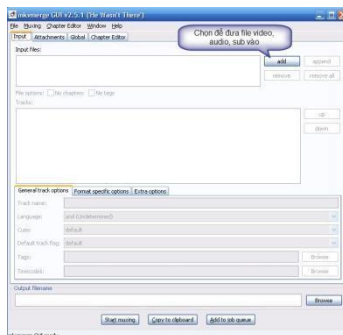
<http://vi.wikipedia.org/wiki/Audacity>

[http://www.youtube.com/watch?v=3TyF\\_vukSVY](http://www.youtube.com/watch?v=3TyF_vukSVY)

<http://www.youtube.com/watch?v=5rQr5YVNqI>

0

## l. MKVToolNix



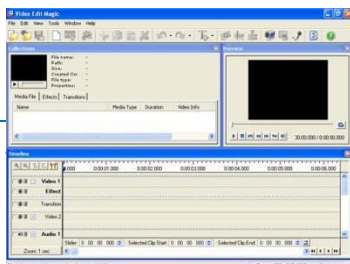
MKVToolNix cho phép lấy phần audio của bộ phim chất lượng kém hơn ghép qua phim chất lượng cao của bạn. Tất cả các yêu cầu trên có thể thực hiện dễ dàng với MKVToolNix. Dù tên là MKVToolNix nhưng

MKVToolNix hỗ trợ hầu hết các định dạng video, audio và sub hiện có, hơn nữa, đây là công cụ miễn phí 100%.

<http://www.bunkus.org/videotools/mkvtoolnix/downloads.html#windows>

<http://forum.taiphannamem.org/index.php?topic=605.0>

## m. Video Edit magic



Khi bạn có nhiều đoạn phim và muốn nối lại với nhau bằng các hiệu ứng



độc đáo hay chèn chữ giới thiệu, Video Edit Magic là một công cụ hữu ích. Phần mềm này chỉ có 12,8 MB nên không yêu cầu máy tính cấu hình cao. Ngoài ra, Video Edit Magic còn có thể chụp hình từ webcam và chèn luôn vào đoạn phim đang chỉnh sửa. Bản dùng thử 30 ngày sẽ chèn đoạn chữ quảng cáo vào sản phẩm cuối cùng. Phần hướng dẫn được thực hiện dưới

dạng flash giúp người dùng dễ dàng theo dõi.

[http://www.download.com.vn/audio+video/video+editor+studio/5020\\_video-edit-magic.aspx](http://www.download.com.vn/audio+video/video+editor+studio/5020_video-edit-magic.aspx)

#### n. Audio To Video Mixer

Công cụ "Audio to Video Mixer" sẽ giúp bạn thực hiện việc lồng âm thanh vào video số chỉ bằng một vài cú click chuột.



<http://www.download.com.vn/timkiem/Audio+to+Video+Mixer/index.aspx>

<http://diendan.yeulaptop.com/phan-mem-long-am-thanh-vao-video-audio-video-mixer-3-0-7-a-15524.html>

## E. Các công cụ xử lý ngôn ngữ tự nhiên

#### o. VLSP

Hệ thống trình diễn một số sản phẩm của nhánh đề tài "Xử lý văn bản" là một phần của đề tài KC01.01/06-10 "Nghiên cứu phát triển một số sản phẩm thiết yếu về xử lý tiếng nói và văn bản tiếng Việt" (VLSP).

Hệ thống tích hợp sản phẩm do Nguyễn Việt Cường và Nguyễn Lê Minh (JAIST) phát triển. Các công cụ xử lý tiếng Việt do các nhóm thực hiện cung cấp. Hệ thống tích hợp sử dụng các công cụ mã nguồn mở: PHP, MySQL, Smarty, Snoopy, WZToolTip, Mudim, Zapatec.

<http://vlsp.vietlp.org:8080/demo/>

#### p. Stanford NLP



The Stanford NLP Group makes parts of our Natural Language Processing software available to everyone. These are statistical NLP toolkits for various major computational linguistics

problems. They can be incorporated into applications with human language technology needs.

<http://nlp.stanford.edu:8080/parser/>

## F. Các công cụ mã hóa và ẩn dữ liệu

### q. as3crypto

Cryptography library for ActionScript 3, including partial TLS 1.0 support <http://crypto.hurlant.com/demo/>

r. Javascript implementation of DES <http://www.tero.co.uk/des/test.php>

s. AspEncrypt Demo 2: Text Encryption/Decryption  
[http://support.persits.com/encrypt/demo\\_text.asp](http://support.persits.com/encrypt/demo_text.asp)

### t. Secret Key Encryption Demo

<http://www.codeproject.com/Articles/28112/Secret-Key-Encryption-Demo>

### u. RSA Demo Applet

<http://cisnet.baruch.cuny.edu/holowczak/classes/9444/rsademo/>

### v. MSU StegoVideo

Unique tool for hiding information in video (VirtualDub filter/standalone executable)

[\*MSU Graphics & Media Lab \(Video Group\)\*](#)

MSU StegoVideo allows hiding any file in a video sequence.

Different popular codecs were analyzed and an algorithm, providing the smallest data loss after compression, was chosen. Convolutional codes with Viterbi decoding are used to correct occurred errors.

[http://compression.ru/video/stego\\_video/index\\_en.html](http://compression.ru/video/stego_video/index_en.html)

**Ngoài ra, cũng có thể sử dụng (thêm) các phần mềm hợp pháp khác phù hợp với công việc.**