

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

Đào Đức Anh - Nguyễn Thành Nhân

Xây dựng hệ thống gợi ý sản phẩm
dựa trên mô hình Auto-Encoder

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN
CHƯƠNG TRÌNH CHÍNH QUY

Tp. Hồ Chí Minh, tháng MM/YYYY

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

Đào Đức Anh - 1712270

Nguyễn Thành Nhân - 1712631

**Xây dựng hệ thống gợi ý sản phẩm
dựa trên mô hình Auto-Encoder**

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN
CHƯƠNG TRÌNH CHÍNH QUY

GIÁO VIÊN HƯỚNG DẪN

ThS. Trần Trung Kiên

Tp. Hồ Chí Minh, tháng MM/YYYY

Lời cảm ơn

Tôi xin chân thành cảm ơn ...

Mục lục

Lời cảm ơn	i
Đề cương chi tiết	ii
Mục lục	ii
Tóm tắt	v
1 Ngôn ngữ	1
2 Trình bày báo cáo	2
2.1 Quy định chung	2
2.2 Bố cục của báo cáo	2
2.3 Bảng biểu, hình vẽ, phương trình	4
2.3.1 Bảng biểu, hình vẽ	4
2.3.2 Phương trình toán học	5
2.4 Viết tắt	5
2.5 Tài liệu tham khảo và cách trích dẫn	6
3 Hướng dẫn sử dụng template	7
3.1 Trích dẫn tài liệu	7
3.2 Chèn mã nguồn	7
3.3 Hình ảnh	8
3.4 Bảng biểu	9
3.5 Công thức	9

Danh mục công trình của tác giả	11
Tài liệu tham khảo	12
A Ngữ pháp tiếng Việt	14
B Ngữ pháp tiếng Nôm	15

Danh sách hình

3.1	Hình ví dụ 1	9
3.2	Hình ví dụ 2	9

Danh sách bảng

3.1	Bảng ví dụ 1	10
-----	------------------------	----

Chương 1

Ngôn ngữ

Ngôn ngữ để viết và trình bày báo cáo khóa luận tốt nghiệp, đồ án tốt nghiệp, thực tập tốt nghiệp (sau đây gọi chung là báo cáo) là tiếng Việt hoặc tiếng Anh. Trường hợp chọn ngôn ngữ tiếng Anh để viết và trình bày báo cáo, sinh viên cần có đơn đề nghị, được cán bộ hướng dẫn (CBHD) đồng ý và nộp cho bộ phận Giáo vụ của Khoa vào thời điểm đăng ký đề tài để xin ý kiến. Báo cáo viết và trình bày bằng tiếng Anh phải có bản tóm tắt viết bằng tiếng Việt.

Chương 2

Trình bày báo cáo

2.1 Quy định chung

Báo cáo phải được trình bày ngắn gọn, rõ ràng, mạch lạc, sạch sẽ, không được tẩy xóa, có đánh số trang, đánh số bảng biểu, hình vẽ, đồ thị.

Nội dung báo cáo được phân thành các chương. Số thứ tự của các chương, mục được đánh số bằng hệ thống số Ả-rập, không dùng số La mã. Các mục và tiểu mục được đánh số bằng các nhóm hai hoặc ba chữ số, cách nhau một dấu chấm: số thứ nhất chỉ số chương, chỉ số thứ hai chỉ số mục, số thứ ba chỉ số tiểu mục.

Báo cáo trình bày sử dụng khổ giấy với việc canh lề như sau: Lề trên 3 cm, lề dưới 2,5 cm, lề trái 3 cm, lề phải 2 cm. Đánh số trang ở giữa bên dưới. Đánh số trang ở giữa bên dưới.

Font chữ dùng trong báo cáo (Times New Roman) với kích cỡ (size) 13-14pt, sử dụng chế độ dẫn dòng (line spacing) chế độ 1.5 lines.

2.2 Bố cục của báo cáo

Nội dung của báo cáo tối thiểu 50 trang khổ A4 và không nên vượt quá 100 trang (không kể các trang bìa, lời cảm ơn, mục lục, tài liệu tham khảo ...) theo trình tự như sau:

- **MỞ ĐẦU** (thường đặt tên là “Giới thiệu”): Trình bày lý do chọn đề tài, mục đích, đối tượng và phạm vi nghiên cứu. Mô tả bài toán mà đề tài giải quyết. Bài toán này có gì hay? Tại sao lại cần giải quyết bài toán này? Bài toán này có gì khó? Có những hướng nào để giải quyết bài toán này? Những hướng giải quyết trước đây có những vấn đề gì chưa giải quyết được? Các câu hỏi nghiên cứu mà đề tài trả lời hoặc những vấn đề mà đề tài sẽ giải quyết. Các đóng góp của đề tài.
- **TỔNG QUAN** (thường đặt tên là “Các công trình liên quan”): Phân tích đánh giá các hướng nghiên cứu đã có của các tác giả trong và ngoài nước liên quan đến đề tài; nêu những vấn đề còn tồn tại (những vấn đề nào mà các công trình khác chưa giải quyết được); chỉ ra những vấn đề mà đề tài cần tập trung, nghiên cứu giải quyết.
- **NGHIÊN CỨU THỰC NGHIỆM HOẶC LÝ THUYẾT** (thường đặt tên là “Phương pháp đề xuất”): Trình bày cơ sở lý thuyết, lý luận, giả thiết khoa học và phương pháp nghiên cứu đã được sử dụng trong đề tài.

Nếu đề xuất hướng giải quyết mới, mô hình mới thì cần mô tả chi tiết cách giải quyết của mình (chi tiết tới mức người khác có thể dựa vào phần này mà cài đặt lại được đúng hoàn toàn phương pháp của mình đề ra).

- **TRÌNH BÀY, ĐÁNH GIÁ BÀN LUẬN VỀ CÁC KẾT QUẢ** (thường đặt tên là “Kết quả thí nghiệm”): Mô tả các kết quả nghiên cứu khoa học hoặc kết quả thực nghiệm. Đối với đề tài ứng dụng có kết quả là sản phẩm phần mềm phải có hồ sơ thiết kế, cài đặt,... theo một trong các mô hình đã học (UML,...).

Thông thường cần mô tả môi trường thí nghiệm trước như sử dụng dữ liệu nào, dùng độ đo nào để đánh giá, môi trường chạy thí nghiệm (cấu hình máy nếu cần phân tích thông tin về thời gian chạy thực

nghiệm). Sau đó, nêu kết quả thực nghiệm, bàn luận và giải thích kết quả.

- **KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN** (thường đặt tên là “Kết luận”): Trình bày những kết quả đạt được, những đóng góp mới và những đề xuất mới, kiến nghị về những hướng nghiên cứu tiếp theo.
- **DANH MỤC TÀI LIỆU THAM KHẢO**: Chỉ bao gồm các tài liệu được trích dẫn, sử dụng và đề cập tới để bàn luận trong báo cáo. Phần này các bạn chuẩn bị 1 file BIB để lưu các tài liệu trích dẫn. Khi các bạn trích dẫn một tài liệu nào đó, LaTeX sẽ tự động thêm vào danh mục tài liệu tham khảo giúp các bạn. Các bạn xem hướng dẫn cách trích dẫn ở chương sau.
- **PHỤ LỤC**: Phần này bao gồm nội dung cần thiết nhằm minh họa hoặc hỗ trợ cho nội dung báo cáo như số liệu, mẫu biểu, tranh ảnh, ... Phụ lục không được dày hơn phần chính của báo cáo. Nếu có công trình công bố thì để vào phần phụ lục này.

2.3 Bảng biểu, hình vẽ, phương trình

Việc đánh số bảng biểu, hình vẽ, phương trình phải gắn với số chương; ví dụ hình 3.4 có nghĩa là hình thứ 4 trong Chương 3. Mọi đồ thị, bảng biểu, hình vẽ lấy từ các nguồn khác phải được trích dẫn đầy đủ.

2.3.1 Bảng biểu, hình vẽ

Đầu đề của bảng biểu ghi phía trên bảng, đầu đề của hình vẽ ghi phía dưới hình.

Thông thường, những bảng ngắn và đồ thị phải đi liền với phần nội dung đề cập tới các bảng và đồ thị này ở lần thứ nhất. Các bảng dài có thể để ở những trang riêng nhưng cũng phải tiếp theo ngay phần nội dung

đề cập tới bảng này ở lần đầu tiên. Các bảng rộng vẫn nên trình bày theo chiều đứng dài 297mm của trang giấy, chiều rộng của trang giấy có thể hơn 210mm. Chú ý gấp trang giấy sao cho số và đầu đề của hình vẽ hoặc bảng vẫn có thể nhìn thấy ngay mà không cần mở rộng tờ giấy. Tuy nhiên hạn chế sử dụng các bảng quá rộng này.

Đối với những trang giấy có chiều đứng hơn 297mm (bản đồ, bản vẽ,...) thì có thể để trong một phong bì cứng dính bên trong bìa sau của báo cáo.

Các hình vẽ phải sạch sẽ bằng mực đen để có thể sao chụp lại; có đánh số và ghi đầy đủ đầu đề, cỡ chữ phải bằng cỡ chữ sử dụng trong báo cáo.

Khi đề cập đến các bảng biểu và hình vẽ phải nêu rõ số của hình và bảng biểu đó, ví dụ “... được nêu trong Bảng 4.1” hoặc “xem Hình 3.2” mà không được viết “... được nêu trong bảng dưới đây” hoặc “trong đồ thị của X và Y sau”.

2.3.2 Phương trình toán học

Việc trình bày phương trình toán học trên một dòng đơn hoặc dòng kép tùy ý, tuy nhiên phải thống nhất trong toàn báo cáo.

Khi ký hiệu xuất hiện lần đầu tiên thì phải giải thích và đơn vị tính phải đi kèm ngay trong phương trình có ký hiệu đó. Nếu cần thiết, danh mục của tất cả các ký hiệu, chữ viết tắt và nghĩa của chúng cần được liệt kê và để ở phần đầu của báo cáo.

Tất cả các phương trình cần được đánh số và để trong ngoặc đơn đặt bên phía lề phải. Nếu một nhóm phương trình mang cùng một số thì những số này cũng được để trong ngoặc, hoặc mỗi phương trình trong nhóm phương trình (5.1) có thể được đánh số là (5.1.1), (5.1.2), (5.1.3).

2.4 Viết tắt

Không lạm dụng việc viết tắt trong báo cáo. Chỉ viết tắt những từ, cụm từ hoặc thuật ngữ được sử dụng nhiều lần trong báo cáo. Không

viết tắt những cụm từ dài, những mệnh đề; không viết tắt những cụm từ ít xuất hiện trong báo cáo. Nếu cần viết tắt những từ thuật ngữ, tên các cơ quan, tổ chức,... thì được viết tắt sau lần viết thứ nhất có kèm theo chữ viết tắt trong ngoặc đơn. Nếu báo cáo có nhiều chữ viết tắt thì phải có bảng danh mục các chữ viết tắt (xếp theo thứ tự ABC) ở phần đầu báo cáo.

Nhắc lại: **không lạm dụng việc viết tắt** trong báo cáo. Khi các bạn sử dụng từ viết tắt, người đọc sẽ phải lật lại những phần đã đọc, để tìm lại xem từ viết tắt đó nghĩa là gì. Việc này sẽ làm chậm tốc độ đọc và sẽ khiến người đọc khó theo dõi báo cáo của bạn hơn. Nếu có thể, hạn chế hoàn toàn việc dùng viết tắt.

2.5 Tài liệu tham khảo và cách trích dẫn

Mọi ý kiến, khái niệm có ý nghĩa, mang tính chất gợi ý không phải của riêng tác giả và mọi tham khảo khác phải được trích dẫn và chỉ ra nguồn trong danh mục Tài liệu tham khảo của báo cáo. Nguồn được trích dẫn phải được liệt kê chính xác trong danh mục Tài liệu tham khảo.

Việc trích dẫn, tham khảo chủ yếu nhằm thừa nhận nguồn của những ý tưởng có giá trị giúp người đọc theo được mạch suy nghĩ của tác giả, không làm trở ngại việc đọc.

Không trích dẫn những kiến thức phổ biến, mọi người đều biết cũng như không làm báo cáo nặng nề với những tham khảo trích dẫn.

Nếu không có điều kiện tiếp cận được một tài liệu gốc mà phải trích dẫn thông qua một tài liệu khác thì phải nêu ra trích dẫn này, đồng thời tài liệu gốc đó không được liệt kê trong danh mục tài liệu tham khảo của báo cáo.

Chương 3

Hướng dẫn sử dụng template

3.1 Trích dẫn tài liệu

Dùng lệnh `\cite` để trích dẫn một hoặc nhiều tài liệu tham khảo. Tài liệu tham khảo có thể là trang web [7, 2], bài báo khoa học [4], sách [5, 3, 1], bài tạp chí [9] hoặc các nguồn tham khảo khác. Lưu ý khi trích dẫn tài liệu tham khảo, cần viết câu sao cho bỏ phần trong cặp ngoặc vuông đi thì câu vẫn đầy đủ ý nghĩa. Ví dụ, thay vì viết “Nghiên cứu [9] chỉ ra rằng ...” thì nên viết “Nghiên cứu của Zhang [9] chỉ ra rằng ...”. Một ví dụ khác, thay vì viết “... như trong công trình nghiên cứu [4].” thì nên viết “... như trong công trình nghiên cứu của Cavnar và Trenkle [4].”

3.2 Chèn mã nguồn

Để chèn mã nguồn, cần dùng package listings [7]:

```
1 \usepackage{listings}
```

Mã nguồn có thể được chèn trực tiếp như sau:

```
1 print "Hello , World!"
```

hoặc chèn thông qua tập tin chứa mã nguồn trong thư mục *SourceCode* như sau:

```
1 #include <iostream.h>
2
3 main()
4 {
5     for (;;)
6     {
7         cout << "Hello World! ";
8     }
9 }
```

Để chèn mã giả, cần dùng package algorithm [**Algorithm**]:

```
1 \usepackage{algorithm}
```

Có thể chèn mã giả vào như sau:

Algorithm 1 My algorithm

```
1: procedure MYPROCEDURE
2:    $stringlen \leftarrow \text{length of } string$ 
3:    $i \leftarrow patlen$ 
4:   top:
5:     if  $i > stringlen$  then return false
6:      $j \leftarrow patlen$ 
7:   loop:
8:     if  $string(i) = path(j)$  then
9:        $j \leftarrow j - 1.$ 
10:       $i \leftarrow i - 1.$ 
11:      goto loop.
12:    close;
13:     $i \leftarrow i + \max(delta_1(string(i)), delta_2(j)).$ 
14:    goto top.
```

3.3 Hình ảnh

Để chèn hình ảnh, cần dùng package graphicx [6]:

```
1 \usepackage{graphicx}
```

Hình 3.1, hình 3.2 là một số ví dụ về chèn hình ảnh.



Hình 3.1: Hình ví dụ 1



Hình 3.2: Hình ví dụ 2

3.4 Bảng biểu

Để tạo bảng biểu, tham khảo thêm tại sharelatex.com [8]. Bảng 3.1 là một ví dụ về bảng. Ngoài ra, có một số tool online ¹ có thể được dùng để tạo bảng biểu một cách trực quan.

3.5 Công thức

Công thức có thể chèn vào trong cùng một dòng như $\sqrt{a^2 + b^2}$ hoặc nằm trên dòng riêng như công thức 3.1.

$$x = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + a_4}}} \quad (3.1)$$

¹Như trang <http://www.tablesgenerator.com/>

Bảng 3.1: Bảng ví dụ 1

Country List			
Country Name or Area Name	ISO ALPHA 2 Code	ISO ALPHA 3 Code	ISO numeric Code
Afghanistan	AF	AFG	004
Aland Islands	AX	ALA	248
Albania	AL	ALB	008
Algeria	DZ	DZA	012
American Samoa	AS	ASM	016
Andorra	AD	AND	020
Angola	AO	AGO	024

Danh mục công trình của tác giả

1. Tạp chí ABC
2. Tạp chí XYZ

Tài liệu tham khảo

Tiếng Việt

- [1] Diệp Quang Ban and Hoàng Van Thung. *Ngữ pháp tiếng Việt*. Nhà xuất bản Giáo dục Việt Nam, 2006.
- [2] DHQG-HCM Trường Đại học Khoa học Tự nhiên. *Quy định và hướng dẫn thực hiện luận văn thạc sĩ*. 2011. URL: http://www.hcmus.edu.vn/index.php?option=com_content&task=blogcategory&id=142&Itemid=506 (visited on 06/06/2015).
- [3] Đinh Điền. *Xử lý ngôn ngữ tự nhiên*. Nhà xuất bản Đại học Quốc gia TP.HCM, 2006.

Tiếng Anh

- [4] Cavnar, William B. and Trenkle, John M. “N-Gram-Based Text Categorization”. In: *In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*. 1994, pp. 161–175.
- [5] Knuth, Donald E. *The T_EXbook*. Addison-Wesley, 1984.
- [6] Online. *LaTeX/Floats, Figures and Captions*. URL: http://en.wikibooks.org/wiki/LaTeX/Floats,_Figures_and_Captions (visited on 06/06/2015).
- [7] Online. *LaTeX/Source Code Listings*. URL: http://en.wikibooks.org/wiki/LaTeX/Source_Code_Listings (visited on 06/06/2015).

- [8] Online. *LaTeX/Tables*. URL: <http://en.wikibooks.org/wiki/LaTeX/Tables> (visited on 06/06/2015).
- [9] Zhang, Kaizhong and Shasha, Dennis. “Simple fast algorithms for the editing distance between trees and related problems”. In: *SIAM Journal on Computing*, Volume 18 Issue 6 (1989), pp. 1245–1262.

Phụ lục A

Ngữ pháp tiếng Việt

Đây là phụ lục.

Phụ lục B

Ngữ pháp tiếng Nôm

Đây là phụ lục 2.