

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN CÔNG NGHỆ TRI THỨC

NGUYỄN SINH VIÊN - 1234567

SỬ DỤNG LATEX TRONG
KHOÁ LUẬN TỐT NGHIỆP

KHOÁ LUẬN TỐT NGHIỆP CỬ NHÂN CNTT

GIÁO VIÊN HƯỚNG DẪN
ĐINH ĐIỀN

KHOÁ 2012-2016

NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

Đây là đề tài có ý nghĩa thực tiễn và hữu ích khi ứng dụng trong thực tế. Sinh viên đã biết kết hợp điểm mạnh của từng mô hình (noisy channel và độ liên kết ngữ cảnh) để cho ra kết quả tốt nhất có thể được. Sinh viên có đề xuất sử dụng ngưỡng động và hàm max trong cách tính độ liên kết ngữ cảnh và đã chứng minh sự hiệu quả của đề xuất này dựa trên kết quả thực nghiệm.

Cách trình bày của khoá luận tốt nghiệp tuân thủ theo cách trình bày chuẩn, khá súc tích, ngắn gọn.

Trong quá trình làm, sinh viên có thể hiện sự cố gắng để hoàn thành khoá luận tốt nghiệp một cách tốt nhất trong khả năng của mình. Tuy đôi khi vẫn còn sắp xếp thời gian chưa hợp lí nhưng nhìn chung sinh viên đã hoàn thành tốt so với yêu cầu của một khoá luận tốt nghiệp.

Tp. HCM, ngày 21 tháng 07 năm 2015

Giáo viên hướng dẫn

Đinh Điền

NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN

Bài toán kiểm lỗi chính tả là một trong những bài toán cơ bản và quen thuộc trong xử lý ngôn ngữ tự nhiên. Giải quyết bài toán này giúp cho người sử dụng máy tính tránh được những lỗi văn bản không đáng có do nhầm lẫn khi gõ trên bàn phím hay do thói quen phát âm chưa chuẩn,...

Trong xử lý ngôn ngữ tự nhiên tiếng Việt đã có những công trình giải quyết bài toán này trong những năm qua, mặc dù vẫn còn lẻ tẻ. Khoá luận tốt nghiệp này đóng góp thêm một lời giải cho bài toán này.

Rõ ràng rằng việc phát hiện ra lỗi có thể đơn giản trong trường hợp tiếng không xuất hiện trong từ điển (lỗi non-syllable) nhưng khi tiếng đó thực sự tồn tại thì việc quyết định đó phải là lỗi sai hay không phụ thuộc nhiều vào ngữ cảnh các tiếng xung quanh. Hơn nữa, bài toán này không chỉ dừng ở mức phát hiện lỗi mà phải đến mức đề xuất phương án sửa lỗi sai đó.

Khoá luận này đã chọn được hướng tiếp cận thích hợp khi sử dụng liên kết ngữ cảnh để xác định lỗi sai. Đồng thời lựa chọn ứng viên thích hợp từ tập ứng viên phát sinh được để đề xuất cách sửa lỗi.

Khoá luận này được phát triển trên cơ sở của nhóm tác giả Nguyễn Thị Xuân Hương và có thêm cải tiến nhờ vào sự quan sát, lý luận và thử nghiệm. Đóng góp chính của khoá luận là thay đổi hàm quyết định trong liên kết ngữ cảnh (dùng hàm max thay vì hàm trung bình nhân) và đưa ra ngưỡng động (thay vì ngưỡng tĩnh) trên cơ sở của lý thuyết Noisy Channel.

Tác giả khoá luận đã dành nhiều công sức thu thập và xây dựng ngữ liệu hợp lý (trên sự hỗ trợ của các cộng tác viên) để bước đầu có thể dùng để thử nghiệm và đánh giá. Kết quả thử nghiệm cho thấy các đề xuất và cải tiến trong khoá luận là đáng ghi nhận và có thể chấp nhận được.

Tác giả khoá luận đã trình bày báo cáo khoá luận gồm 6 chương và 2 phụ lục rõ ràng và hợp lý. Dĩ nhiên rằng, báo cáo hiện giờ tốt hơn nhiều so với báo cáo ban đầu sau khi gặp người phản biện. Điều đó cho thấy tác giả đã lắng nghe để hoàn chỉnh khoá luận này về mặt báo cáo và chương trình.

Điểm hạn chế của khoá luận nằm ở chỗ giả định lỗi sai xuất hiện chỉ tại một tiếng khi xem xét. Trên thực tế lỗi sai có thể xuất hiện từ hai tiếng liền kề trở lên và giữa các lỗi sai có tác động qua lại lẫn nhau.

Ngoài ra, do hướng tiếp cận dựa trên thống kê khiến cho những từ hiếm gặp hoặc thường sai do thói quen của đại đa số người không thể phát hiện lỗi được (chẳng hạn với cụm từ “vô hình trung”).

Dù vậy, tác giả đã thể hiện sự cố gắng, nền tảng kiến thức và các kỹ năng của mình để hoàn thành khoá luận tốt nghiệp đáp ứng yêu cầu một khoá luận tốt nghiệp cử nhân Công nghệ thông tin.

Khóa luận đáp ứng yêu cầu của Khóa luận cử nhân CNTT.

Tp. HCM, ngày 21 tháng 07 năm 2015

Giáo viên phản biện

Nghiêm Quốc Minh

Lời cảm ơn

Tôi xin chân thành cảm ơn ...

Khoa Công Nghệ Thông Tin
Bộ môn Công nghệ tri thức

ĐỀ CƯƠNG CHI TIẾT

Tên đề tài: Sử dụng LaTeX trong Khoá luận tốt nghiệp
Giáo viên hướng dẫn: Đinh Điền
Thời gian thực hiện: 01/01/2000-01/01/2001
Sinh viên thực hiện: Nguyễn Sinh Viên - 1234567
Loại đề tài: Tìm hiểu công nghệ (có hoặc không ứng dụng minh hoạ), Xây dựng ứng dụng, ...

Nội dung đề tài: mô tả chi tiết nội dung đề tài, yêu cầu, phương pháp thực hiện, kết quả đạt được, ...
Kế hoạch thực hiện: mô tả chi tiết thời gian của các giai đoạn thực hiện và phân công công việc của từng thành viên trong nhóm

Xác nhận của GVHD

Ngày ... tháng ... năm 2015

Đinh Điền

Nguyễn Sinh Viên

Mục lục

Lời cảm ơn	i
Đề cương chi tiết	ii
Mục lục	iii
Tóm tắt	vi
1 Ngôn ngữ	1
2 Trình bày luận văn	2
2.1 Tiểu mục	2
2.2 Bảng biểu, hình vẽ, phương trình	3
2.3 Viết tắt	4
2.4 Tài liệu tham khảo và cách trích dẫn	4
2.5 Phụ lục của luận văn	5
3 Hướng dẫn sử dụng template	6
3.1 Trích dẫn tài liệu	6
3.2 Chèn mã nguồn	6
3.3 Hình ảnh	7
3.4 Bảng biểu	8
3.5 Công thức	8

Tài liệu tham khảo	10
A Ngữ pháp tiếng Việt	11
B Ngữ pháp tiếng Nôm	12

Danh sách hình

3.1	Hình ví dụ 1	7
3.2	Hình ví dụ 2	8

Danh sách bảng

3.1	Bảng ví dụ 1	8
-----	------------------------	---

Tóm tắt

Tóm tắt khóa luận: trình bày tóm tắt vấn đề nghiên cứu, các hướng tiếp cận, cách giải quyết vấn đề và một số kết quả đạt được.

Chương 1

Ngôn ngữ

Ngôn ngữ để viết và trình bày luận văn là tiếng Việt hoặc tiếng Anh. Trường hợp chọn ngôn ngữ tiếng Anh để viết và trình bày luận án, học viên cao học (HVCH) cần có văn bản đề nghị, được cán bộ hướng dẫn (CBHD) đồng ý và nộp cho phòng Đào tạo Sau đại học (phòng ĐT SDH) vào thời điểm đăng ký đề tài luận văn để xin ý kiến phê duyệt của Thủ trưởng cơ sở đào tạo (CSDT). Luận văn viết và trình bày bằng tiếng Anh phải có bản tóm tắt luận văn viết bằng tiếng Việt.

Tóm tắt luận văn: Tóm tắt luận văn phải in theo kích thước 140 x 210 mm (khổ A4 gấp đôi). Tóm tắt luận văn được trình bày nhiều nhất trong 24 trang in trên hai mặt giấy, cỡ chữ Times New Roman 11 của hệ soạn thảo Winword hoặc phần mềm soạn thảo Latex đối với các chuyên ngành thuộc ngành Toán. Mật độ chữ bình thường, không được nén hoặc kéo giãn khoảng cách giữa các chữ. Chế độ dẫn dòng là Exactly 17pt. Lề trên, lề dưới, lề trái, lề phải đều là 1.5 cm. Các bảng biểu trình bày theo chiều ngang khổ giấy thì đầu bảng là lề trái của trang. Tóm tắt luận án phải phản ánh trung thực kết cấu, bố cục và nội dung của luận án, phải ghi đầy đủ toàn văn kết luận của luận án. Mẫu trình bày trang bìa của tóm tắt luận văn (phụ lục 1).

Chương 2

Trình bày luận văn

Luận văn 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. Luận văn được in trên một mặt giấy A4 (210 x 297 mm) và không vượt quá 80 trang A4, không tính phần phụ lục (nếu có). Cỡ chữ Times New Roman 13 của hệ soạn thảo Winword hoặc tương đương. Mật độ chữ bình thường, không được nén hoặc kéo giãn khoảng cách giữa các chữ; dẫn dòng đặt ở chế độ 1,5 lines; lề trên 3,5cm; lề dưới 3 cm; lề trái 3,5cm; lề phải 2 cm. Số trang được đánh ở giữa, phía cuối mỗi trang giấy. Tất cả các chương, phần, mục, tiểu mục phải dùng chữ số tự nhiên, không được dùng số la mã. Mẫu trình bày trang bìa (phụ lục 2), trang phụ bìa (phụ lục 3) và mục lục (phụ lục 4).

2.1 Tiểu mục

Các tiểu mục của luận văn được trình bày và đánh số thành nhóm chữ số, nhiều nhất gồm bốn chữ số với số thứ nhất chỉ số chương (ví dụ: 4.1.2.1. chỉ tiểu mục 1 nhóm tiểu mục 2 mục 1 chương 4). Tại mỗi nhóm tiểu mục phải có ít nhất hai tiểu mục, nghĩa là không thể có tiểu mục 2.1.1 mà không có tiểu mục 2.1.2 tiếp theo.

2.2 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 lấy từ các nguồn khác phải được trích dẫn đầy đủ. Nguồn được trích dẫn phải được liệt kê chính xác trong danh mục Tài liệu tham khảo. 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 luận văn. 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 văn bản luận văn. 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”.

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

2.3 Viết tắt

Không lạm dụng việc viết tắt trong luận văn. Chỉ viết tắt những từ, cụm từ hoặc thuật ngữ được sử dụng nhiều lần trong luận văn. Không viết tắt những cụm từ dài, những mệnh đề; không viết tắt những cụm từ ít xuất hiện trong luận văn. 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 luận văn có nhiều chữ viết tắt thì phải có bảng danh mục các chữ viết tắt (xếp theo thứ tự ABC) ở phần đầu luận văn.

2.4 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 luận văn. Không trích dẫn những kiến thức phổ biến, mọi người đều biết cũng như không làm luận văn nặng nề với những tham khảo trích dẫn. 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. 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 luận văn. Cách sắp xếp danh mục tài liệu tham khảo xem phụ lục 5. Việc trích dẫn là theo số thứ tự của tài liệu ở danh mục tài liệu tham

khảo và được đặt trong ngoặc vuông, khi cần có cả số trang, ví dụ [15, tr.314-315]. Đối với phần được trích dẫn từ nhiều nguồn tài liệu khác nhau, số của từng tài liệu được đặt độc lập trong từng ngoặc vuông, theo thứ tự tăng dần, ví dụ [19], [25], [41], [42].

2.5 Phụ lục của luận văn

Phần này bao gồm nội dung cần thiết nhằm minh họa hoặc hỗ trợ cho nội dung luận văn như số liệu, mẫu biểu, tranh ảnh,... Phụ lục không được dày hơn phần chính của luận văn.

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 [4, 9], bài báo khoa học [1], sách [2, 8, 7], bài tạp chí [6] hoặc các nguồn tham khảo khác.

Các tiểu mục của luận văn được trình bày và đánh số thành nhóm chữ số, nhiều nhất gồm bốn chữ số với số thứ nhất chỉ số chương (ví dụ: 4.1.2.1. chỉ tiểu mục 1 nhóm tiểu mục 2 mục 1 chương 4). Tại mỗi nhóm tiểu mục phải có ít nhất hai tiểu mục, nghĩa là không thể có tiểu mục 2.1.1 mà không có tiểu mục 2.1.2 tiếp theo.

3.2 Chèn mã nguồn

Để chèn mã nguồn, cần dùng package listings [4]:

```
\usepackage{listings}
```

Mã nguồn có thể được chèn trực tiếp như sau:

```
print "Hello ,_World! "
```

hoặc chèn thông qua tập tin chứa mã nguồn trong thư mục “Source-Code” như sau:

```
#include <iostream.h>
```

```
main()  
{  
    for ( ;; )  
    {  
        cout << "Hello _World! _" ;  
    }  
}
```

3.3 Hình ảnh

Để chèn hình ảnh, cần dùng package `graphicx` [3]:

```
\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 [5]. Bảng 3.1 là một ví dụ về bảng.

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

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 sau:

$$x = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + a_4}}} \tag{3.1}$$

Tài liệu tham khảo

Tiếng Anh

- [1] 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.
- [2] Knuth, Donald E. *The T_EXbook*. Addison-Wesley, 1984.
- [3] Online. *LaTeX/Floats, Figures and Captions*. URL: http://en.wikibooks.org/wiki/LaTeX/Floats,_Figures_and_Captions (visited on 06/06/2015).
- [4] Online. *LaTeX/Source Code Listings*. URL: http://en.wikibooks.org/wiki/LaTeX/Source_Code_Listings (visited on 06/06/2015).
- [5] Online. *LaTeX/Tables*. URL: <http://en.wikibooks.org/wiki/LaTeX/Tables> (visited on 06/06/2015).
- [6] 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.