**MÔ TẢ ĐỀ TÀI KHÓA LUẬN TỐT NGHIỆP**

# Thông tin giảng viên hướng dẫn

|  |  |
| --- | --- |
| Họ tên: | **ThS. Nguyễn Thị Anh Thư** |
| E-mail: | thunta@uit.edu.vn |
| Điện thoại: | 090 737 9067 |
| Website: |  |

# Tên đề tài

|  |  |
| --- | --- |
| Tiếng Việt: | **Khai thác cấu trúc mạng xã hội ứng dụng vào lĩnh vực phân tích tâm lý người dùng trên Wiki.** |
| Tiếng Anh: | **Exploting Social Network Structure for Person-to-Person Sentiment Analysis on Wiki.** |

# Giới thiệu, phạm vi và giới hạn đề tài

## Giới thiệu

***Tóm tắt.***

Quan điểm đánh giá giữa người này với người khác được sử dụng phổ biến trong các bài diễn văn, bài luận, cuộc đàm thoại, … Do đó, việc dự đoán quan điểm người dùng là yếu tố quan trọng được ứng dụng trong: *việc xây dựng danh tiếng khi tranh cử, xây dựng các liên kết xã hội và khảo sát lấy ý kiến công chúng*.

Việc đánh giá quan điểm này có thể được phân tích riêng biệt bằng cách sử dụng các mạng xã hội có dấu (*mạng có 2 loại liên kết positive và negative*) hoặc phân tích tâm lý dựa trên nội dung văn bản. Tuy nhiên, các phương pháp cổ điển này đã bỏ qua tương tác phong phú giữa ngôn ngữ và bối cảnh xã hội. Phương pháp mới được đề xuất bởi *Robert* có thể nắm bắt được các tương tác như vậy bằng cách tổng hợp các thông tin từ các mạng xã hội có dấu như: *mối quan hệ giữa các người dùng trong mạng xã hội và các văn bản đánh giá liên quan*.

Xây dựng module (*prototype*) giải bài toán dự đoán quan điểm người dùng trên mạng xã hội Wiki.

***Từ khóa tiếng Việt.*** Phân tích tâm lý, cấu trúc mạng xã hội, mạng xã hội có dấu.

***Từ khóa tiếng Anh.*** Sentiment analysis, social network structure, signed social network.

## Phạm vi và giới hạn

* Lĩnh vực nghiên cứu: khai thác dữ liệu trên mạng xã hội.
* Phạm vi nghiên cứu: mạng xã hội có dấu, cấu trúc đồ thị mạng, phân tích dữ liệu text, mô hình máy học.
* Ngôn ngữ lập trình: C++, Python, Matlab.
* Database: Wikipedia Requests for Adminship (*with text*).

# Nội dung thực hiện, và phương pháp

**Nội dung 1**: Nghiên cứu bài toán dự đoán quan điểm của người dùng trên mạng xã hội.

Phương pháp: Đọc các bài báo khoa học liên quan và tổng hợp tài liệu.

Kết quả dự kiến: Định nghĩa bài toán (*input, output*), ứng dụng của bài toán, khó khăn và thách thức, các công trình nghiên cứu liên quan, framework, database, độ đo đánh giá.

Thời gian: 2 ~ 3 tháng.

**Nội dung 2**: Nghiên cứu và xử lý dữ liệu.

Phương pháp: Tiến hành download dữ liệu, đọc phần mô tả dữ liệu và tìm hiểu về định dạng dữ liệu. Sau đó, tiến hành cài đặt xử lý dữ liệu.

Kết quả dự kiến: Module xử lý dữ liệu.

Thời gian: 1 ~ 2 tháng.

**Nội dung 3**: Rút trích đặc trưng và xây dựng mô hình máy học.

Phương pháp: Dựa vào bài báo khoa học, tiến hành cài đặt rút trích đặc trưng và xây dựng mô hình máy học cho bài toán.

Kết quả dự kiến: Module rút trích đặc trưng; module huấn luyện và dự đoán.

Thời gian: 2 ~ 3 tháng.

**Nội dung 4**: Chạy thực nghiệm và đánh giá.

Phương pháp: Tiến hành cài đặt thực nghiệm trên bộ dữ liệu của Wiki.

Kết quả dự kiến: Kết quả thực nghiệm và bảng phân tích đánh giá.

Thời gian: 1 ~ 3 tháng.

# Kế hoạch thực hiện

Thời gian thực hiện: 6 ~ 11 tháng.

# Kỹ năng/Kỹ thuật về ngôn ngữ lập trình, Cơ sở dữ liệu, …

* Ngôn ngữ lập trình: C++, Python, Matlab.
* Thư viện: SNAP.
* Database: Wikipedia Requests for Adminship (*with text*).
* Link:
  + Thư viện: <http://snap.stanford.edu/about.html>
  + Data: <http://snap.stanford.edu/data/wiki-RfA.html>

# Kết quả dự kiến

* Một Application (*prototype*).
  + Input: Mạng xã hội có dấu và văn bản text liên quan.
  + Output: Quan điểm của người dùng A với người dùng B là tích cực / tiêu cực / trung tính (*positive / negative/ neutral*).

# Tài liệu tham khảo

[1] R. West, H. S. Paskov, J. Leskovec, C. Potts, “Exploiting Social Network Structure for Person-to-Person Sentiment Analysis”, in *Transactions of the Association for Computational Linguistics (TACL), 2,* 2014, eprint arXiv:1409.2450.

[2] Anh-Thu Nguyen-Thi, Phuc Quang Nguyen, Thanh Duc Ngo, Tu-Anh Nguyen-Hoang, “Transfer AdaBoost SVM for Link Prediction in Newly Signed Social Networks using Explicit and PNR Features”, in *Proceedings of The Nineteenth International Conference on Knowledge-Based and Intelligent Information & Engineering Systems*, 7-9/9/2015, Singapore.

[3] J. Leskovec, D. Huttenlocher, J. Kleinberg, “Predicting positive and negative links in online social networks”, in *WWW*, 2010, p. 641-650.

[4] J. Leskovec, D. Huttenlocher, J. Kleinberg, “Signed networks in social media”, in *Proc. 28th CHI*, 2010.