**TỔNG HỢP QUAN ĐIỂM DỰA TRÊN MÔ HÌNH THỐNG KÊ VÀ ỨNG DỤNG VÀO KHAI PHÁ QUAN ĐIỂM TRONG VĂN BẢN TIN TỨC TIẾNG VIỆT**

**Vũ Xuân Sơn**

*Khóa QH-2007-I/CQ , ngành Công nghệ thông tin*

**Tóm tắt khóa luận tốt nghiệp:**

Khai phá quan điểm trên miền tin tức là một lĩnh vực mới, nhận được nhiều sự quan tâm trong những năm gần đây, và đánh dấu một bước phát triển trong khai phá văn bản (text mining). Khi mà khai phá văn bản không chỉ là phân tích về mặt “***câu chữ***”, “***ngữ pháp***” văn bản mà thực sự hướng tới việc phân tích ngữ nghĩa, giúp máy tính thật sự *“****hiểu****”* nội dung văn bản “***nói gì***” và “***quan điểm***” của người viết như thế nào (ví dụ: khen/chê trong văn bản đó).

Khóa luận tập trung nghiên cứu phương pháp và xây dựng mô hình thống kê cho tổng hợp quan điểm trên miền ứng dụng tin tức tiếng Việt nhằm ứng dụng vào máy tìm kiếm quan điểm trên miền dữ liệu tin tức tiếng Việt. Với đầu vào là một danh từ chỉ tên thực thể người dùng quan tâm, hệ thống tiến hành gửi truy vấn lên các máy tìm kiếm (Google, Yahoo..) và lấy về các trang tin có chứa bình luận của người dùng. Với tập các trang tin thu thập được, hệ thống tiến hành tổng hợp quan điểm và trả về kết quả tổng hợp cho người dùng.

Khóa luận tiến hành xây dựng thử nghiệm áp dụng mô hình đề xuất trên miền dữ liệu là các bình luận từ trang tin VnExpress. Khóa luận cũng đã cho ra sản phẩm VnNewsOpinion được giới thiệu và chạy thử nghiệm trên trang <http://sonvx.net/vietnews>. Đánh giá thực nghiệm cho kết quả theo thang điểm là 5, cho thấy giá trị đáp ứng trung bình và chất lượng tổng hợp đạt mức điểm khả quan gần 4. Kết quả này cho thấy mô hình đề xuất là đúng đắn và có thể triển khai thực tế.

***Từ khóa:*** KPQĐ, khai phá quan điểm, opinion mining