

HỆ THỐNG TRẢ LỜI TỰ ĐỘNG BẰNG TIN NHẮN SMS AGENTS VÀ CÁC GIẢI PHÁP THỰC NGHIỆM

Phạm Việt Bình - Phùng Trung Nghĩa (*Khoa Công nghệ thông tin - ĐH Thái Nguyên*)

1. Đặt vấn đề

Hình thức trả lời tự động bằng tin nhắn SMS đã khá quen thuộc tại Việt Nam. Tuy nhiên vẫn chưa có một phương pháp hiệu quả để xây dựng hệ thống SMS agent hoạt động ổn định, hiệu quả, và rẻ tiền nhất là trong trường hợp cùng lúc có rất nhiều người cùng truy cập. Bài báo này trình bày về các giải pháp thực nghiệm nhằm tối ưu hoá hoạt động của hệ thống SMS agents tại Khoa Công nghệ thông tin, Đại học Thái Nguyên.

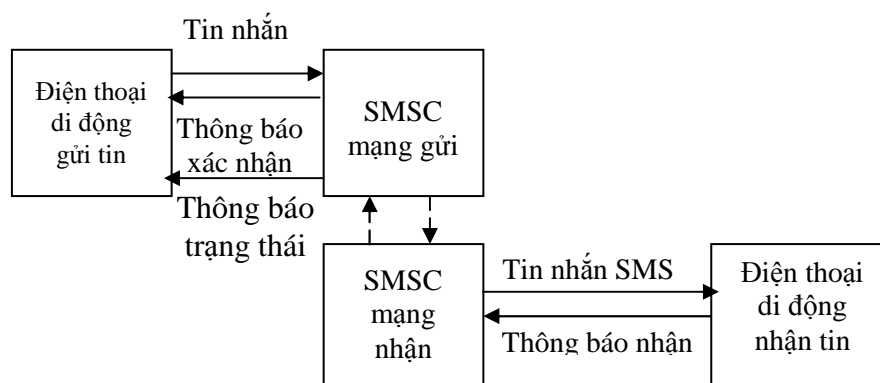
2. Một số kiến thức liên quan

2.1. Dịch vụ SMS

SMS (Short Message Service) là kỹ thuật cho phép gửi và nhận tin nhắn giữa các điện thoại di động. Như tên gọi, SMS có nghĩa là dịch vụ nhắn tin ngắn, lượng dữ liệu có thể mang trong một tin nhắn SMS là rất hạn chế. Một tin nhắn SMS có thể mang nhiều nhất 140 bytes, do vậy nó có thể chứa được 160 kí tự nếu sử dụng mã hoá kí tự 7 bit hoặc 70 kí tự nếu sử dụng mã hoá kí tự Unicode UCS2 16 bit [6].

2.2. Trung tâm nhắn tin (SMSC)

Một SMSC chịu trách nhiệm xử lý tất cả các tin nhắn SMS trong mạng thông tin di động, Một tin nhắn SMS có thể phải trung chuyển qua nhiều SMSC cho tới đích. Chức năng chính của một SMSC là định tuyến cho tin nhắn SMS tới máy nhận tin [6].



Hình 1. Mô hình gửi nhận tin nhắn SMS qua SMSC điển hình

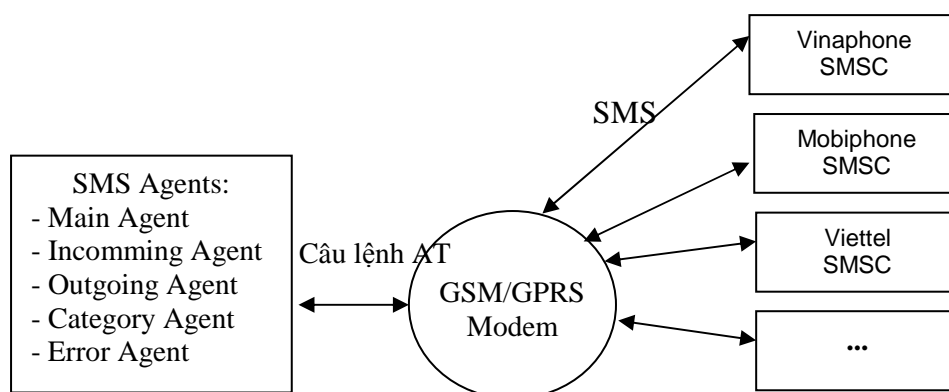
2.3. GSM/GPRS Modem

GSM/GPRS Modem là thiết bị modem không dây kết nối máy tính với mạng điện thoại di động. GSM/GPRS modem cũng như hầu hết các thiết bị giao tiếp với máy tính qua cổng COM có thể được làm việc thông qua các câu lệnh AT [4].

3. Các giải pháp thực nghiệm

3.1. Tổng quan về hệ thống

Để hỗ trợ tốt công tác quản lý của Khoa, giúp sinh viên và gia đình có thể tra cứu các thông tin cần thiết về quá trình học tập của sinh viên, chúng tôi đã xây dựng hệ xây dựng một hệ thống trả lời SMS tự động thông minh bằng tin nhắn (gọi là SMS agents).



Hình 2. Mô hình hệ thống trả lời tin nhắn tự động SMS agents

Một số đặc điểm, chức năng của hệ thống:

- Để kết nối máy PC với mạng điện thoại di động, chúng tôi sử dụng GSM/GPRS Modem Wismo của hãng InterCEL.

- Hệ thống được xây dựng theo cơ chế đa tác tử (multi-agents), tức là khi hệ thống hoạt động cùng lúc sẽ có nhiều agent là các chương trình chạy liên tục và song song đồng thời.

- Hệ thống có thể kết nối với nhiều nguồn dữ liệu khác nhau, có thể bổ sung cơ sở dữ liệu để cung cấp các dịch vụ trả lời thông tin mới trong lúc đang hoạt động.

- Để hệ thống không bị quá tải dẫn đến bị treo khi có quá nhiều yêu cầu truy cập trong một thời điểm, chúng tôi đã xây dựng các bộ đệm hàng đợi lưu tin nhắn.

- Hệ thống có khả năng đoán nhận yêu cầu của người sử dụng để đưa ra thông báo lỗi về cú pháp câu hỏi khi người sử dụng nhập câu hỏi sai cú pháp.

- Hệ thống có khả năng tự phân loại người sử dụng để có thể gửi các thông báo tới nhóm người sử dụng phù hợp.

3.2 Xây dựng cơ chế đa hàng đợi cho hệ thống

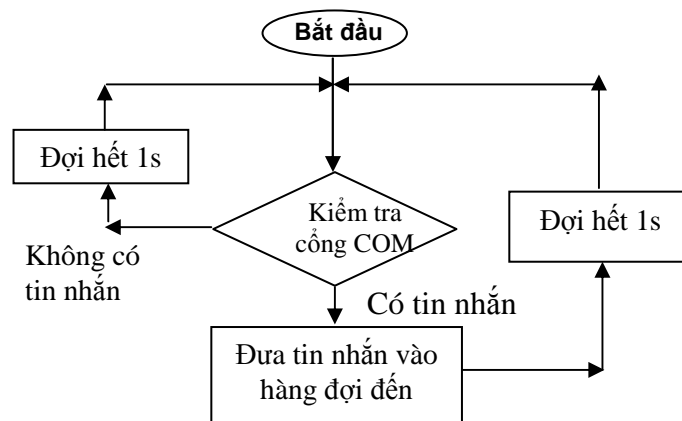
Các thiết bị SMSC, GSM/GPRS Modem, cổng COM của máy tính mặc định đều cung cấp cơ chế hàng đợi đệm. Ở đây chúng tôi xây dựng thêm một hàng đợi cục bộ bằng phần mềm giúp SMS Agents có khả năng kiểm soát luồng dữ liệu vào ra, tránh xảy ra ùn tắc thông tin do cổng COM của máy tính hoặc Modem đang bị chiếm dụng khi nhiều tin nhắn cùng đến một lúc.

Ở đây chúng tôi sử dụng giải pháp sử dụng 3 hàng đợi gồm hàng đợi đến (incoming queue), hàng đợi đi (outgoing queue) và hàng đợi lỗi (error queue), tương ứng sẽ có 3 agent chạy độc lập, song song trên máy tính đảm nhận việc xử lý dữ liệu tại các hàng đợi.

3.3. Hoạt động của các Agent

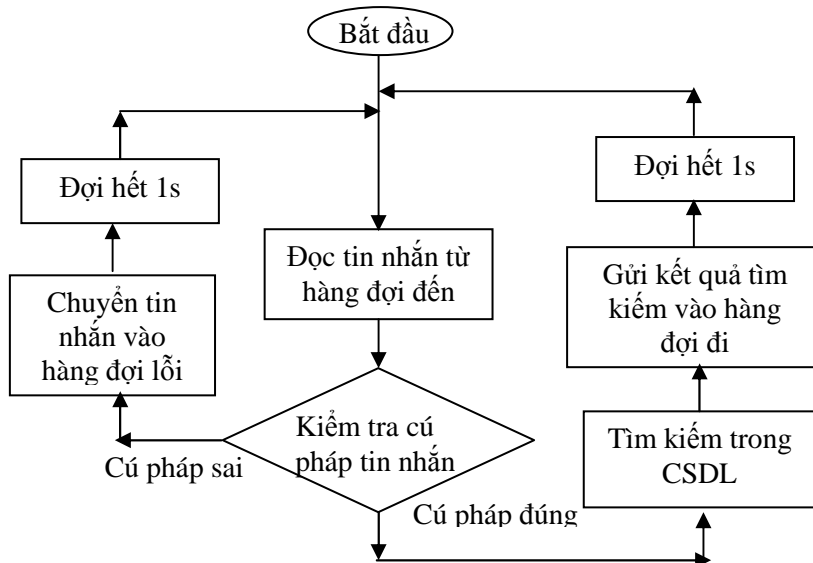
Agent là các chương trình chạy liên tục, độc lập với nhau. Mỗi agent sẽ được khởi động khi hệ thống khởi động và chạy liên tục với một vòng lặp vô tận nhờ một bộ đếm thời gian (một đối tượng được hỗ trợ trong hầu hết các ngôn ngữ lập trình Windows như Visual Basic 6.0). Riêng Category Agent sẽ chỉ được khởi động vào một thời điểm (mặc định chọn là 3.00 hàng ngày). Khi kết thúc hệ thống sự kiện kết thúc sẽ tác động đến tất cả các bộ đếm thời gian để kết thúc tất cả các agent.

Incomming Agent: Incomming Agent liên tục kiểm tra trạng thái cổng COM sau khoảng thời gian 1s. Khi phát hiện tin nhắn được gửi từ GSM/GPRS modem tới cổng COM, tin nhắn không được xử lý ngay mà đưa vào đầu vào hàng đợi đến.



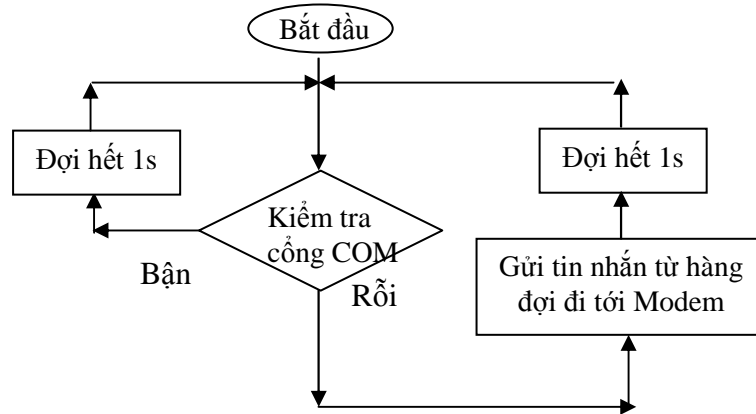
Hình 3. Lưu đồ thuật toán Incomming Agent

Main Agent: Main Agent liên tục đọc tin nhắn từ hàng đợi đến sau 1s, kiểm tra cú pháp tin nhắn: nếu đúng cú pháp sẽ thực hiện tìm kiếm trong cơ sở dữ liệu tương ứng và gửi kết quả tìm kiếm vào hàng đợi đi; nếu sai cú pháp sẽ được chuyển sang hàng đợi lỗi.



Hình 4. Lưu đồ thuật toán Main Agent

Outgoing Agent: Outgoing Agent kiểm tra trạng thái cổng COM sau mỗi khoảng thời gian 1s. Nếu trạng thái là bận sẽ đợi đến lần kiểm tra tiếp theo. Nếu trạng thái là rỗi sẽ gửi tin nhắn từ đầu ra của hàng đợi đi tới GSM modem qua cổng COM.



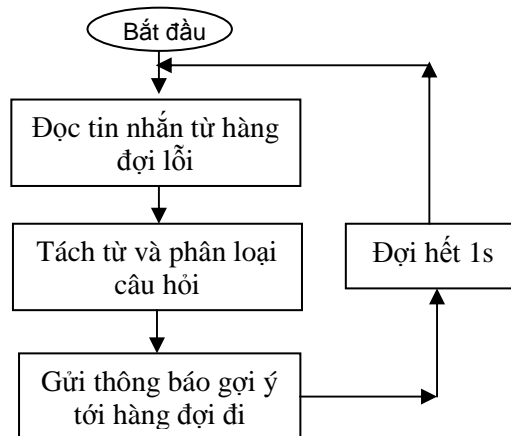
Hình 5. Lưu đồ thuật toán Outgoing Agent

Error Agent: Các hệ thống trả lời tin nhắn hiện nay đều dựa trên cú pháp câu hỏi “đóng”, tức là cú pháp được quy định trước. Khi người sử dụng gửi đến một tin nhắn lỗi không đúng cú pháp, thông thường các hệ thống sẽ đưa ra một thông báo lỗi và gợi ý về cú pháp câu hỏi đóng. Tuy nhiên trong trường hợp hệ thống trả lời thông tin nhiều lĩnh vực thì thông báo gợi ý về câu hỏi đóng có thể rất dài và gây khó khăn, phiền toái cho người sử dụng.

Ở đây chúng tôi sử dụng giải pháp khi nhận được câu hỏi lỗi không đúng cú pháp sẽ thực hiện phân loại câu hỏi theo lĩnh vực và chỉ gợi ý về cú pháp đóng của lĩnh vực được phân loại [5].

Phương pháp phân loại câu hỏi được thực hiện qua các bước sau:

- B1: Xây dựng bộ ngữ liệu gồm các từ “chuyên ngành” đối với từng lĩnh vực thông tin
- B2: Đánh trọng số cho các từ trong bộ ngữ liệu dựa trên xác suất xuất hiện câu hỏi có từ đó trong từng lĩnh vực và một số điều kiện ràng buộc khác.
- B3: Tách từ trong câu hỏi theo thuật toán LRMM, tính tổng trọng số các từ được tách ứng với từng lĩnh vực và phân loại câu hỏi về lĩnh vực có tổng trọng số lớn nhất.




Hình 6. Lưu đồ thuật toán Error Agent

Category Agent: Category Agent sẽ phân loại người sử dụng với các quyền khác nhau. Người sử dụng có đăng ký sẽ được phân loại theo yêu cầu khi đăng ký. Người sử dụng không đăng ký (guest) sẽ được phân loại dựa theo cơ chế suy đoán. Nếu mật độ và số lượng tin nhắn người của người sử dụng gửi tới hệ thống yêu cầu một lĩnh vực nào đó lớn hơn một ngưỡng thì người sử dụng sẽ được tự động phân loại vào nhóm tương ứng. Ngưỡng được xác định qua thực nghiệm. Category Agent chỉ được kích hoạt một lần trong ngày vào thời điểm hệ thống ít bận nhất (theo thực nghiệm là 3h00 hàng ngày).

Sau khi được phân loại mỗi khi có thông báo của Khoa hay lịch công tác, giao việc, người quản trị sẽ cập nhật thông báo và hệ thống sẽ gửi đến danh sách người sử dụng nằm trong nhóm tương ứng.

4. Kết luận

Hệ thống SMS agents tại Khoa Công nghệ thông tin, Đại học Thái Nguyên sử dụng phương pháp lập trình SMS truyền thống nhưng có nhiều sáng kiến cải tiến để tối ưu hoá hệ thống. Những cải tiến đó đã chứng tỏ được hiệu quả trong quá trình hoạt động thực nghiệm của hệ thống. Đây sẽ là một công cụ quan trọng giúp lãnh đạo Khoa, các bộ phận chức năng thực hiện hiệu quả hơn các công tác quản lý hành chính, quản lý cán bộ và sinh viên trong khoa, đồng thời cũng là một công cụ hữu ích để sinh viên và gia đình có thể dễ dàng tra cứu các thông tin về quá trình học tập của sinh viên trong khoa 

Tóm tắt

Bài báo này trình bày về hệ thống trả lời thông tin tự động thông minh bằng tin nhắn SMS tại Khoa Công nghệ thông tin, Đại học Thái Nguyên và các giải pháp thực nghiệm để tối ưu hóa hệ thống.

Summary

This paper introduces the intelligent automatic answering system at the Faculty of Information Technology, Thai Nguyen University and the implementation solutions to optimize the system.

Tài liệu tham khảo

- [1] Nguyễn Văn Tu, Từ và vốn từ tiếng Việt hiện đại, NXB Đại học & THCN, Hà Nội, 1978.
- [2] AT command Interface guide, Wavecom Coporation, 8/2002
- [3] Dinh Dien, Hoang Kiem, N V Toan, Vietnamese word segmentation. The sixth natural language processing, Tokyo, Japan, 2001.
- [4] InterCEL Stanalone modem reference, InterCEL Pty Ltd, <http://www.intercel.com.au>
- [5] Phạm Việt Bình, Phùng Trung Nghĩa, Một phương pháp phân loại câu hỏi tiếng Việt không dấu, chuyên đề nghiên cứu thuộc đề tài B2006-TN07-01, 2007.
- [6] SMS standards, IEEE, <http://www.ieee.org>
- [7] Ricardo Beaza-Yates & Berthier Ribeiro-Neto, Modern Information Retrieval, Addison Press, Anh, 1999.