

## **Đánh giá các công việc đã làm (6/1/2017)**

### **Project: Xây dựng hệ thống hỏi đáp luật giao thông đường bộ**

#### **1. Đánh giá độ tương tự**

- *Cách thức đánh giá:* DB gồm 255 câu hỏi, lấy ra 100 câu hỏi. Với mỗi câu hỏi trong bộ 100 câu lấy ra đặt lại 5 câu có nội dung tương tự. Sử dụng các câu hỏi tương tự để tiến hành truy vấn trong bộ DB. Kết quả trả về có xuất hiện các câu hỏi gốc thì được đánh giá cao.
- Kết quả đạt được thấp (<30%)
- DB được lưu trên server muốn truy vấn phải có kết nối internet => cần phải cải thiện.

#### **2. Tiền xử lý**

- Xử dụng thư viện Vntokenizer, Vntagger.
- Câu hỏi đầu vào được đưa lên server xử lý.

#### **3. Lemur**

- *Indexing:* xử lý với stopwords list, cần phải tách từ trước khi tiến hành indexing.
- *Retrieval:* chưa biết chọn tham số, chưa biết cấu hình. Hạn chế về tốc độ.
- *Yêu cầu thêm:*
  - ✓ Cài đặt trên server
  - ✓ Thực hiện trên điện thoại bằng mạng wifi, 3G.

#### **4. CSDL luật giao thông**

- *Gồm:*
  - ✓ 24 nghị định
  - ✓ 44 thông tư
- *Yêu cầu thêm:*
  - ✓ Làm rõ hệ thống tổ chức, cấu trúc
  - ✓ Xác định xem các văn bản lấy về đã bao phủ hết các lĩnh vực liên quan đến luật giao thông chưa: về xử phạt, về hướng dẫn thi bằng lái, mua bán, sang tên đổi chủ...
- *Vấn đề gặp phải về lưu trữ:* các văn bản pháp luật sửa đổi thì có thể tìm được văn bản gốc bị sửa đổi, còn các văn bản gốc không biết đã bị sửa đổi bởi văn bản mới nào.

#### **5. CSDL câu hỏi:**

- Gồm 255 câu hỏi về xử phạt đối với người tham gia giao thông được lấy về trên mạng.