Mô tả công việc giai đoạn 1: Đây là giai đoạn tiền xử lý các trang văn bản tiếng Việt. Bằng các giải thuật xử lý ảnh. Tiếp theo sau là bước tách dòng, tách đoạn, chuẩn hóa các từ đã tách (1-gam).

Lưu đồ công việc:

Đánh giá kết quả:

Chi tiết:

1. Tiền xử lý ảnh
   1. Giải thuật a
   2. Giải thuật b
   3. Giải thuật c
2. Phân đoạn
   1. Giải thuật a
   2. Giải thuật b
   3. Giải thuật c
3. Chuẩn hóa
   1. Giải thuật a
   2. Giải thuật b
   3. Giải thuật c