

# A Guide for Project

---

Build from scratch a sentiment analysis system that analyzes the reviews of one product (or more) of your interest.

## **Giai đoạn 1: Phân tích nội dung trang Web chứa nhận xét về sản phẩm**

- Chọn một (hoặc nhiều) sản phẩm được quan tâm làm chủ đề
- Chọn một (hoặc nhiều) trang Web (trang bán hàng, blog, forum) để thu thập nhận xét về sản phẩm được quan tâm
- Sử dụng/tự cài đặt công cụ tự động rút trích nội dung văn bản liên quan đến nhận xét sản phẩm từ trang Web mục tiêu.
- Sử dụng/tự cài đặt công cụ tự động tiền xử lý nội dung văn bản rút trích được, có thể xử lý ở mức toàn bộ nhận xét, câu, từ,... tùy theo kế hoạch tương ứng của giai đoạn huấn luyện.

Sản phẩm của giai đoạn này là một chương trình có đặc tả như sau

- Input: trang HTML chứa nhận xét sản phẩm
- Output: tập hợp các nhận xét đã tiền xử lý
- Các bước phân tích cần được tích hợp thành một workflow thống nhất và tự động thay vì phải gọi nhiều chương trình con một cách thủ công.

Ngày vấn đáp sản phẩm (dự kiến): Thứ bảy, ngày 03/11/2018.