NLIE

Native Language Identification Engine

Lee Levi and Yael Avinun

May 19th, 2015

Problem Description

Companies with support and sales websites and services rely on user location or manual selection to direct users to service in the relevant language.

This process has the potential to be automated and more precise, and therefore speed up processes as well as increase sales.

Problem Background

* Size: 100,000+ companies of all sizes

* Severity: Medium. Decreases sales and customer retention, providing a less personal approach.

* Consumers: Any company that offers online sales, support, or multi-language services

Current Solution

✤ Manual selection of language preference

✤ User-location detection

✤ User-initiated request for specific language-speaking agent

Our Solution - General Overview

✤ An API for companies that have multi-language online sales and support, which identifies in real-time, from English text, what the native language of the customer is.

✤ Business Model: Pay-per-service. Free to detect english/non-english, Premium to detect specific languages (priced depending on the number of languages).
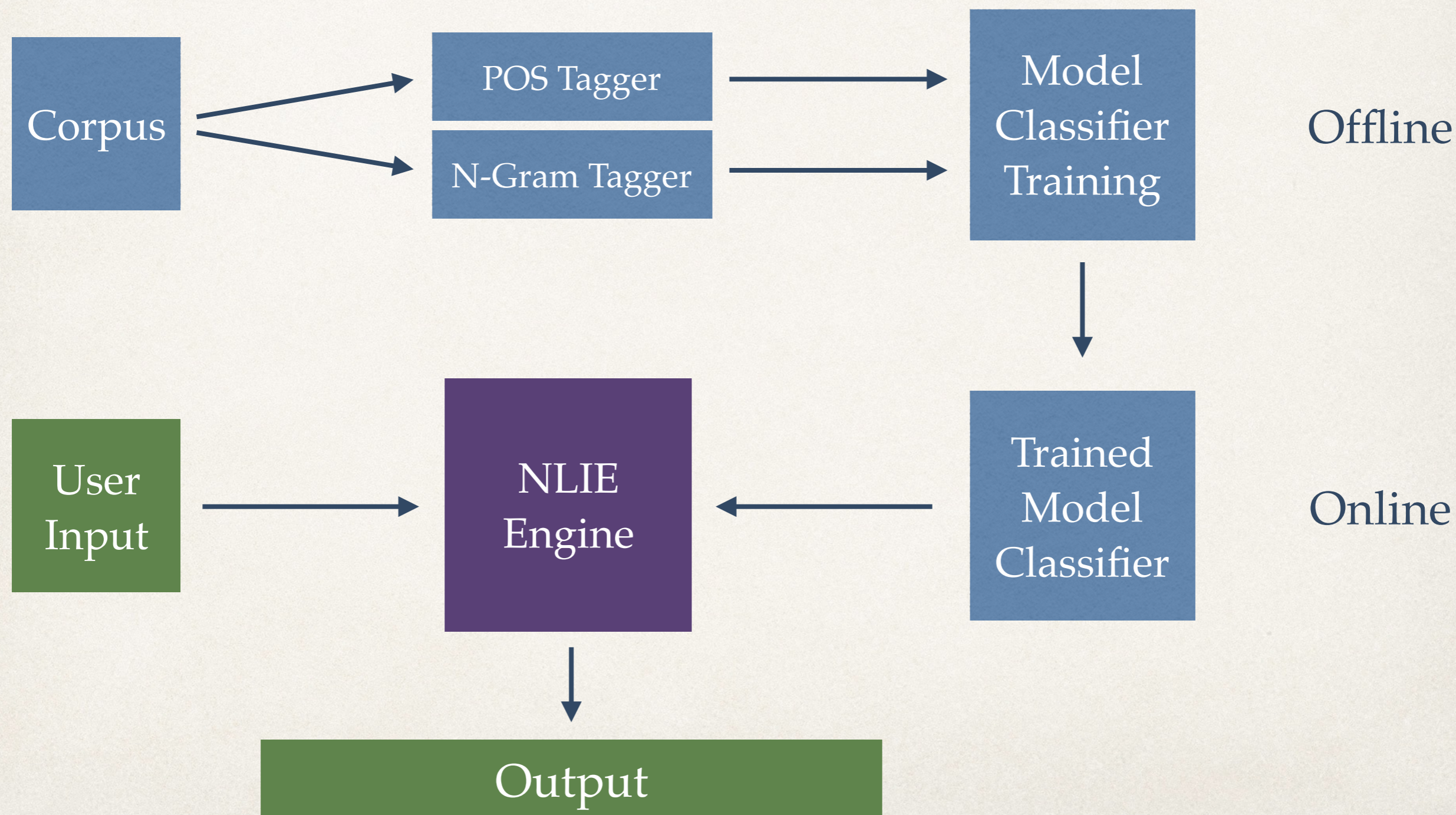
Our Solution - GUI

✤ The GUI will be a website for management of the API, where the company will be provided with the code and documentation, as well as (potentially) analytics of the company's use.

Our Solution - Data

✤ Lang-8 English Learner Corpus

Our Solution - High-Level Design

Our Solution - Off-the-Shelf Libraries

✤ NLTK

✤ SciKit-Learn

✤ Pickle

Our Solution

1. Milestones:

   1. Offline Work (until 31.5) : Choose a machine learning approach, implement and test the abilities of the classifier

   2. API Implementation (until 7.6): Create an API for the engine that a company could use on top of their own code.

   3. GUI (until 21.6): Implement a multi-user website which includes documentation, billing, information, and potentially, user-specific analytics.