Methodology

What have we done so far?

Language model:

We built a language model using the brown corpus

Smoothing

UNK word smoothing (less than 5 words)

Interpolation (smoothing using the development set to choose between unigrams,

bigrams and trigrams - 0.14, 0.25, 0.61)

SEMPRE Integration

Basic samples

Android application:

Fully tested the MS Computer Vision API

Created an application that can query the API

Preprocessing of the data

What are your data

Features used