

Involving Machine Learning and Artificial Intelligence in Automatic Localization on Mozilla Verbatim, Pootle.

Sudheesh Singanamalla, Rashid Khan.

Abstract Project Proposal to Mozilla for Google Summer of Code 2014

Localization plays a major role in Mozilla in spreading its vision of an open and literate web. Localizers are major contributors in making the open source products from Mozilla like Firefox and websites available in 30+ languages, this involves a lot of data being generated in the translation followed by a review process of these translated string. The project aims to use existing database of translations for a particular string and use rating factors and other rating methods to figure out the best translation to a particular string. Also this project can use automatic translation since the data generated by the users over many years can be used for mining and then processing for translation of strings.

This project could also be used in localization of the documentation present on MDN into different languages for the world to use. A feed can be built onto MDN requesting the to be localized content as paragraphs of data and then localizing them using the automatic localization features.

PROPOSAL:

We would like to propose this project, as 2 GSoC subprojects that could be taken under Mozilla.

1. Implement API services on existing Mozilla verbatim Pootle.
2. Use the API services along with Big Data and AI in automatic translation.

BENEFITS:

1. Translation time can be saved.
2. More time is spent in review process since translation of these strings happens automatically.

OBJECTIVES:

Short term goal : To make a machine learning integration with Mozilla's localization platform.

- Successfully be able to use all possible sources of data to translate all strings into atleast 3 different languages automatically. Existing possibilities with the languages we can speak and localize into , from English are
 - Hindi
 - French
 - Arabic
 - Urdu
 - Telugu

Long term goal : To export the project into a full fledged stand alone application with links

to various localization platforms like Transifex, Pootle etc., thus enhancing the localization features of many open source software projects.