Text Classification Framework

By Olivia Wu, Teresa Wan, Terry Li

Contents

1. Domain
2. Generality VS Specificity
3. Project structure
4. Plugin interfaces

Domain / Generality

Domain

1. Purpose: Text Classification & Entity Analysis using Google NLP
2. Data API: NY Times / Wikipedia / Guardian
3. Visualization API: Multiple Time-series charts / Word Cloud

Generality

1. Framework: Classification Using NLP
2. Data Plugin: Using web API fetch data and extract text
3. Visualization Plugin: Customize graph / visualization

Project Structure & Plugin Interface

Project Structure
– Similar to recitation 09: decouple Backend & GUI

– Using service loader to register plugins

Plugin Interface

– Data Plugin: runWebAPIRequest() / processData()

– Visualization Plugin: visualize()

TextVisualizationFramework \<interface>

- currentDataPlugin

- currentVisualizationPlugin

- registeredPlugins

+ registerDataPlugin

+ registerVisualizationPlugin

+ getTextInfo

+ visualizeData

TextVisualizationFramework

- currentDataPlugin

- currentVisualizationPlugin

- registeredPlugins

+ registerDataPlugin

+ registerVisualizationPlugin

+ getTextInfo

+ visualizeData

implements

has a

has a

DataPlugin \<interface>

- result

- runWebAPIRequest

- processData

+ getProcessedData

VisualizationPlugin \<interface>

+ visualize