Pythia

Employing Lexical and Semantic Features for Sentiment Analysis

Ioannis Katakis	Harokopio University of Athens imktks@gmail.com
Iraklis Varlamis	Harokopio University of Athens varlamis@hua.gr
George Tsatsaronis	Technical University Dresden george.tsatsaronis@biotec.tu-dresden.de
Presented by: Alina Petrova	Technical University Dresden alina.petrova@gmail.com

pythia

What is Pythia ?

Pythia is an online service for sentiment analysis and word sense disambiguation.

It performs:

- Word sense disambiguation using popular WSD techniques
- Sense-level sentiment analysis using SentiWordNet
- Sentence-level sentiment analysis using different classification models

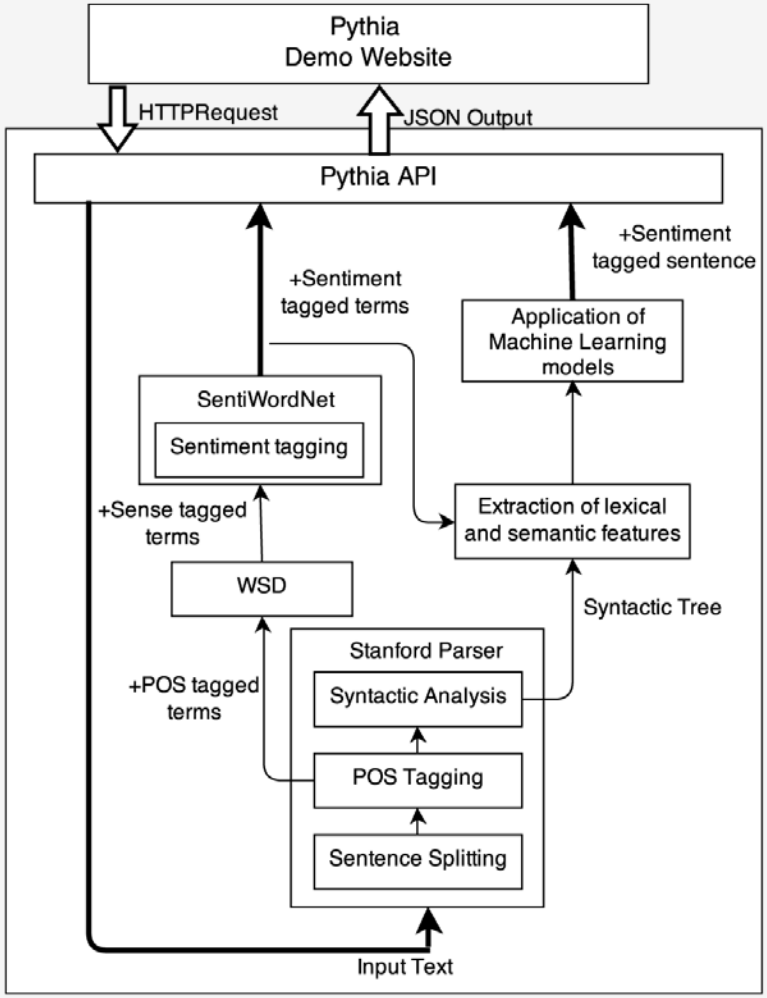
pythia

Contribution

- Web Service that performs sentiment analysis and word sense disambiguation
  - Different combinations of WSD methods and feature sets for sentiment classification
- REST API for using the service via HTTP requests
- Graphical interface – GUI for easier interaction with the service
  - http://omiotis.hua.gr/pythia

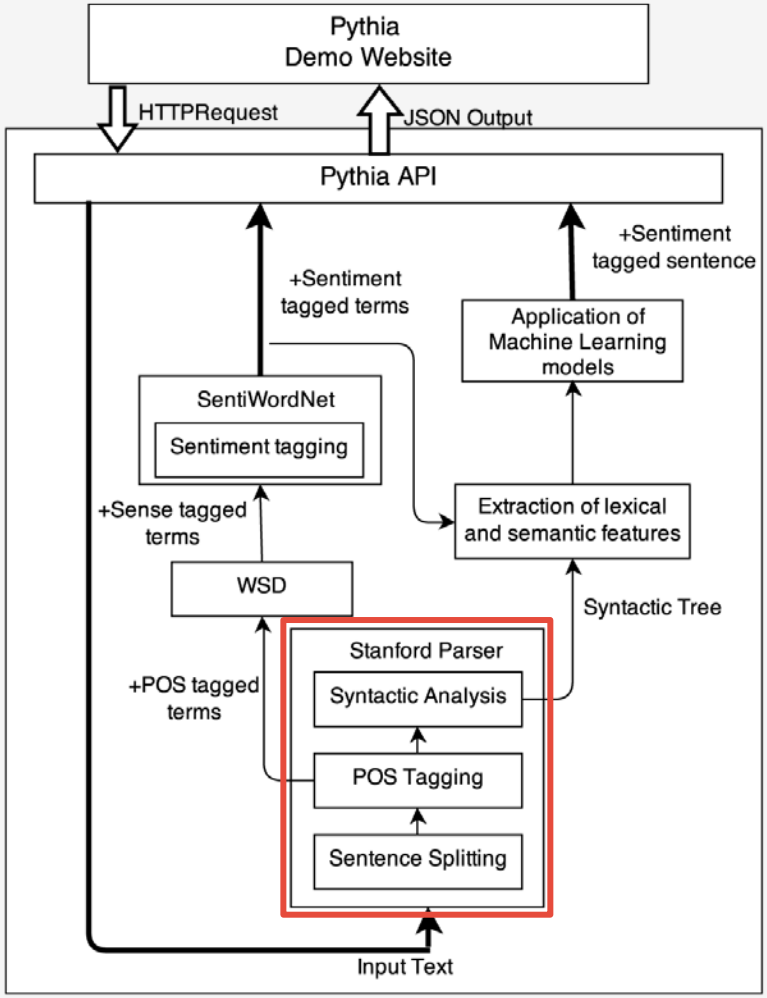
pythia

Architecture



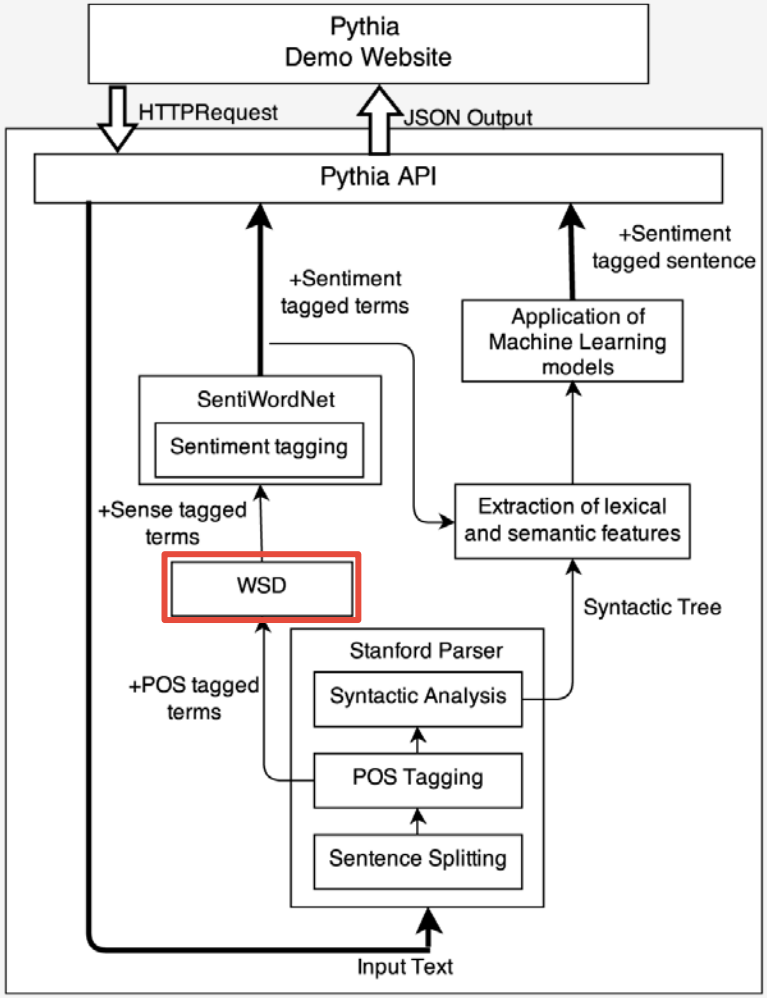
pythia

Architecture - Syntactic Parser



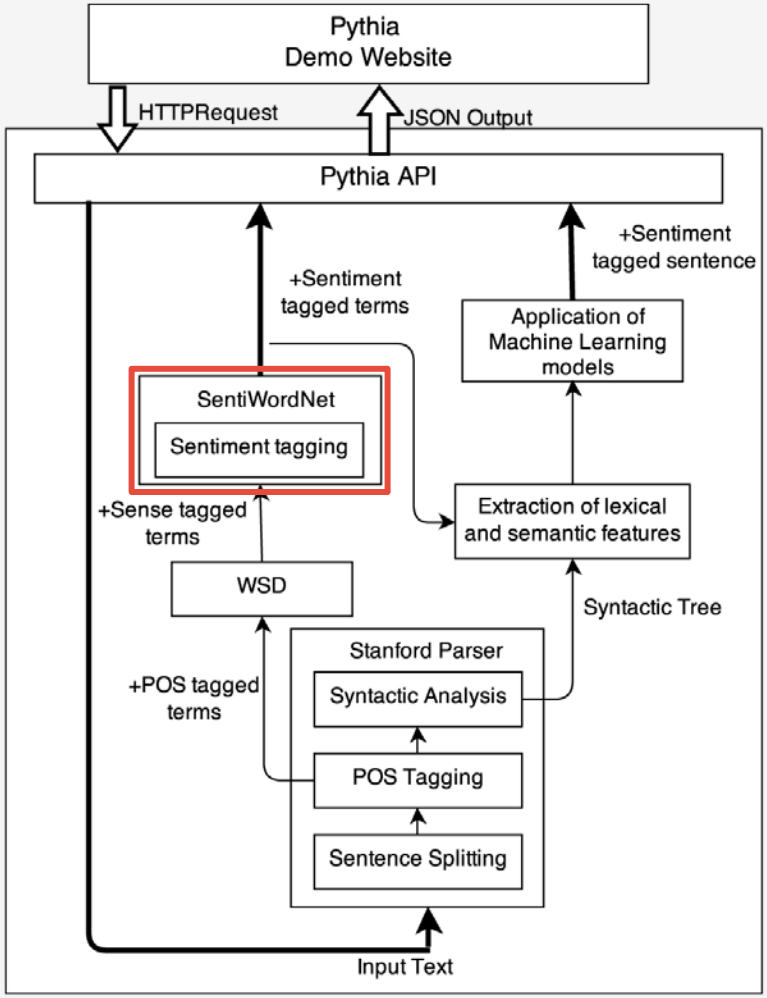
pythia

Architecture - Word sense disambiguation



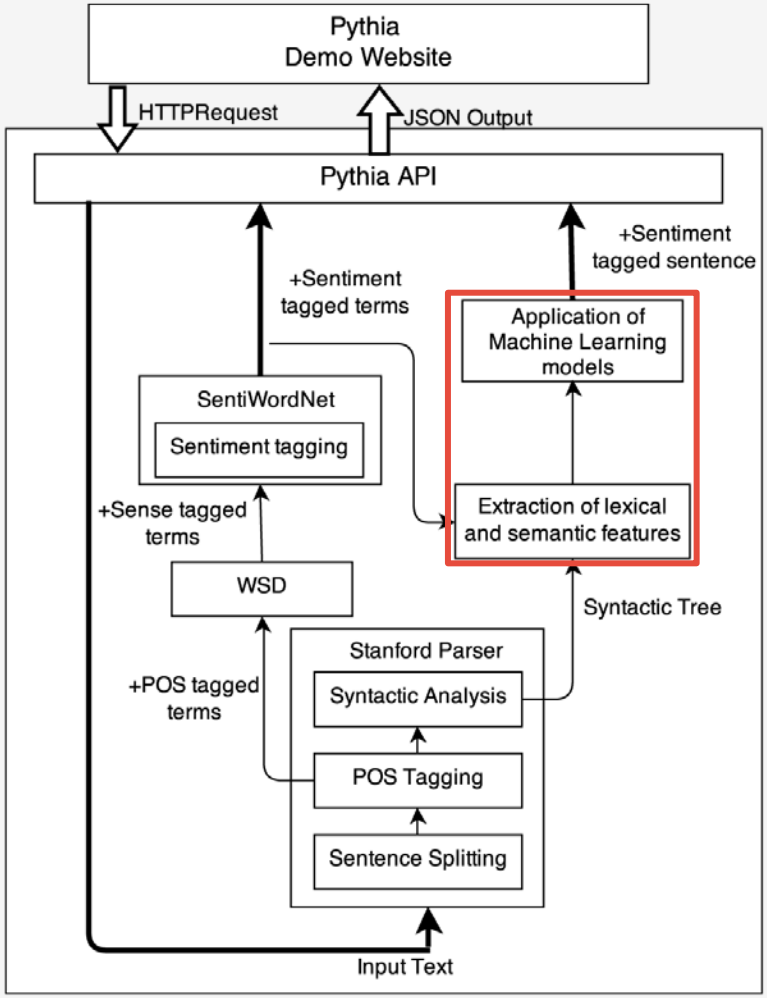
pythia

Architecture - Sense level sentiment analysis



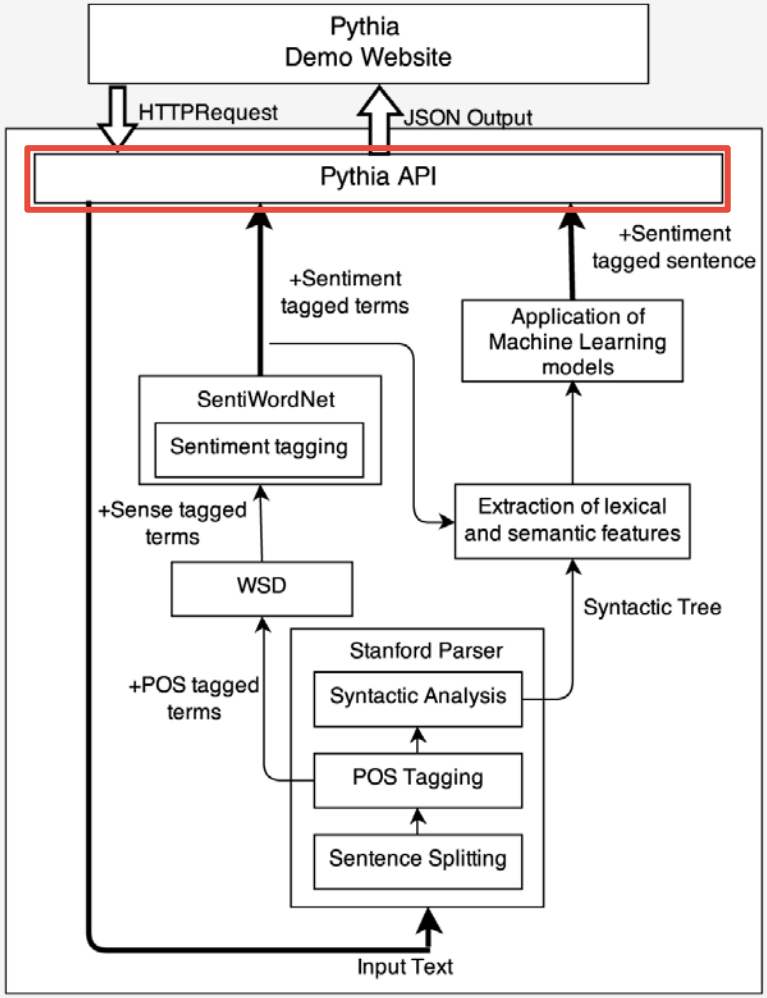
pythia

Architecture - Sentence level sentiment analysis



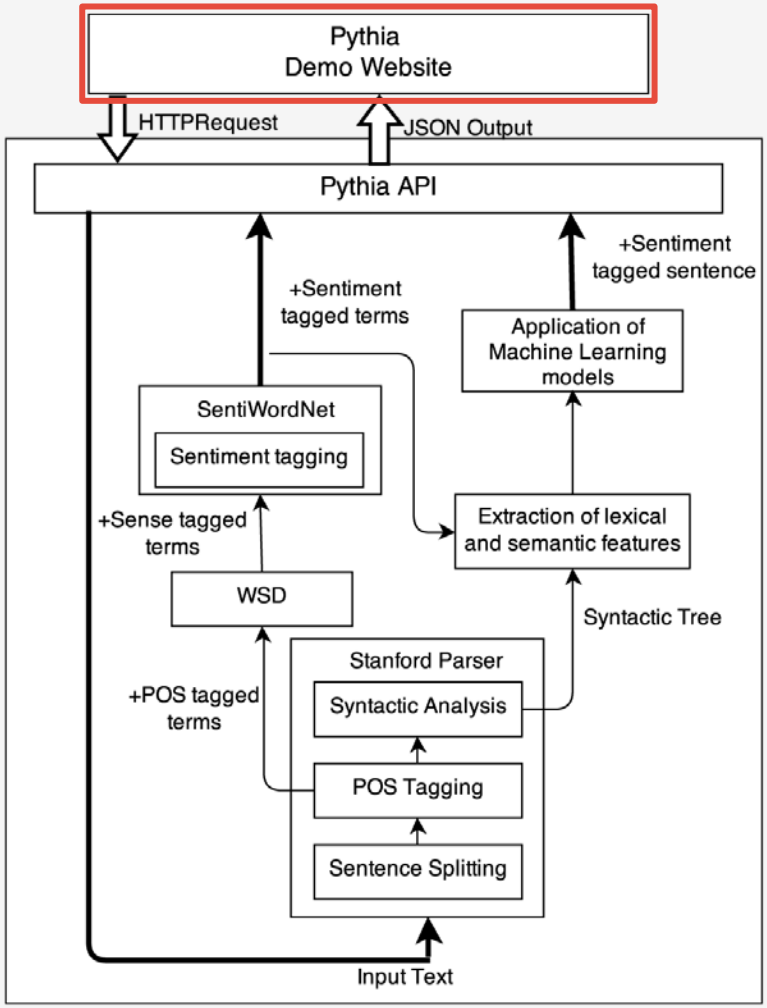
pythia

Architecture - API



pythia

Architecture - GUI



pythia

Tools and resources

- Back-end
  - WordNet
  - Stanford Parser
  - SentiWordNet
- Front-end
  - Java - Jersey Framework
  - HTML
  - JavaScript - jQuery, Chart, rangy, Trip, jQuery-slimScroll, iCheck
  - CSS

pythia

Pythia's Back-end

- WSD methods:
  - First Sense (FS)
  - Weighted Degree (WDEG)
  - Integer Linear Programming (ILP)
- Classifiers:
  - Support Vector Machines (SVM)
  - Logistic Regression
  - Naive Bayes

pythia

WSD Methods

- First Sense (FS)

  the most popular sense for each word according to WordNet
- Weighted Degree (WDEG)

  algorithm which computes the weighted sum of the edges for each node
- Integer Linear Programming (ILP)

  method that addresses the problem of word-sense disambiguation as a linear programming problem (maximization of pairwise sense similarity)
  - Panagiotopoulou, V., Varlamis, I., Androutsopoulos, I., & Tsatsaronis, G. (2012). Word sense disambiguation as an integer linear programming problem. In Artificial Intelligence: Theories and Applications (pp. 33-40). Springer Berlin Heidelberg.

pythia

Sentence sentiment classification

Classification on a movie reviews dataset using:

- 3 classifiers:
  - SVM
  - Logistic Regression
  - Naive Bayes
- 5 set of features:
  - Semantic features (4)
  - Char n−grams (11,923)
  - Term n−grams (214,342)
  - All n−grams (225,475)
  - All Features (225,515)

pythia

Pythia's Front-end

- API (omiotis.hua.gr/pythia/api.html)
  - REST
  - HTTP GET and POST requests
  - Different endpoints for each word sense disambiguation method
  - Response data in JSON format
- Demo GUI (omiotis.hua.gr/pythia)
  - User friendly interface based on JavaScript that exposes the service capabilities
  - How-to-use animation clip for easier customization of the service

pythia

Evaluation

- The evaluation was performed on the movie reviews dataset (Pang & Lee 2005) as employed by Socher et al 2013
- Our best sentence-level sentiment analysis model achieved accuracy up to **80.73%**

ML	Semantic Features (40)	Char $n-$grams (11,923)	Term $n-$grams (214,342)	All $n-$grams (225,475)	All Features (225,515)
SVM	68.26	73.35	**80.11**	79.01	80.04
Log. Regression	**68.43**	69.07	77.31	78.65	79.01
Naive Bayes	64.66	**75.35**	74.32	**79.81**	**80.73**

- The combination of all features, semantic and lexical, leads to the best results.

Conclusion

- Pythia is a flexible system consisting of individual subsystems
- It has the ability to employ different components in order to achieve better performance
- Its different components may affect negatively the overall performance of the system

Demo time…

omiotis.hua.gr/pythia

Thank you!