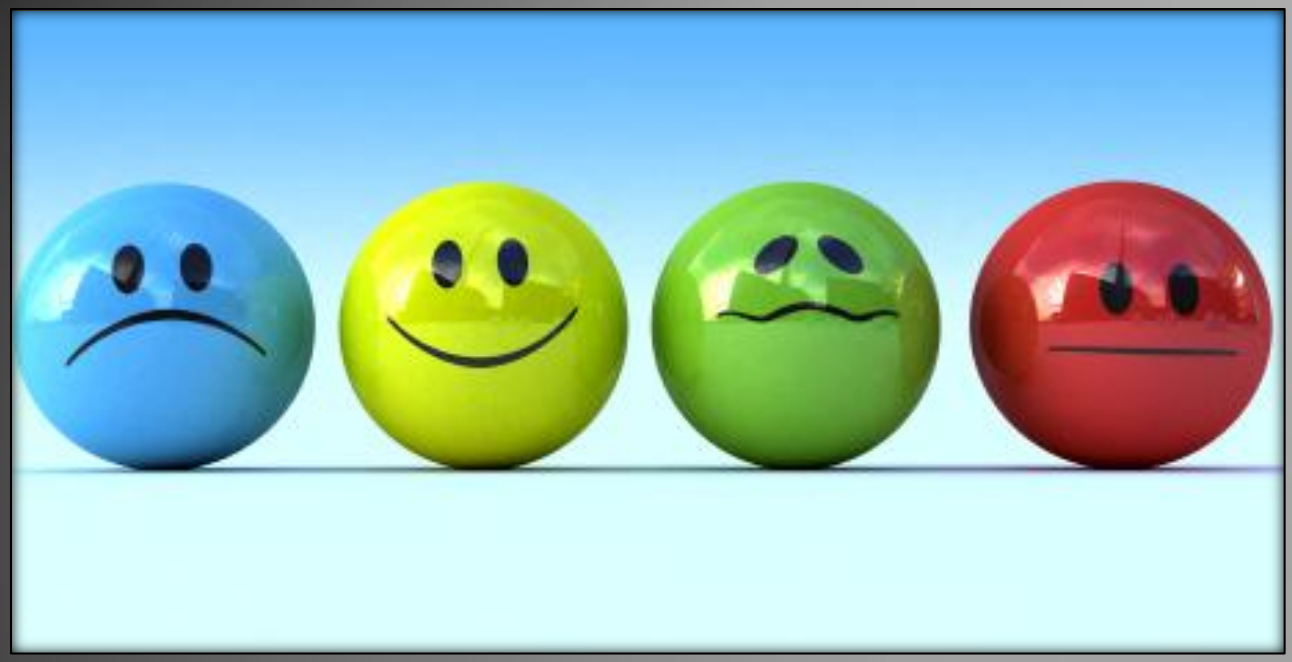


Sentiment Analysis system for Enterprise tracking and rating services (For Persian language)

Created By:
Fatemeh Amiri
Mohammadhassan Khodashahi

Mentor:
Dr. Simon Screrri

universitätbonn

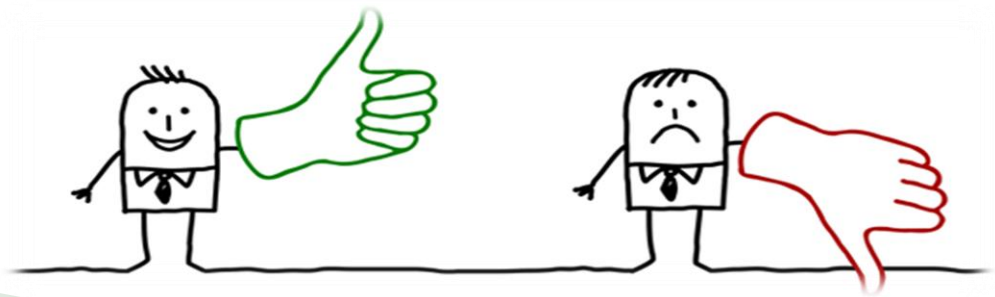
contents

- Motivation
- Introduction
- Programming languages
- Applications
- Approach
  - ✓ Data collection
  - ✓ Processing resource
    - Tokenizer
    - Sentence splitter
    - Pos tagger
    - Jape rules
- Testing
  - ✓ Performance testing (1)
  - ✓ Performance testing (2)
- Conclusion and future works
- References

Motivation

- Online shopping become more preferable for customers.
- E-business has improved in different aspects.

- User feedbacks and comments on different products have an important roll on other users' decision.
- Sentiment analysis works on this amount of data to give the expected information for different Enterprises to help them improved.

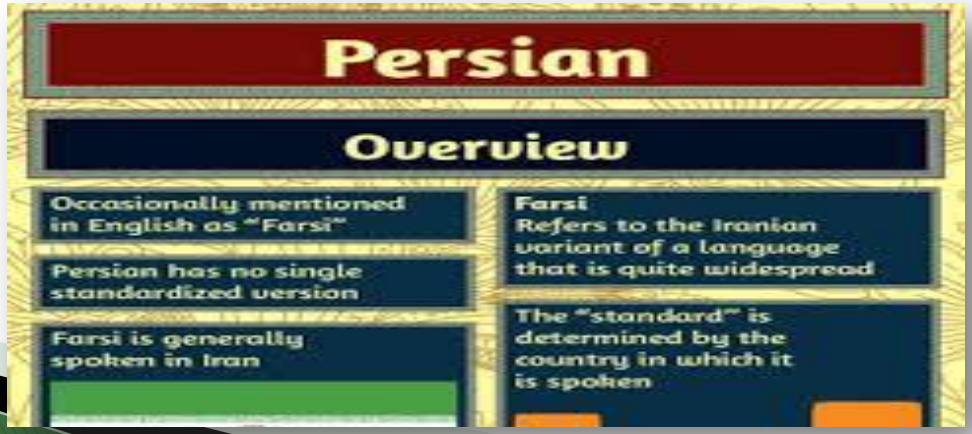


Introduction

- Definition (Wikipedia): Sentiment analysis (opinion mining) refers to the use of natural language processing and text mining to identify and extract subjective information in source materials.

- The lexicon-based approach works on semantic orientation of words and phrases in the document (Turney 2002).

- Why for Persian language?! Number of Persian websites and users increasing, but no such research for Persian API and sentiment analysis.



Programming languages

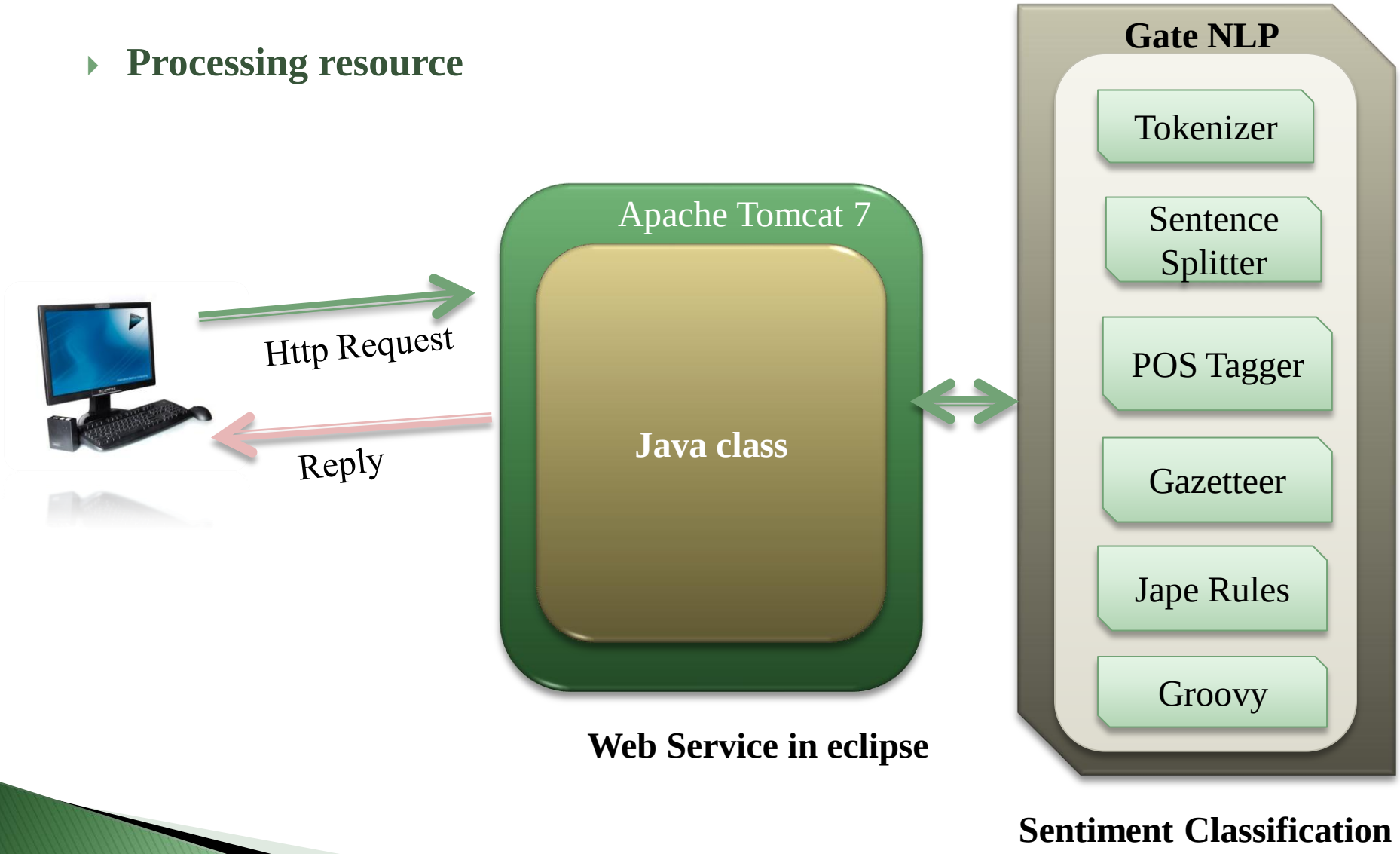
Applications

Approach

- **Data collection:**
  - Gathering 6500 Persian adjectives in a web application
    www.computerssl.com
  - Asking Persian people to vote about their sense on each word or phrase
  - Calculate the average voting for each word, to define the sentiment of them.

Approach…

- **Processing resource**



Approach…

- **Tokenizer:** splits the text into very simple tokens
- **Sentence splitter:** Fragments the text into its sentences
- **Gazetter:** our Gazeteer is our gathered adjectives and phrases with their sentiment
- **Jape Rules:** To identify regular expressions we have formulated as grammar base
  - **Word level**

    “بی” “Bi” and “نا” “Na” + noun => negative word (بی معرفت, نامرد)

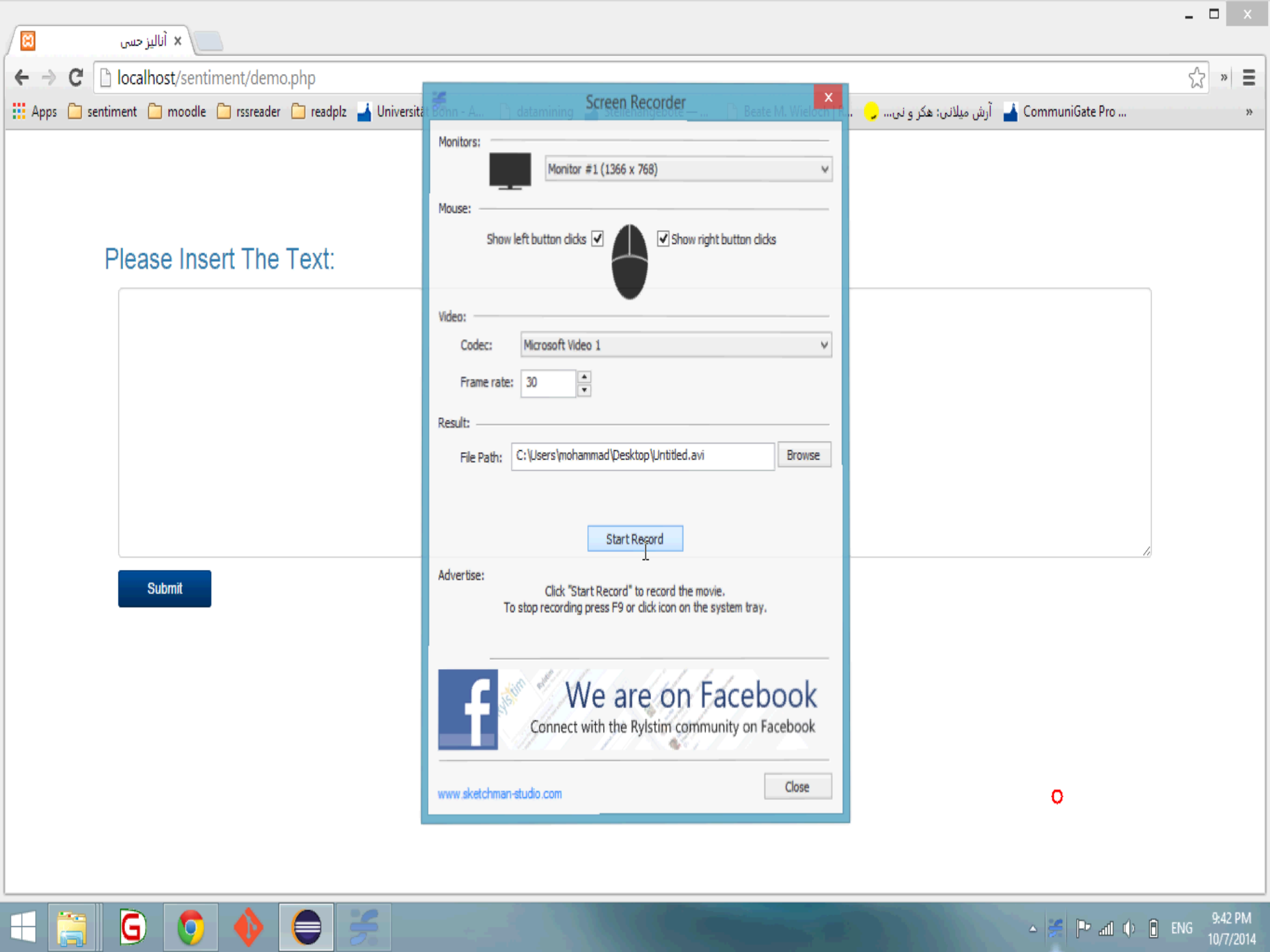
  - “ ن” + Verb => negative verb (نمی فهمد)
  - **Sentence level**

    “ He is not a lair”

    “That film had a lot of famous actors but it couldn’t attract people’s

Selected Processing resources

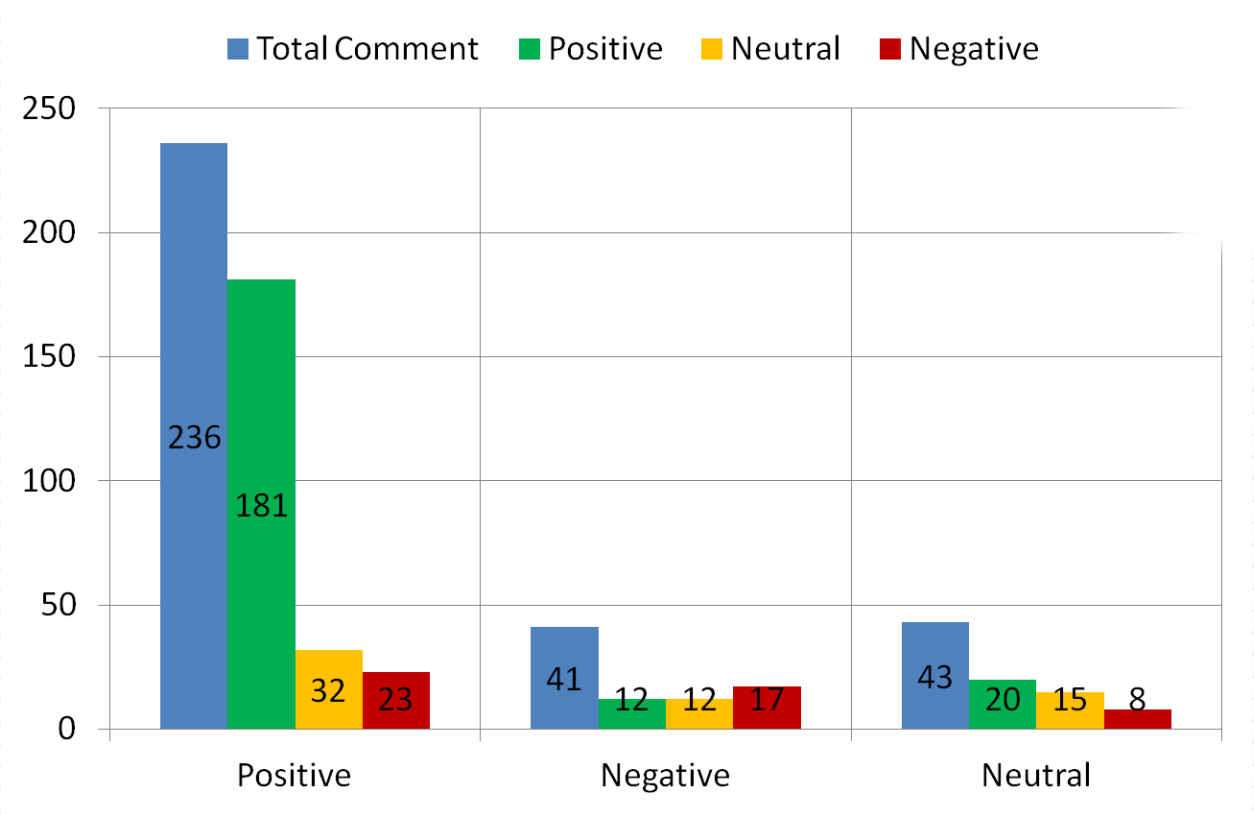
!	Name	Type
	Tokenizer	Persian Tokenizer
	Sentence_splitter	Persian Sentence Splitter
	POs	Persian POS Tagger
	Gazetteer	Hash Gazetteer
	Negation	JAPE Transducer
	Negword	JAPE Transducer
	Negverb	JAPE Transducer
	Sentences	JAPE Transducer
	Negverb_in_sentences	JAPE Transducer
	Groovy	Groovy scripting PR


آنالیز حسی
localhost/sentiment/demo.php
Apps
sentiment
moodle
rssreader
readplz
CommuniGate Pro ...
Please Insert The Text:
Submit
Screen Recorder
Monitors:
Monitor #1 (1366 x 768)
Mouse:
Show left button clicks
Show right button clicks
Video:
Codec:
Microsoft Video 1
Frame rate:
30
Result:
File Path:
C:\Users\mohammad\Desktop\Untitled.avi
Browse
Start Record
Advertise:
Click "Start Record" to record the movie.
To stop recording press F9 or click icon on the system tray.
We are on Facebook
Connect with the Rylstim community on Facebook
www.sketchman-studio.com
Close
ENG
9:42 PM
10/7/2014

Testing

- **Performance test (1):**

Using a Persian web site : www.iran-booking.com



Testing…

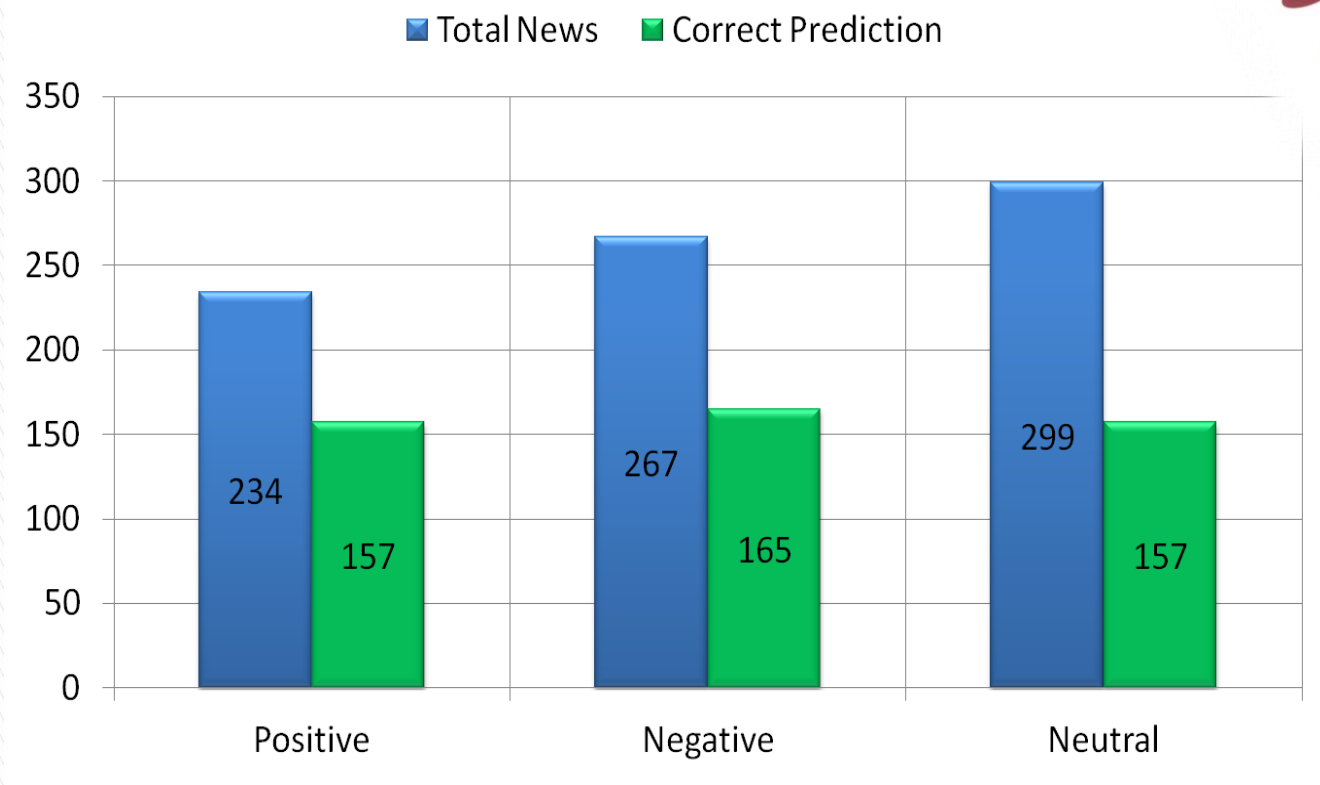
- **Performance test (2)**

Gathering about 5000 news from most famous Persian news websites depends on Alexa rating.

Choose 800 news randomly and asked Persian people to vote on them about their sense.

Performance test (2)…

1200 votes have been gathered on 800 news.



Conclusion and Future works

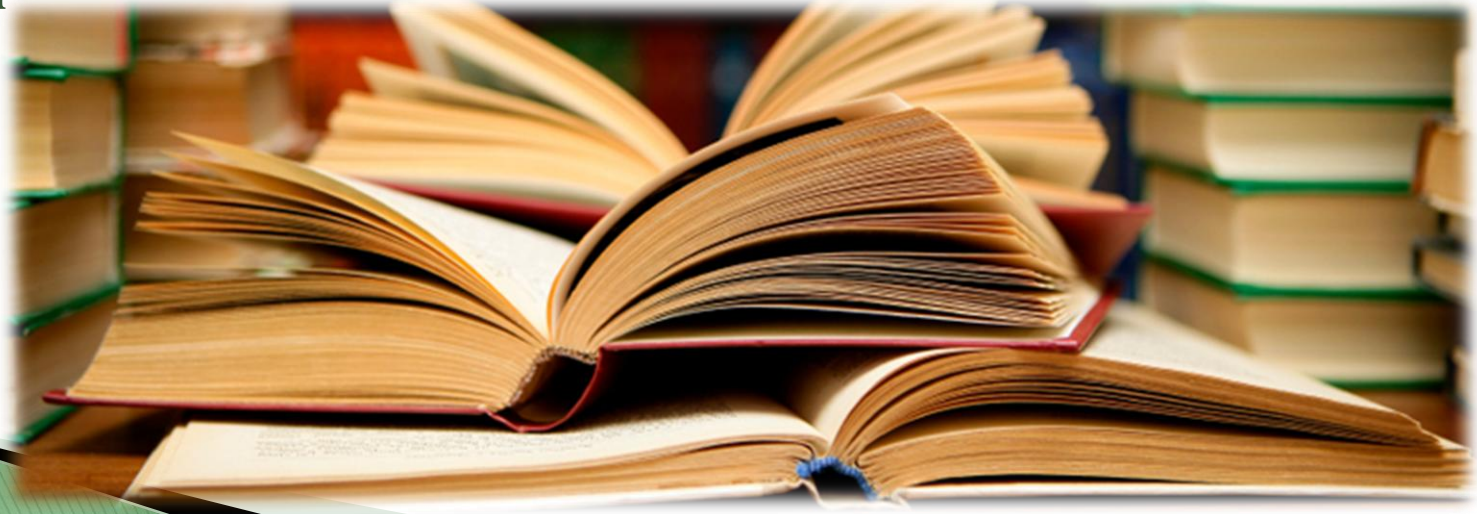
- For the first Persian API we receive to 65.7% accuracy till now

How to improved the accuracy?

- more jape rules especially in the form of “BUT” “Although” and …
- Encouraging more Persian people to vote on the adjectives and phrases to have a better lexicon data base.
- Working on informal form too that’s common in tweets and comments.

References

- **[1]** A SVM-Based Method for Sentiment Analysis in Persian Language (Mohammad Sadegh Hajmohammadi, Roliana Ibrahim)
- **[2]** https://gate.ac.uk/
- **[3]** "Majid Sezavar" http://sazvar.student.um.ac.ir/
- **[4]** http://www.coreservlets.com/Apache-Tomcat-Tutorial/tomcat-7-with-eclipse.html
- **[5]** http://javapapers.com/web-service/java-web-service-using-eclipse



Thanks for your attention…