

Arabic Sentimental Lexicon Creation and Finding Related Entities

Mini Project – **ESWC 2014**

Mentors : **Elena Simperl & Isabelle**

Presented by – Malik Saad

Khalik Hussain, Takeshi Masuda

The Noora Group

Date 06 Sept 2014

Introduction

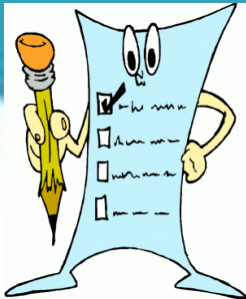


Introduction

- Before Web
 - Ask friends and family
 - Surveys
- After Web
 - Blogs, Online Discussion Forums



+



Sentiment Analysis

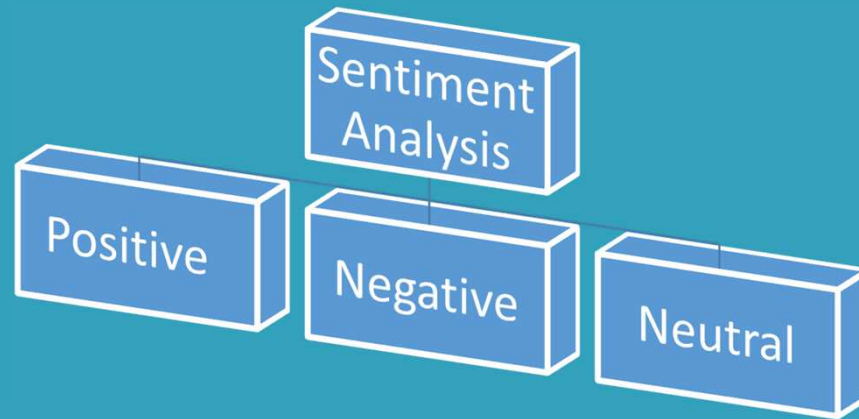
Automatically detecting sentiments from text.

Examples

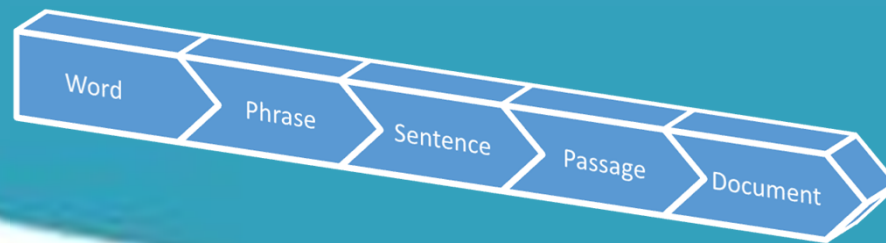
This camera
has a **poor**
battery

Kalamaki is a
beautiful
place to visit

Sentiment Analysis (2)



Text Granularity Level



General Approach

- Use a Sentimental Lexicon (in support with NLP or Machine Learning).
- Sentimental Lexicon – provide polarities of subjective words
 - SentiWordNet
 - General Inquirer
 - Opinion Lexicon, etc.

Example – SentiWordNet (SWN)

```
a 00005107 0.5 0 uncut#7 full-length#2 complete; "the full-length play"
a 00005205 0.5 0 absolute#1 perfect or complete or pure; "absolute loyalty"; "absolute silence"; "absolute truth"; "absolute alcohol"
a 00005473 0.75 0 direct#10 lacking compromising or mitigating elements; exact; "the direct opposite"
a 00005599 0.5 0.5 unquestioning#2 implicit#2 being without doubt or reserve; "implicit trust"
a 00005718 0.125 0 infinite#4 total and all-embracing; "God's infinite wisdom"
```

Challenges for Sentiment Analysis

- Sentiment-Topic Association
- Noisy & Informal Text
- Contextual Ambiguities
- Unavailability of Tools for Arabic Text

Sentiment Lexicon Creation for Arabic Language Text

Sentiment Lexicon Creation from Arabic Language Text

Arabic Lexicon Creation

1. Stock Market Textual Data (200 documents) downloaded from Web
2. Sentiment Lexicon in English (SentiWordNet)

Processing Pipeline

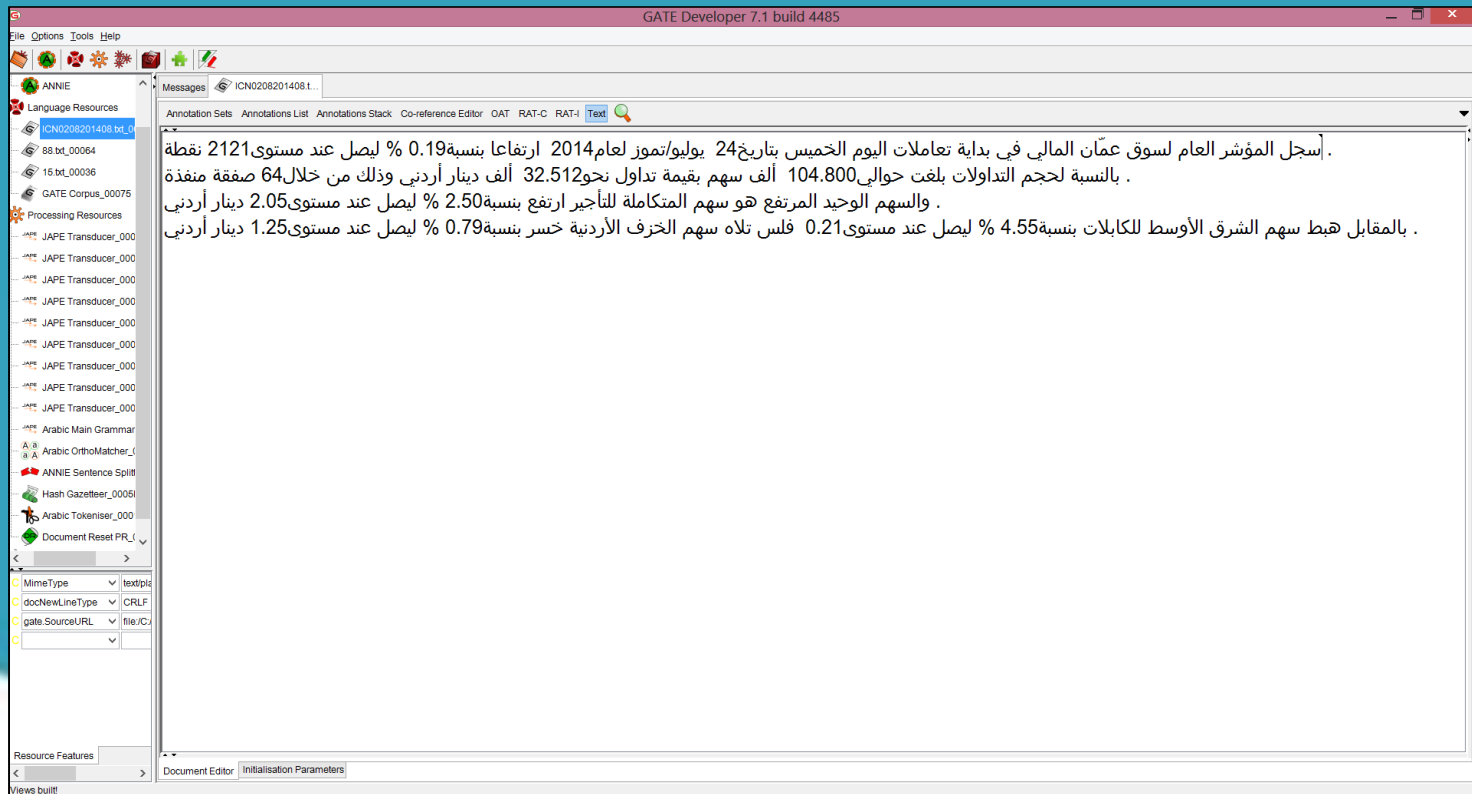
1. Download Financial Data
2. Pre-processing
 - POS Tagging
 - Stop Word Removal
3. Locate verbs and adjectives
4. Translate them to English equivalents using Google Translate APIs
5. Stem these words
6. Use SentiWordNet to find their positive and negative score
7. Write Arabic word and its positive/negative score as output in a text file

Related Entity Extraction

1. Locate the sentimental Arabic words in the document and use window based approach to find related entities (nouns in our case)
2. Create a list of related entities for each sentimental word in the data collection
3. Filter it manually to remove incorrectly tagged nouns by POS tagger.

Screen Shots

Figure 1: Arabic Text



Screen Shots

Figure 2:

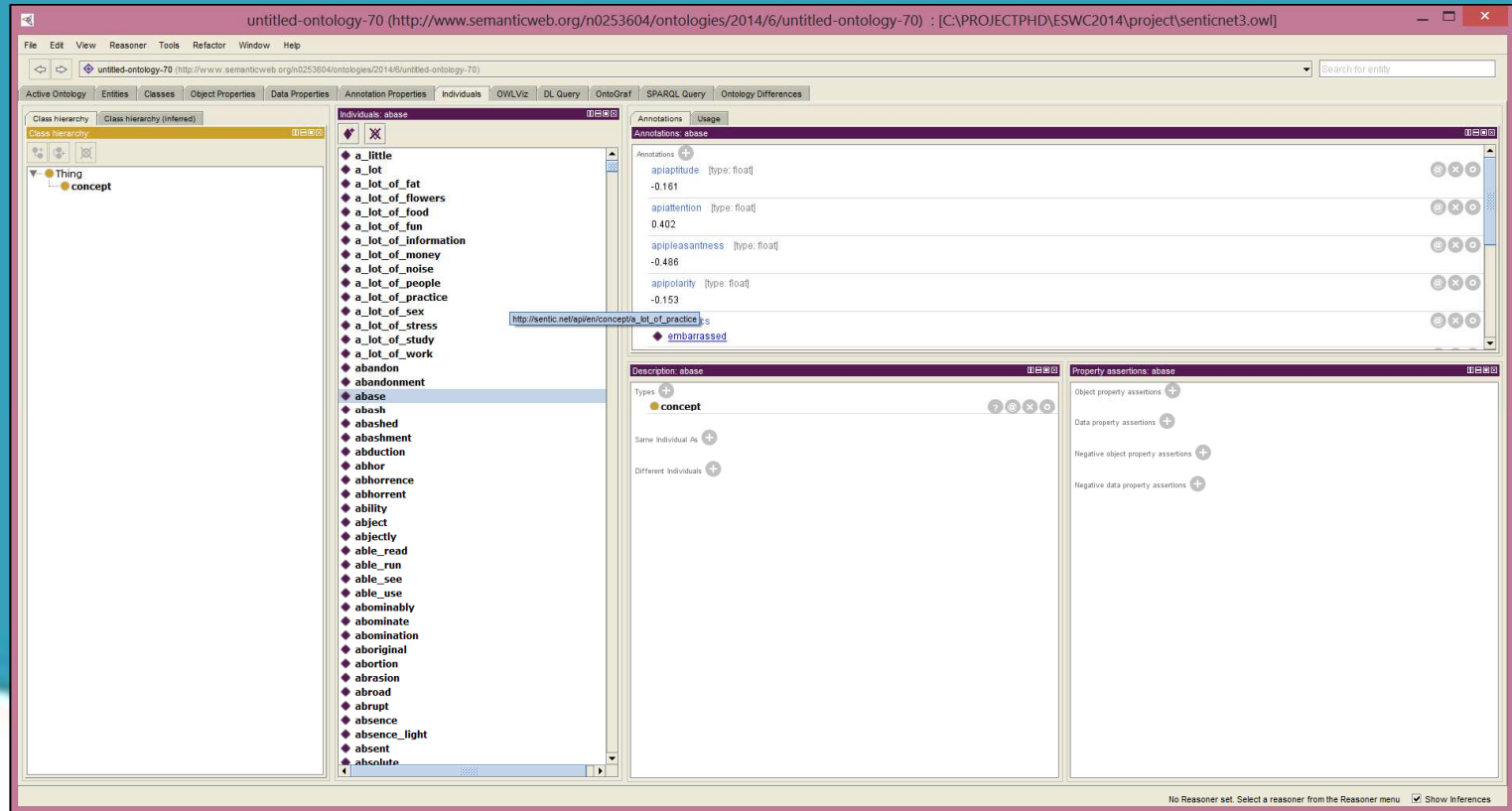
```
form_pattern02
Phase:der_adj
Input: word1 Token
//note that we are using Lookup and Token both inside our rules.
Options: control = appelt

Rule: pick
(
  (
    {Token.string== "شركة"}
    {Token.string== "وشركة"}
    {Token.string== "سهم"}
    {Token.string== "وسهم"} |
    {Token.string== "قطاع"}
    {Token.string== "مؤشر"}
    {Token.string== "المؤشر"}
  ):port1

  (
    {word1.Type=="Noun"}
    ({word1.Tvoe=="Noun"})*
  )
)
```

Screen Shots

Figure 3: English Sentimental Lexicon



Summary

1. Sentiment lexicon for Arabic does not exist
2. Creating Arabic Sentimental Lexicon from Arabic Financial Text and English lexicon
3. Finding related entities co-occurring with the sentimental words

Thanks