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

- Sentiment Analysis
- Positive
- Negative
- Neutral

Word → Phrase → Sentence → Passage → Document
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)
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.
أسجل المؤشر العام لسوق عمّان العالي في بداية تعاملات اليوم الخميس بتاريخ 24 يوليو/تموز لعام 2014 ارتفاعاً بنسبة 0.19% ليصل عند مستوى 212.1 نقطة.

بالنسبة لحجم التداول، بلغ حوالي 104.8 مليون سهم بقيمة بروكير 32.5 مليون دينار أردني، وذلك من خلال 84 صفقة معقدة.

والمستوى المرتفع هو هم المكثاف الأعلى أرتفاع بنسبة 2.5% ليصل عند مستوى 2.05 دينار أردني.

بال مقابل، هبط سهم الشرق الأوسط للكلات بنسبة 4.55% ليصل عند مستوى 0.21 دينار أردني، سهم الخرف الأردنية بنسبة 0.79% ليصل عند مستوى 1.25 دينار أردني.
Screen Shots

Figure 2:
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