

Big Data and Arabic Sentiment Analysis



Presented by:
Dr Jassim Haji
Director of IT, Gulf Air



Social Media Explosion

- Billions of users in social media/networks
- Never in history had we this amount of opinionated data
- Word of mouth can make or break your business



Social Media Importance to Business

- Crucial to monitor your customer feedback and public perception



- When you know how your customer feel & think:
 - Take corrective/proactive action if the customer is upset
 - Exploit satisfied customers and good publicity

Social Media Importance to Business

- You can monitor your competitors



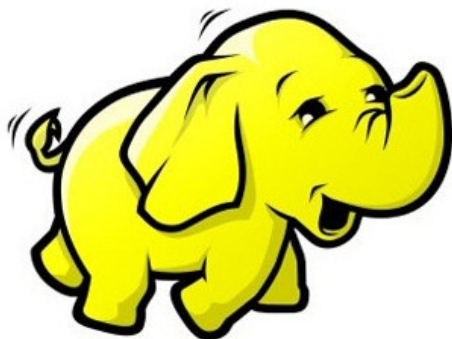
- You can track market situation



- The tool to achieve this is sentiment analysis and opinion mining

Big Data & Social Media Monitoring

- Big Data is playing a critical role in social media monitoring
- Automates the feed, makes it structure for processing
- Scalable to handle huge amount of data processing
- Gulf Air is using Cloudera's open-source software on virtual servers (leveraging existing Gulf Air's cloud environment)
- Gulf Air developed the skills required for this implementation internally without relying on external parties



Sentiment Analysis/Opinion Mining

- Determining feelings from text: (attitudes, opinions, emotions)



- Complex field, subset within Natural Language Processing
- With social media, sentiment analysis has gained importance
- Very subjective, depends on topics/language/ region/context, no single solution to fit all



The Arabic Language

- Arabic: most prevalent language in the Middle East and North Africa

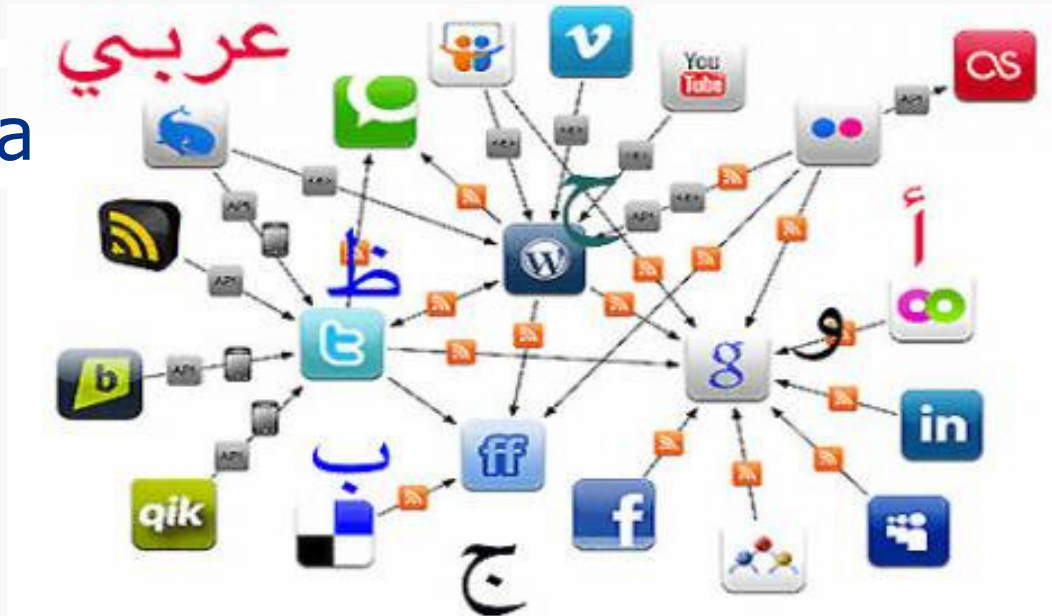


- 422 million native speakers
- Rich but complex language and grammar, with formal, colloquial and dialect varieties.



Arabic Social Media

- More than 80% of posts in social media about Gulf Air are in Arabic
- Very few providers offer Arabic Sentiment Analysis, with limitations and inaccuracies
- Available solutions don't satisfy Gulf Air's needs



Arabic Sentiment Analysis Approach

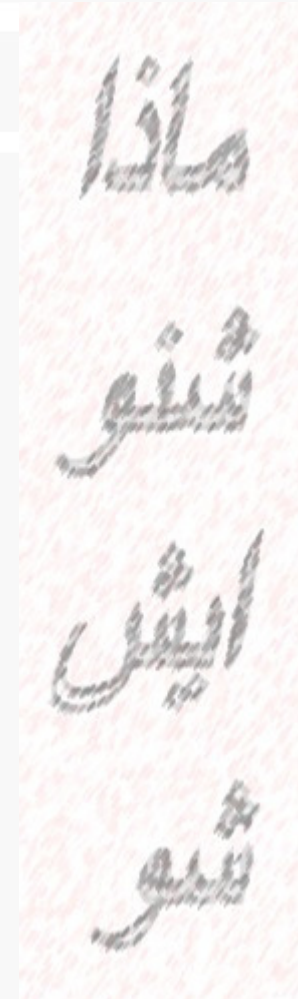
- Gulf Air adopted Lexicon-based approach
- A language is comprised of:
 - Lexicon (list of words)
 - Rules for combining words to form meaning
- Lexicon and rules between words were defined to determine meaning and sentiment

Lexicon

English	Arabic
Gulf Air	طيران الخليج
Flight	رحلة
Seat	مقعد
Airport	مطار
On time	على الوقت
Comfy	مريح

Scope of Arabic Sentiment Analysis

- Formal Arabic
- Local dialects of the region
- Informal and colloquial words and phrases frequently used
- English Translation
- Arabic tweets (Twitter) are currently being analyzed (short and concise text)



Arabic Sentiment Analysis Maturation

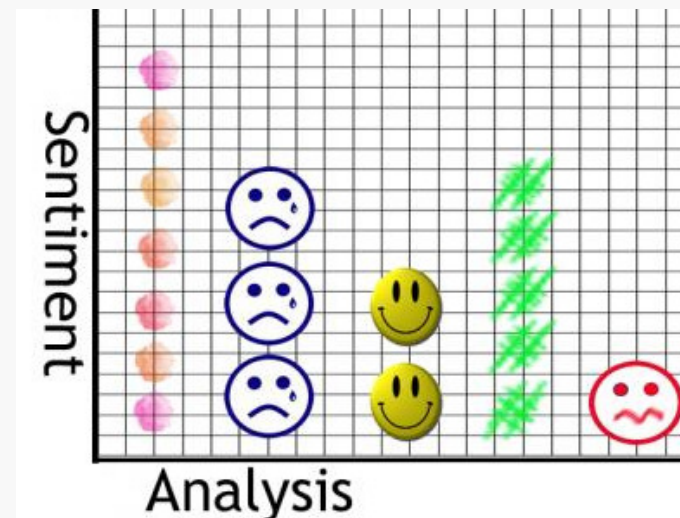
- Gulf Air has rebuilt the Arabic language in the sentiment analysis logic
- Just like a child learning a language, this logic has to learn and mature to increase accuracy of results



Arabic Sentiment Analysis Maturation



- Test and learn approach used:
 - Continuously adding new words
 - Improving/changing the rules
- Results are regularly reviewed to find opportunities for improvements
- Accuracy achieved as of now is 84%



Business Benefits to Gulf Air

- Timely reporting of hot topics and areas of controversy about Gulf Air
- Public/customers' sentiment reporting
- Quick corrective action on negative situations to increase customer satisfaction
- Trend analysis of sentiments and topics over time
- Quick reaction to competition and market/demand shifts



Gulf Air's Future Plans

- Extend the Sentiment Analysis Logic to handle the English language
- Expand to other forms of social media, like Facebook
- Cover newspapers' articles and reviews in travel sites (complex text structure)



Q & A

سين

جيم



Gulf Air is the 1st airline and one of few organizations to develop and use Arabic Sentiment Analysis with Big Data