Techniques and Applications for Sentiment Analysis

Communications of the ACM | April 2013 | Vol. 56 | nº 4
DOI:10.1145/2436256.2436274

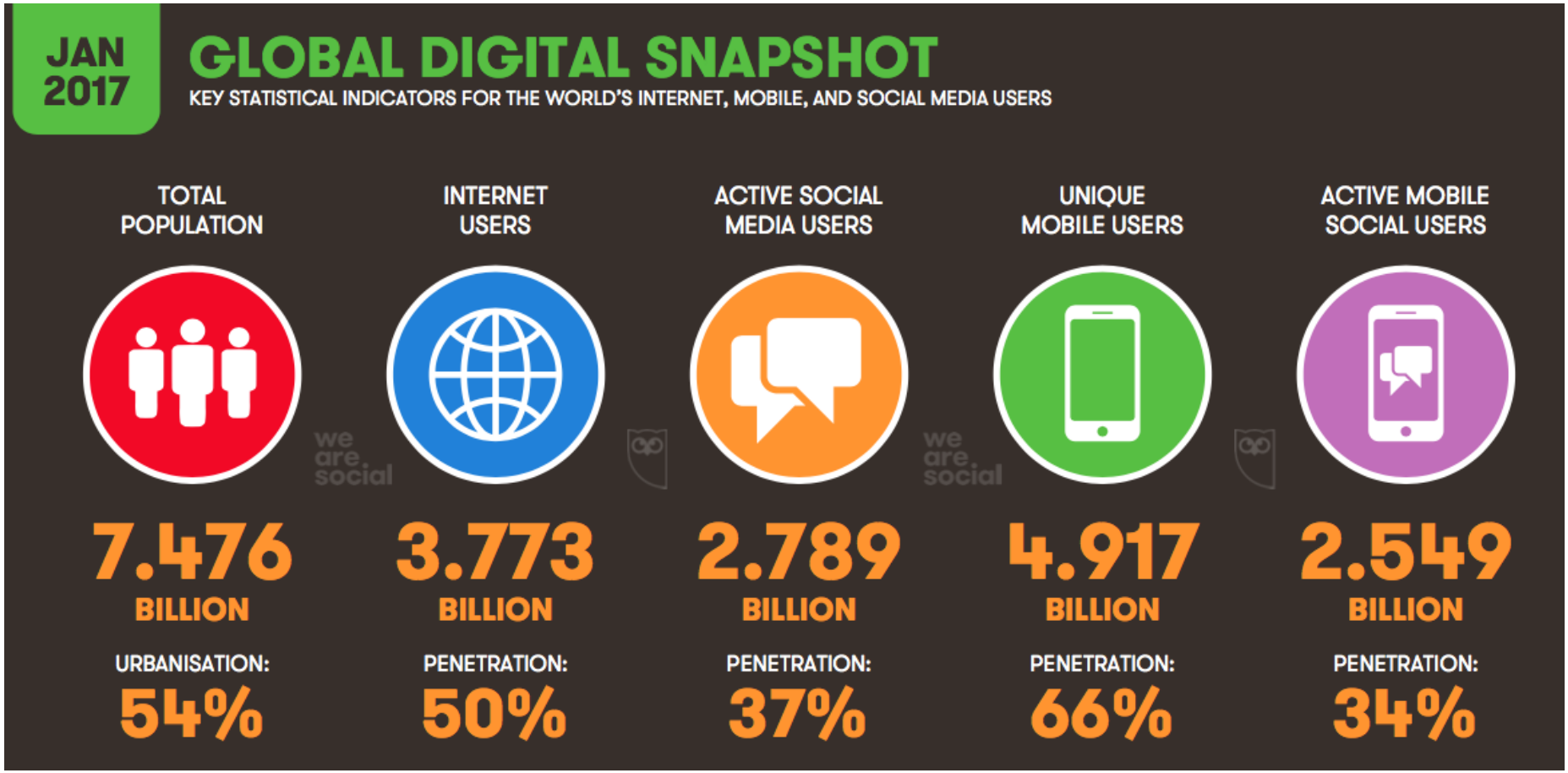
Ronen Feldman
Hebrew University
ronen.feldman@huji.ac.il

Airton Bordin Junior
airtonbjunior@gmail.com
Prof. Dr. Nádia Félix Felipe da Silva
nadia@inf.ufg.br

Roteiro

- Introduction
  - Context
  - Problem
  - Objective
- Sentiment Analysis
- Sentiment Lexicon
- Applications
- Research Issues
- Conclusion

Context


JAN 2017
GLOBAL DIGITAL SNAPSHOT
KEY STATISTICAL INDICATORS FOR THE WORLD'S INTERNET, MOBILE, AND SOCIAL MEDIA USERS
TOTAL POPULATION
INTERNET USERS
ACTIVE SOCIAL MEDIA USERS
UNIQUE MOBILE USERS
ACTIVE MOBILE SOCIAL USERS
we are social
we are social
7.476 BILLION
3.773 BILLION
2.789 BILLION
4.917 BILLION
2.549 BILLION
URBANISATION: 54%
PENETRATION: 50%
PENETRATION: 37%
PENETRATION: 66%
PENETRATION: 34%

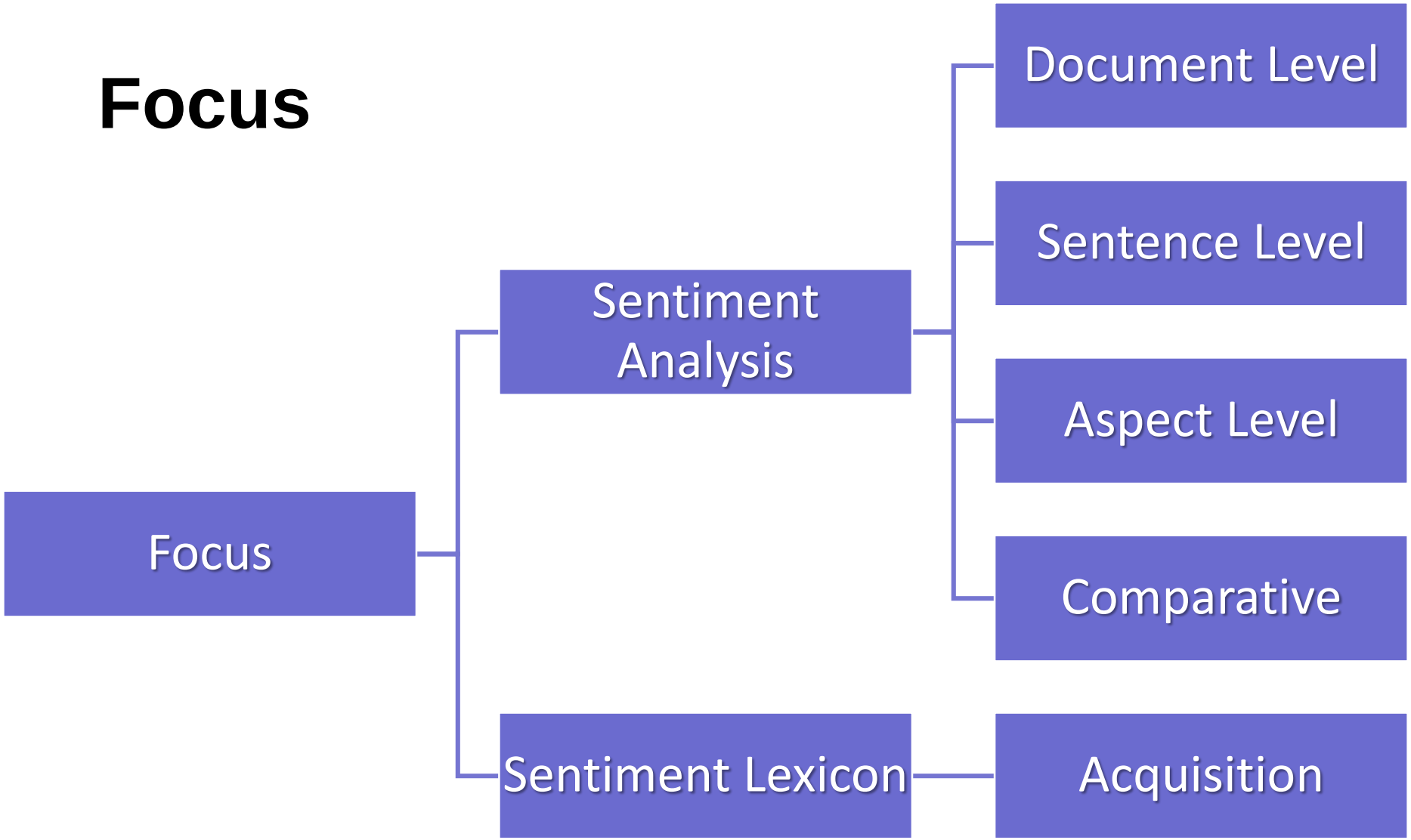
Objective

- Present the main research problems related to Sentiment Analysis (SA) and some of the techniques used to solve them
- Review some of the major application areas where sentiment analysis is being used today

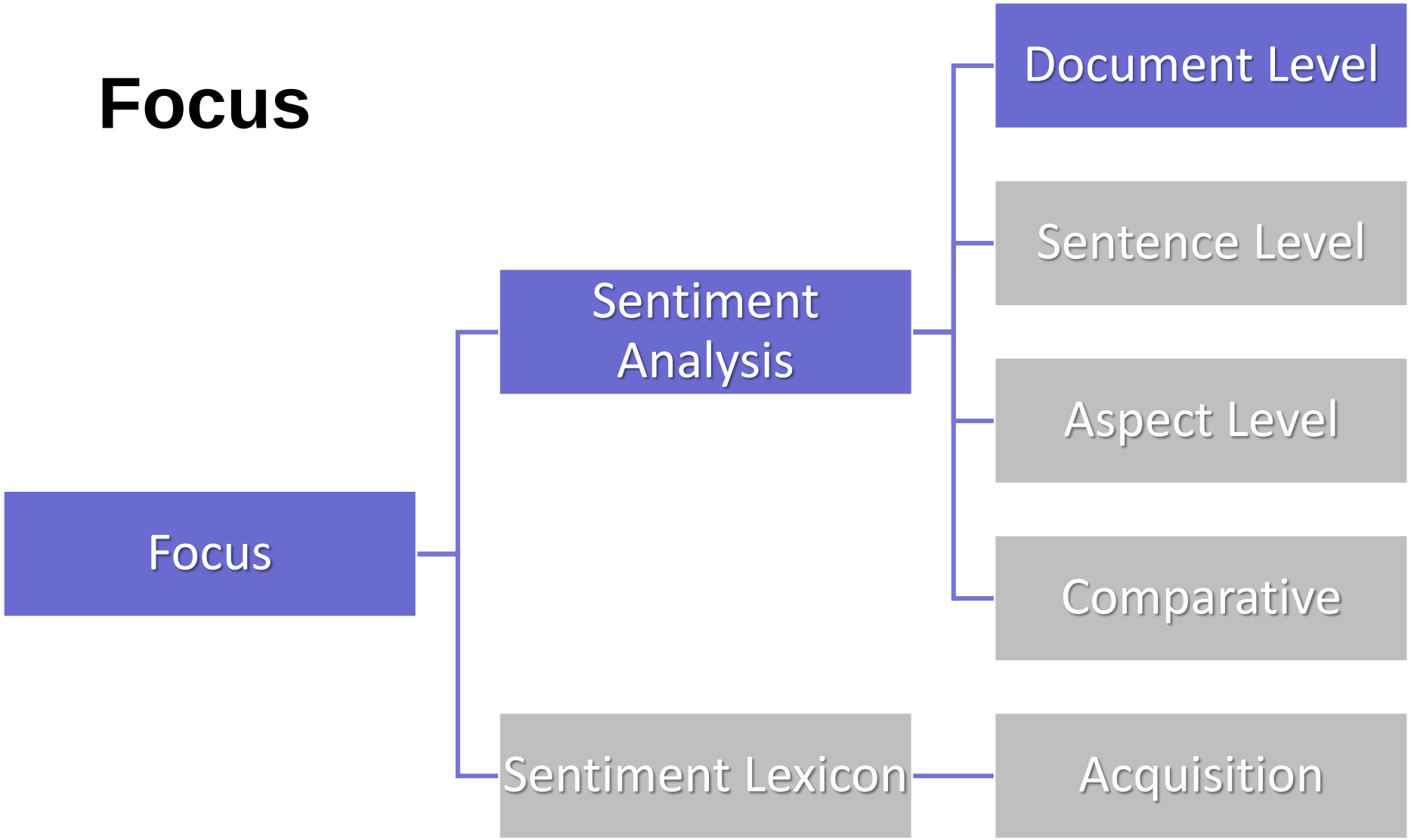
Objective

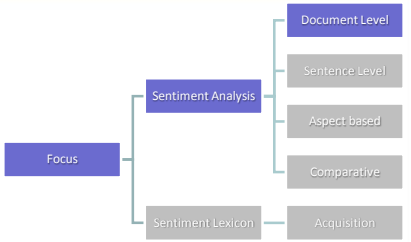
- Present the main research problems related to Sentiment Analysis (SA) and some of the techniques used to solve them
- Review some of the major application areas where sentiment analysis is being used today

Focus


Focus
Sentiment Analysis
Document Level
Sentence Level
Aspect Level
Comparative
Sentiment Lexicon
Acquisition

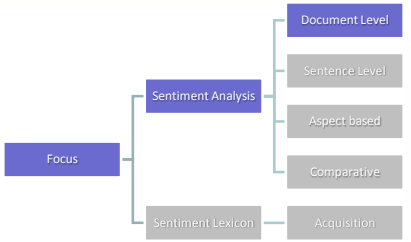
Focus


Focus
Sentiment Analysis
Document Level
Sentence Level
Aspect Level
Comparative
Sentiment Lexicon
Acquisition



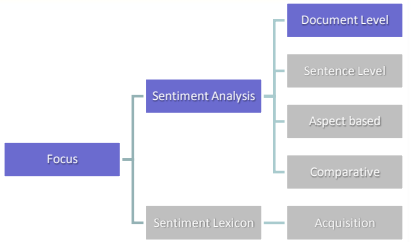
SA – Document Level

- Assumed that the document contains an opinion on one main object expressed by the author of the document
- Two main approaches
  - Supervised learning
  - Unsupervised learning

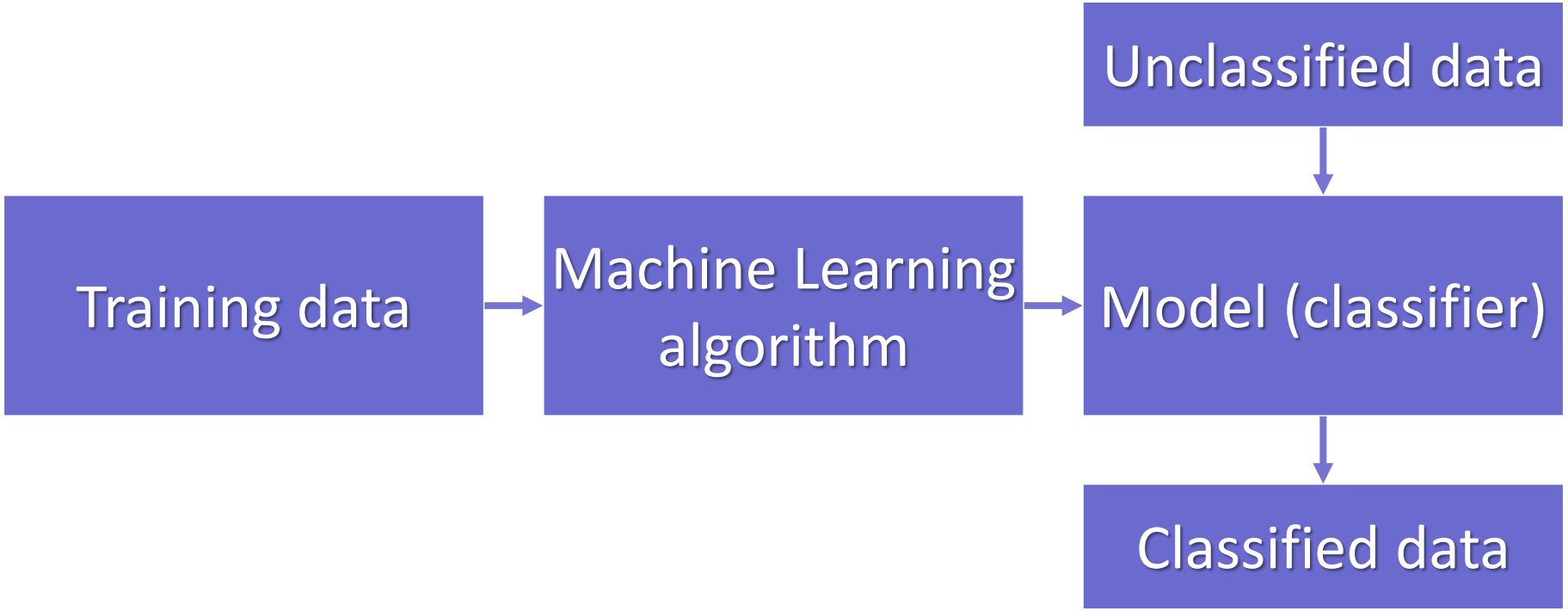


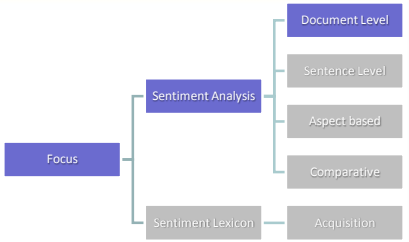
SA – Document Level

- Supervised learning
  - Finite set of classes
  - Training data is available
  - Classes
    - Positive/negative/neutral
    - Numeric scale (stars)
  - SVM, KNN, Naïve Bayes, Logistic Regression, Genetic programming



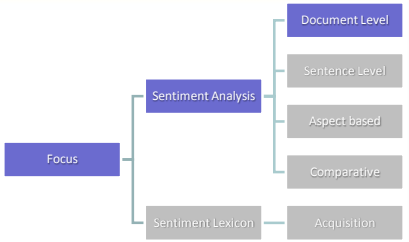
SA – Document Level





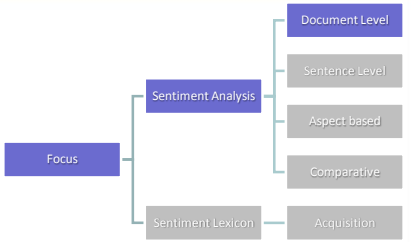
SA – Document Level

- Important representations
  - Bag of words
  - TFIDF
  - Part of Speech (PoS)
  - Sentiment Lexicons



SA – Document Level

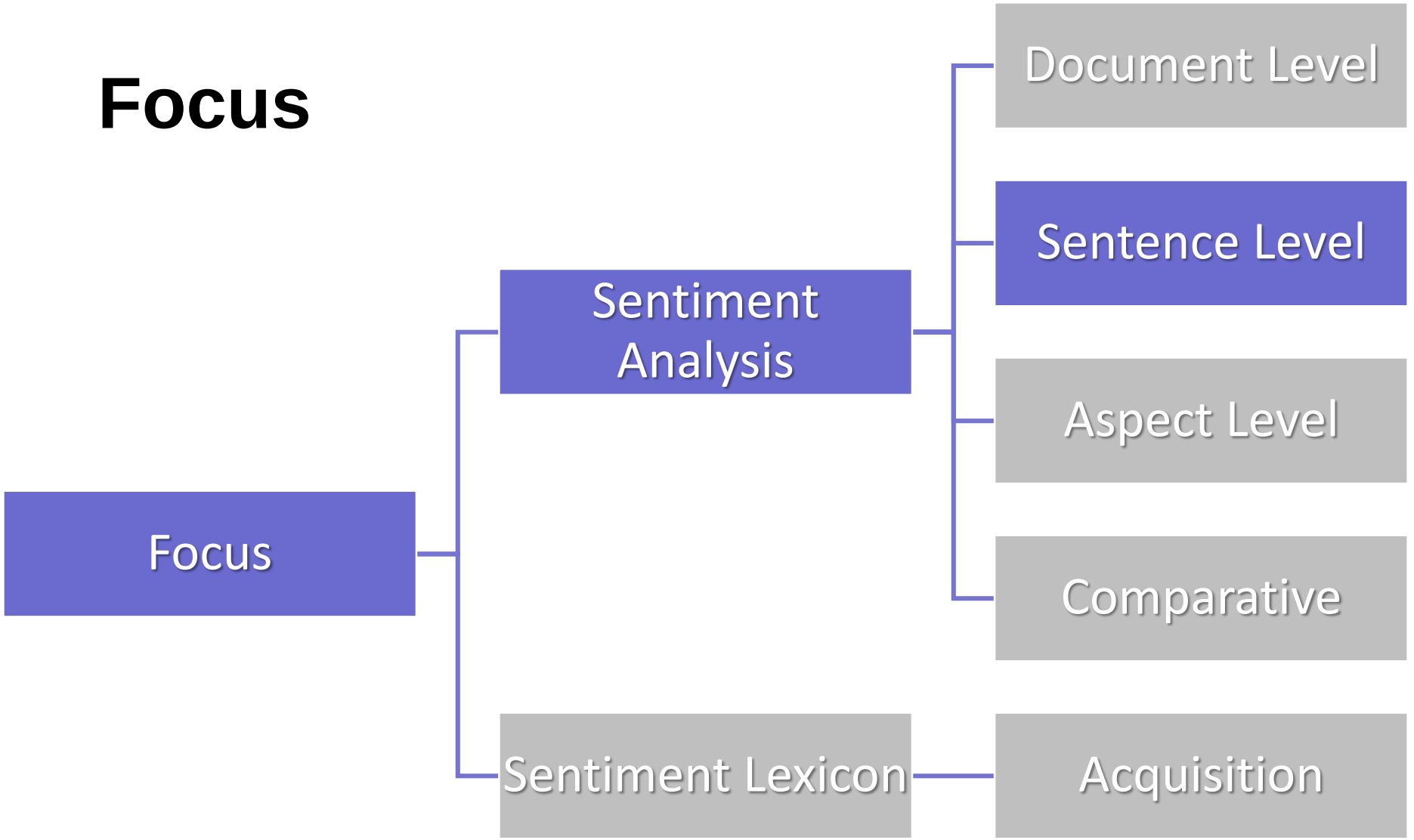
- Unsupervised learning
  - Semantic orientation (SO) specific phrases
  - PMI (Pointwise Mutual Information) of the phrase with two sentiment words
  - PMI(P, W)
    - Statistical dependence between phrase P and word W based on their co-occurrence in a corpus

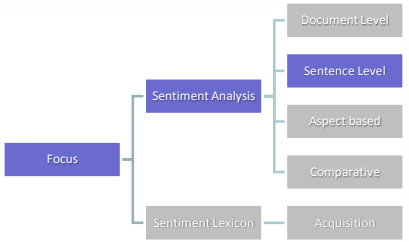


SA – Document Level

- Unsupervised learning
  - The two words used in Turney (2002) are ‘excellent’ and ‘poor’
  - The SO measures whether P is closer in meaning to the positive word (‘excellent’) or the negative word (‘poor’)

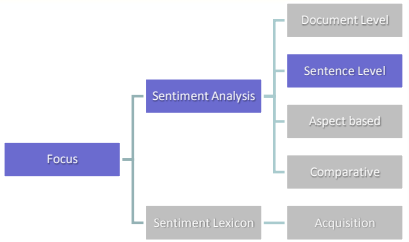
Focus


Focus
Sentiment Analysis
Document Level
Sentence Level
Aspect Level
Comparative
Sentiment Lexicon
Acquisition



SA – Sentence Level

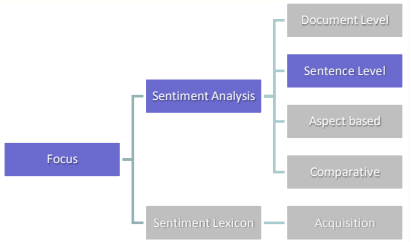
- Assume
  - We know the identity of the entity discussed in the sentence
  - There is a single opinion in each sentence



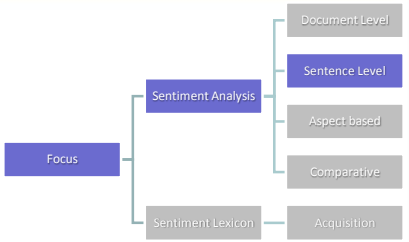
SA – Sentence Level

- Before analyzing the polarity of the sentences we must determine if the sentences are subjective or objective
  - Only subjective sentences will be analyzed

SA – Sentence Level



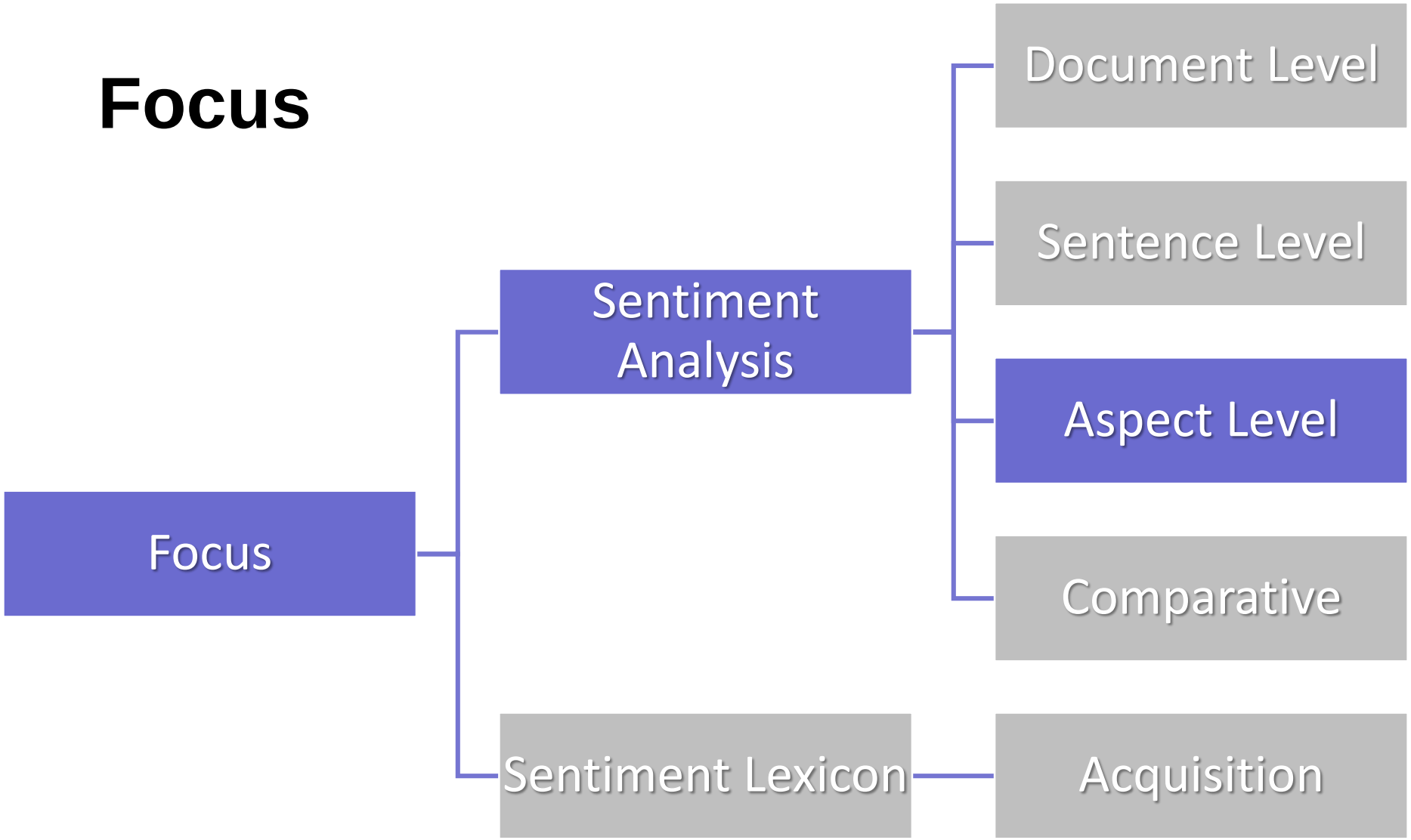
- Most approaches are based on supervised learning
- Unsupervised approach is similar of Turney (2002)

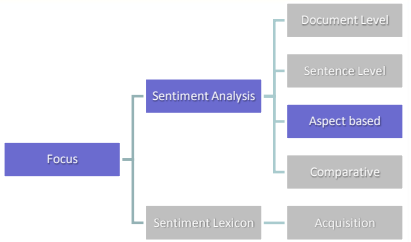


SA – Sentence Level

- Handle different types of sentences by different strategies
- Sentences that need unique strategies
  - Conditional
  - Question
  - Sarcastic

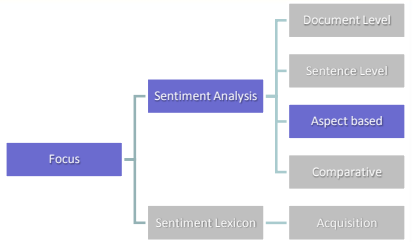
Focus


Focus
Sentiment Analysis
Sentiment Lexicon
Document Level
Sentence Level
Aspect Level
Comparative
Acquisition



SA – Aspect Level

- In many cases, people talk about entities that have many aspects (attributes) and they have a **different opinion about each of the aspects**
- Often happens in reviews about products or in discussion forums



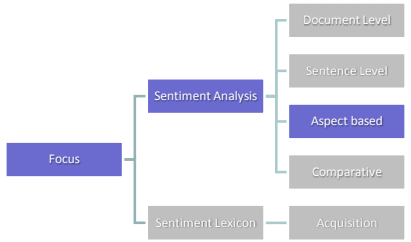
SA – Aspect Level

★★★☆☆ **I want to love it, I really do. But I can't.**

By waetherman on November 15, 2011

Verified Purchase

"As a long-time Kindle fan I was eager to get my hands on a Fire. There are some great aspects; the device is quick and for the most part dead-simple to use. The screen is fantastic with good brightness and excellent color, and a very wide viewing angle. But there are some downsides too; the small bezel size makes holding it without inadvertent page-turns difficult, the lack of buttons makes controls harder, the accessible storage memory is limited to just 5GB."



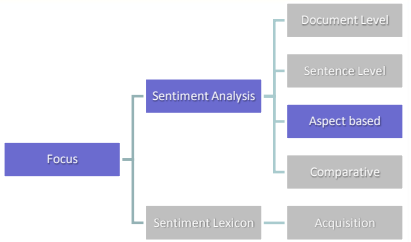
SA – Aspect Level

★★★☆☆ **I want to love it, I really do. But I can't.**

By waetherman on November 15, 2011

Verified Purchase

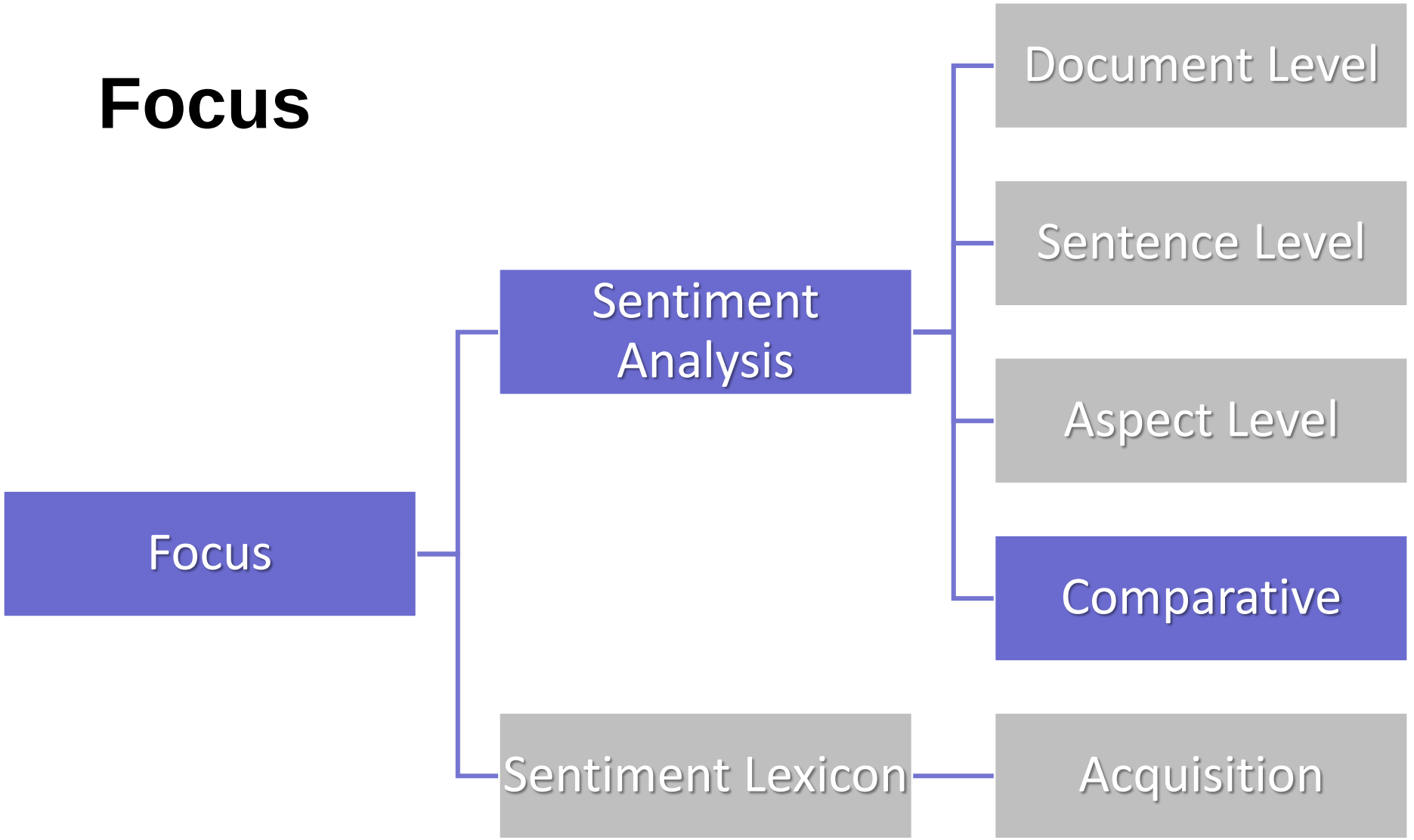
"As a long-time Kindle fan I was eager to get my hands on a Fire. There are some great aspects; the ***device*** *is* ***quick*** *and for the most part* ***dead-simple to use****. The* ***screen*** *is* ***fantastic*** *with* ***good brightness*** *and* ***excellent color****, and a very* ***wide viewing angle****. But there are some downsides too; the* ***small bezel*** *size makes* ***holding*** *it without inadvertent page-turns* ***difficult****, the lack of* ***buttons*** *makes* ***controls harder****, the accessible* ***storage memory*** *is* ***limited*** *to just 5GB."*

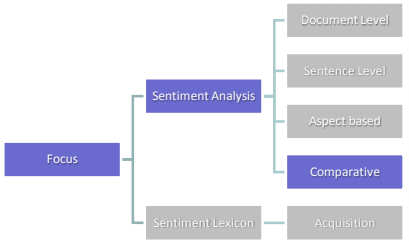


SA – Aspect Level

- Can view the problem of aspect identification as an information extraction problem
  - Use a tagged corpus to train a sequence classifier such as a Conditional Random Field (CRF)
- Implicit aspects

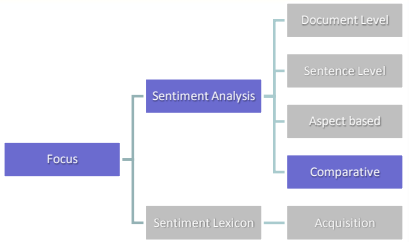
Focus


Focus
Sentiment Analysis
Document Level
Sentence Level
Aspect Level
Comparative
Sentiment Lexicon
Acquisition



SA – Comparative

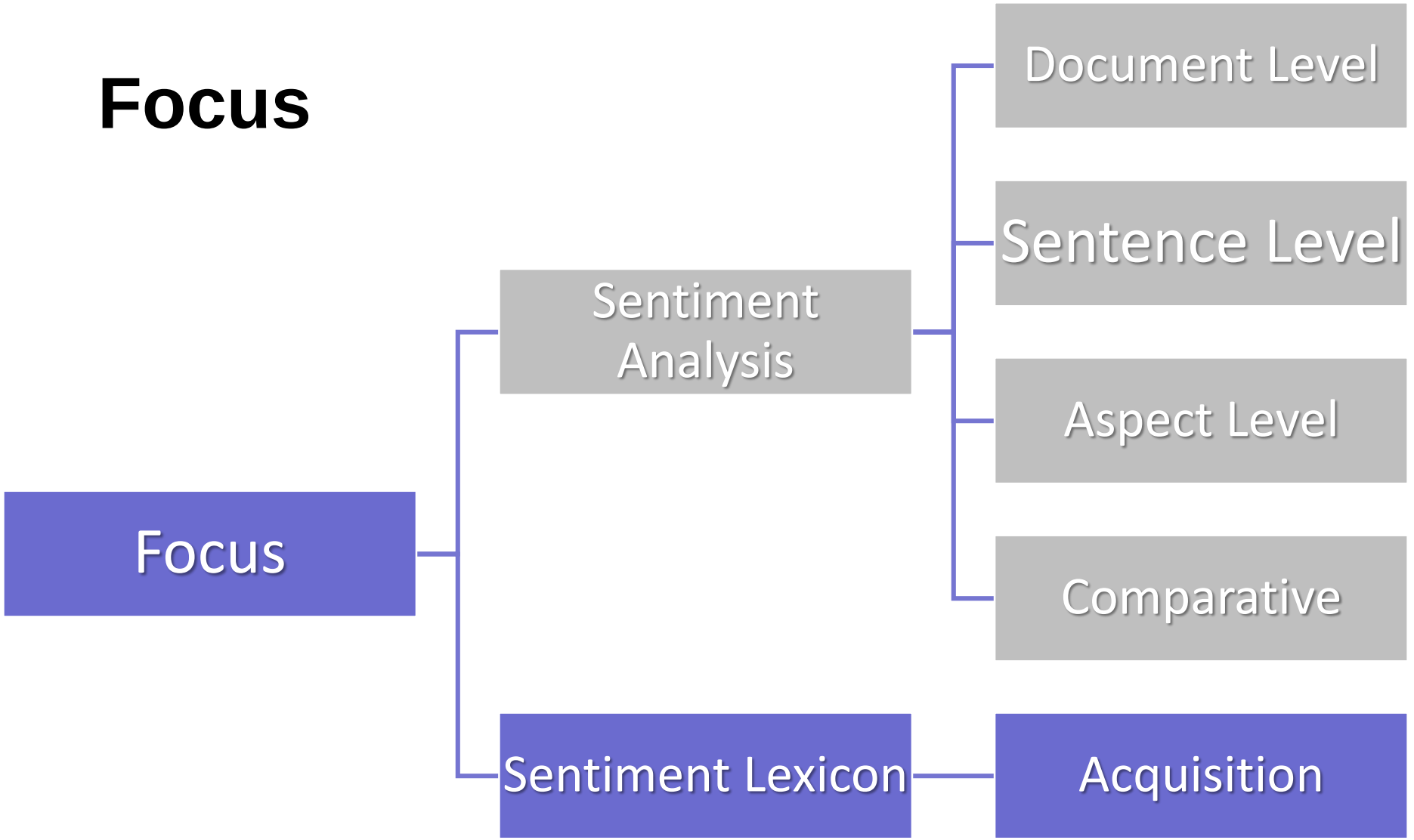
- In many cases users do not provide a direct opinion about one product but instead provide comparable opinions
- The goal of SA in this case is to identify the sentences that contain comparative opinions, and to extract the preferred entity(-ies) in each opinion.

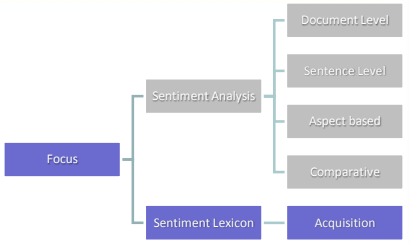


SA – Comparative

- Comparative adjectives adverbs
  - ‘more,’ ‘less,’ and words ending with –er (for example, ‘lighter’)
- Superlative adjectives and adverbs
  - ‘most,’ ‘least,’ and words ending with –est (for example, ‘finest’)
- Additional phrases
  - ‘favor,’ ‘exceed,’ ‘outperform,’ ‘prefer,’ ‘than,’ ‘superior,’ ‘inferior,’ ‘number one’

Focus

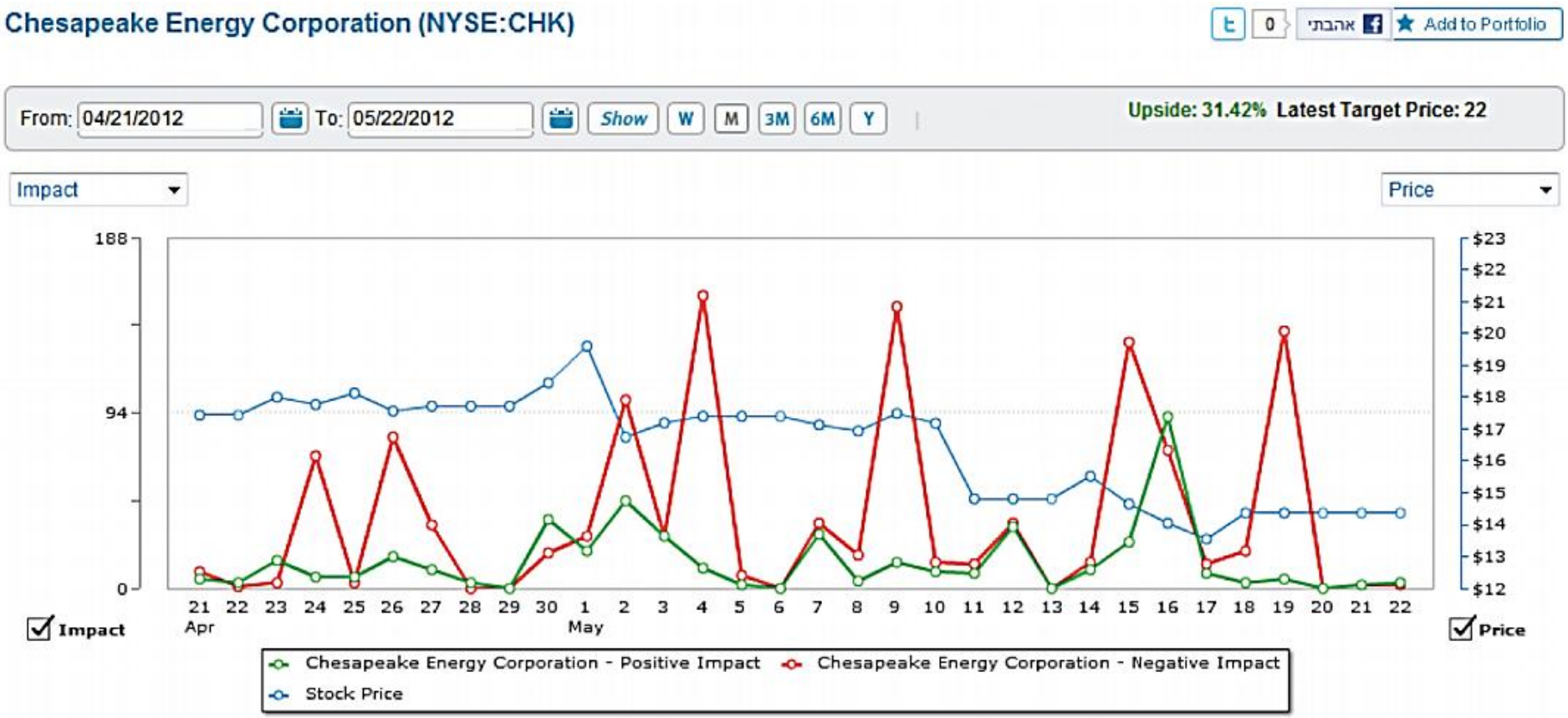

Focus
Sentiment Analysis
Document Level
Sentence Level
Aspect Level
Comparative
Sentiment Lexicon
Acquisition



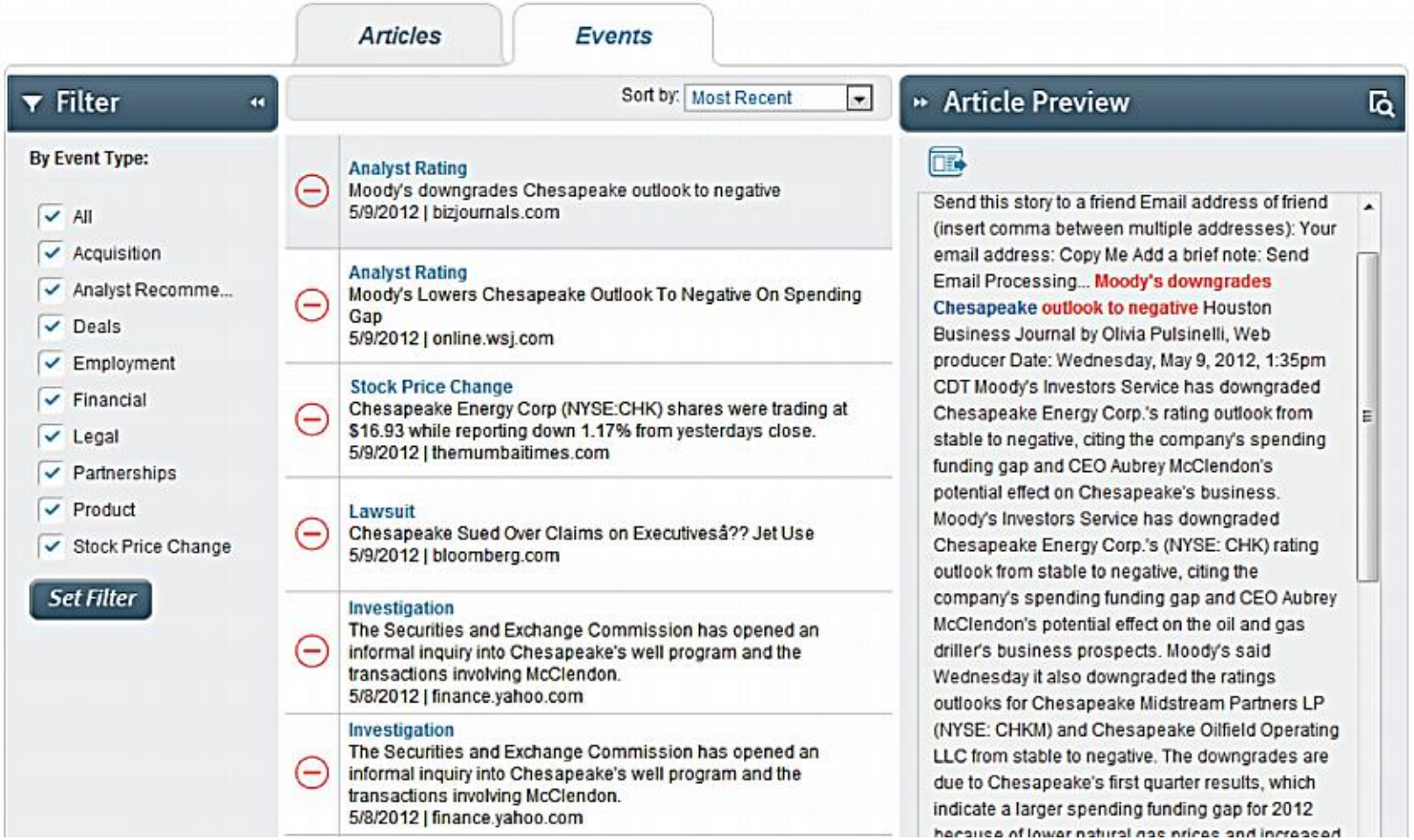
Sentiment Lexicon

- Most crucial resource for most sentiment analysis algorithms
- Acquisition
  - Manual
  - Dictionary
    - Sentiwordnet
  - Corpus
    - Sentiment consistency

Applications



Applications



Research Issues

- Automatic entity resolution
- Sarcasm
- Noisy texts
- Sentiment to objective statements

Conclusion

- Reviewed some of the main research problems within the field of SA
- Discussed ways to solve each of these problems
- Described some of the major applications
- Provided a few major open challenges

Techniques and Applications for Sentiment Analysis

Communications of the ACM | April 2013 | Vol. 56 | nº 4
DOI:10.1145/2436256.2436274

Ronen Feldman
Hebrew University
ronen.feldman@huji.ac.il

Airton Bordin Junior
airtonbjunior@gmail.com
Prof. Dr. Nádia Félix Felipe da Silva
nadia@inf.ufg.br

Introduction

“The king suite was spacious, clean, and well appointed. The reception staff, bellmen, and housekeeping were very helpful. Requests for extras from the maid were always provided. The heating and air conditioning functioned well; this was good as the weather was variable. The sofa bed was the best I’ve ever experienced. The king size bed was very comfortable. The building and rooms are very well soundproofed. The neighborhood is the best for shopping, restaurants, and access to subway. Only “complaint” has to do with high-speed Internet access. It’s only available on floors 8–12.”

Introduction