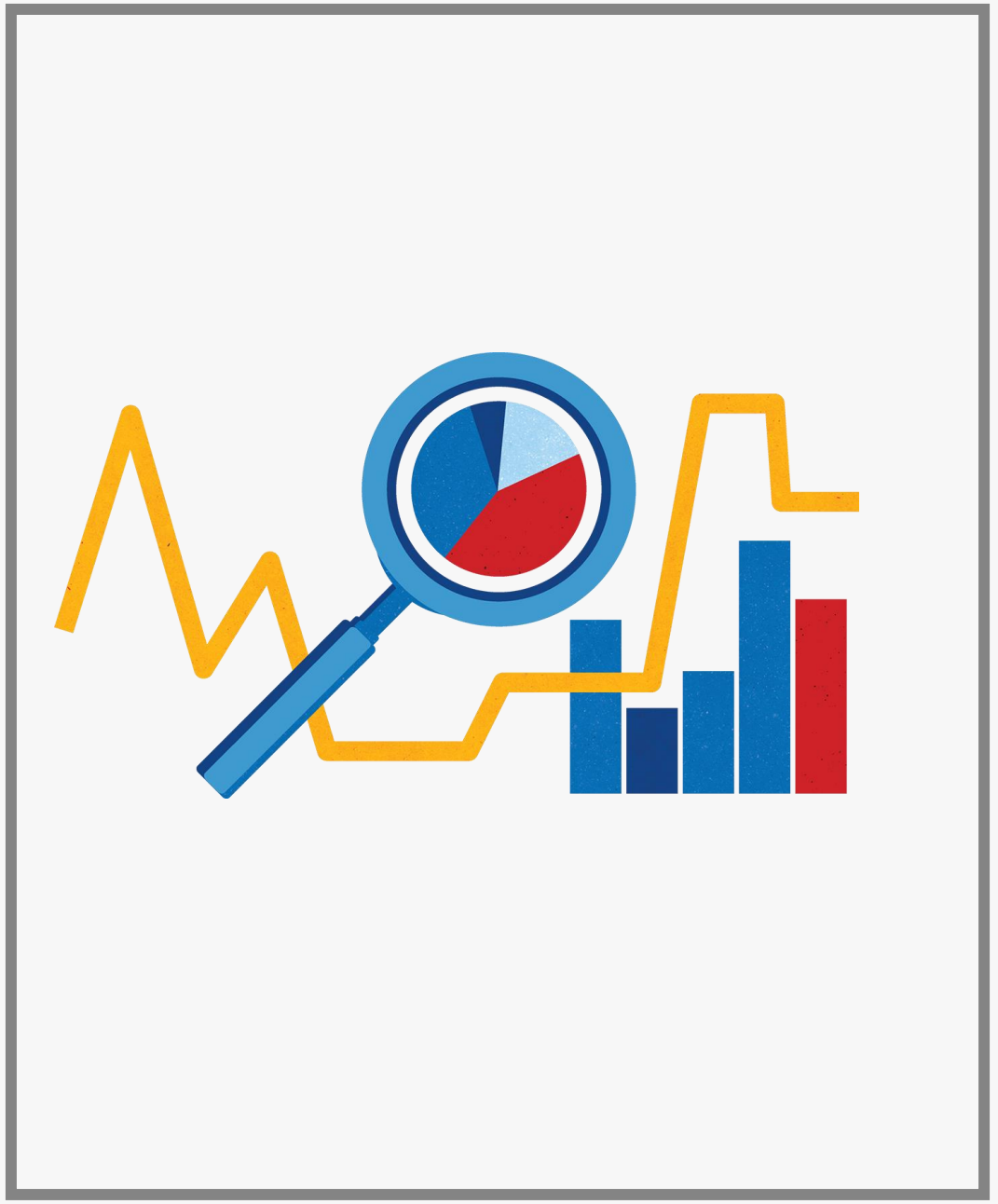
By
Dr. Hussein Hazimeh

Lebanese University
Faculty of Information 1

Data Science Departement

2nd year – Data Analysis

March – 2020 – Chapter 5



What is Sentiment?

- Sentiment = feelings
  - Attitudes
  - Emotions
  - Opinions
- Subjective impressions, not facts

What is Sentiment?

- Generally, a binary opposition in opinions is assumed
- For/against, like/dislike, good/bad, etc. Some
- sentiment analysis jargon:
  - “Semantic orientation”
  - “Polarity”

What is Sentiment Analysis?

- Using NLP, statistics, or machine learning methods to extract, identify, or otherwise characterize the sentiment content of a text unit
- Sometimes refered to as *opinion mining,* although the emphasis in this case is on extraction

Questions SA might ask

- Is this product review positive or negative?
- Is this customer email satisfied or dis-satisfied?
- Based on a sample of tweets, how are people responding to this ad campaign/product release/news item?
- How have bloggers' attitudes about the president changed since the election?

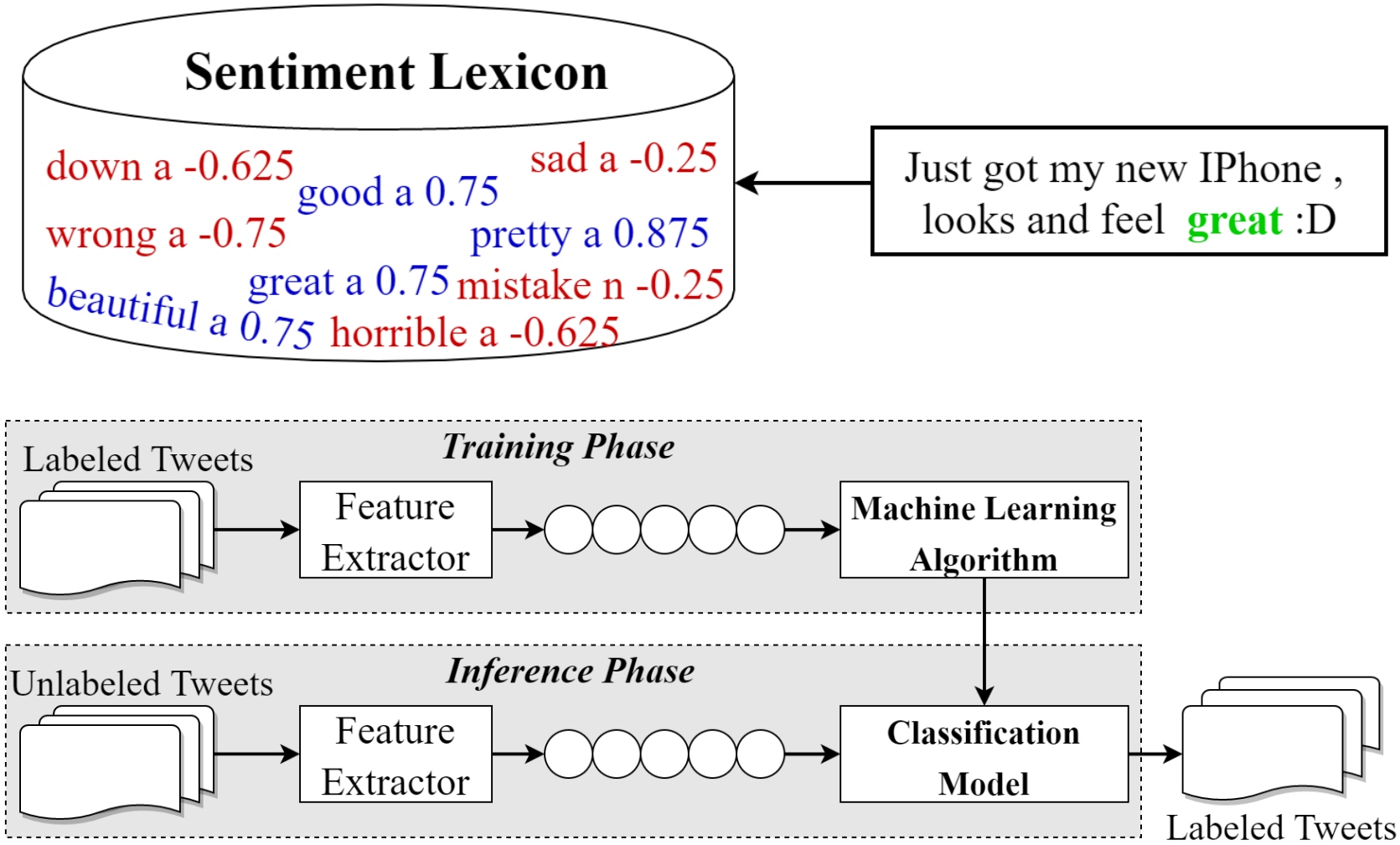
BEST FOR You
ORGANICS COMPANY

Sentiment analysis Levels

- Document Level
  - Large amount of data at once
  - Extract sentiment with/without entities
  - Extract lexical features
- Sentence Level
  - Subjective sentence
  - Objective sentence
- Aspect Level
  - Phrases are featured product oriented
  - The idea is that pos/neg opinion consists of a sentiment and an aim

Sentiment Classification

- Semantic Orientation Lexicon
  - Corpus-based technique
  - Dictionary-based technique
- Lexicon base
  - Ranked bag of words
- Machine learning base
  - Supervised
    - Need training set
    - classification
  - Unsupervised
    - Clustering



Other related tasks

- Information extraction (discarding subjective information)
- Question answering (recognizing opinion- oriented questions)
- Summarization (accounting for multiple viewpoints)

Other related tasks

- “Flame” detection
- Identifying child-suitability of videos based on comments
- Bias identification in news sources
- Identifying (in)appropriate content for ad placement

Applications in Business Intelligence

- Question: “Why aren't consumers buying our laptop?”
- We know the concrete data: price, specs, competition, etc.
- We want to know subjective data: “the design is tacky,” “customer service was condescending”
- Misperceptions are also important, e.g. “updated drivers aren't available” (even though they are)

BEST FOR *You*
ORGANICS COMPANY

Applications in Business Intelligence

- It is very difficult to survey customers who *didn't* buy the company's laptop
- Instead, you could use SA to

A) search the web for opinions and reviews of this and competing laptops. Blogs, Epinions, amazon, tweets, etc.

B) create condensed versions or a digest of consensus points

BEST FOR *You*
ORGANICS COMPANY

Cross domain applications

- Insights and applications from SA have been useful in other areas
  - Politics/political science
  - Law/policy making
  - Sociology
  - Psychology

Political SA

- Numerous applications and possibilities
- Analyzing trends, identifying ideological bias, targeting advertising/messages, gauging reactions, etc.
- Evaluation of public/voters' opinions
- Views/discussions of policy
- More on this in lecture 3

SA and Sociology

- Idea propagation through groups is an important concept in sociology (cf. Rogers 1962, *Diffusion of Innovations*)
- Opinions and reactions to ideas are relevant to adoption of new ideas
- Analyzing sentiment reactions on blogs can give insight to this process
- E.g. Kale et al (2007), *Modeling trust and influence in the blogosphere using link polarity*

SA and Psychology

- » Potential to augment psychological investigations/experiments with data extracted from NL text
- » Dream sentiment analysis (Nadeau et al., 2006)

In general,

- Humans are subjective creatures and opinions are important. Being able to interact with people on that level has many advantages for information systems.

How SA is different

- Comparatively few categories (positive/negative, 3 stars, etc) compared to text categorization
- Crosses domains, topics, and users
- Categories not independent (opposing or regression-like)
- Characteristics of answers to opinion- based questions are different from fact- based questions, so opinion-based IE differs from trad IE

Challenges in SA

- People express opinions in complex ways
- In opinion texts, lexical content alone can be misleading
- Intra-textual and sub-sentential reversals, negation, topic change common
- Rhetorical devices/modes such as sarcasm, irony, implication, etc.

A letter to a hardware store*

“Dear <hardware store>

Yesterday I had occasion to visit <your competitor>. The had an excellent selection, friendly and helpful salespeople, and the lowest prices in town.

You guys suck.

Sincerely,”

*an apocryphal example

BEST FOR YOU
ORGANICS COMPANY

What to classify

- There are many possibilities for what we might want to classify:
  - Users
  - Texts
  - Sentences (paragraphs, chunks of text?)
  - Words
  - Tweets/updates

Classifying words/short phrases

- The building blocks of sentiment expression
- Short phrases may be just as important (or moreso) as words:
  - “lowest prices”
  - “high quality”
- We need an approach to deal with these before moving on to other classification tasks

Polarity keywords

- There seems to be *some* relation between positive words and postive reviews
- Can we come up with a set of keywords by hand to identify polarity?

Smileys

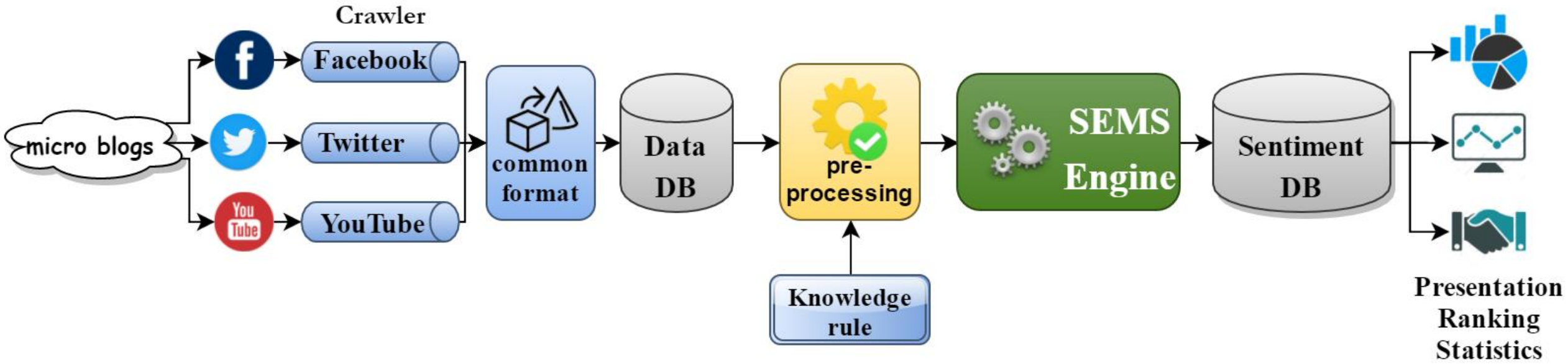
- A common approach for working with tweets and short text updates
- Very little text to work with
- Sentiment most succinctly represented with emoticons/smileys

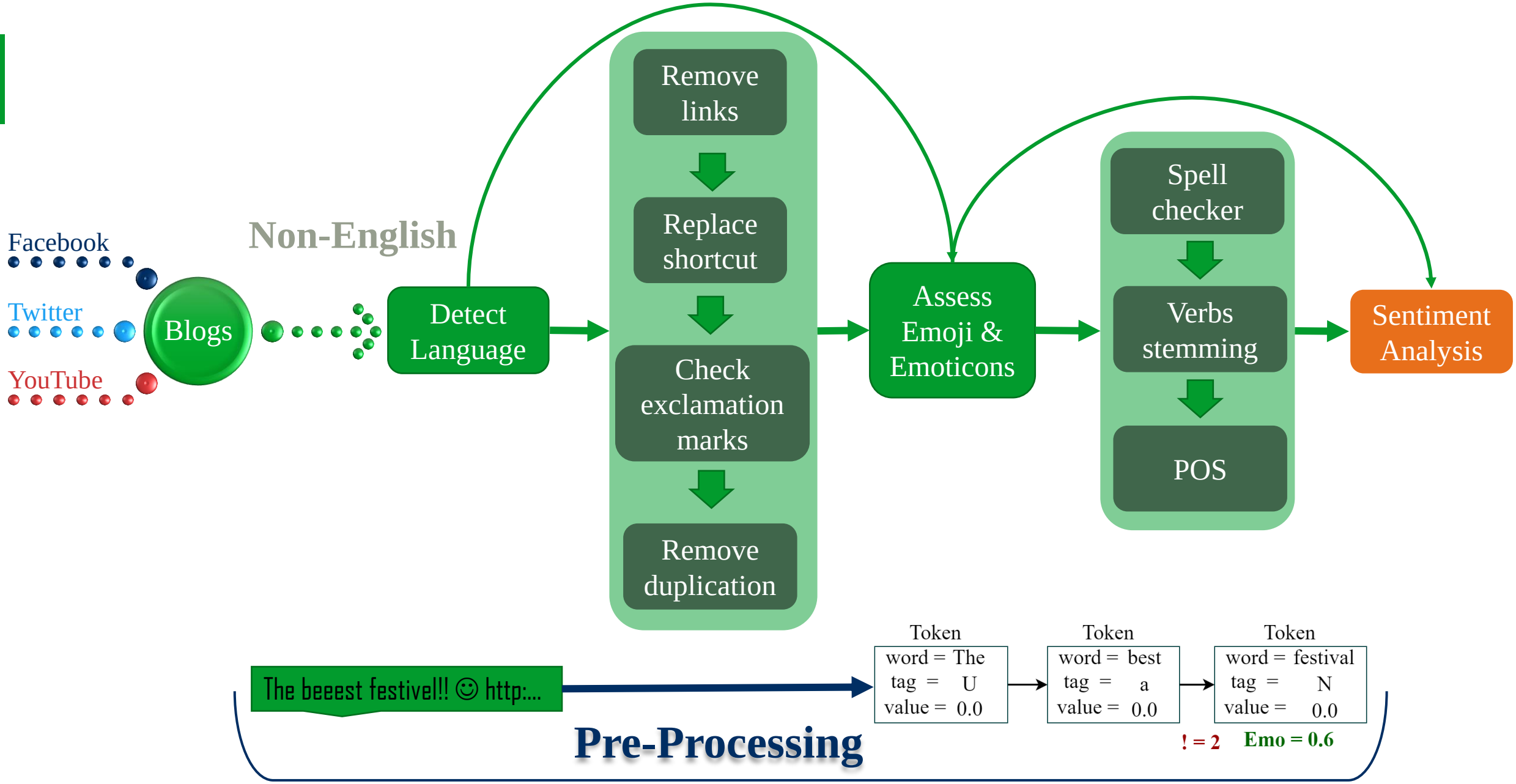
Sentiment analysis on events – case study

BEST FOR You
ORGANICS COMPANY

SEMS Architecture

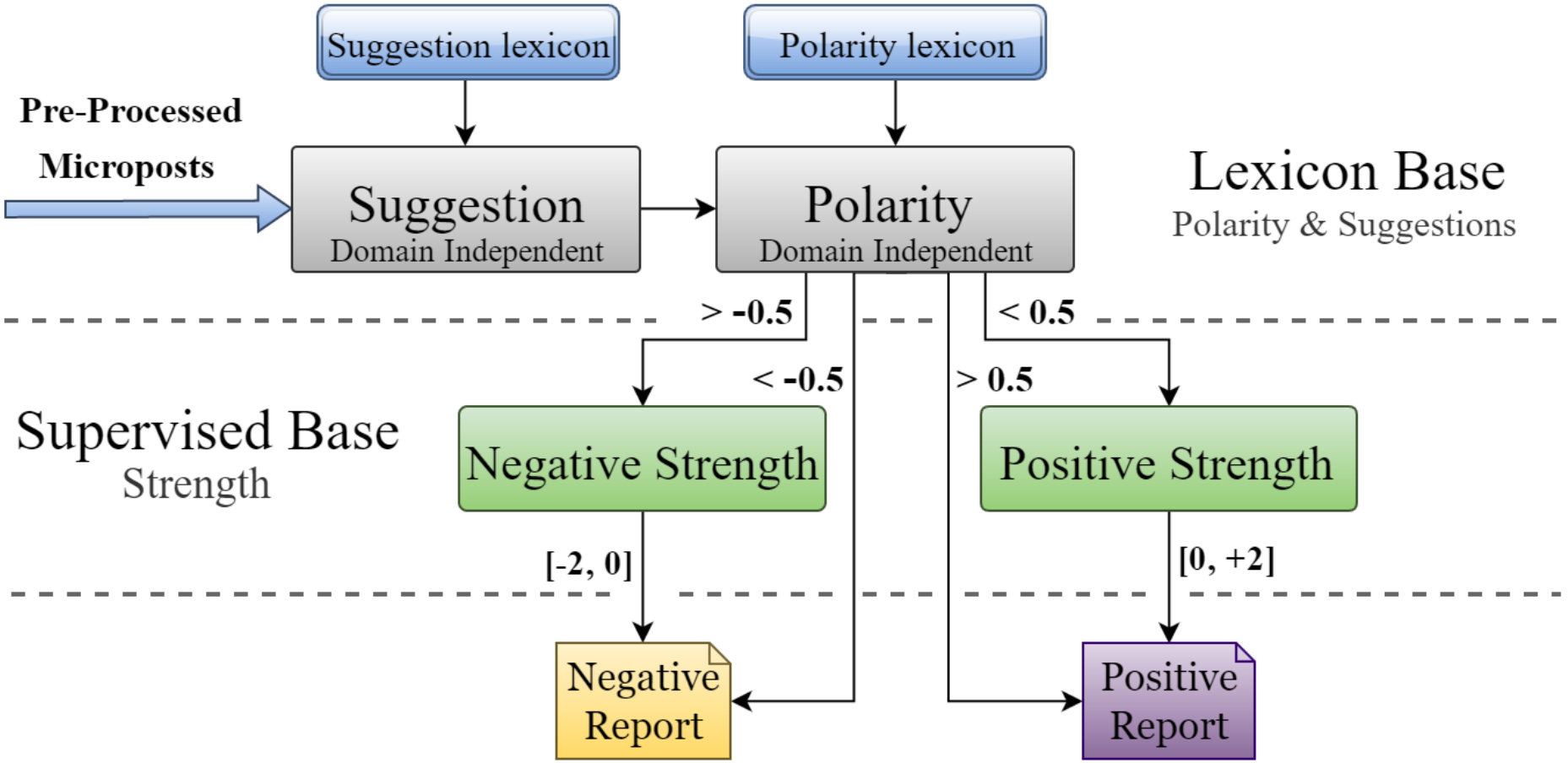
» The whole architecture from the three different sources until reports

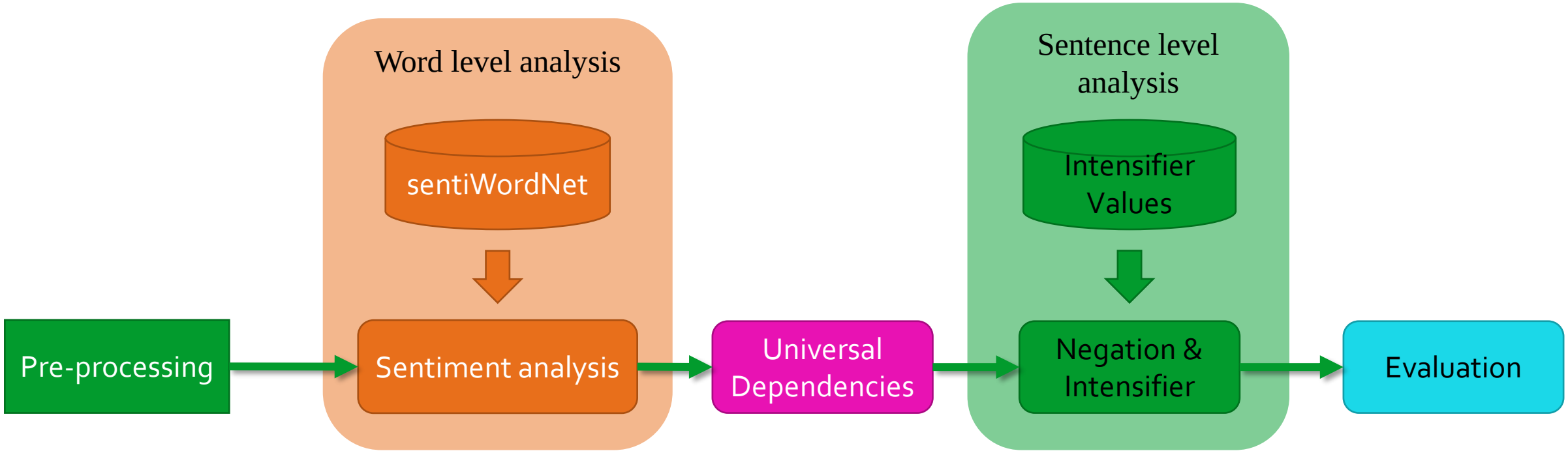




BEST FOR You
ORGANICS COMPANY

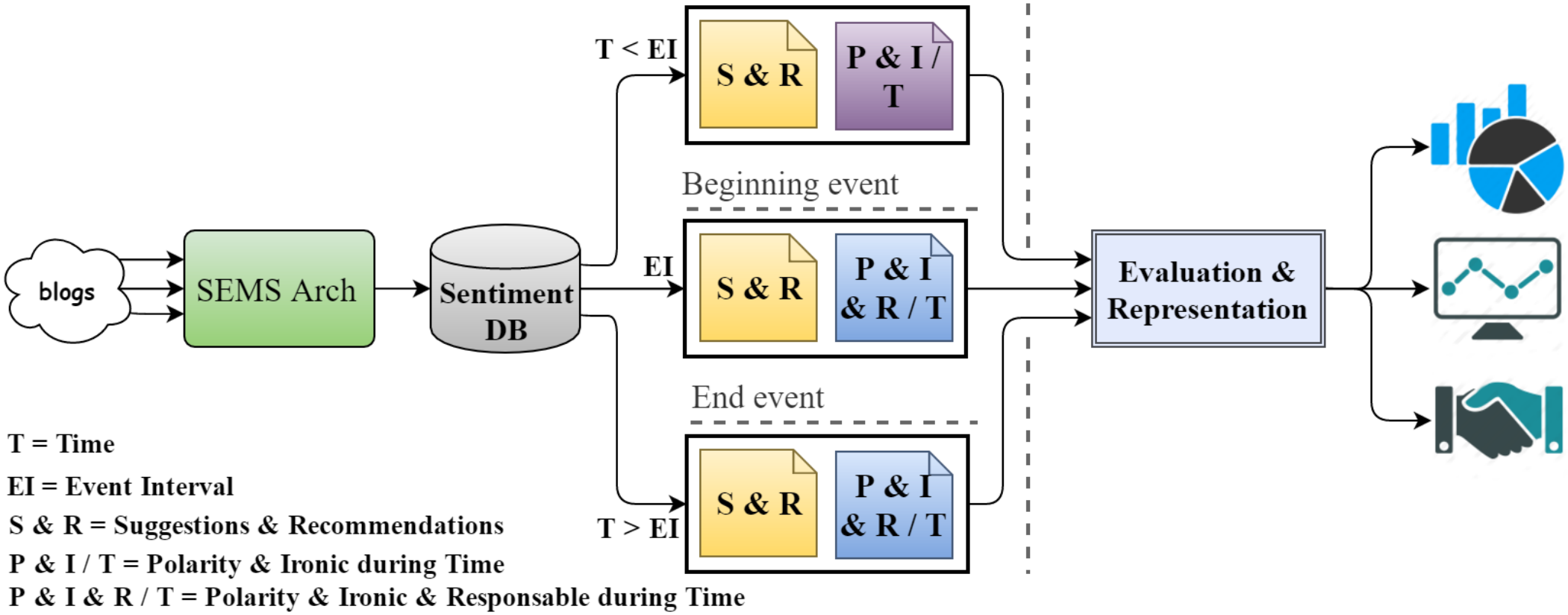
SEMS Engine

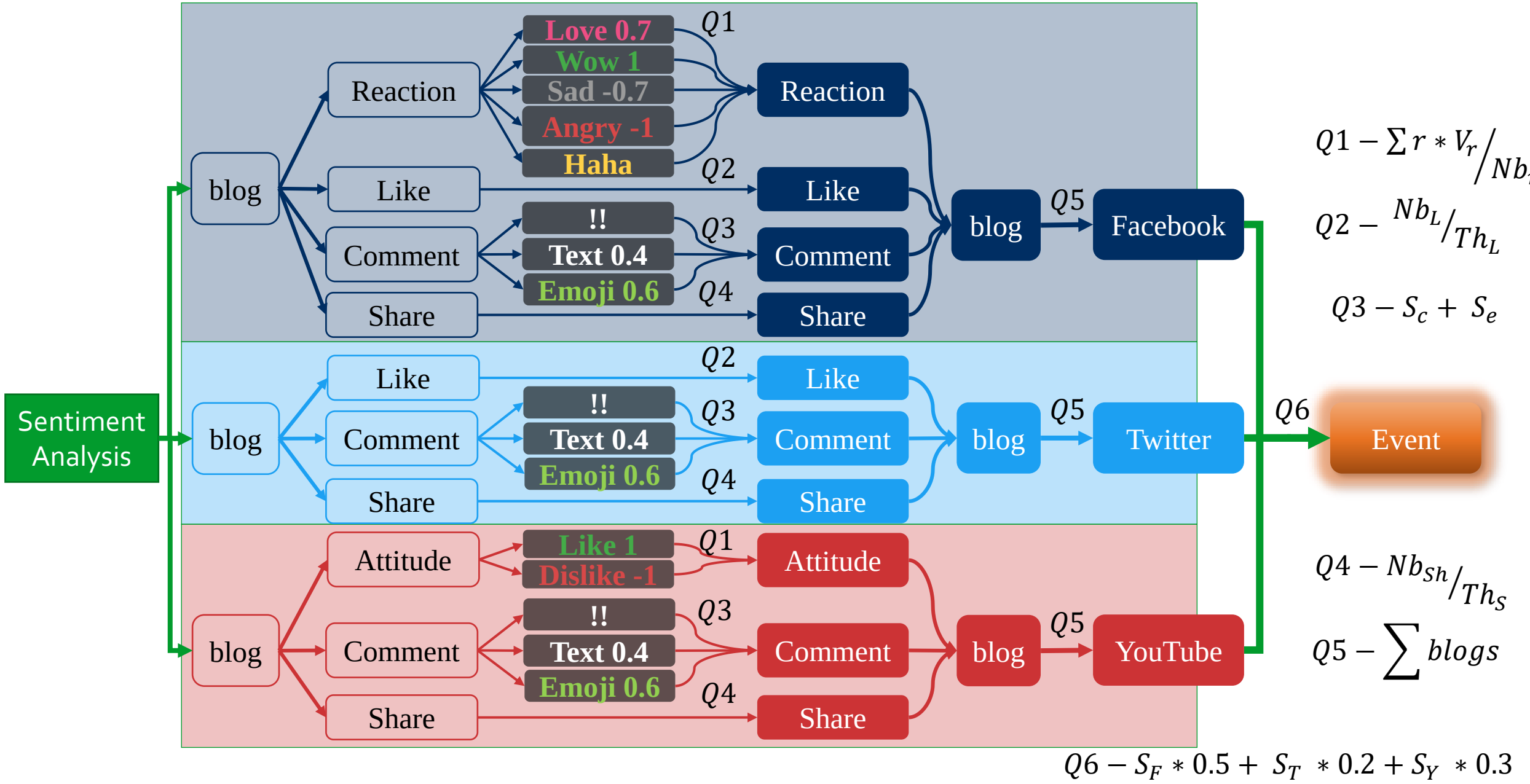




Sentiment Analysis

Validation & Presentation




Sentiment Analysis
blog
Reaction
Like
Comment
Share
Love 0.7
Wow 1
Sad -0.7
Angry -1
Haha
!!
Text 0.4
Emoji 0.6
Q1
Q2
Q3
Q4
Reaction
Like
Comment
Share
blog
Q5
Facebook
blog
Like
Comment
Share
!!
Text 0.4
Emoji 0.6
Q2
Q3
Q4
Like
Comment
Share
blog
Q5
Twitter
Q6
Event
blog
Attitude
Comment
Share
Like 1
Dislike -1
!!
Text 0.4
Emoji 0.6
Q1
Q3
Q4
Attitude
Comment
Share
blog
Q5
YouTube
$Q1 - \sum r * V_r / Nb_r$
$Q2 - Nb_L / Th_L$
$Q3 - S_c + S_e$
$Q4 - Nb_{Sh} / Th_S$
$Q5 - \sum blogs$
$Q6 - S_F * 0.5 + S_T * 0.2 + S_Y * 0.3$