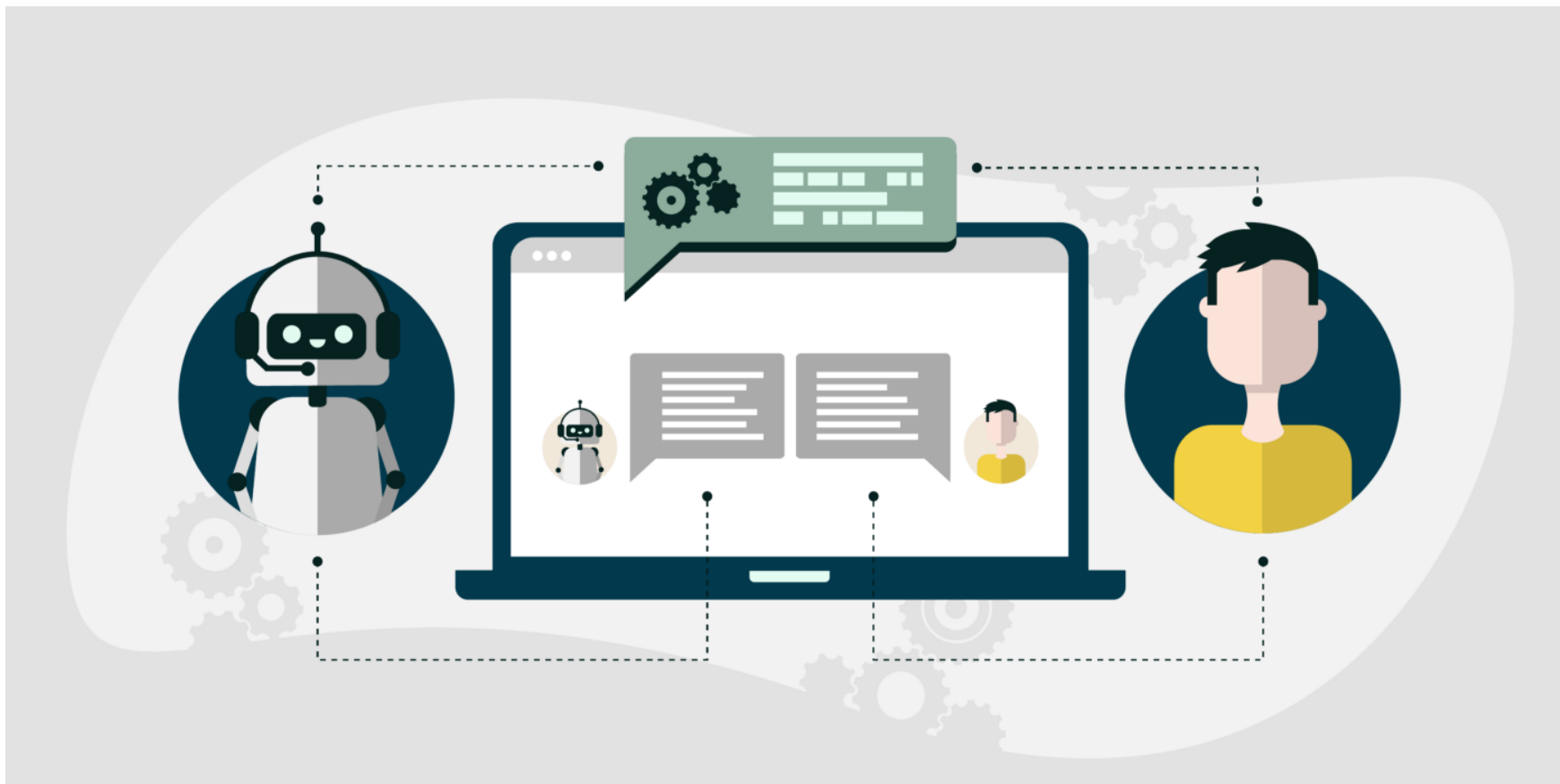Refs : https://www.searchenginejournal.com/facebook-messenger-chatbot-lead-generation/260959/#close

Week III Introduction to Natural Language Processing
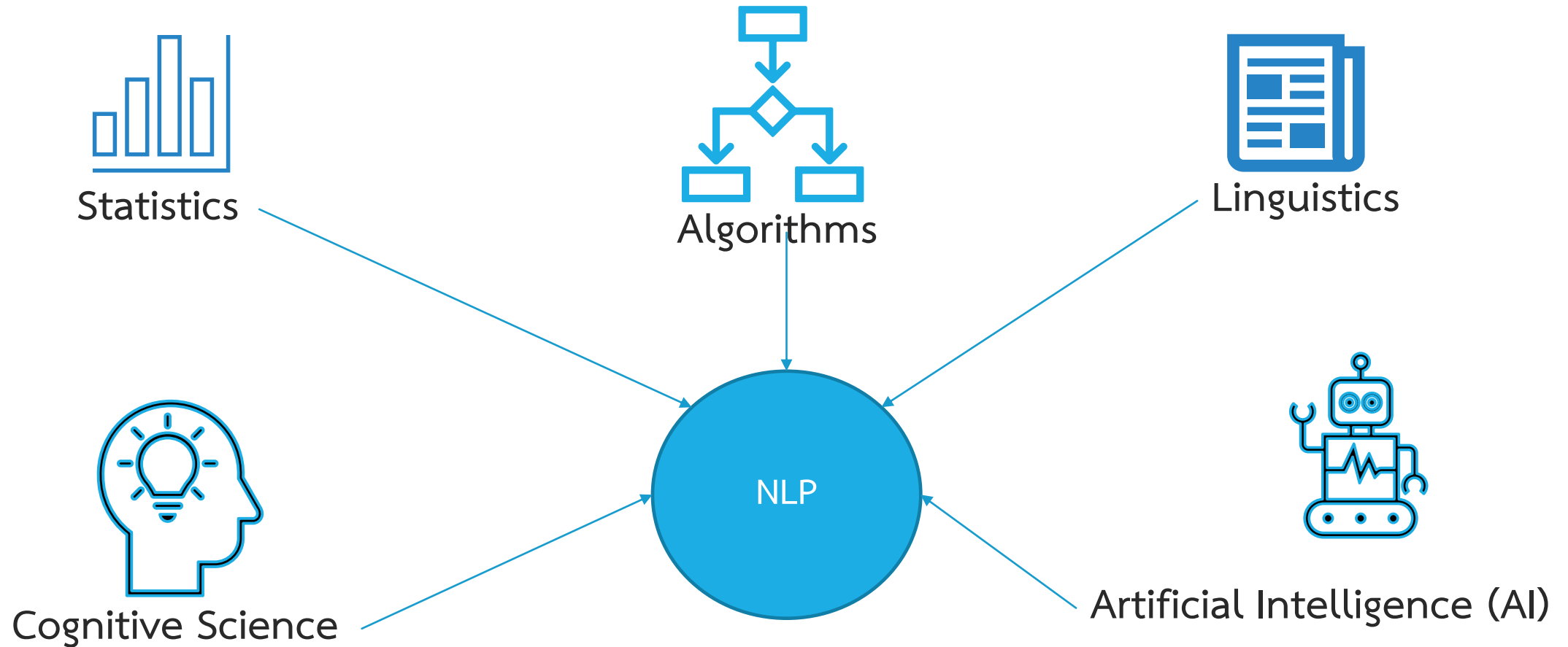
Natural Language Processing (NLP)

Natural Language Processing (NLP) = การประมวลผลภาษาธรรมชาติ

NLP is technique for automating linguistic tasks [3]

Refs : https://medium.com/@natratanonkanraweekultana/natural-language-processing-nlp-40af75482b62

Natural Language Processing (NLP)

Statistics

Algorithms

Linguistics

Cognitive Science

NLP

Artificial Intelligence (AI)

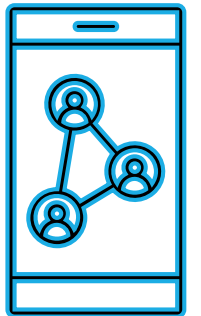Refs : https://github.com/ekapolc/NLP_2021/blob/main/slides/L01_Intro_2020.pdf
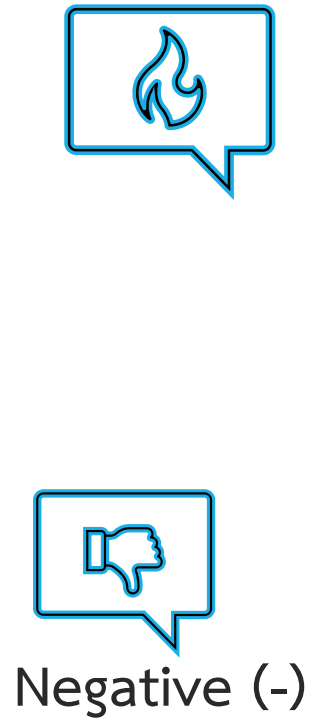
Applications of Natural Language Processing

Text Analytics : discovering insights from text data. [3]

Artificial Intelligence products that help automate language tasks. [3]

Sentiment Analysis

Positive (+)

Negative (-)

Topic Modeling

Word Cloud

Social Monitoring : Social Listening

Refs : https://www.mandalasystem.com/

Benefits of NLP

NLP allow us to understand a large amount of text without reading it all. [3]

Refs : https://avocadorunner.wordpress.com/2014/11/07/natural-language-processing-nlp-and-big-data-by-sagence/
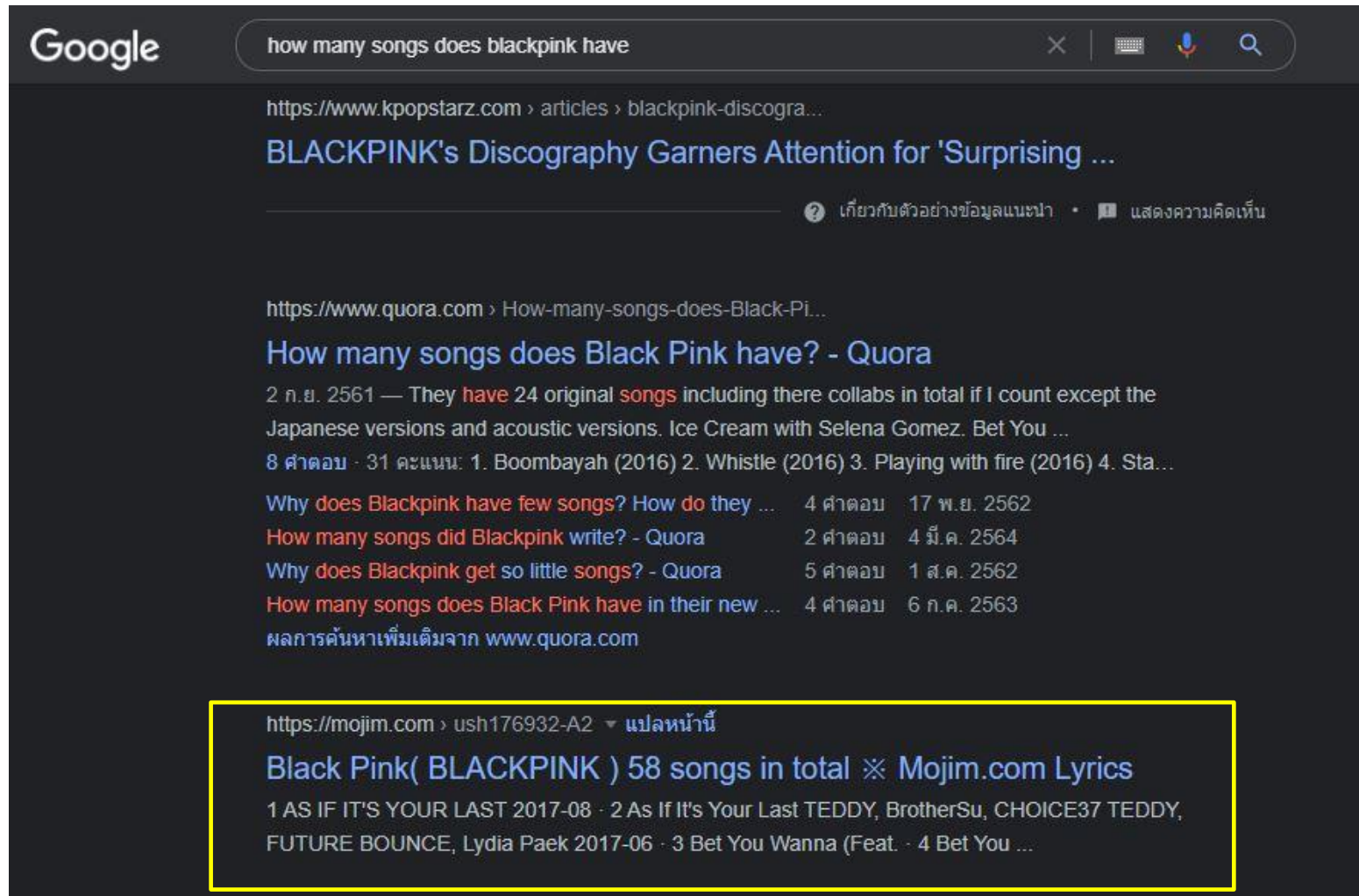
AI application from NLP

Ham spam filter

Search Engine with question answering system

Refs : https://www.google.com/

Instant Writing Assistant

- ENGAGEMENT

Sciantists discovered a substance that could prove that there is life on Venus. The head of the Buendía family is both good and bad. He's good because he cares about his family and that makes him a good man. But he's bad because he makes impulsive decisions that affect his family and that can sometimes be bad for them. It would be better for his family if he could not be as bad.

good → okay

The word **good** is often overused. Consider using a more specific synonym to improve the sharpness of your writing.

(?) Learn more

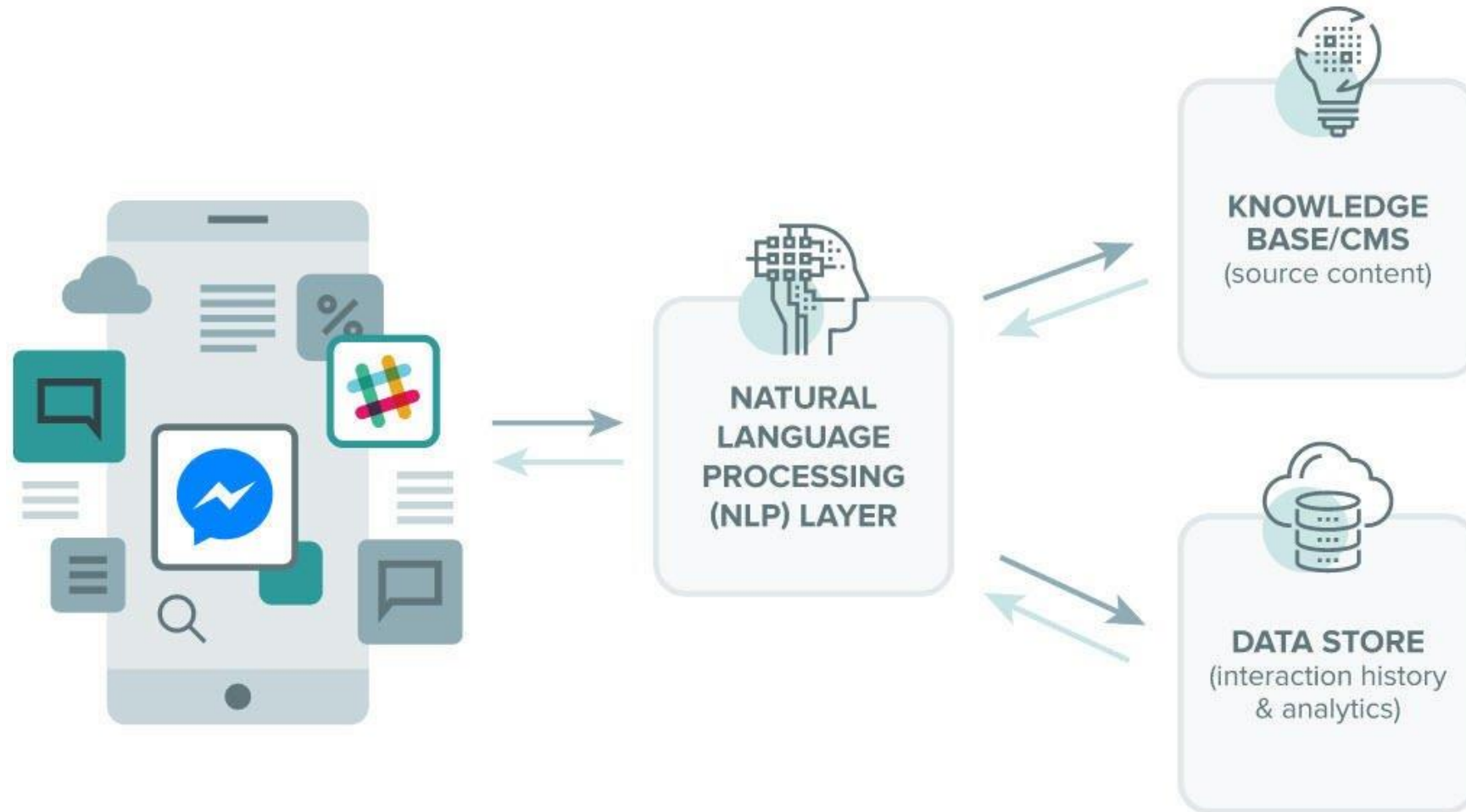Refs : https://www.grammarly.com/blog/student-writing-skills/

Chatbot

Refs : https://www.rainmaker.in.th/introduce-line-bot-designer/

Chatbot architecture

Refs : https://www.wordstream.com/blog/ws/chatbots

More Applications

- Transcriber
- Automated tutor
- Virtual psychologist/therapist
- Line urgency detection
- Book summarization
- Report generator
- Etc. …

Refs : https://chatbotsmagazine.com/https-medium-com-hellohaptik-chatbots-in-banking-examples-best-use-cases-and-the-future-1106432c3a25
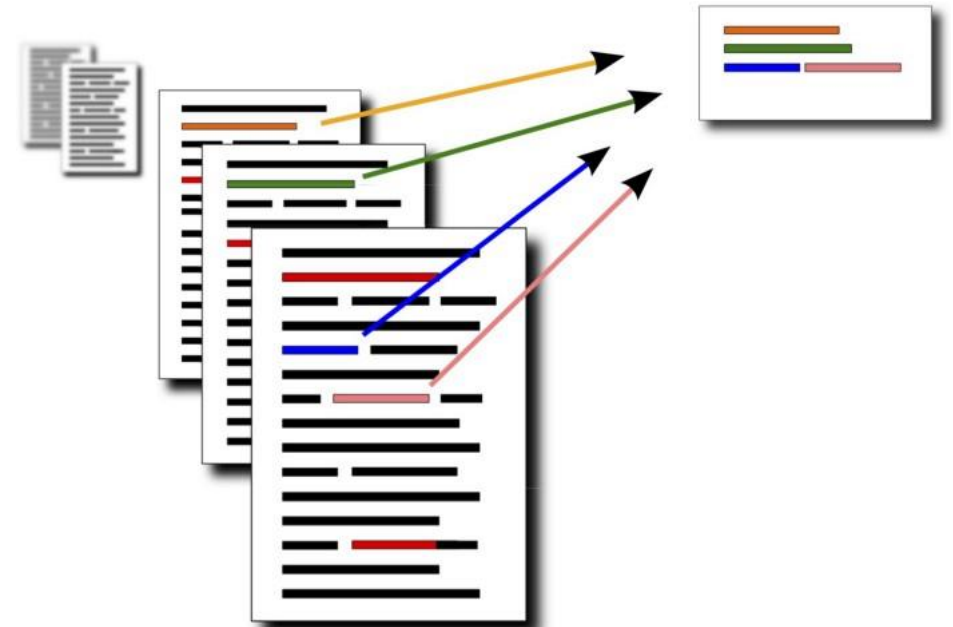
Text processing

Linguistics provides a theoretical framework for understanding the language data. [3]
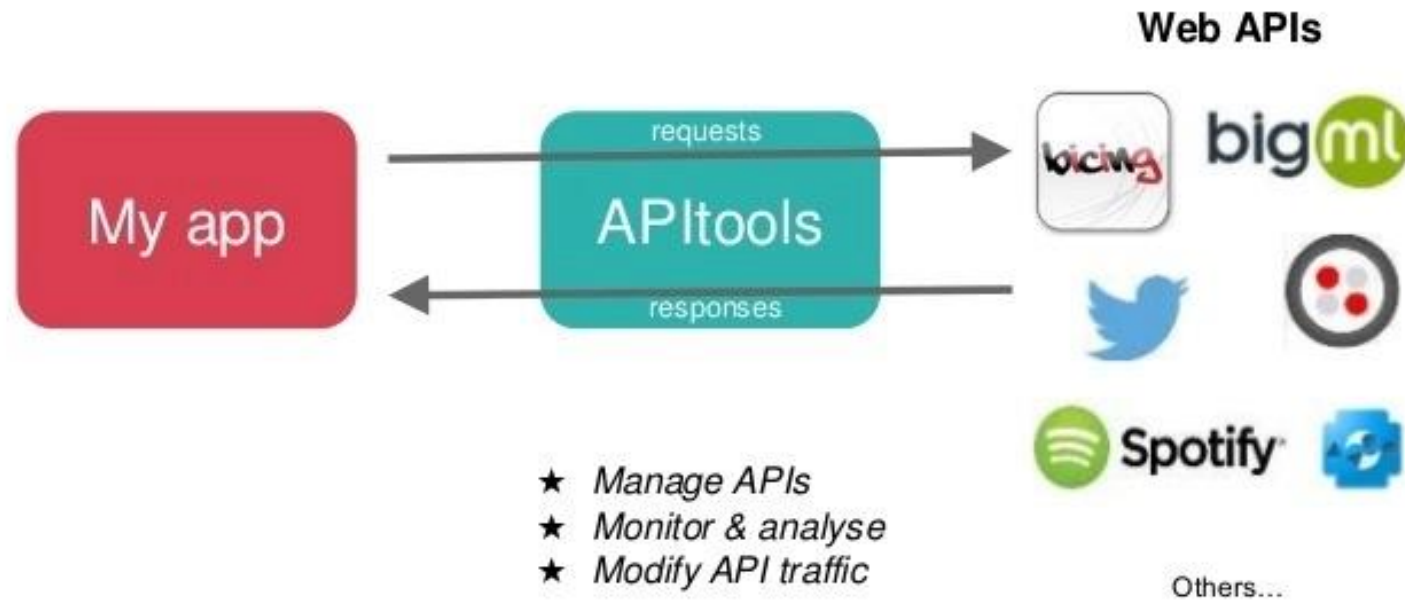
Technology

- Automatic text processing [4]

  - Collect the data from various platforms

  - Word segmentation, text cleaning, and normalization

- Machine Learning [4]

  - The intelligence algorithm of NLP

Refs : https://towardsdatascience.com/automatic-text-summarization-simplified-3b7c10c4093a

API call

How does it work ?

My app → requests → **APItools** → Web APIs

responses ←

★ Manage APIs
★ Monitor & analyse
★ Modify API traffic

Others...

Get in touch! hello@apitools.com   @ManfredBo   APITOOLS

Refs : https://www.quora.com/What-is-an-api-call

Raw data

#อร่อยไปแดก https://t.co/xL0lOxBljp

1553019815517769728 : 🍶🌟 brownie cereal ✅ กระปุกละ 79฿

มี 3รสชาติ -ดาร์กช็อคฯ -ไมโล

-Oatly dark choc คลีน100% (ไร้แป้งนมเนย) 💪❤️ #แล้วหนูอ่ะ #นวยขายยา #โหนกระแส #saveG #บุพเพสันนิวาส2 #BigMouth #ถูกและดี #ราคาถูก #

1553017805859360768 : น่ากินมากๆเลยงับ #อร่อยไปแดก https://t.co/vciViNlmQ0

1553016392751886336 : ฝากด้วยจ้า #pizza #pizzaday #pizzadelivery #pattaya #pattayafood #pattayacity #อร่อยบอกต่อ #อร่อยไปแดก H

1553016271762927617 : โต๊ะจีนกันหน่อยยย😊

📍ซ้งโภชนา #กัลปพฤกษ์

ครั้งแรกเลยกับที่นี่ ร้านอาหารจีนเก่าแก่

ออริจินัลเค้าอยู่เจริญกรุง

อร่อยเลยนะ ส่วนตัวชอบออส่วน สลัดกุ้ง ปลากระพงนึ่งซีอิ๊ว ใครอยากหาร้านพาผู้ใหญ่ไปกินที่นี่ได้เลย เค้ารับจัดงานแต่งด้วยน้า @aroii #อร่อย #อร่อยไปแดก

#อร่อยบอกต่อ https://t.co/YxVqXVbiWa

1553015619376664576 : สัปดาห์นี้คือหมดตั้งแต่ช่วง1ทุ่ม ขออภัยหากท่านใดออเดอร์ไม่ทันนะคะ ไม่โกรธกันนะ #อร่อยบอกต่อ #อร่อยไปแดก #อร่อย #ของกิน

1553015165439713280 : 🦑✨Plawann

ปลาวาฬบ้าบิ่นมะพร้าวน้ำหอม✨🦑

กล่องเล็ก 55 บาท

กล่องใหญ่ 100 บาท

Stop words

- Words that do not add much to the meaning of the sentence. [4]

  - English: is, on, over, the, for, a, out

  - Thai: กว่า กัน กับ ขอ คือ จะ สุด ว่า ได้ ใน

- There exists no standard list of stop words. [4]

Checklist for text data cleaning

- What do we want to do with symbols, numbers, and non-language part?

- Do we want to remove stop words?

- Do we want to lowercase the data? Jennie -> jennie

- Do we want to stem the words? searched, searching -> search [4]

Word segmentation

```
[ ] sefr_cut.load_model(engine='tl-deepcut-ws1000')

    loading model.....
    Success
```

```
[ ] print(sefr_cut.tokenize(['สวัสดีประเทศไทย','ลุงตู่สู้ๆ']))
    print(sefr_cut.tokenize(['สวัสดีประเทศไทย']))
    print(sefr_cut.tokenize('สวัสดีประเทศไทย'))

    [['สวัสดี', 'ประเทศ', 'ไทย'], ['ลุงตู่', 'สู้', 'ๆ']]
    [['สวัสดี', 'ประเทศ', 'ไทย']]
    [['สวัสดี', 'ประเทศ', 'ไทย']]
```

Refs : https://colab.research.google.com/drive/1xA2rzYVnVWwxy6oFkISiG63x-

5u1gwa1?usp=sharing&fbclid=IwAR0paHckUf-aWVXH147YoDFnQ5LjaZBUuH30odJg4ZcZE9devUaJ-

n7eKJs#scrollTo=BC0OVEvsd0ew

References

1. https://www.cc.gatech.edu/classes/AY2020/cs7650_spring/slides/1_intro.pdf

2. https://www.ibm.com/cloud/learn/natural-language-processing

3. https://attapol.github.io/compling/

4. https://www.youtube.com/watch?v=p83LW16b09Y&t=773s

5. https://www.cs.cornell.edu/courses/cs2112/2019fa/labs/lab07/Slides.pdf