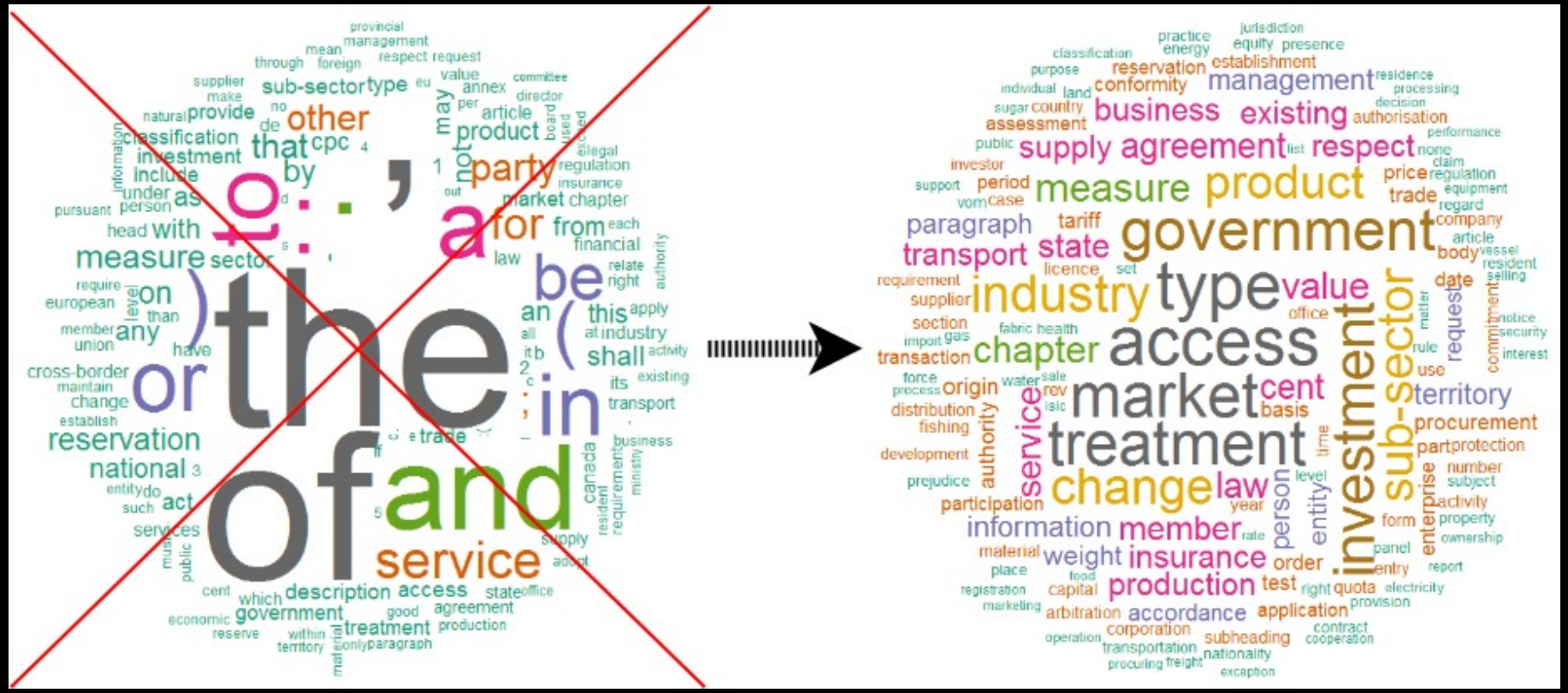
KEYWORD EXTRACTION

A brief introduction with open-source Python tools

DEFINITION

- Keyword extraction is the automated process of extracting the **most relevant and important** words and phrases (keywords, key phrases) from text



USE CASES

- Social media monitoring
- Brand monitoring
- Customer service / support (e.g. respond to urgent customer queries)
- Customer feedback (e.g. tagging incoming customer review / survey response)
- Business intelligence
- Search engine optimization (SEO)
- Product analytics
- Knowledge management

BUSINESS VALUES

- Save time and costs
  - Big data era: more than 290 billion emails sent and received on a daily basis, and half a million tweets posted every single minute, where 80+% of the data generated is unstructured (text, audio, video)
  - Keyword extraction is a powerful tool to free your team of tedious manual processing
  - Distill big data to capture the most important info just in seconds

- Obtain actionable insights
  - It provides you with actionable insights that you can use to make better business decisions
  - Help you understand your data and your customers for a data-driven business strategy
    - *E.g. "What percentage of customer reviews are saying something related to Price?*
    - *The aspects of your product that need to be improved, and what customers are saying about your competition, among other things"*

ADVANTAGES OF KEYWORD EXTRACTION

- Keyword extraction models are easy to set up and implement
- Automatically index data, or generate keyword tag clouds
- Scalability
  - Automating this task gives you the freedom to concentrate on other parts of your job
- Use consistent criteria
  - KE is based on rules and predefined parameters to be free of inconsistency that is common for manual analysis
- Enables real-time analysis
  - Get insights about what's being said about your product as they happen

KEYWORD EXTRACTION APPROACHES

- Statistical, linguistic, graph-based, machine learning or a hybrid scheme

- Open-source Python libraries and packages for keyword extraction (see notebook)
  - RAKE NLTK
  - Scikit-Learn with TF-IDF
  - PKE
  - FlashText

REFERENCES

1. A Comprehensive Guide to **Keyword Extraction** analysis: what it is, how it works, use cases: https://monkeylearn.com/keyword-extraction/
2. A list of NLP GitHub Repos for Keyword Extraction (in star ranking order): https://github.com/topics/keyword-extraction
3. FlashText: https://github.com/vi3k6i5/flashtext
4. RAKE-NLTK: https://github.com/csurfer/rake-nltk
5. PKE: https://github.com/boudinfl/pke#implemented-models
6. SpaCy: https://spacy.io/ and https://spacy.io/usage/spacy-101
7. Textacy: https://github.com/chartbeat-labs/textacy
8. Scikit-Learn with TF-IDF: https://github.com/kavgan/nlp-in-practice/tree/master/tf-idf