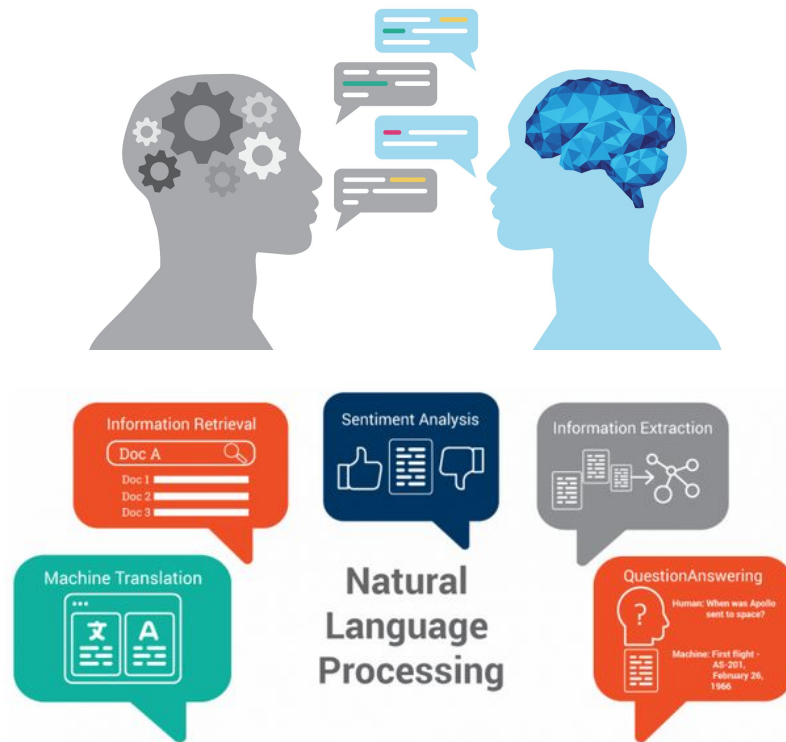

Introduction to Natural Language Processing for Reputation Tracking

Nicolay Rusnachenko

What is Natural Language Processing (NLP)

1. **Subfield** of CS, Linguistics, AI
2. **Focused on** the interactions between computers and human language
3. **Rapid information growth** makes studies in automatization of language processing quite important



Natural Language Processing for Text Generation

Text Generative

- Machine translation
- Text Summarization
- Text Simplification
- Question-Answering
- Dialogue-Systems (Chat-Bots)



Natural Language Processing for Information Retrieval

IR, Information Retrieval

- Named Entities Recognition (NER)
- Relation Extraction (RE)
- Sentiment Analysis

The collage illustrates various NLP applications:

- Google Docs Suggestion:** A screenshot of a Google Docs document with the text "It's awesome that snails can sleep for three years." A sidebar suggests replacing "awesome" with "fantastic" or "incredible". A note states: "The word **awesome** is often overused. Consider using a more specific synonym to improve the sharpness of your writing."
- Email Sentiment Analysis:** An email from Pat Marlowe (Director of IT, Green Bay Packers) to support@technology.com. The text is analyzed for sentiment, with phrases like "Very disappointed with the support...", "We are not happy with the progress on this case.", and "This problem was a critical failure at the time and we expected a timely RCA." highlighted in red.
- Document Entity Extraction:** A document snippet with various entities and issues highlighted in colored boxes: "Lack of progress", "Critical issue", "Frustration", "Urgency", "Profanity", "Production Issue", "Feedback", "Churn Risk", "Documentation Gap", "Escalation Request", "Confusion", "Revenue Impact", "Keywords", "RCA", "Patch (v3.2)", "AWS EC2", "CIO", "Agent signals", "Politeness", "Empathy", "Lack of progress", "Assessment", "Attention Score", "Poor", "Agent Interaction", "Sentiment Score", "Account Health", "Predictions", "Likely to escalate", "Churn risk".
- Network Graph:** A network graph showing relationships between various entities. Nodes include Salman, Yemen Republic, Afghanistan, Near East, Washington, EU, Saudi Arabia, Egypt, Crimea, ISIL, USA, Russia, Israel, America, Obama, Iran, Author, Persian Gulf, Moscow, Bashar Asad, Syria, Ukraine, and Washington. Red dashed lines indicate relationships between nodes.

Natural Language Processing Subtasks

1. Generation

- Summarization
- Simplification
- Question-Answering
- Dialogue-Systems (Chat-Bots)

2. Information Retrieval

- Named Entities Recognition (NER)
- Relation Extraction (RE)
- Sentiment Analysis

NLP Cases for Reputation Tracking Use Cases

Based on News and Reviews:



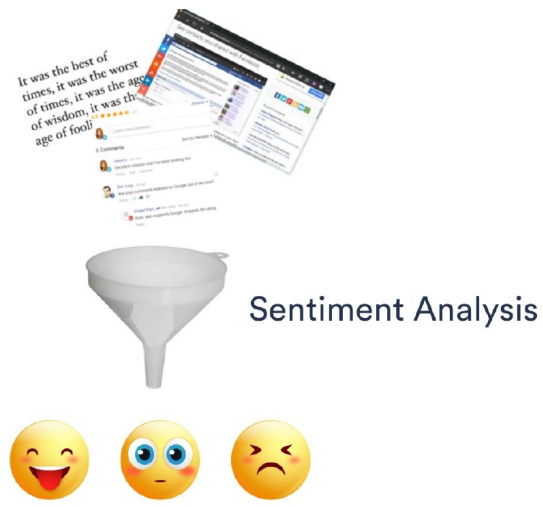
IR, Information Retrieval (Required)

- Named Entities Recognition (NER)
- Relation Extraction (RE)

How it works?

Text Classification:

Predict sentiment class for a given review/news/text.



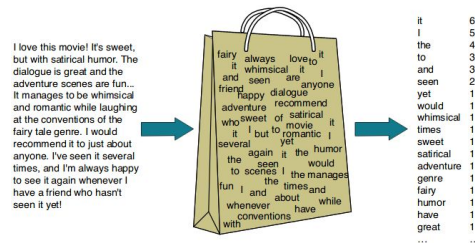
Methods:

- **Rule-based** — process and treat input text by using a predefined patterns
- **Machine-learning based** — adopt model which

Machine Learning methods in Sentiment Analysis

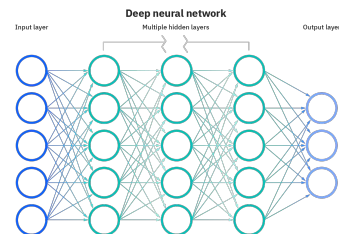
Conventional Methods and Classifiers:

- Bag-of-words representation of texts.



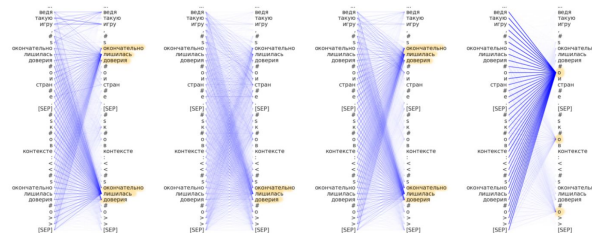
Neural Networks (2011 and later):

- Embeddings
- Allows to capture connection between words



Language Models and Transformers (since 2018):

- Pretrained on a large set of texts.
- Could be simultaneously adopted for many tasks
- Able to track weight and importance of words



Conclusion

- Reputation tracking is considered to be a part of the Information Retrieval problem from a mass-media texts (Sentiment Analysis)
- Machine Learning finds a significant application for NLP due to the complexity of natural language text structures
- Recent advances and appearance of Language Models causes a significant breakthrough in most NLP tasks.

Thank you for your attention!