LUND UNIVERSITY

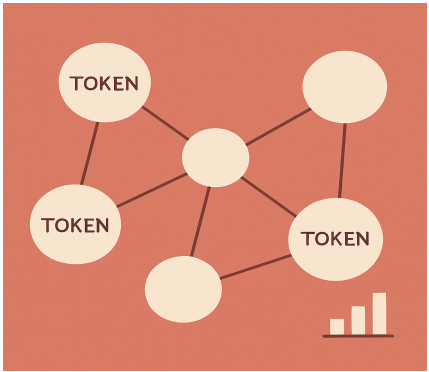
Module 3: Analyzing text content with natural language processing

Lesson 3.2: Part-of-speech and named entity recognition

nils.holmberg@iko.lu.se

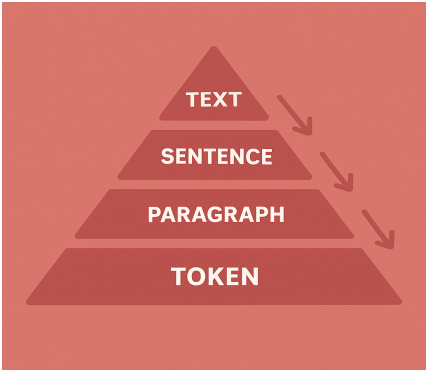
Token Relationships and Knowledge Graphs

- Tokens build text meaning
- Aspect sentiment targets topics
- Knowledge graphs map entities
- Relationships clarify linguistics
- Drives sentiment, recommendation, chatbots



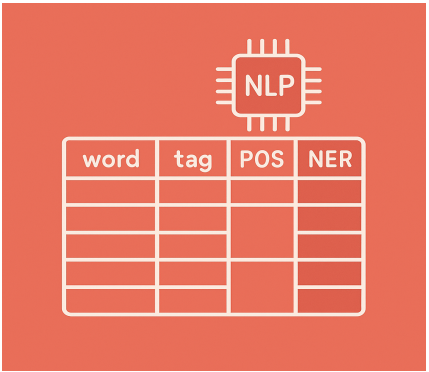
Units of NLP Analysis

- Analysis spans texts to tokens
- Texts provide narrative context
- Paragraphs, sentences segment processing
- Tokens carry POS attributes
- Granular units scale analysis



Applying SpaCy NLP Models to Dataframes

- SpaCy offers pretrained models
- Dataframes align text outputs
- Models add token entity metadata
- Batch processing boosts efficiency
- Enables classification, summaries, sentiment

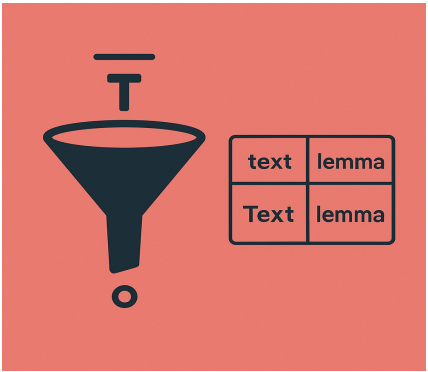


Iterating Over SpaCy Documents

- Sentence dataframes structure analysis
- Docs expose linguistic annotations
- Iteration extracts sentence attributes
- Blends structured and NLP insights
- Supports summarization, search contexts

Text Normalization and Token Attributes

- Normalization standardizes tokens
- Lemmatization finds base forms
- Token attributes reveal linguistics
- Consistency improves downstream tasks
- Crucial for multilingual noisy text



Inferring Named Entity Recognition (NER)

- NER identifies names organizations
- Entities deliver structured insights
- Applications: categorization fraud sentiment
- Powers personalization and search
- Builds knowledge graphs QA

Inferring Part-of-Speech Tagging (POS)

- POS tags grammatical roles
- Captures syntactic structure
- Supports translation and generation
- Improves sentiment topic accuracy
- Enables dependency coreference tasks