

GROOPER TRAINING



COURSE: DATA EXTRACTION FOR UNSTRUCTURED DOCUMENTS

EXPECTATIONS: Learn how to extract data from unstructured documents using Grooper's natural language processing. Students will explore user-assisted machine learning using the TF/IDF algorithm, and how to apply paragraph detection and flow-based collation methods to simulate human reading.

INTENDED AUDIENCE: You work in any industry with a requirement to extract data from unstructured documents.

PREREQUISITES: Architect training or 1+ years' experience working in Grooper.

COURSE DETAILS:

DELIVERY FORMAT: Remote Online Training

COURSE DURATION: 3 days

WHAT YOU'LL LEARN:

- Natural language processing using the Field Class data extractor
- User-assisted machine learning with the TF/IDF algorithm
- How to train your data, and when to stop training
- Grooper's paragraph detection
- How to use the Data Type extractor for unstructured documents
- Flow-based collation methods



KEY CONCEPTS:

- Dealing with bad OCR
- Fuzzy RegEx
- Dealing with multiple names, parties, etc.
- Lexicon training



FINAL EXAM:

- Students will build data models to create structure from unstructured data
- Students will demonstrate their understanding of the TF/IDF algorithm by training document sets to extract data via the Field Class extractor
- Students will workshop simple to complex document sets across multiple industries and successfully target required data elements

