

# GROOPER TRAINING



## COURSE: DATA EXTRACTION FOR STRUCTURED & SEMI-STRUCTURED DOCUMENTS

**EXPECTATIONS:** Discover different methods for configuring data extraction on structured and semi-structured documents. Learn data modeling on document sets, using the Data Type extractor to target, collate, and populate results.

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

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

### COURSE DETAILS:

**DELIVERY FORMAT:** Instructor-led training in Oklahoma City, OK

**COURSE DURATION:** 3 days

Contact your sales representative for registration information.

### WHAT YOU'LL LEARN:

- Intermediate regular expressions to pattern match data structures
- Hierarchical data modeling and inheritance
- Collation methods to take advantage of special relationships and data manipulation
- Data sectioning to target repeating sections of similar information on a single document
- Intermediate extractor design to take advantage of data filtering, pre and post-processing



### KEY CONCEPTS:

- Lexicons
- Data extraction
- Fuzzy matching



### FINAL EXAM:

- Analyze extraction methods to determine how to target data on structured and semi-structured documents
- Configure data models to demonstrate an understanding of hierarchical data modeling
- Successfully target and extract data from complex data sets

