



Data Extraction for Semi-Structured Documents

EXTRACT VALUABLE DATA AND INSIGHTS FROM PDFS, WORD DOCUMENTS, AND MORE

Structured data

Every analyst dreams of clean, structured data where each datum "lives" at a specific place in a spreadsheet or database, quick and easy to find.

Business intelligence (BI) tools use data "addresses" to identify, reach and extract relevant data for an analyst's queries.

Unstructured data

Unstructured text documents, such as tweets and news articles, also contain valuable data. But these data, like companies and product names, don't have "locations" in the same way that a spreadsheet cell has rows and columns.

Analysts must use natural language processing (NLP) software to identify, extract and structure the data within unstructured documents.

Semi-structured data

Meanwhile, some files contain structured data in seemingly unstructured formats.

SEMI-STRUCTURED DOCUMENTS:

- Contracts
- Regulatory updates
- Research papers
- Financial documents
- Market reports
- News articles
- SEC filings
- Requests for Proposal (RFPs)

Lexalytics is the industry leader in translating text into profitable decisions through the Lexalytics Intelligence Platform, a complete, modular data analytics solution.

Data analysts turn to Lexalytics to gain more value from large volumes of structured, unstructured and semi-structured text documents.

Think about legal contracts, financial documents, or even this datasheet. These are **written in text** but have **structured sections** organized under headings or in tables and lists.

Each section adds a layer of **context** to the words within. For example, it's useful to know that a contract mentions a particular company and a dollar amount. But it's even more valuable to know that these are mentioned in the context of "service level agreement."

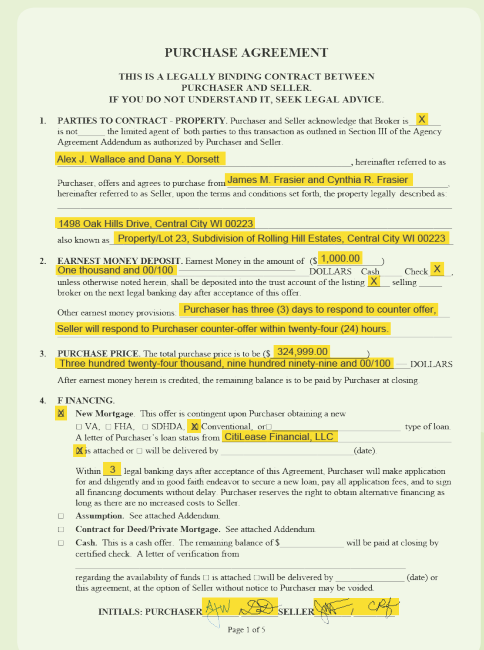


Figure 1 | Example of a semi-structured document: real estate purchase agreement

CONTACT US TODAY



Applications of data extraction services:

- Help customer support agents resolve issues more quickly
- Analyze data from stock market reports to identify trending companies
- Stay up-to-date with regulatory updates and changes
- Flag input errors and suspicious financial recommendations for an auditor to review
- Extract relevant information from EHRs to improve clinical decision making and revenue cycle management
- Aggregate similar contracts to comply with Public Contracts Regulations

Extract
VALUABLE DATA
and
GAIN CONTEXTUAL INSIGHTS
from semi-structured documents

BI tools are built to handle structured data but stumble with unstructured text because each sentence looks like one very large datum. NLP tools excel at analyzing unstructured text, but don't account for the importance of structure. **Semi-structured documents**, like contracts, are left in a "dead zone" where BI and NLP tools both fall short.

VALUABLE DATA TO EXTRACT:

- Recommendations
- Customer requirements
- Subscription details
- Disclaimers
- Deadlines
- Stock ticker symbols
- Contract roles
- Age ranges
- Medical codes
- Products
- Cash amounts
- Order numbers
- Illnesses
- Companies

Unlock the full value of your documents

Lexalytics solves this by combining a **semi-structured data parser** with NLP.

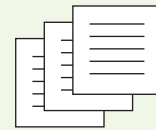
We use machine learning to understand the underlying structure of financial, medical and legal documents. Then we extract structured data directly and use NLP to transform unstructured text into structured based on each clients' needs.

Our approach unlocks the full value of these documents and opens up new analytics applications in regulatory compliance and other fields.

CONTACT US TODAY

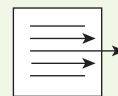


4-STEP DATA EXTRACTION SERVICES



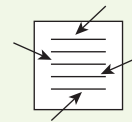
1 Gather documents

You have valuable data locked away in complex, semi-structured text documents.



2 Evaluate structure

We use our semi-structured data parser to evaluate the underlying structure of your semi-structured documents.



3a Extract structured data

We automatically extract already-structured data and insert it into your database, warehouse or other storage system.



3b Process unstructured data

We use natural language processing to transform unstructured text into structured data.



4 Integrate or act

Add this capability into your products for customers to leverage, or improve your decision-making.

Figure 2 | Lexalytics uses semi-structured data parsing and NLP to transform previously-inaccessible information into usable data.