LEXALYTICS

48 NORTH PLEASANT ST. UNIT 301
AMHERST MA 01002 USA

sales@lexalytics.com
1-800-377-8036
www.lexalytics.com

DATASHEET

Semi-Structured Data Parsing

IDENTIFY, EXTRACT AND ANALYZE DATA FROM MEDICAL, FINANCIAL, AND LEGAL DOCUMENTS

Semi-structured documents contain structured data in seemingly unstructured formats.

Most tools fall short at analyzing these documents because they overlook important data or fail to account for the influence of structure on context.

Lexalytics combines a semi-structured data parser with natural language processing to solve this problem.

Text documents generally contain structured and unstructured data. **Structured data** points "live" at defined locations that analytics tools use to locate them. **Unstructured data** requires natural language processing to identify and structure for further analysis.

But not all documents are cleanly structured or unstructured. Some contain **unstructured text inside structured elements**, including tables and headers, which add important context and nuance to the data. Meanwhile, other documents have **structured elements hidden in unstructured text**, such as lists written as sentences.

Think about legal contracts, financial documents and medical records. These are written in text but have structured sections that add context to the words within. For example, it's useful to know that a land deed mentions a company, a person and a bank. But it's even more valuable to know that these are brought up in the context of "Lender," "Borrower" and "Trustee."

Natural language processing (NLP) systems excel at analyzing unstructured text but don't account for how structure of a document influences the data within it. **Business intelligence** (BI) tools are built to analyze structured data but fall short with unstructured text because each sentence looks like one very large datum. **Semi-structured documents** are left in a "dead zone" where both BI and NLP tools fall short. Lexalytics solves this problem by combining a semi-structured data parser with natural language processing.

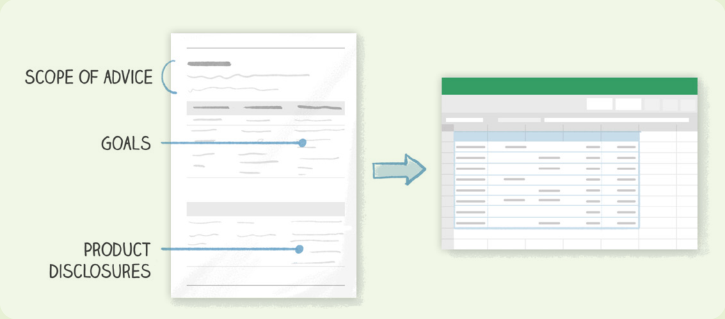
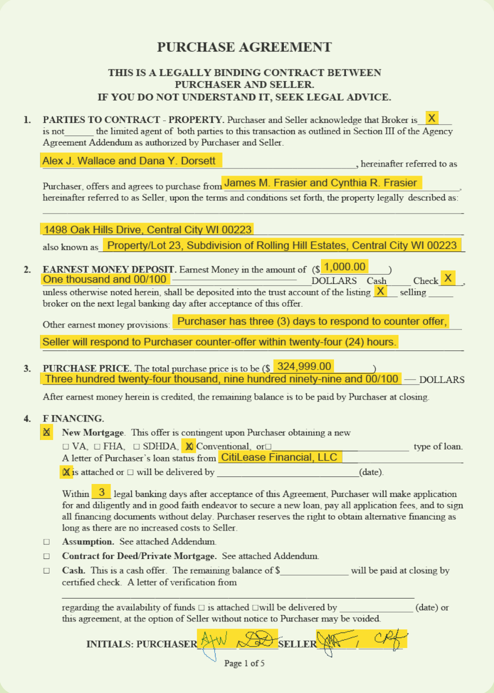


Figure 1 | *We use semi-structured data parsing to identify, extract and structure data from financial documents*



PURCHASE AGREEMENT

THIS IS A LEGALLY BINDING CONTRACT BETWEEN
PURCHASER AND SELLER.
IF YOU DO NOT UNDERSTAND IT, SEEK LEGAL ADVICE.

1. **PARTIES TO CONTRACT - PROPERTY.** Purchaser and Seller acknowledge that Broker is X is not______ the limited agent of both parties to this transaction as outlined in Section III of the Agency Agreement Addendum as authorized by Purchaser and Seller.

   Alex J. Wallace and Dana Y. Dorsett, hereinafter referred to as

   Purchaser, offers and agrees to purchase from James M. Frasier and Cynthia R. Frasier, hereinafter referred to as Seller, upon the terms and conditions set forth, the property legally described as:

   1498 Oak Hills Drive, Central City WI 00223

   also known as Property/Lot 23, Subdivision of Rolling Hill Estates, Central City WI 00223

2. **EARNEST MONEY DEPOSIT.** Earnest Money in the amount of ($ 1,000.00)
   One thousand and 00/100 DOLLARS Cash_____ Check X, unless otherwise noted herein, shall be deposited into the trust account of the listing X selling _____ broker on the next legal banking day after acceptance of this offer.

   Other earnest money provisions: Purchaser has three (3) days to respond to counter offer, Seller will respond to Purchaser counter-offer within twenty-four (24) hours.

3. **PURCHASE PRICE.** The total purchase price is to be ($ 324,999.00)
   Three hundred twenty-four thousand, nine hundred ninety-nine and 00/100 — DOLLARS

   After earnest money herein is credited, the remaining balance is to be paid by Purchaser at closing.

4. **F INANCING.**

   ☒ **New Mortgage**. This offer is contingent upon Purchaser obtaining a new
   □ VA, □ FHA, □ SDHDA, ☒ Conventional, or□ ____________ type of loan.
   A letter of Purchaser's loan status from CitiLease Financial, LLC
   ☒ is attached or □ will be delivered by ____________(date).

   Within 3 legal banking days after acceptance of this Agreement, Purchaser will make application for and diligently and in good faith endeavor to secure a new loan, pay all application fees, and to sign all financing documents without delay. Purchaser reserves the right to obtain alternative financing as long as there are no increased costs to Seller.

   □ **Assumption.** See attached Addendum.

   □ **Contract for Deed/Private Mortgage.** See attached Addendum.

   □ **Cash.** This is a cash offer. The remaining balance of $____________ will be paid at closing by certified check. A letter of verification from ____________ regarding the availability of funds □ is attached □will be delivered by ____________ (date) or this agreement, at the option of Seller without notice to Purchaser may be voided.

INITIALS: PURCHASER AJW ______ **SELLER** ______ / CRF

Page 1 of 5

Figure 2 | *Example of a semi-structured document: real estate purchase agreement*

LEXALYTICS

48 NORTH PLEASANT ST. UNIT 301
AMHERST MA 01002 USA

sales@lexalytics.com
1-800-377-8036
www.lexalytics.com

DATASHEET

Applications of Semi-Structured Data Parsing

- Extract and structure data from financial Statement of Advice documents for auditors to review
- Extract relevant information from EHRs to improve clinical decision making and revenue cycle management
- Flag input errors and suspicious financial recommendations for an auditor to review
- Stay up-to-date with regulatory updates and changes in healthcare diagnostic and billing codes

DATA TO EXTRACT:

- *Medical codes*
- *Contract roles*
- *Stock ticker symbols*
- *Illnesses*
- *Disclaimers*
- *Subscription details*
- *Deadlines*
- *Age ranges*
- *Products*
- *Addresses*
- *Order numbers*
- *Disclosures*

SEMI-STRUCTURED DOCUMENTS:

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

Our semi-structured data parser is a set of python code that we can “teach” to understand the structure of a document, such as land deeds, Medicaid updates, or SEC filings. Sometimes this is as simple as writing a few software rules. Other times, we train machine learning models and combine them with rules.

Either way, the **parser learns how to break down each document down into its component sub-structures**, including tables, headers and lists. Then we use API calls to our **Salience text analytics engine** to identify, extract and structure all of the textual and alphanumeric data within those elements into a usable format.

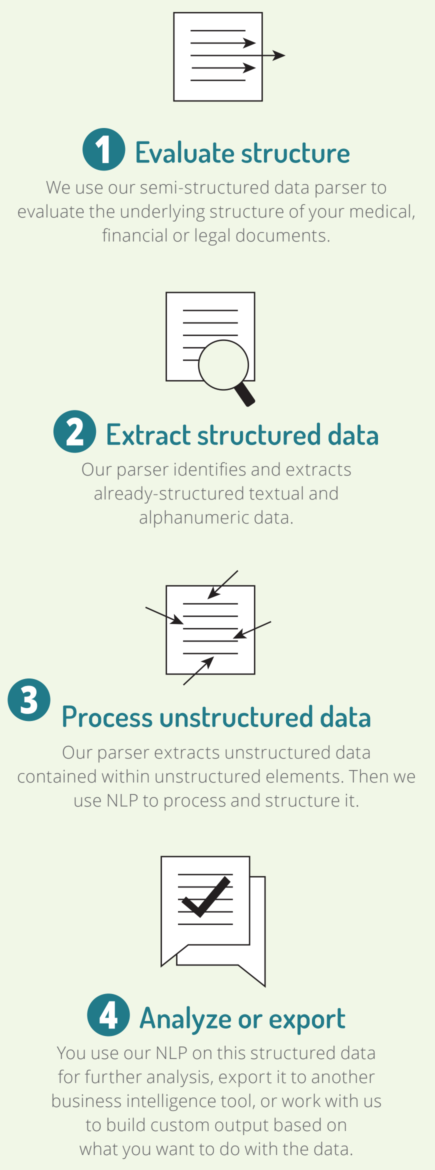
Once the data is prepared, you can **do whatever you want with it**: use our NLP engine for further analysis, export it into another business intelligence tool, or work with us to build some sort of custom output.

By combining our technologies in this way, we account how the structure of a document influences the data within it. This approach **unlocks the full value of semi-structured documents** and opens up new possibilities in regulatory compliance and other fields.

FURTHER READING:

✓ *Semi-Custom Applications*
✓ *AI for Regulatory Compliance*
lexalytics.com/resources

SEMI-STRUCTURED DATA PARSING PROCESS



1 Evaluate structure

We use our semi-structured data parser to evaluate the underlying structure of your medical, financial or legal documents.

2 Extract structured data

Our parser identifies and extracts already-structured textual and alphanumeric data.

3 Process unstructured data

Our parser extracts unstructured data contained within unstructured elements. Then we use NLP to process and structure it.

4 Analyze or export

You use our NLP on this structured data for further analysis, export it to another business intelligence tool, or work with us to build custom output based on what you want to do with the data.

Figure 3 | *Lexalytics uses semi-structured data parsing and NLP to transform previously-inaccessible information into usable data*