

Contract Analysis

Presented by –
Harika Reddy Puli



1

Business
Problem



2

Project
Objective



3

Process
Flow



4

Text
Conversion



5

Feature
Extraction



6

Data
Visualization

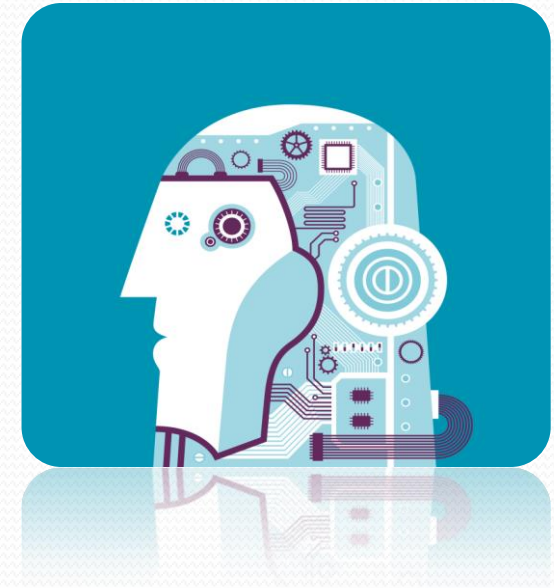


7

Future
Scope

Business Viewpoint

- Contracts play a vital role in any business process and their success
- Organizations possess thousands of contracts and many of them are paper based making it difficult to get relevant information and make informed decisions
- A contract is often complex and lengthy, full of legal terms and unfamiliar concepts, and it can be difficult to keep track of its contents, unless regular reviews are carried out.
- And to change any clause it requires lot of manual effort.
- In the days of uncertainty that lies ahead, it is of utmost importance to understand the key terms, clauses and any obligations present in a contract.



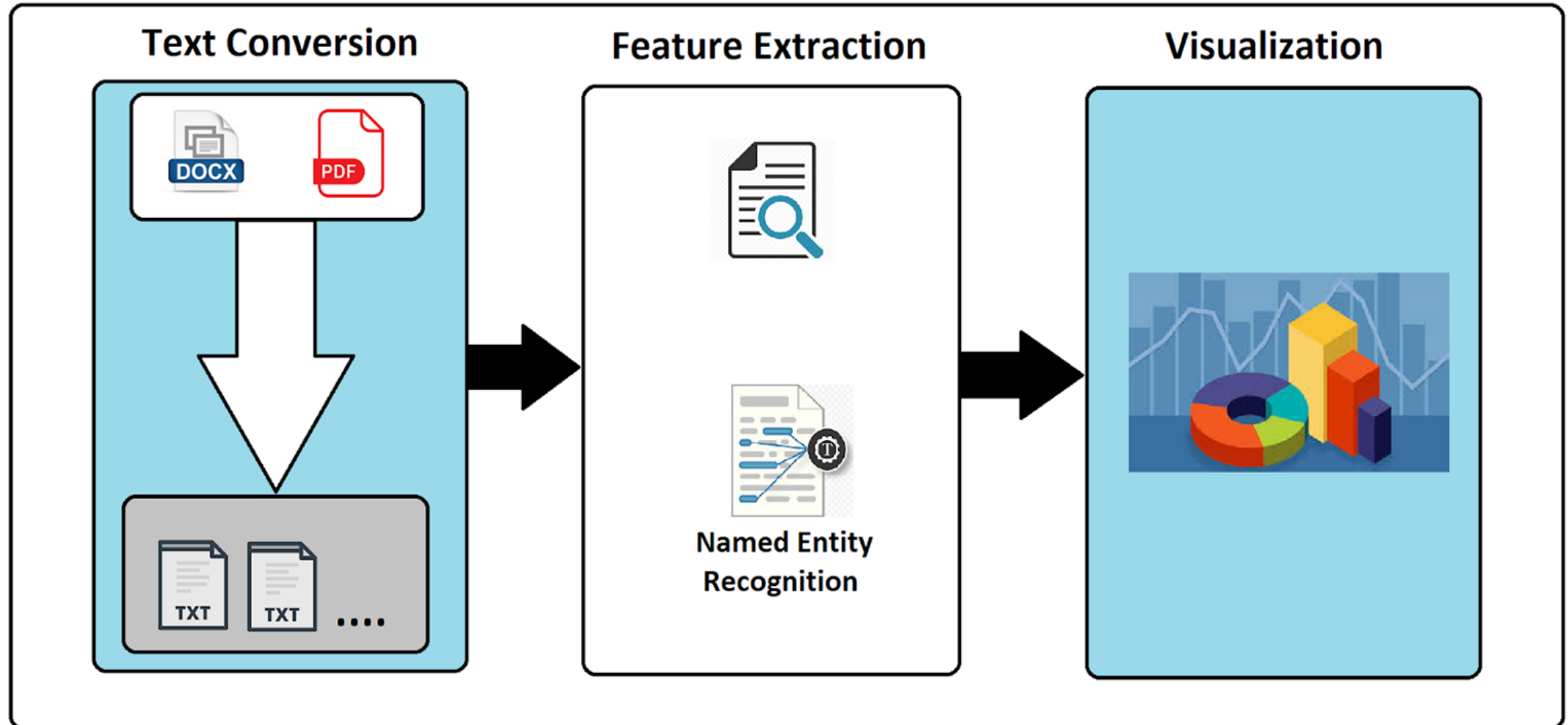
Contract Analysis

- Contract analysis is a process used to examine and track information within a contract.
- This trackable information could be financial or related to the status of the contract, its clauses, and changes in company policy
- Contract analysis starts to become even more powerful when the analysis can be extended to multiple contracts
- Organizations analyze their contracts to improve their strategy, risk management and also their contracting process.
- For example, “How many of their contracts contain Force Majeure clause? Who are the liable parties involved? “ And many

Force Majeure

- Force Majeure in French means “Superior Force”
- In business circles, "force majeure" describes those uncontrollable events (such as war, labor stoppages, act of god, natural disasters) that are not the fault of any party and that make it difficult or impossible to carry out normal business.
- If a Force Majeure clause is mentioned, then it states which parties involved are liable and if the event is beyond control then what are its termination processes.
- At the start of the Covid-19 pandemic, this became a key issue as companies raced to understand the impact on their contracts and business.
- More information on how important is to mention Force Majeure clause and how covid19 has impacted it can be found in the References

Process Flow



Text Conversion

SCHEDULE D FORM OF STATEMENT OF WORK

Statement of Work Number

PREAMBLE

This Statement of Work ("SOW"), effective as of 01st June 2019 (the "SOW Effective Date"), is entered into by and between **ABC Corporation** and **Latentview Analytics Corporation** ("Vendor") pursuant to the terms and conditions set forth in the "Master Agreement".

[WHEREAS, ABC Corporation issued and Vendor responded to a request for proposal ("RFP" entitled") Disability RFP/RFI Inquiry Self-Service to which Vendor affirmatively responded and stated that it is fully capable and willing to provide Business intelligence implementation services services, as more fully described in this SOW (the "Services").]



SCHEDULE D FORM OF STATEMENT OF WORK Statement of Work Number PREAMBLE 01st June
nt of Work the Master Agreement shall take precedence in every instance unless this Sta
o the Master Agreement the SOW number stated above a description of the Services to whi
by Vendor to ABC Corporation in accordance with the Performance and Delivery Schedule s
o for any proposed changes to the Services and/or Deliverables and shall send such Chang
oactivity from any nuclear fuel or from any nuclear waste from the combustion of nuclea
ed the Operator shall be relieved of its obligations to provide the Services but shall
rne by the Authority (the Operator having taken reasonable steps to mitigate the Cost).
ovisions and the Authority shall pay the Supplier an amount calculated and certified in
idating the trusted source system for all the data elements a. b. Streamline data ext
ode 10 Monthly New offsets by month monthly report 11 Monthly STD claims with LTD cov
ates vi. Risk Issues register vii. Management plans (scope schedule cost quality comm
c. Information Glossary Access to the infrastructure to build selfserve module which i

- The parental format of the contracts will be in .DOCX or in .PDF format.
- With the help of text conversion engine, all the data will be converted to .TXT format.
- This engine also performs some data preprocessing steps

Libraries Used

- **OS** – provides a portable way of using operating system-dependent functionality.
 - Python method `os.listdir()` returns a list containing the names of the entries in the directory given by path.
 - We have used it to load all the pdf & word files (contracts) into a list.
- **Docx** – is a Python library for creating and updating Microsoft Word (.docx) files.
 - `document()` - creates an instance of a word document
 - `document(path-to-the-doc)` - We have used it to create an instance of a word document
- **Re** - used to work with Regular Expressions.
 - `Re.sub()` – to replace patterns/sub-strings in a string.
 - We have used it to remove extra white spaces in the documents, after converting them into .txt format

Contd.

➤ PyPDF2 – used for performing tasks like extracting document specific information, manipulating the pages of PDF files.

- `open('example.pdf', 'rb')` - creates a file object
- `PyPDF2. PdfFileReader` - creates a pdf reader object
- `.numPages` - prints the number of pages in the pdf file
- `.getPage(n)` - creates a page object
- `.extractText()` - extracts text from a chosen page object
- `.close()` – closes a file object

➤ Openpyxl – used to read and write excel files.

- `Workbook()` - creates a workbook (object)
- `.active` - Gets the workbook active sheet object from the active attribute i.e creates a worksheet.

Feature Extraction (Rule Based Approach)

- Cleaning the text data
- Converting unstructured text into a structured form
- Searching for entities force majeure and pandemic in the cleaned data

file_name	Force_Majeure	Pandemic_or_disease
Contract_FM_1.txt	Yes	Yes
Contract_FM_10.txt	Yes	No
Contract_FM_11.txt	Yes	No
Contract_FM_12_pdf.txt	Yes	No
Contract_FM_13.txt	Yes	No
Contract_FM_14_pdf.txt	Yes	No
Contract_FM_15_pdf.txt	Yes	No

Named Entity Recognition

- Named-entity recognition (NER) is the process of automatically identifying the entities discussed in a text
- It is probably the first step towards information extraction that seeks to locate and classify named entities in text into pre-defined categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages, etc.
- NER is used in many fields of Natural Language Processing.
- By running an existing NER model,

Firing Mr. **Strzok PERSON**, however, removes a favorite target of Mr. **Trump PERSON** from the ranks of the **F.B.I. GPE** and gives Mr. **Bowdich PERSON** and the **F.B.I. GPE** director, **Christopher A. Wray PERSON**, a chance to move beyond the president's ire.

spaCy

- spaCy is an open-source library for advanced Natural Language Processing in Python.
- spaCy uses neural network models for parsing, tagging & entity recognition
- The spaCy allows you to train NER models by both updating an existing spacy model to suit the specific context of your text documents and also to train a fresh NER model from scratch.
- It can be used to build information extraction or natural language understanding systems, or to pre-process text for deep learning.
- Pre-trained entities - 'person', 'organization', 'location'
- Instead of using a existing pre-trained spacy model and update it, we are creating an empty model
- Entities extracted using NER are
 - Liability of the parties involved
 - Notice Type
 - Termination Period



Custom NER Training

- For training a spacy model, data needs to be annotated in this form
- Created annotated data using different variations of entities appearing in contracts

```
TRAIN_DATA = [  
    ("text1", {"entities": [(start, end, "Liability")]}),  
    ("text2", {"entities": [(start, end, "Liability")]}),  
    ("text3", {"entities": [(start, end, "Notification_type")]}),  
    ("text4", {"entities": [(start, end, "Termination_period")]}),  
    ("text5", {"entities": [(start, end, "Liability")]}),  
    ("text6", {"entities": [(start, end, "Notification_type")]}),  
]
```

- Training Steps:
 - Created a blank spacy english model
 - Disabled all pipes in the model other than 'ner'
 - Added the new labels 'Liability', 'Notification_type' & 'Termination_period'
 - Updated the model using the annotated data
 - Ran the update for a number of iterations till loss becomes almost zero
 - Checked the accuracy of the model by applying it on the contracts
 - Changed drop rate and number of iterations and tried different combinations for better accuracy

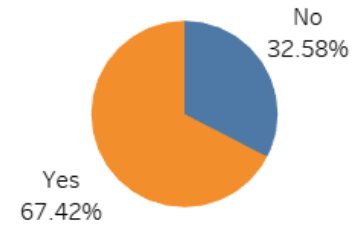
Feature Extraction (Sample Output)

txt	file_name	force_majeure	disease	majeure_col	Liability	Notification_Type	Termination_Period
SCHEDULE	Contract_	Yes	No	FORCE MAJEUR	Neither Party	email message or(ii)	90
SCHEDULE	Contract_	Yes	No	FORCE MAJEUR	No Party	written	30
SCHEDULE	Contract_	Yes	No	FORCE MAJEUR	No Party	telephoneemail or an	90
SCHEDULE	Contract_	Yes	No	FORCE MAJEUR	Neither the Authority nor the	telephone (b) mailed	90

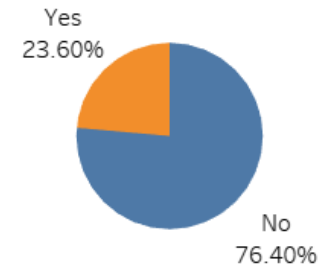
assembly.7.1.5 riot commotion strikes go slows lock outs or disorder unless solely restricted to employees of the Supplier or of I
or threats of terrorism.7.2 Consequences of Force Majeure Event7.2.1 Neither the Authority nor the Operator LIABILITY sha
his Contract to the extent that performance of their respective obligations (excluding payment obligations) is prevented by an Ev
after the Effective Date.7.2.4 If an Event of Force Majeure results in a loss or damage to the Facility then Operator shall rectify
extent required by the Authority PROVIDED that any Cost of rectification (less any insurance proceeds received by the Operato
by the Authority (the Operator having taken reasonable steps to mitigate the Cost).7.2.6 The Operator shall be entitled to
y Charge during the period of interruption caused by the Event of Force Majeure.7.2.7 The Contract Period shall be extended b
eriod of interruption caused by an Event of Force Majeure.1.2.6 OR 1.2.77.2.8 All notices requests demands and other
orce Majeure Event shall be given by (a) telephone (b) mailed NOTIFICATION_TYPE by reputable overnight courier and receip
1 notice or other communication shall have been directed or (c) sent by facsimile transmission with receipt of oral confirmation t
received7.3 Optional Termination Payment and ReleaseIrrespective of any extension of time if an Event of Force Majeure occu
period of 180 TERMINATION_PERIOD days either the Authority or the Operator may give to the other a notice of termination.If
g Force Majeure then Operator should not have a termination right he is being paid.which shall take effect 28 days after the givi

Data Visualization

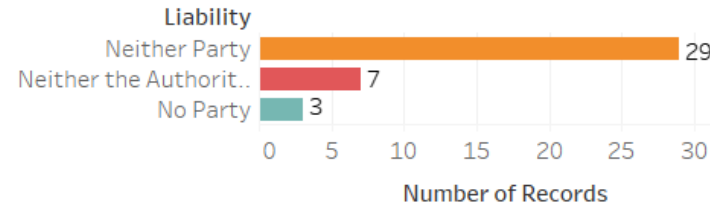
Contracts containing Force Majeure clause



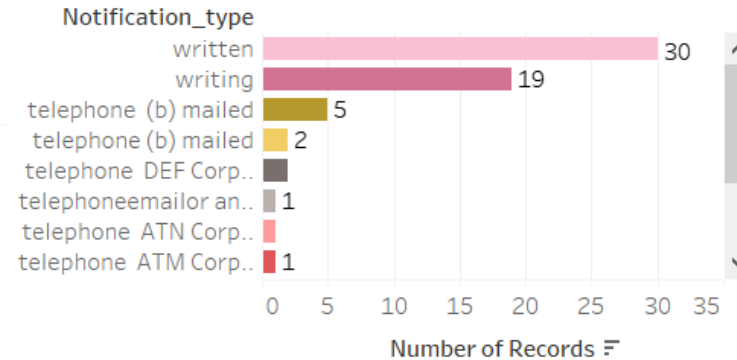
Contracts containing Pandemic key words



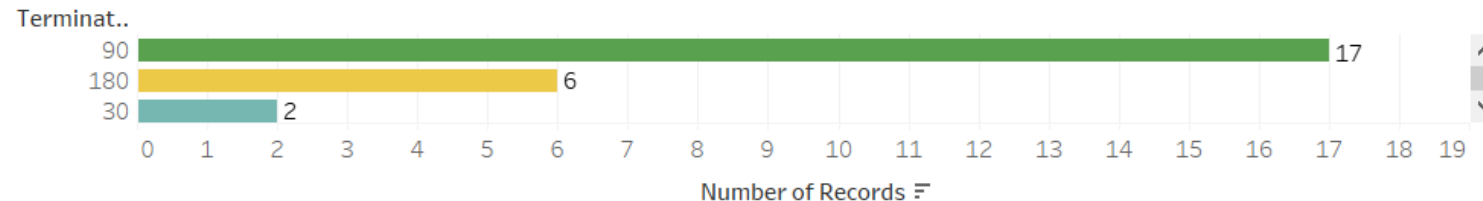
Liability



Notification Type



Termination Period

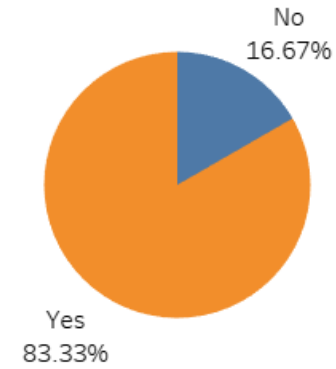


Model Performance

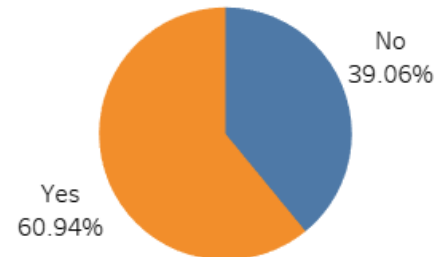
Force Majeure Clause



Pandemic Key Words



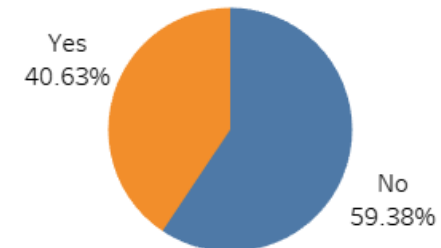
Liability



Notification Type

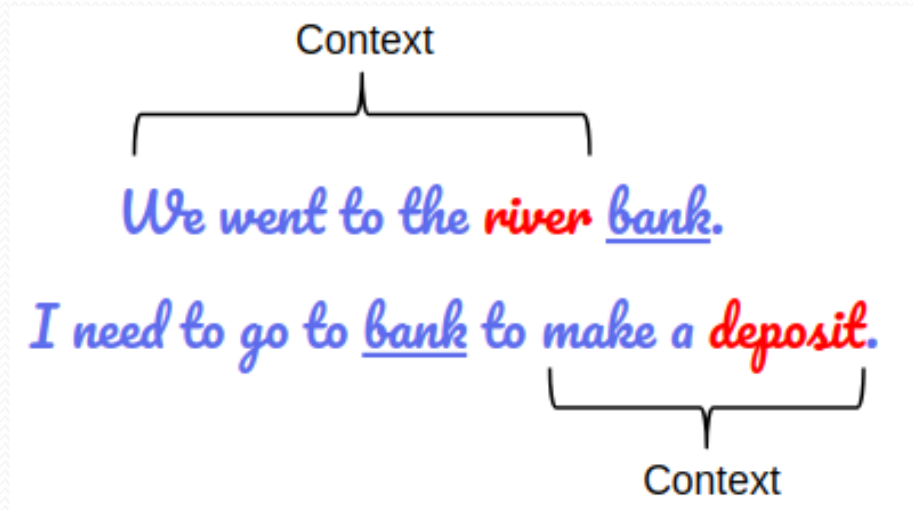


Termination Period



Future Scope (BERT)

- BERT stands for **B**idirectional **E**ncoder **R**epresentations from **T**ransformers.
- It is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context.
- A BERT model is already pre-trained on a large corpus of unlabeled text which includes the entire Wikipedia(that's 2,500 million words!) and Book Corpus (800 million words).
- This is the reason behind the success of BERT, because the model is trained on such huge corpus, it starts to pick up deeper meanings and understands how the language works.



References

- <https://www.lexology.com/library/detail.aspx?g=ob7ad00a-4bf9-4212-aa25-34119b3cfo89>
- <https://www.shearman.com/Perspectives/2020/05/COVID-19-Impact-on-Commercial-Contracts>
- <https://towardsdatascience.com/named-entity-recognition-with-nltk-and-spacy-8c4a7d88e7da>
- <https://towardsdatascience.com/custom-named-entity-recognition-using-spacy-7140ebbb3718>
- <https://www.analyticsvidhya.com/blog/2019/09/demystifying-bert-groundbreaking-nlp-framework/>