

Cognitive Text Extraction Engine (CTEE)

Tech Mahindra's cognitive text extraction engine (CTEE) powered by IBM Watson for business SMEs helps to define and extract domain-specific information from unstructured data sources.

CTEE framework enables one to annotate, train, and extract information using user-defined syntactical patterns. Text extraction engine enables machine learning to read into unstructured data sources and extract relevant domain-specific data points. CTEE is a built system to read bulk source documents at the same time and publish the results

OUR SOLUTION

Tech Mahindra's CTEE solution is enabled with near real-time machine learning capabilities to read unstructured data from a variety of documents such as annual reports/proxy statements, news feeds, etc. Our solution leverages machine learning and natural language classifiers to train for understanding the linguistic nuances, meanings, and relationships in specific industries. Human intervention could be reduced up to more than 35% of routine jobs. It is a highly scalable solution that offers high-cost optimization to the organization.

Workflow

- Extraction of specific information from a free form, user-generated text, and patterns in a human identifiable and syntactical format
- Identify complicated patterns and relationships. CTEE leverages the frameworks available in IBM Watson, Tensorflow and, Pytorch.
- Define entities and syntactical relationships to build an iterative model using Watson Knowledge Studio platform.
- Leverage the capabilities of cognitive analytics and machine learning to create contextual insights
- Potential to reach more than 70% accuracy in the first year

Solution Technology

- **IBM Watson Cognitive Services**
 - Natural Language Understanding
 - Watson Knowledge Studio
 - Discovery
- **Front End**
 - NodeJS SDK

KEY CHALLENGES

Domain SMEs have to manually read and comprehend a bulk of unstructured documents to identify and capture specific information. For instance, if an SME needs to review audit relevant details from a set of documents, it will involve tedious and time-consuming effort to review multiple pages to sort through and obtain requisite information. This also creates room for manual errors and leads to incorrect or inaccurate data points being captured.

BENEFITS

- High level of accuracy and false positives identification to trigger the manual process
- Significant reduction in turnaround time with automation capability
- Reduced need for human intervention significantly eliminates human error in the process
- Additional training is a lot easier once the domain corpus is built sufficiently

TECHM NXT.NOW ADVANTAGE

- Diversified global player with 20+ years of data, analytics, and AI practice
- 11,000+ data, analytics, and AI associates with domain led consulting focus
- Global footprints across 55+ countries
- Partner-enabled ecosystem with COEs across all these technologies
- Comprehensive frameworks, processes and tools to move PoCs to production



To know more, reach us at
Wave2.Communications@techmahindra.com



**Tech
Mahindra**
Connected World. Connected Experiences.