# INFOCEPTS

# Text Analytics for Leading Life Sciences Company



## Summary

Improved insights into clinical trial data for a life sciences company through the use of text analytics.

### Industry

Healthcare

#### **Technologies**

MicroStrategy, Apache Mahout, Open NLP

## The Challenge

A leading life sciences company wanted to reduce the time to review clinical trials for its users, sometimes thousands of records, while evaluating risk.

### The Solution

- Designed a data mart for clinical trial data
- Built text mining engine using OpenNLP to decrease the amount of language and generate business insights from clinical trial text data
- Sentence Detection Tokenization Parsing Document Categorization
- Augmented for subject-specific dictionary implementation

# The Benefits

- Decreased verbosity of textual data
- Generated Insights from unstructured data
- Created a clinical trial phase classification
- Increased screening of data by 20%

# **Text Summarization Engine**







