Product Sheet: Text Analytics

CATEGORY	FEATURE	WHAT IS IT?	WHY IS IT INCLUDED?	HOW DOES IT WORK?
Unstructured Source Integration	Public Data Integration	• Processes unstructured text from news, regulatory, corporate and other public web sources	• Turns public unstructured data sources into a manageable, actionable data resource for client use cases and business processes	• Integrates directly with RSS feeds • Supports imported documents in PDF, HTML and more source formats • Uses a cloud-hosted schema-less NoSQL data lake to scale and store immense volumes of unstructured data
	Private Data Integration	• Allows a direct, secure import of private documents for analysis	• Turns your private unstructured data sources into a manageable, actionable data resource for client use cases and business processes • Supports the data privacy expectations of enterprise-scale organizations	• Uses a private cloud-hosted schema-less NoSQL data lake to scale and store immense volumes of unstructured data
	Document Preparation	• Allows for the import of scanned and other non-machine readable documents • Detects the language used throughout a document	• Prepares documents for additional downstream processing activities required for use cases and business processes	• Uses Optical Character Recognition (OCR) capabilities to prepare scanned documents for further processing • Leverages computer vision capabilities to isolate tables for specific processing activities
Document Classification	Document Classification	• Interprets individual documents to provide a configurable classification • Analyzes a group of documents to identify anomalies or outliers	• Classifies documents to support a variety of business use cases and downstream processing requirements • Classification and grouping of documents more efficiently supports document digitization programs and enterprise document search requirements	• Utilizes an array of techniques such as Word2vec, Doc2vec, TF-IDF, and Latent Dirichlet Allocation (LDA) to support document classification requirements
Language Modeling & Training	Specialized Language Modeling	• Rapidly train new models to fit your source, client, or use case • Trained models achieve 90%+ accuracy across a variety of news, financial and corporate documents, including legalese-filled case documents and legal filings	• Different models can be applied to different types of document sources, supporting unique, fine-tuned analysis across legal documents, financial documents, news articles and more	• Uses general and specialized neural-language models to better support disparate sources of information
	Model Feedback & Tuning	• Provide feedback to the model for continued model improvement in support of higher accuracy requirements	• Allows user influence for language model tuning to enable continuously improving accuracy of data identification and extraction across source documents	• Leverages machine learning capabilities to continuously improve language models for greater accuracy when processing the language of specific source documents

KINGLAND

Product Sheet: Text Analytics

CATEGORY	FEATURE	WHAT IS IT?	WHY IS IT INCLUDED?	HOW DOES IT WORK?
Data Identification & Extraction	Entities & People	• Uses trained language models to identify, tag and extract entities and people within strings of texts, tables and other semi-structured text • Pre-existing fully-trained models identify entities and people throughout most sources and language patterns • Customizable models can be added, configured and trained in a matter of days or weeks to support unique use cases Supports a variety of customizable business rules to extract and prepare data for downstream processing	• Identifies entities and people within unstructured text for exploration within the user interface • Extracts entities and people from unstructured text for integration with downstream data activities and business processes	• Executes custom identification that works with parts of speech, lexemes and syntactic parse • Understands context to identify pronouns, family names, entity names, abbreviations and more
	Events	• Uses a mixture of trained and configurable language models to identify, tag and extract events within strings of texts, tables and other semi-structured text • Pre-existing models identify a variety of corporate action and risk related events within a document, such as name changes, merger/acquisitions, director/shareholder changes, penalty/fines, expulsions/suspensions, insolvency and more. • Models for new events can be added, configured and trained in a matter of days or weeks to support unique use cases • Supports a variety of customizable business rules to extract and prepare data for downstream processing	• Events provide valuable insights specific to your use cases and business processes • Triangulate events, entities and people to develop context within your unstructured source documents	• Executes custom identification that works with parts of speech, lexemes and syntactic parse • Understands context to identify events that are attributable back to specific entities, people and data attributes

KINGLAND

Product Sheet: Text Analytics

CATEGORY	FEATURE	WHAT IS IT?	WHY IS IT INCLUDED?	HOW DOES IT WORK?
Data Identification & Extraction	Data Attributes	• Uses a mixture of trained and configurable language models to identify, tag and extract additional data attributes within strings of texts, tables and other semi-structured text • Pre-existing models identify a variety of popular data attributes within a document, such as dates, locations, currency, time and more • Models for new data attributes can be added, configured and trained in a matter of weeks (longer for complex data attributes) • Supports a variety of customizable business rules to extract and prepare data for downstream processing	• Supports additional use cases and business processes related to specific, unique data attributes extracted from unstructured source documents • Associate data attributes to entities, people and events to develop context within your unstructured source documents	• Executes custom identification that works with parts of speech, lexemes and syntactic parse • Understands context to identify data attributes that are attributable back to specific entities, people and events
	Tabular Data	• Identifies and extracts data nested within tables in unstructured source documents • Pre-existing table processing models identify entities, people and events, as well as a variety of popular data attributes within tables • Models for custom tables with specific accuracy requirements can be configured and trained in a matter of weeks (longer for complex tables) • Supports a variety of customizable business rules to extract and prepare table data for downstream processing	• Supports data identification and extraction beyond natural language, deriving value out of the content within tables in an unstructured source document • Associate table data to entities, people and events to develop context within your source documents	• Executes an ensemble of algorithms and heuristics to provide context while extracting entities, people, events and data attributes from tables within unstructured source documents • Understands context to identify table data that is attributable back to specific entities, people and events

KINGLAND

Product Sheet: Text Analytics

CATEGORY	FEATURE	WHAT IS IT?	WHY IS IT INCLUDED?	HOW DOES IT WORK?
Text Analytics UI	Document SpeedRead	• Summarizes all relevant entities, people, events and data attributes within the text, enabling faster access to specific sections, words or headlines of a document	• Allows efficient document exploration within a secure web application • Improves the speed of reading and comprehension to streamline manual data operational activities for a variety of client use cases and business processes	• Web application UI is surfaced securely for browser-based access by users • Intuitive and modern UI enables efficient usability with regards to manual data operations tasks
	Explore & Discover UI	• Explores and searches all unstructured source content within the data lake, filtering content and documents within a specific time range and for focused organizations, people and events	• Provides a real-time, manageable overview of all unstructured data sources within the data lake • Allows users to explore the data lake content, narrow document scope and improve operational efficiency for a variety of client use cases and business processes	• Schema-less data lake supports complex search and filtering in near real-time Web application UI is surfaced securely for browser-based access by users, enabling efficient usability for manual data operations tasks
	Advanced Analytics Dashboard	• Summarizes all unstructured content within the data lake in a variety of dashboards and UI widgets to provide valuable insight into the trends of organizations, people and events mentioned within content	• Provides analytical meta-data useful to administrative users • Exposes trends of data mentioned within unstructured content in support of specific business processes and use cases	• Uses widget-based charts and dashboards driven by scalable APIs • Web application UI is surfaced securely for browser-based access by users, enabling efficient usability for monitoring data trends within the unstructured source content processing
	Administration & Configuration UI	• Enables UI-based configuration of key sources and events, enabling flexibility for a variety of client use cases and business processes • New unstructured data sources and events can be onboarded and tailored for use cases in a matter of days	• Enables new unstructured data sources to be onboarded quickly for evolving business processes and use cases • Exposes a human-readable model for configuring new events, which are processed across source content to highlight actionable insights	• Web application UI is surfaced securely for browser-based access by administrators, enabling efficient onboarding of sources and events for specific use cases across unstructured content • Configuration information for sources and events is leveraged by the back-end Text Analytics engine to process new unstructured content for insights important to client use cases

KINGLAND

Product Sheet: Text Analytics

CATEGORY	FEATURE	WHAT IS IT?	WHY IS IT INCLUDED?	HOW DOES IT WORK?
Exports, Alerts & Integrations	Data Delivery Module	• Subscribes the user to user-selected entities, people and/or events processed within unstructured source content in the data lake • Supports system-to-system integration	• Refines and tailors the processed content to just the entities, people, events and data attributes important to your use cases • Enables downstream system integrations for improved business processes	• Tailored, use case-specific entity and people lists can be integrated through back-end integrations for refining the processed content • Downstream consumption of content and data is available through batch, REST APIs and message queues
	Notifications & Alerts	• Provides notifications when subscribed entities, people and/or events are detected within processed source content	• Ensures users and downstream systems are immediately aware of events and insights processed within source content for the entities and people that matter most to your business	• Supports notifications directly within the application or in downstream systems via message queuing architecture • Email notifications can be scheduled or supported dynamically
Cloud & DevSecOps	Cloud & DevSecOps	• Scalability, security, flexibility and a mix of best practices including continuous integration and continuous delivery (CI/CD) and an API-driven architecture optimized for cloud	• Backs the overall Text Analytics SaaS solution with the security and development best practices that enterprise-scale organizations expect of their software solutions • Cloud optimization enables the solution to scale to a near infinite amount of source content and processing for the most demanding of use cases	• Best-of-breed DevSecOps technologies are brought together to automate the build/secure/test/deploy development lifecycle • Cloud optimization in AWS ensures scalability and security needed for even the most complex and demanding use cases and performance requirements

KINGLAND