

CLOUD TEXT ANALYTICS®



180 Admiral Cochrane Drive
Suite 305
Annapolis, MD 21401
www.orbistechnologies.com



Overview

The Orbis Cloud Text Analytics® platform allows users to search, discover and rapidly connect artifacts across large data sets.

The Orbis CTA® product platform is grounded in a detailed understanding of natural language processing (NLP) and incorporates advanced distributed processing techniques. The result is a highly scalable and extremely accurate text analytics platform extensible to almost any domain.

Product Highlights

The CTA® platform provides highly accurate cloud-based text analytics services solely dedicated to extracting detailed information and cross-correlating Big Data artifacts such as documents, reports, emails, and webpages.

- Machine-learning-based natural language processing (NLP) technology to identify and extract named entities and hidden semantic relationships.
- Implemented using an industry standard framework so that existing client text analytics components can be rapidly integrated for greater information-extraction precision specific to client's highly specialized domains.
- Incorporates evidential reasoning algorithms to correlate artifacts based on content.
- Easy to use web-based dashboard that enables end users to build an interactive environment based on their needs and work requirements.

Innovative Features

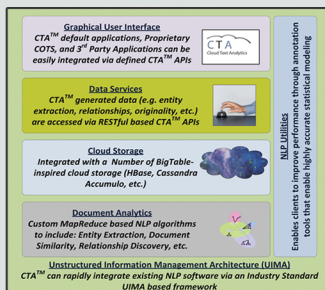
The Orbis CTA® platform performs extraction and analysis of unstructured, semi-structured, and structured data.

- **Easy to Use User Interface:** A web-based dashboard application integrated into HTML5 browser interface as well as the open source Ozone Widget Framework (OWF).
- **Web Service Enabled Data Services and Application Programming Interfaces (APIs):** CTA® provides a collection of data services (e.g. entity extraction, relationship discovery, information originality, and indexing) via Representational State Transfer (RESTful) APIs to enable clients to organically develop new applications or use CTA® data in existing applications.
- **NoSQL Cloud-Based Data Storage:** CTA® populates various NoSQL (e.g. MarkLogic, Apache Cassandra, HBase, Accumulo, CloudBase) and Resource Description Framework (RDF) databases to provide a scalable data service to meet any Big Data processing environment.
- **Affordable Sustainment:** An easy to sustain Big Data product, clients use CTA® utilities to tune algorithms and configure CTA® for new data sets.
- **Foreign Language Processing:** Multiple foreign language plug-ins are available that allow for the processing of non-English documents and the correlation of foreign language documents to English documents.

CTA® User Interface (UI) Components

Orbis CTA® enables clients to rapidly develop custom text-analytics applications by providing a defined application programming interface (API) to the output of the analytics platform. In addition, CTA® also has a dashboard that enables users to search, view and analyze their corpus. Key functionality includes:

- **Semantic Search:** Users search over instances of data, named entities, relationships, or semantic concepts. For example, a specific named entity (e.g. “Joe”) associated with a class of data (e.g. “Person”). The result set is a collection of artifacts with at-a-glance information:
 - time and data associated with the artifact creation;
 - important relations (entities in the target document that are related to a searched-for entity);
 - sample of data from within the artifact (text or image);
 - visual representation of the amount of original content found in the document.
- **Document Viewer:** Users review, modify, and inspect extraction results allowing for highlighting extracted text, an inventory of correlated entities and graphical representation of the underlying semantics illustrating the association between entities.
- **Image Viewer:** Users select image artifacts to view along with any extracted soft-biometric information. This feature also allows users to find similar images.
- **Related Artifacts and Analytics:** Users are shown a graph of artifacts closely related to a selected artifact, as well as the evidence of correlation from Orbis’ advanced NLP and evidential reasoning algorithms. This allows the user to discover relevant artifacts in the corpus that are related to the search term, even if the search entity or term is not present in the artifact.
- **Workspace:** A workspace for users to record the information discovered via the document-review process. Users can copy portions of a relationship graph directly from the Document Viewer to the Workspace. The CTA® Workspace tracks the lineage and pedigree of the entities so that the user can see the source document for each individual entity. Results are exported via common industry formats to incorporate into other analyst tools, presentations, or work products.



Client Use Example-Turn Key Big Data

A large government organization wanted to integrate existing text analytics software tools with CTA® to create a Big Data platform capable of processing tens of millions of documents and reports. The effort had several key technical challenges that made CTA® the best product for the project. Select examples include:

Use of the Legacy Data Architecture was Required:

Clients typically have an existing data architecture that must be used. The CTA® configuration was modified to adopt the existing client data architecture in order to integrate seamlessly into the existing client data and processing architecture.

Use of Third Party Application Developers

The client has an existing team of third party applications developers to build applications using CTA® data service APIs. Orbis provided a collection of data services that non-Orbis developers used to build new applications.

Unique Domain Data with Unique Vocabularies

The client has its own specific collection of domain reports with lexicon and vocabulary that are distinctive. The CTA® utilities allow clients to tune the machine learning algorithms to recognize these unique situations.

Q&A about CTA®

Q: What makes CTA® better than other commercial entity extraction tools?

Most of the entity extraction products can extract with a similar degree of accuracy. However, CTA® goes beyond entity extraction to provide a collection of core document services (e.g. sentence and entity tokenization), Big Data analytics (e.g. cross-document correlation), search services (keyword, concept-based, and faceted) and an end user dashboard application (search, document viewer and enterprise document clustering).

Q: How long does it take to integrate CTA® into our existing environment?

Using CTA® utilities, the product can typically be configured for a new client environment in a few days.

Q: Do I need Orbis to support the product after it is installed?

No. Many clients request Orbis to support the program to build new analytics. However, the product can be supported via a software utility that can be licensed as part of CTA®.

For more information, contact info@orbistechnologies.com.