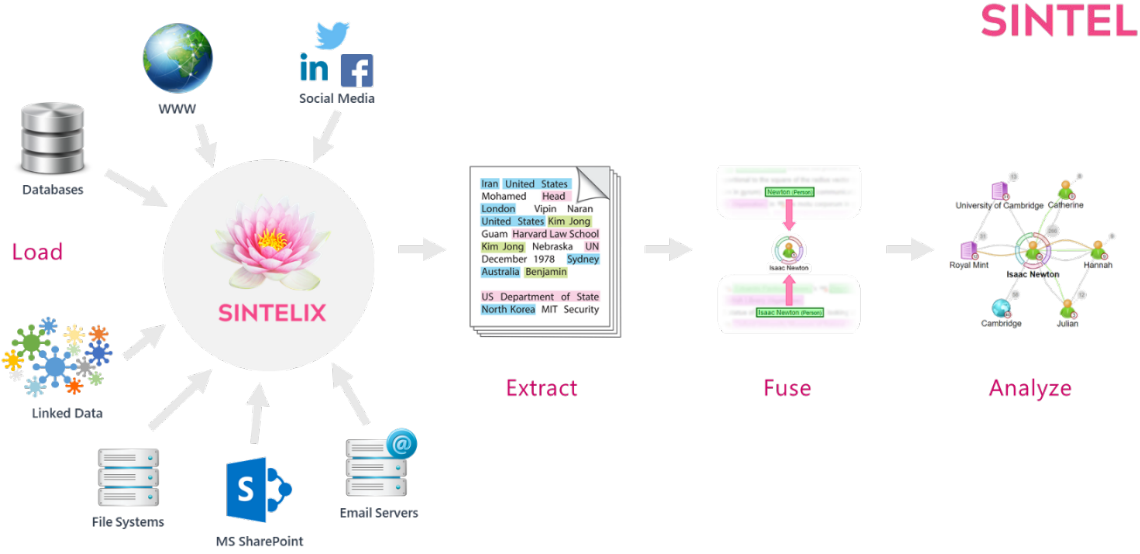
Sintelix Whitepaper



Abstract

The Sintelix text intelligence solution thrives on data – providing an end to end solution from collection to analysis. Sintelix is the leading, fully integrated, text analytics solution used by law enforcement, defence and intelligence customers worldwide including the UK MOD, US DoD. Data is converted into searchable, actionable information – by extracting entities, events, relationships and networks.

A broad range of visualisation tools are available including tables, maps, charts and timelines. Many of Sintelix's out-of-the-box capabilities are world-leading and every aspect of Sintelix can be configured to the customer's needs from the user interface. Sintelix provides automation of time and resource-intensive activities, enabling unstructured data to be leveraged at speed and scale. Sintelix can be used to create structured data from human readable data. This structured data supports multidimensional analysis of open source data and internal data holdings. Sintelix's API and its 125+ connectors can deliver data to conventional structured data systems.

Capabilities

Open Source Data Collection, Social Media Scraping

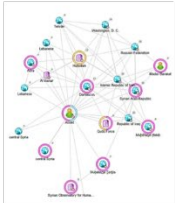
OCR, Normalization Structure recognition, metadata

Entity Extraction Relationship Extraction (Multilingual)

Geotagging

Multimodal Search & Discovery, Alerts, Similarity Search

Visualization Tools

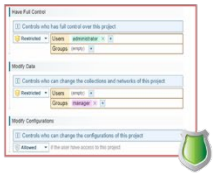


Document Tagging & Classification

Topic Discovery, Tracking Spreadsheet for Text

Workstations, Workgroups, Enterprise

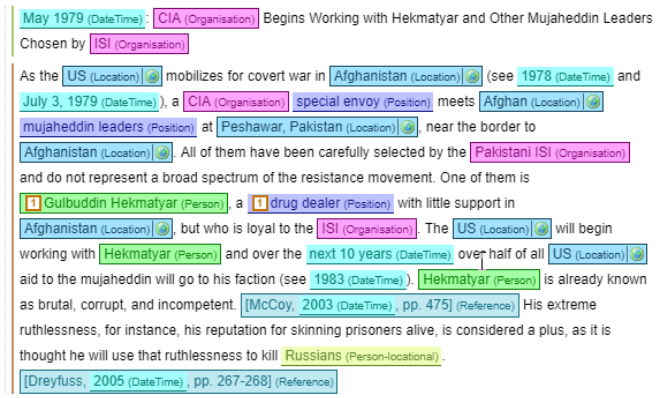
Multilevel Security



One Platform, Many Applications

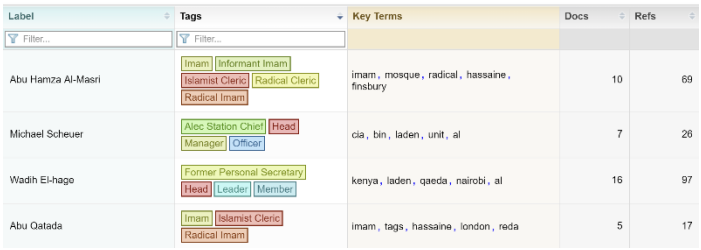
Ingest any Document

- Ingest over 1500 file types
- Email, PST, OST files, server connection
- Open and Dark Web Harvesting
- Social Media … Twitter, Facebook etc
- OCR – Optical Character Recognition
- Audio & video transcription



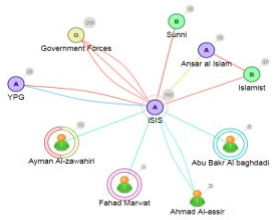
Extract Information & Networks

- Extract entities, relationships and properties
- Geotagging and clickable maps
- Classification, Tagging & Predictive Coding
- Entity Resolution & Community Detection
- Topic Extraction and Mapping
- World best scores
- Immediate value and easy configuration



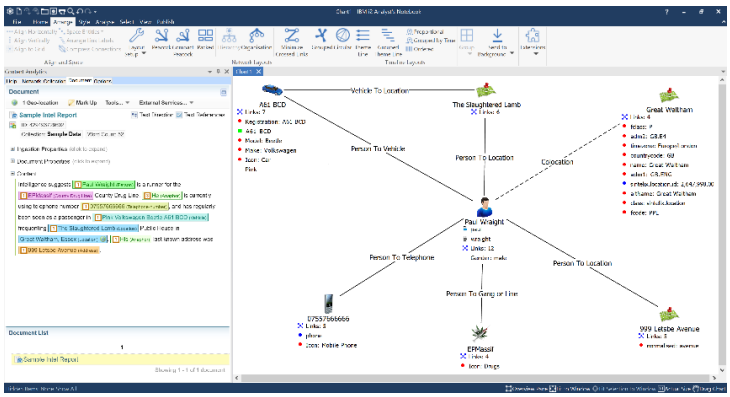
Explore, Analyse & Report

- Advanced Contextual Search
  - on text, metadata, entities, search result clustering, synonyms, favourites, history
- Alerts - emailed direct to you
- Document Taxonomies & BCS Trees
- Network visualization and editing
- Drill down to documents
- Open export to multiple formats
- Configurable “Spreadsheet for text” capability
- Reporting via an advanced Wiki



Technology, Integration

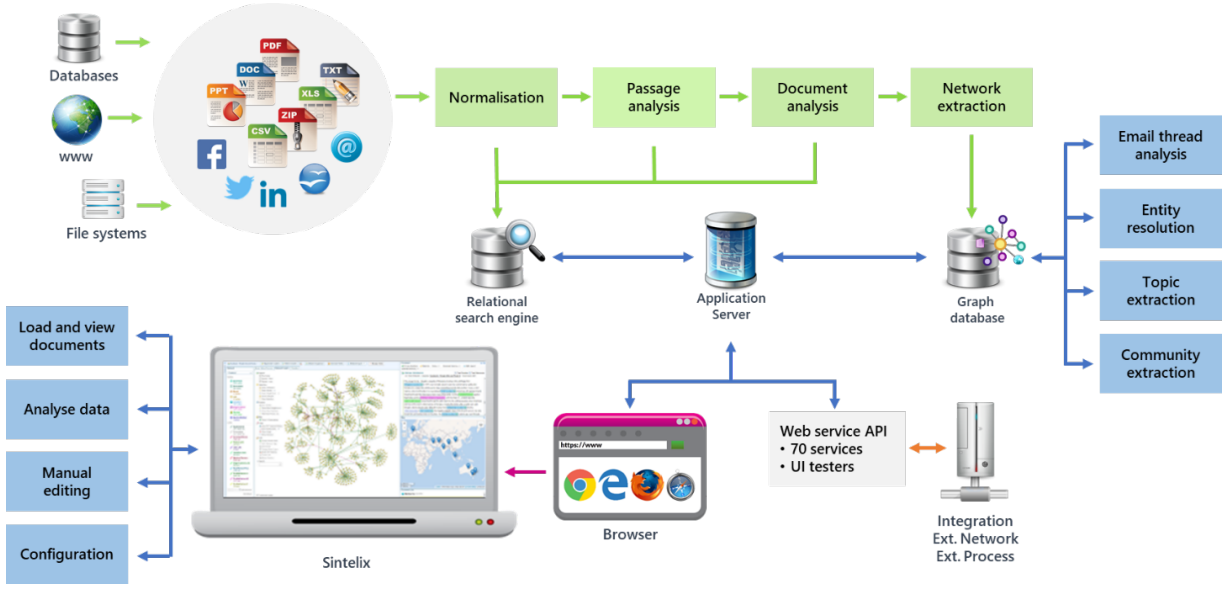
- Java server with Web browser clients
- Multiuser HTML 5 UIs + over 70 web services
- Integration to IBM i2, SharePoint, 125+ platforms
- Scalable to multiple terabytes of data
- Full claims-based security
- LDAP, Active Directory; single sign on
- No software limit to data sizes
- No software limit to network sizes



Speed (8-core processor)

- 30 pages of text per core per second
- 2.6 million pages per core per day
- 80 million entities per core per day

Architecture & Deployment



Processing Chains. Sintelix can be thought of as a system for collecting, normalising and structuring unstructured data. The system accepts data in over 1,500 different file formats. The first stage of input processing is normalisation input data formats into a standard "document model". After normalisation, the document processing pipeline and the network generating pipeline (diagram, above) rapidly extract information and construct a network representing all the extracted information and relationships.

Document collections store normalised text data and source data, if required. Sintelix enriches the normalised documents with metadata, entities, relationships and classifications. Enriched Sintelix documents can be searched via simple keyword search or by using the advanced features of Sintelix's search interface. Collections and networks scale independently in the system making it possible to handle big data.

Networks combine the information Sintelix extracts from documents to expose the explicit and implied linkages. Sintelix Networks store the information that drives visualisations: tables, link charts, timelines and maps. Sintelix's advanced entity resolution capability allow networks to be fused with information derived from incoming data.

Data Storage, Scalability. Sintelix contains two main stores: for documents and facts derived from them and for networks derived from structured and unstructured data input. Sintelix's storage is based on the highly regarded Lucene search engine, which has been extended to handle big data sets. The document-fact store houses the normalised and source documents, together with the references, entities, relationships and tags derived from them. The network store is a graph database and provides source-supported data ready for analysis via Sintelix visualisation tools.

Sintelix requires no support from any third-party database, enabling it to be deployed quickly and easily and provide value immediately. The system can absorb processor cores indefinitely, providing linearly increasing input processing speed and capacity. Scaling is only limited by underlying hardware. The primary data objects in Sintelix are **document collections**, **networks** and **configurations**.

Configurations. Sintelix is highly configurable and contains built-in development environments to make the development each type of configuration rapid and testable.

Workgroup and Enterprise Capabilities. Sintelix "projects" are a versatile mechanism for operating multiple workspaces and workgroups. They allow shareable workspaces to be isolated, backed up, restored

and exported. Projects have their own security settings and may contain many user-defined document collections, networks and configurations. Sintelix is equipped a fully configurable claims-based security capability.

Deployment. Sintelix deploys on either **Windows** or **UNIX** family operating systems including Linux and OS X. It is a client-server system; the client is your web browser and the server is written in **Java** and runs on the Java Virtual Machine (**JVM**). Sintelix can be accessed from any modern platform or via **APIs** and operates as easily in the **cloud** as **on premise**. **The minimum memory requirement is 2 GB and two cores**.

Conclusion

Sintelix is unparalleled in providing a complete platform that handles the end-to-end unstructured data needs of Defence, Intelligence, and Law Enforcement. Sintelix automates manual work and its highly performant capabilities can be combined to create new efficient approaches and high-productivity workflows. Commonly, vendors only provide fragments of the desired overall process with limited use cases causing continued challenges. Deploying Sintelix avoids the cost, delay and risk of integrating its internal components.

Where others create silos for single use cases and integrate multiple solutions from different providers, resulting in ongoing complexity and cost. Sintelix offers immediate value, use case independence, and open data import/export – so maximising value and return on investment.

Sintelix continues to rapidly grow – with customer requests directly influencing future developments. We delight in progressively improving Sintelix to accommodate clients evolving needs and provide steadily increasing return on investment.