

**Sinara's** Universal Data Extractor (UDE) is a highly configurable software application that allows both financial market data providers and data consumers to identify and extract key data items from a variety of data sources such as news stories, including LSE's RNS.

Although many sources (e.g. news stories) are making increasing use of tagging and metadata through schemes such as XML, XHTML, etc, most data items in such sources are still often in free text within the body of the document. UDE utilises advanced technology and intelligent algorithms to perform a deep level of data extraction from such content.

The UDE software processes data feeds and provides extracted content in a variety of forms. A workflow application included with UDE can be used to quickly verify extracted content.

UDE can be tailored to individual customer requirements, making it a useful tool for a range of data extraction requirements

**BENEFITS** **Time-saving/cost-effectiveness** - For customers whose original data sources start out as text, the task of identifying and extracting the key information from these sources is time-consuming and repetitive. Sinara have demonstrated that by deploying UDE, customers can save much of this effort and cost, which in turn allows them to offer faster or more complete data to their clients or trading systems.

**Accuracy** – the UDE system runs constantly, processing the most up-to-date announcements and providing accurate and instantaneous output from those. The UDE is capable of up to 95% successful automatic extraction. Any verification or correction can be done via an easy-to-use and configurable graphical front-end application that comes bundled with the UDE system.

**Flexibility** – the UDE can be configured to meet exact requirements on a per customer basis. Only data required by the customer will be extracted, removing the need to filter out any peripheral or unwanted content. UDE can be configured to find specific numerical values, dates and text strings. A comprehensive Data Dictionary enables UDE to identify different terms used to define the same type of data item (e.g. in a Company Profit & Loss Statement, the value for "Group Turnover" may be shown as Turnover, Sales, Revenue from Sales, etc).

**Standardisation** – all data extracted from content is converted into standard formats and stored in a database, allowing for easy comparison of like data items.

**INTEGRATION** The UDE architecture allows it to be easily integrated into a customer's existing systems. The UDE software runs on standard Microsoft technology including Windows and SQL Server.

**Input** – the UDE includes a number of input feed handlers that process content from a variety of sources such as XML-based feeds, streaming data feeds, flat files or PDF files. New input handlers can be produced for other formats as required. Content can either be submitted in batches of files or as a real-time feed.

Output – the UDE can provide output to either the UDE integrated Windows Workflow application, which allows for the viewing and optional correction of extraction results, or directly to external systems such as databases, csv files or spreadsheets.

## APPLICATION AREAS

UDE's flexibility makes it a useful tool for financial market data vendors seeking to improve their data capture operations, and also for hedge funds, traders or broker/dealers who want to enhance the performance of their automated trading algorithms.

### Financial Market Data Vendors

Many data vendors need to gather and process a wide variety of source data in order to produce their own data products. This source data can be delivered in a number of formats (e.g. text, flat files, data feeds, PDF documents, XML, etc). The task of identifying and extracting the key information from these sources can be both time-consuming and costly. UDE can be configured to extract specific data fields from the source data to produce the core data in a single format required by the vendor. The UDE output can then be exported into the vendor's own workflow systems to verify the content before passing into their own products.

Vendors who deploy UDE can save a lot of manual effort and cost, which in turn allows them to offer faster or more complete data to their clients.

### Trading companies

News-based content is an increasingly important part of trading strategies, which are often influenced by news stories such as results announcements. UDE can extract important data from these announcements quickly and accurately, to make key values (e.g. turnover, profit, dividend payments, etc) available to a client's trading algorithms.

## COMPANY BACKGROUND

Sinara is a London-based software house delivering innovative and reliable solutions to leading organisations in the financial and market data sectors.

We have built up a successful track record as a financial systems developer – working for banks, exchanges and financial institutions. Our speciality is in developing systems that are considered to be technically difficult – for example, real-time systems, complex database requirements, interfaces to other systems, etc. We also offer consultancy and development services in financial MIS/workflow systems, where we can combine our business and analysis skills with the ability to build browser-based systems linked to corporate or dedicated databases.

Sinara works with a wide range of both established and leading-edge technologies. Our choice of technology for a specific project will always be based upon the most appropriate and reliable tools that will meet our clients' needs. Our current skills include the following:

**Platforms:** Windows (2008/2010,XP,.NET); UNIX (Linux, Solaris)

**Web:** HTML/DHTML, XML, Ajax, ASP, ADO, IIS, Apache, Perl

**Languages:** C/C++/C#, Java, VB

**Databases:** SQL Server, Transact-SQL, Oracle, PL-SQL, OCI/ODBC/JDBC

**Communications:** TCP/IP , IP Multicast, FTP, HTTP, SMTP, SSL