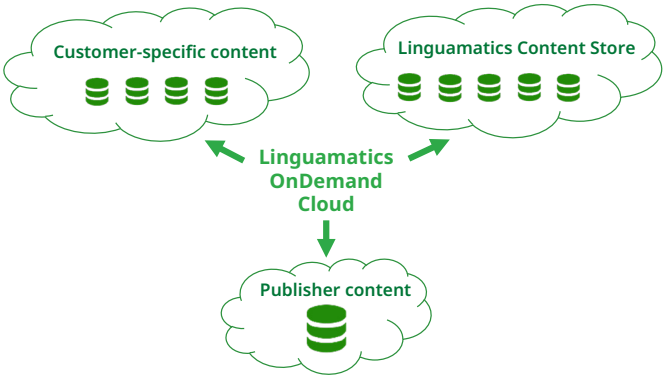
Cloud-Based Text Mining From Linguamatics OnDemand Cloud

Linguamatics OnDemand Cloud allows you to access the capabilities and benefits of our award-winning natural language processing (NLP)-based text mining platform via the cloud

The platform includes software, access to content and vocabularies/thesauri, all managed by Linguamatics. Accessing the platform via the cloud provides a cost-effective, easily accessible, high performance text mining system to rapidly extract facts and relationships from a wide range of content sources.

Figure 1: OnDemand Provides Access To A Wide Variety Of Content, Regardless Of Location



KEY FEATURES AND BENEFITS
Fast speed to insight—identify facts and relationships easily and quickly with patented NLP text mining software.
Fast ROI, get up and running quickly—no need to train internal staff to install and set up the software/hardware.
Low operating costs, reduced total cost of ownership (TCO).
24/7 access to up-to-date standard content sets, e.g. MEDLINE®, FDA Drug Labels, Patents, etc.
Reduce capital expenditure—no new hardware required.
Instant and secure service.
Flexibility to accommodate your own custom content, e.g. internal documents.
Easy integration with in-house applications and workflows.
Up-to-date vocabularies and thesauri.
Great technical support and timely software updates.

The options

There are different service levels of OnDemand available depending on your requirements for hardware, operational support, and integration with applications and workflows.

ONDEMAND

This service provides secure (authentication and confidentiality) access to the Linguamatics text mining platform via your web browser. It includes the option to select a range of ready-to-use content sources to text mine—MEDLINE® comes as the default with OnDemand. The content can be used in conjunction with a set of domain-specific vocabularies and thesauri, and a library of out-of-the-box powerful queries, to extract key information. There is flexibility for users to develop their own queries and vocabularies specifically for the projects they are working on. Linguamatics' expert technical staff are available for assistance in developing queries as well as any other technical support required. The Linguamatics operations team manages the installation and maintenance of the software and ensures the indexes are kept up to date.

ONDEMAND PLUS

This service includes everything in OnDemand, but adds the following:

- the ability to integrate applications or workflows through a RESTful Web Services application programming interface (WSAPI);
- user-specified content sets (e.g. internal data, content licensed from a third party, etc.);
- the ability to include your own custom ontologies to use with, and in addition to, the content sets (e.g. you may have created your own ontologies); and
- an extended range of RAM/processor/disk space options and support facilities.

Which option do I choose?

Here is a quick summary of each service provided. More information is provided in the table of comparison on the next page.

ONDEMAND

Text mining access to a standard set of off-the-shelf content sources.

ONDEMAND PLUS SILVER

Includes everything in OnDemand, plus the ability to integrate applications and workflows using the Web Services API. and to text mine a small amount of your own custom content.

ONDEMAND PLUS GOLD

Includes everything in Silver, with more powerful hardware and additional operational support.

ONDEMAND PLUS PLATINUM

Includes everything in Gold, with the ability to mine a larger amount of your own custom content and double the amount of operational support. At this level, you can purchase additional resources (RAM or disk space) to increase the power of the server.

Table of comparison for OnDemand and OnDemand Plus options

	ONDEMAND	ONDEMAND PLUS		
		SILVER	GOLD	PLATINUM
RAM (GB)	n/a	32	64	64
CUSTOM INDEX SPACE (TB)	n/a	1	2	4
PROCESSOR CORES	n/a	4	8	8
OPERATIONS SUPPORT (HOURS/WEEK)	n/a	1	2	4
TECHNICAL SUPPORT	Yes	Yes	Yes	Yes
STANDARD ONTOLOGIES	Yes	Yes	Yes	Yes
ACCESS TO STANDARD CONTENT • MEDLINE® • Other standard content (purchased separately)	Yes	Yes	Yes	Yes
API ACCESS	No	Yes	Yes	Yes
CUSTOM ONTOLOGIES	No	Yes	Yes	Yes
CUSTOM CONTENT AND CUSTOMER ACCESS TO INDEXING	No	Yes	Yes	Yes
INDEXING ASSISTANCE	No	Yes	Yes	Yes
CUSTOM SEARCH PORTAL	No	Yes	Yes	Yes

Description of features

RAM (GB)	Total RAM-based memory for the server. Relevant for running concurrent tasks.
CUSTOM INDEX SPACE (TB)	Disk space available to store source files and index data for customers' own indexes: This is separate to the disk space used by the standard OnDemand content sets (MEDLINE®, FDA Drug Labels, Patents, etc.).
PROCESSOR CORES	Number of processor cores available. Relevant for parallel querying and indexing to aid performance.
OPERATIONS SUPPORT (HOURS/WEEK)	Operations support, up to a specified number of hours per week, for system administration including configuration and indexing of custom content and ontologies.
TECHNICAL SUPPORT	Standard Linguamatics technical support.
STANDARD ONTOLOGIES	Access to Linguamatics' ready-to-use standard ontologies for OnDemand (please see the ontologies fact sheet or https://www.linguamatics.com/products/ontologies for more information).
STANDARD CONTENT (PURCHASED SEPARATELY)	Patents, Patent Abstracts, FDA Drug Labels, ClinicalTrials.gov, NIH Grants, PubMed Central Open Subset, OMIM (visit https://www.linguamatics.com/products/content-store for the most up-to-date list of content).
CUSTOM ONTOLOGIES	Customer-specific ontologies included when building custom indexes.
CUSTOM CONTENT AND CUSTOMER ACCESS TO INDEXING	Customers use their own source data to create additional indexes. Ability for users to create their own custom indexes, subject to Custom Index Space for the chosen OnDemand Plus option.
API ACCESS	External applications can make API calls to the server to programmatically drive Linguamatics workflows.
INDEXING ASSISTANCE	Assistance in creating indexes from Linguamatics' Operations Support team.
CUSTOM SEARCH PORTAL	Customers can use a simplified search interface customized to their requirements.

OnDemand content options

MEDLINE®

MEDLINE® contains journal citations and abstracts for biomedical literature from around the world. MEDLINE® is the US National Library of Medicine (NLM) premier bibliographic database, containing over 23 million references to journal articles in life science with a concentration on biomedicine.

FDA ADVERSE EVENTS REPORTING SYSTEM (FAERS)

FAERS is rich source of ready-to-use safety surveillance data to extract information about safety concerns

reported by users and clinicians. It is typically used to monitor and discover safety issues of a marketed drug product. The data is managed and maintained by the FDA, and made available for public download.

CLINICALTRIALS.GOV

ClinicalTrials.gov is a registry of federally and privately supported clinical trials conducted in the United States and around the world. It contains information about medical studies in human volunteers. Most of the records in ClinicalTrials.gov describe clinical trials (also called interventional studies). ClinicalTrials.gov also includes records describing observational studies and programs providing access to investigational drugs outside of clinical trials (expanded access). Studies listed in the database are conducted in all 50 US states, as well as over 210 countries.

FDA DRUG LABELS

The data in this index is sourced from the DailyMed site hosted by the US Food and Drug Administration (FDA). It contains high quality information about marketed drugs. This information includes up-to-date and accurate FDA drug labels (package inserts) that describe the composition, form, packaging and other properties of the drug products in detail.

NIH GRANTS

NIH Grants provides data on research projects funded by the US National Institutes of Health (NIH), the Centers for Disease Control (CDC), the FDA and the Department of Veterans Affairs (VA), their abstracts, and publications and patents citing support from these projects. The data are separated into four major categories of files: Projects, Project Abstracts, Publications and Patents. The data is sourced from ExPORTER site (owned by NIH).

PUBMED CENTRAL OPEN SUBSET

This is a subset of the total collection of articles in PubMed Central (PMC) made available for text mining. PMC is an archive of biomedical and life science journal literature at the NLM. PMC Open Subset is an electronic archive of full-text journal articles under the Creative Commons License, offering more liberal redistribution and re-use of the content than traditional copyrighted work (includes permission to text mine the content).

FULL-TEXT ARTICLES

Access to full-text journal articles via Copyright Clearance Center's RightFind XML for Mining. This integrated solution provides the ability to mine the full text of the article rather than the abstract, ensuring that you don't miss vital data, discoveries and assertions that are only published in the full text.

PATENTS

This includes a complete set of full-text patents from USPTO, EPO and WIPO. The data is provided with a uniform structure to allow consistent searching across all sources, regardless of their origin. The indexes are organized as a complete set or individual authority, or subdivided into era (last year, last five years, last 20 years, etc.).

PATENTS ABSTRACTS

This includes a complete set of patent abstracts (and additional citation information) from all patent agencies. Again, the data is provided with a uniform structure to allow consistent searching across all sources, regardless of their origin, and the indexes are organized as a complete set or subdivided into era (last year, last 5 years, last 20 years etc.).

OMIM

Online Mendelian Inheritance in Man® (OMIM) is a comprehensive catalogue of human genes, and genetic conditions and traits, with particular focus on the molecular relationship between genetic variation and phenotypic expression. Curated at Johns Hopkins University, OMIM has data on over 15,000 genes and 5,000 phenotypes, and provides a powerful resource for mining genotype-phenotype relationships, for target identification, personalized medicine and pharmacogenomics.

CONTACT US

+44 (0)1223 651 910 (U.K.) | +1 617 674 3256 (U.S.)

nlp@iqvia.com

linguamatics.com