


Admin UI Documentation

- [Introduction](#)
- [Overview](#)
 - [Organisations](#)
 - [Knowledge Bases](#)
 - [Adding a knowledge base](#)
 - [Logs](#)
 - [Services](#)
- [User Management](#)
- [Knowledge Base Drop Down](#)
- [Sources](#)
 - [Source Schedule](#)
 - [Source operations](#)
- [The Mind](#)
 - [AI set up](#)
 - [Synonyms](#)
 - [Semantics](#)
 - [Language Changes](#)

Introduction

In this document, we will discuss areas of the admin interface, and explain in detail the functionality of each area. This document should serve as a *guide* to those that are unfamiliar with SimSage.

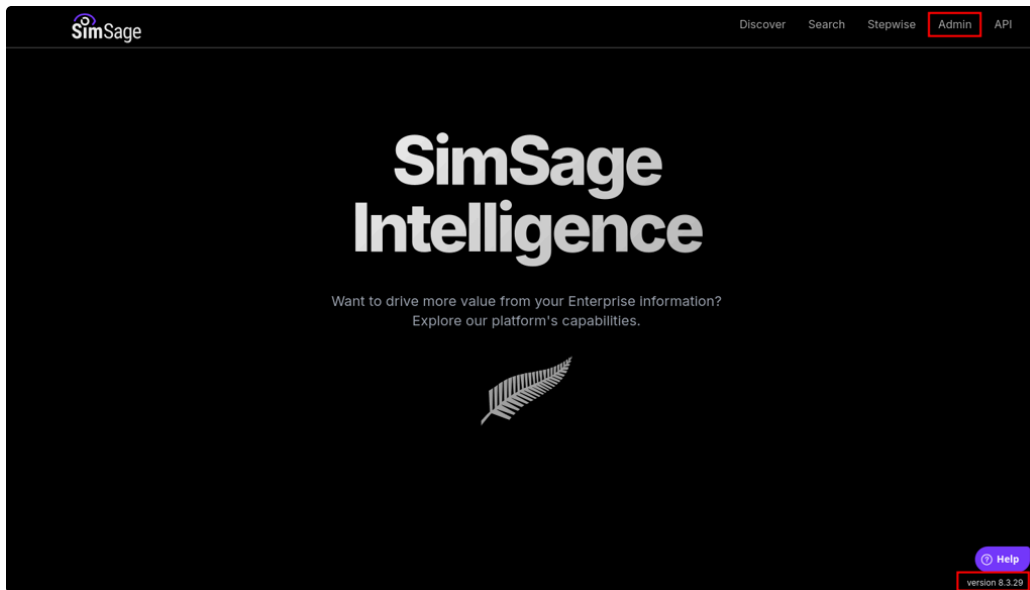
 This document will not provide a *technical understanding* of SimSage. See this document as a walk-through of SimSage.

This is a guide to using the SimSage Admin UI.

Your Admin UI URL will be provided to you by SimSage. This URL is usually

<https://<your-company>.simsage.ai/>

Once opened you should see a sign-in screen as shown below. It is possible that your SimSage instance runs “on prem” (i.e., inside your organisation), or is ip-white-listed, and has access restrictions. Consult your organisation about the right way to access your instance.



SimSage instance landing screen

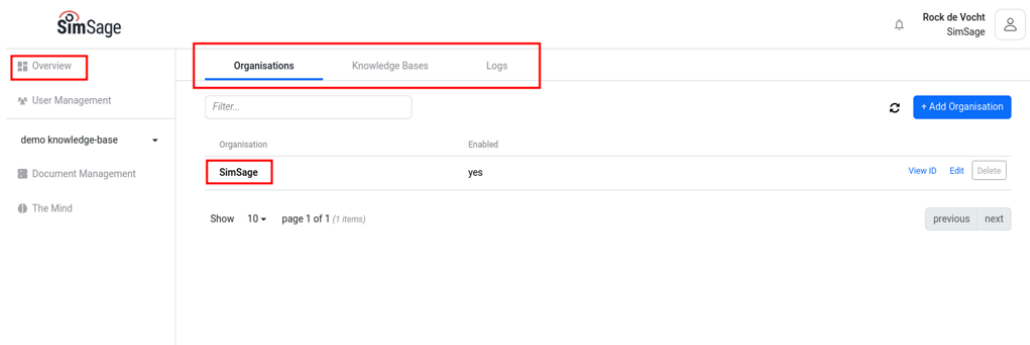
This page shows the available products in the top right part of the page. A “Help” widget is shown below that connects to the SimSage helpdesk. A version number is also shown, informing you what version of the platform you are using. This guide was written starting for version 8.3 of SimSage.

Once you click the Admin link as shown, you will be redirected to your Single Sign On provider. Your sign in experience should reflect that of your company.

SimSage by default is set up to use Single Sign On. This means you should be able to click the “Sign in” button shown and be redirected to an appropriate identity provider. This provided will be set up by SimSage for you. Once you’ve verified your identity and have sufficient access privileges you should see the first screen.

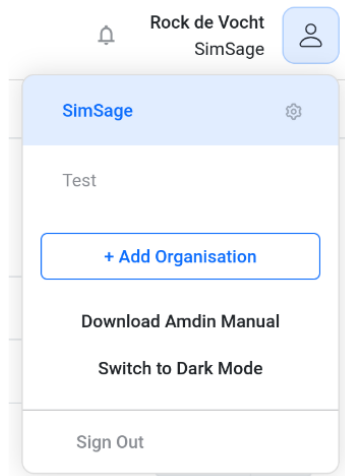
Overview

When you first *login* to SimSage, you will be greeted with the following page:



Overview, showing the selected organisation to be “SimSage”

The *far top right* of the webpage shows a *prompt*, which displays who is currently logged into SimSage (in this case 'Rock de Vocht'). By clicking the user icon, you will invoke a drop down menu - giving options to *logout*, *add/select organizations*, or *open the settings of a organization* (indicated by the cog shown below):



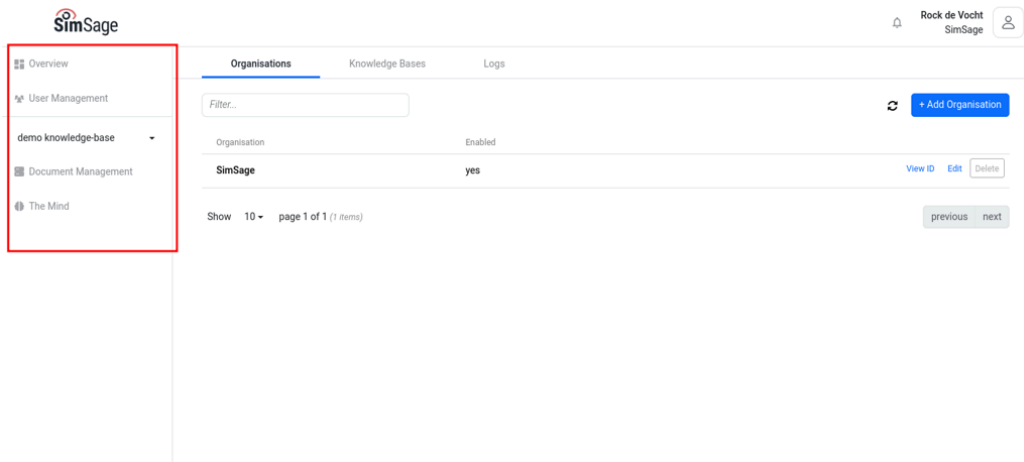
Organisations

Organisations are the top level multi-tenant containers for data. We designed SimSage to be massively scalable. However, on normal instances, we advise the use of one and one only organisation. Our example organisation here is called “SimSage”.

Knowledge-bases are containers for knowledge, in that they group different sources together, and present all information in a single pool of knowledge. Each organisation can have multiple knowledge-bases. The accounting department might have a different knowledge-base from the Human Resources department.

Currently, we are in the 'Overview' page, as indicated by the *contrasting* coloring seen in this vertical navigation bar. Additionally, we can see that there are two other options:

1. User Management (Button)
2. Select Knowledge Base (vertical drop-down menu)



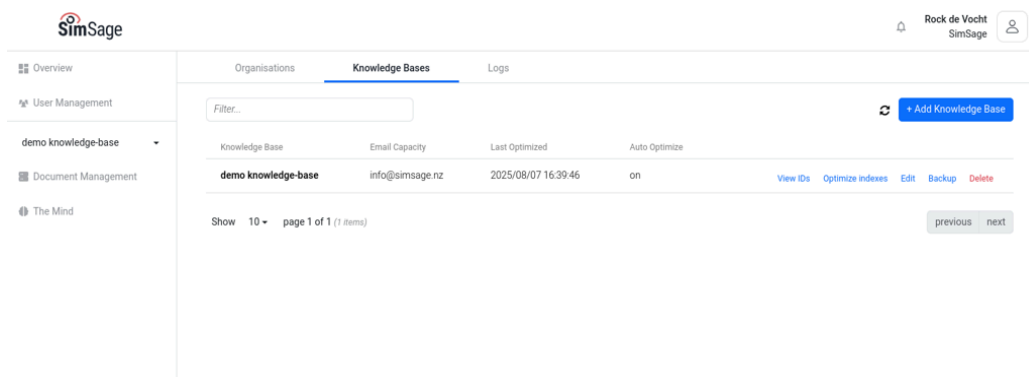
the main navigation menu, with the knowledge-base selector drop-down

'User Management' will show a pane, that defines the permissions that a given user has on our platform, these permissions can be *edited* if you have the right to do so, among other functionality like updating credentials, resetting the password and deleting a user if necessary.

The *knowledge base drop-down* shown above is where you can navigate from one knowledge base to another.

Knowledge Bases

Finally, the main stage of the webpage shows three main tabs to select from, by default, 'Organisations' is selected. Selecting the knowledge-base tab enables us to create new knowledge-bases or select the active knowledge-base to operate on.



Knowledge-bases view

Each knowledge-base has a series of operations associated with it.

1. View IDs - brings up a pop-up dialog showing all the IDs associated with your knowledge-base for remote communications
2. Optimize Indexes - shows a pop-up dialog for manually overriding the index optimization process

3. Edit - pops-up the edit knowledge-base dialog where you can change aspects of your knowledge-base
4. Backup - pops-up a backup menu for creating partial back ups of your knowledge-base
5. Delete - pops-up a confirmation dialog if you wish to delete your knowledge-base. You won't be able to delete a knowledge-base that has any active sources in it. You will need to delete the sources of your knowledge-base first using Document Management of the left-hand-side menu.

The current tab shows all the knowledge bases available to you, supplemented by a number of options.

i By default, most pages will show a maximum of 10 entity instances at a time - though a drop down is available to change this.

As shown, you have options to *filter* through KBs, add them, delete and truncate. We can also see a **filter...** textbox, which you guessed it - allows you to filter through KBs if many are available.

To the right of this we can see the the *refresh button* and immediately next to this the '+ Add Knowledge base' option.

The refresh button will reload all knowledge bases present on screen, including new instances that perhaps were not visible initially.

Adding a knowledge base

If you wish to create a new knowledge base, navigate to '+ Add Knowledge Base' and select his option, you will be presented with the following:

New Knowledge Base

General Index Optimization Schedule

Name

Email Queries

Security ID

Max number of queries (per day) (0 = no limits)

Max analytics retention period (months) (0 = no limits)

☒ Knowledge Base

☐ Capacity Warnings

Cancel Save

Add Knowledge Base dialog

The properties shown above set the *parameters* of a given knowledge base.

i A name and email *must* be provided to create a valid knowledge base

Logs

This page shows all logs made on services provided my SimSage. These are technical system logs that show the operation of all the components that make up SimSage. But first, lets clarify what we mean by a *service*.

Knowledge Bases **Logs** Organisations

All services ▼ Logs This hour Last 2 hours Last 12 hours Last 24 hours Type All Info Debug Error Warn Refresh off 5 seconds 10 seconds

Debug KB 2023/11/17 12:09:45
knowledge base service: GetCrawlerListCmd(org=018bc8f5-2f80-58c3-2a38-43f7033a1624,kb=018bd8da-ec01-4d56-b044-49c673933415,crawlerList=1), took 76 ms

Debug Auth 2023/11/17 12:09:45
auth service: GetUsersCmd(userList.size=15,filter=""), took 40 ms

Debug KB 2023/11/17 12:09:46
knowledge base service: GetCrawlerListCmd(org=018bc8f5-2f80-58c3-2a38-43f7033a1624,kb=018bd8da-ec01-4d56-b044-49c673933415,crawlerList=1), took 83 ms

Debug Auth 2023/11/17 12:09:46
auth service: GetUsersCmd(userList.size=15,filter=""), took 36 ms

Debug Auth 2023/11/17 12:09:47
auth service: GetUsersCmd(userList.size=15,filter=""), took 40 ms

Debug KB 2023/11/17 12:09:47
knowledge base service: GetCrawlerListCmd(org=018bc8f5-2f80-58c3-2a38-43f7033a1624,kb=018bd8da-ec01-4d56-b044-49c673933415,crawlerList=1), took 186 ms

Debug KB 2023/11/17 12:09:48
knowledge base service: GetCrawlerListCmd(org=018bc8f5-2f80-58c3-2a38-43f7033a1624,kb=018bd8da-ec01-4d56-b044-49c673933415,crawlerList=1), took 85 ms

Debug Auth 2023/11/17 12:09:48
auth service: GetUsersCmd(userList.size=15,filter=""), took 40 ms

Debug KB 2023/11/17 12:09:49
knowledge base service: GetCrawlerListCmd(org=018bc8f5-2f80-58c3-2a38-43f7033a1624,kb=018bd8da-ec01-4d56-b044-49c673933415,crawlerList=1), took 75 ms

Debug Auth 2023/11/17 12:09:49
auth service: GetUsersCmd(userList.size=15,filter=""), took 41 ms

Debug KB 2023/11/17 12:09:50
knowledge base service: GetCrawlerListCmd(org=018bc8f5-2f80-58c3-2a38-43f7033a1624,kb=018bd8da-ec01-4d56-b044-49c673933415,crawlerList=1), took 85 ms

Services

SimSage's functionality is split into what we call services. SimSage services serve different purposes. We have services that communicate and copy data from outside systems (Crawlers), we have services that process text and mark-up parts of speech and determine context (Language). Here is a list of the services and their general purpose.

Service	Purpose
Authentication	Handles accounts in SimSage, including logging in via MS accounts/Google
Conversion	Document conversion node, takes PDFs, DOCX, DOC files and extracts text from these.

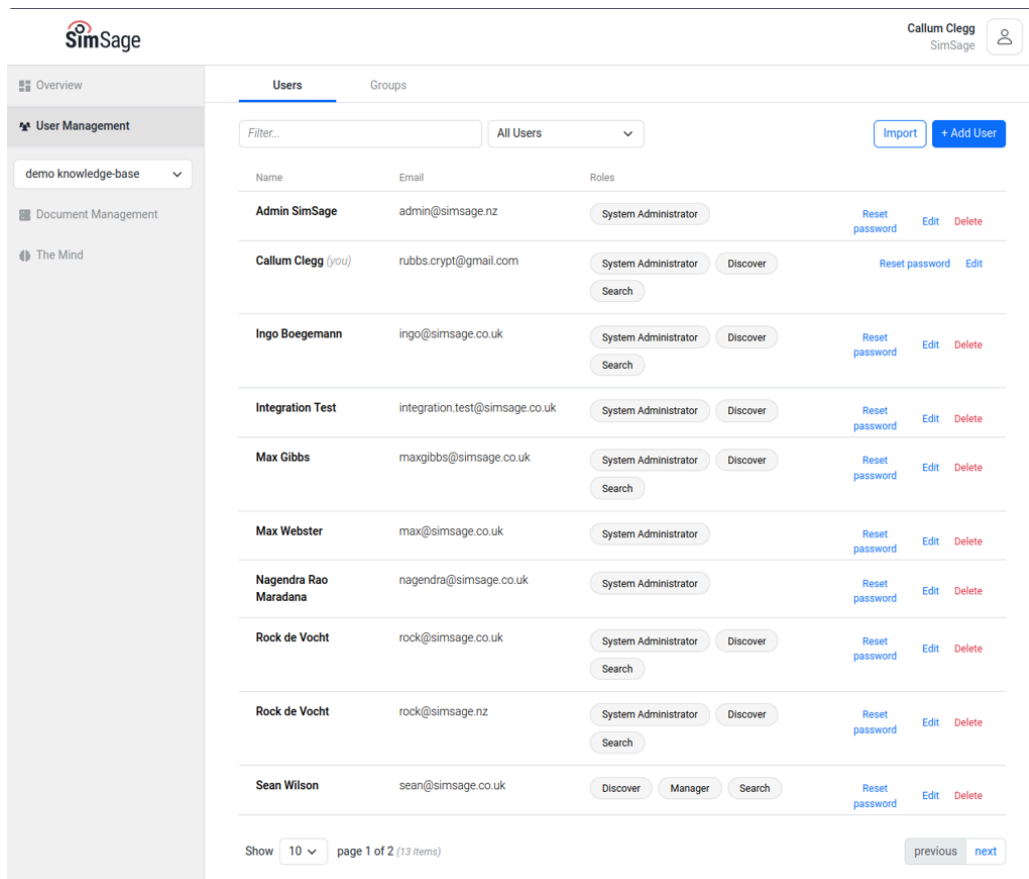
Crawler	Handles crawling of data across cloud services, manage external crawlers and data ingestion.
Document Service	SimSage DMS (not used currently) and Image Conversion services
Knowledge base service	Knowledge-base operations
Language	Natural Language Processing, convert text into marked-up text for searching
Index	Index creation, creates inverted semantic indexes, created by the Language service.
Search	SimSage Search engine
Stats	Statistics gathering and coalating
Discover	Discover services
StepWise	StepWise workflow operations and UI management
Web facade	RESTful web layer for external communications and API access

In the *Logs* page, there is a number of different way to *filter* information to get specifics that you may be searching for. These options include the following:

- By Service
- By Date (2/12/24 hours)
- Type - i.e., info, debug, warn, error
- Refresh rate - off, 5 or 10 seconds

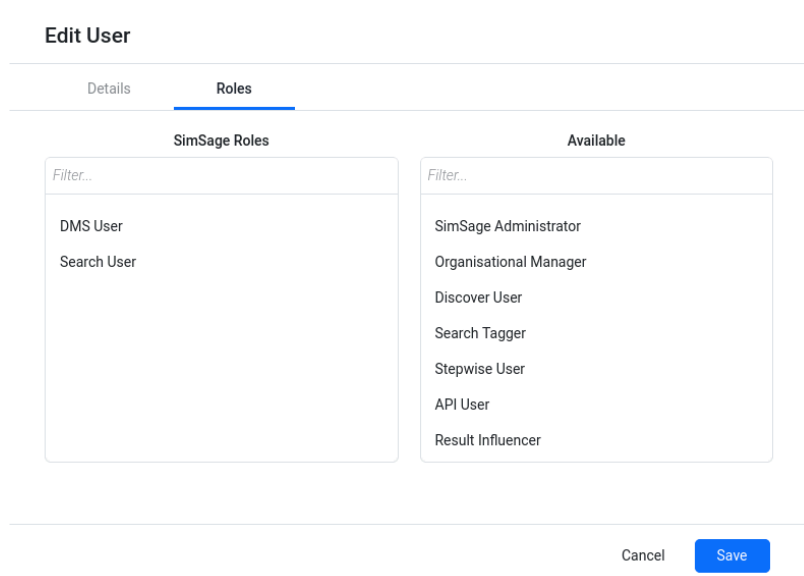
User Management

Managing users and groups is straightforward using the User Management menu, located on the left hand-side of the UI.



User Management UI

Users consist of SimSage roles (what they are allowed to do inside SimSage) and groups (what they can see when searching). Groups can be manually created, and are automatically created by crawlers when they see security groups associated with remote documents / records.



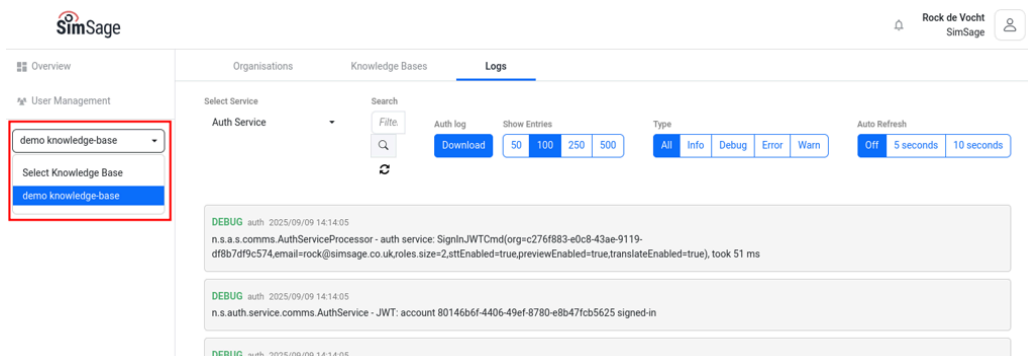
The Edit User dialog

The standard SimSage roles are

Role	Description
DMS User	Access/log in to SimSage DMS, currently not used
Search User	Access/log in to SimSage Search
SimSage Administrator	Access/log in to SimSage admin
Organisation Manager	Access/log in to SimSage admin for one and one organisation only
Discover User	Access/log in to SimSage Discover
Stepwise User	Access/log in to SimSage Stepwise workflow
API User	User account can be used to make API calls for programmatic access
Search Tagger	User can add #tag to search results in the search UI
Result Influencer	User can up-vote / down-vote search results in the search UI for a given set of keywords

Knowledge Base Drop Down

You must select a knowledge-base in order to set up sources, add synonyms, set up AI etc. A knowledge-base is where it all happens in SimSage. You can have multiple knowledge-bases. Our Search UI, by default, points to one of your knowledge-bases (and can switch between if there are multiple). Discover, and StepWise too can switch between knowledge-bases.

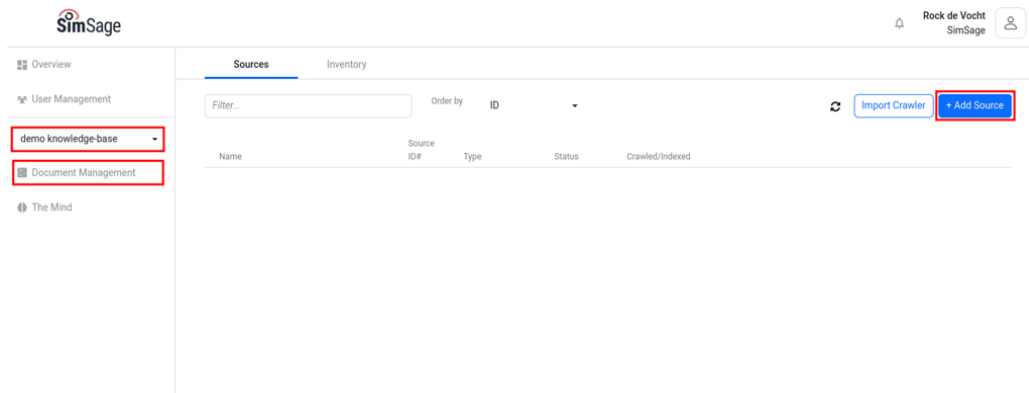


select your knowledge-base first

Sources

Sources are connections to remote systems. SimSage can connect to many different types of remote systems. To mention a few: DropBox, MS SharePoint, MS OneDrive, MS Exchange, Microsoft file-shares, Google-drive, Box, Confluence, Web sites, Jira, etc.

Each remote system has its own unique requirements for set up. SimSage provides documentation for each through the admin UI.



Knowledge-base selected, Document Management selected, then +Add Source

We will briefly show what is involved in setting up a source here.

The 'Add Source' form is displayed with several tabs: General, Metadata, ACLs, MimeTypes, Rules, Check Document, and Schedule. The 'General' tab is active. It contains a 'Crawler Type' dropdown menu (highlighted with a red box) with a list of options including Alfresco Crawler, ArcGIS Crawler, AWS Crawler, Box Crawler, Confluence Crawler, Database Crawler, and Discourse Crawler. Next to it is the 'Crawler Name' field (highlighted with a red box) containing the text 'HR files'. Below these are several input fields: 'Number of files' (set to 1), 'Source Weight' (set to 1), 'crawler pod id (0, 1, 2, ...)' (set to 0), and 'Expected File Count (0 = any)' (set to 0). At the bottom, there are several checkboxes for configuration: 'Remove un-indexed files' (checked), 'Use default built-in relationships' (checked), 'Use speech-to-text (videos, audio transcripts)' (unchecked), 'Newest Documents first' (unchecked), 'Allow anonymous access to these files' (checked), 'Store the binaries of each document' (checked), 'External source' (unchecked), 'Translate foreign languages' (unchecked), 'Enable document image previews' (unchecked), 'Use optical-character-recognition (OCR)' (unchecked), 'Transmit external logs' (unchecked), and 'Only Index Document Metadata' (unchecked). The form has 'Close' and 'Save' buttons at the bottom right.

adding a new source

A source is of a single type, you can only set the type once before saving the source after which the type of a source can no longer be changed (but you can delete a source if you created it in error).

The source also needs a name (Crawler Name). This name is for you to identify what it is you're connecting to.

There are a series of TABS along the top of this pane. These are General, Metadata, ACLs, MimeTypes, Check Document, and Schedule.

There will be at least one more tab added after the General tab once you select a source-type. This panel contains the specific information needed to set up that source.

There are two types of sources in SimSage. Internal sources, and external sources. Internal sources run from the SimSage platform and assume they can connect to the source you're setting up. For instance, Dropbox, would be such a source. Dropbox is available from over the Internet, and with the right network set-up, SimSage can connect to Dropbox and crawl its data.

Here is what the different settings mean

setting	meaning
Remove Un-seen files	Most remote systems (e.g. Microsoft file-shares) will not tell you when a file has been removed. One way of removing files after looking through a whole set, is to mark each file as having been "seen". When we look through the filesystem, we call that a "run" through the files. At the end of a run, any files that haven't been seen during a run, have presumably been deleted. This only works for remote systems that need us to look at all files each time. Other systems, such as Sharepoint 365, will actually tell us what has been removed, and what has changed. You can leave this switched on in those cases, as this option is then ignored.
Use default built-in relationships	SimSage contains around 2.1 million default relationships / synonyms between words. Switching this off, ignores these built-in relationships, and solely uses the user-defined relationships set in "the Mind". Only switch this option off, if you really know what you're doing.
use speech-to-text (video, audio transcripts)	SimSage comes with a built in transcription service. This service will convert English video and audio files into a standard format and push them through a neural network to generate

	transcribed text. This service is much slower than document conversion services due to using a neural network. Video and audio files therefore go into their own queue.
Newest Documents first	If switched on, this source will sort its documents in any results by date first, instead of relevance first.
Allow anonymous access to these files	A switch for testing or for sources that don't have any security (e.g. web-sites). This switches off all ACL checking for the given source. This is handy for testing (as it can just as easily be switched back).
Store the binaries of each document	Some parts of SimSage (e.g. the Stepwise workflow system) require that the original documents remain in their original form on the platform for later processing/redaction etc. This is default "on" at present. This is also (presently) required for re-processing all documents. However, re-processing all documents will not require this come version 8.4 of SimSage (yet to be released). Leave this switch on for now.
Enable document image previews	Small preview images are shown on the left hand-side of search results where possible. The first pages of documents are fetched as part of this process and turned into small JPEG preview images, aiding visual recognition of your documents.
use optical-character-recognition (OCR)	OCR is the process of taking images, and scanning them for text. Any office document is scanned for text using Tesseract OCR if this switch is enabled. This will slow down your document processing significantly. Any text found in these images is then included in the body content of that document.

External Source	External sources are crawlers that run on different machines, outside the SimSage cluster and communicate with SimSage. External crawlers are needed inside corporate environments with vulnerable systems (e.g. Microsoft fileshares). Such systems cannot be opened to the Internet for SimSage to access. An external crawler is a small program that picks up files, and securely sends them to SimSage for processing. Switching this option on will switch this source to "listen for external crawler communications".
Transmit external logs	Only used if External Source is switched on. This tells the external crawler to send its internal logs to SimSage. This is handy for seeing what is happening remotely without having to view logs on the remote system first.
Translate foreign languages	Only available if enabled on the platform. Use a neural network to provide to English translations for a variety of languages.
Only Index Document Metadata	If enabled, the main body content of documents and records is ignored and not processed, only metadata (such as title, author, created-date etc.) are processed.

External sources are “external” to SimSage. We provide crawlers written in different technologies to run on different servers/machines/computers/laptops to fetch data from a protected system, or a computer system inside an Enterprise/Organisation, and securely pass this information to your SimSage instance.

By default, all sources are Internal. There is a switch in the General tab for setting the type of source.

Add Source

General **Dropbox Crawler** Metadata ACLs MimeTypes Check Document Schedule

Client Id (aka. app key) *

Client Secret *

Dropbox oids key
edac4b0b-641a-3781-7566-8c66fcaafb48

Dropbox Redirect URI
https://demo.simsage.ai/api/crawler/dropbox-oids-code/
edac4b0b-641a-3781-7566-8c66fcaafb48
Copy

Start folder

(separate folders by comma)

- Each folder must be part of the root folder and not contain any sub-folders.
- Each folder name must start with '/'.

[Dropbox Setup Guide](#)

Close Save

Example Dropbox TAB showing the set-up guide for Dropbox and its details

Source Schedule

Each source has a schedule of operations. Its schedule exists of a 24x7 window where you can set when a crawler (the thing that runs and connects to the data set up in a source) is allowed to talk to the remote system (e.g., Dropbox).

Add Source

General Dropbox Crawler Metadata ACLs MimeTypes Check Document **Schedule**

☒ Schedule Enabled

All times in GMT (now Tue, 14:42)

	00:00	01:00	02:00	03:00	04:00	05:00	06:00	07:00	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00	
Mon																									
Tue																									
Wed																									
Thu																									
Fri																									
Sat																									
Sun																									

Clear All Select All

Close Save

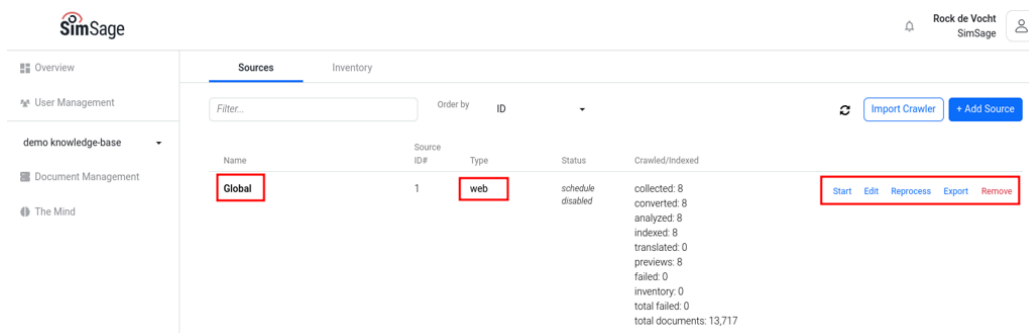
Enable / Disable, and click cells or times, or days of the week to set your source's schedule

A crawler starts a new “run” when it arrives at the next available window in the 24x7 grid. The crawler will keep running until it finishes its run through all the data. The crawler will pause if it reaches a time-slot where it isn't allowed to run, and resume at the next available window. The crawler will do this until it has finished processing all there is to process.

Once finished, the crawler will not start again, until it sees a slot where it isn't allowed to run, followed by a slot where it is allowed to run.

A full schedule (i.e, all 24x7 slots selected) results in a crawler running once every week, where the week starts midnight Sunday going into Monday morning. At present the crawler 24x7 grid is in UTC time and the "time now" is displayed in UTC at the top of the schedule grid.

Source operations



The screenshot shows the SimSage interface. On the left is a sidebar with navigation links: Overview, User Management, demo knowledge-base (selected), Document Management, and The Mind. The main area is titled 'Sources' and contains a table. The table has columns: Name, Source ID#, Type, Status, and Crawled/Indexed. A single row is visible with Name 'Global', Source ID# '1', Type 'web', Status 'schedule disabled', and a list of statistics: collected: 8, converted: 8, analyzed: 8, indexed: 8, translated: 0, previews: 8, failed: 0, inventory: 0, total failed: 0, total documents: 13,717. To the right of the table is a row of action buttons: Start, Edit, Reprocess, Export, and Remove. Above the table is a filter input, an 'Order by' dropdown set to 'ID', and two buttons: 'Import Crawler' and '+ Add Source'.

Name	Source ID#	Type	Status	Crawled/Indexed
Global	1	web	schedule disabled	collected: 8 converted: 8 analyzed: 8 indexed: 8 translated: 0 previews: 8 failed: 0 inventory: 0 total failed: 0 total documents: 13,717

Source details and operations

The example above shows a source called "Global" (the Crawler Name) of Type "web" (an intranet or internet web crawler), and a series of actions.

1. Start - the the source if it isn't already running. You will get an error if it is already running. This is to start the source again if it has finished. The source will not start if the schedule is off for the time when you click this button.
2. Edit - bring back the edit pane to change details of your source. Be careful, some details, such as locations of files can cause existing files to be removed (because the can "disappear" by changing such details)
3. Reprocess - a handy utility that forces SimSage to look at each stored file again without using the crawler, and to put each file through the SimSage processing pipeline again. This is handy for when you've changed synonyms and other language customizations.
4. Export - create a small JSON export of the source for safe storage and re-importing to another SimSage platform, or this platform at a later date. Passwords will not be exported!
5. Remove - destroy this source and all its files. This will ask you first using a pop-up if you are sure.

The Mind

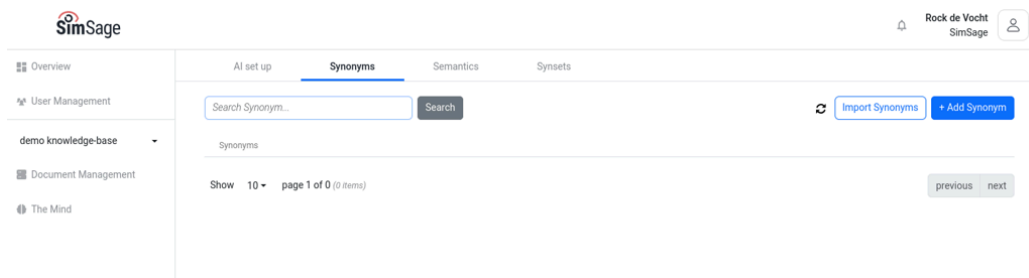
In this area of the documentation, we will explore *the mind*. The mind, mysterious as it sounds, is where you set up LLM connections (e.g., OpenAI, Google Gemini, Ollama), and customize language. We will discuss language customization next.

AI set up

The first section in the mind is the AI set up. Here you can configure LLMs to work with SimSage. Please refer to [SimSage AI settings](#) (SimSage AI settings).

Synonyms

Synonyms are words that are “equivalent” to each other. Synonyms don’t have to be “synonyms” in the strict linguistic sense. They are generally words you’d like to associate that help you find results faster and easier. Linguistic concepts such as Meronyms (part relationships, like “wheel” is part of a “cart”), and other associations are more than welcome here.



The mind, Synonyms

You can import your own synonyms. SimSage contains around 2.5 Million predefined relationships. These are relationships of general every-day English. By importing your own synonyms, you override, and improve SimSage to help with your own unique Organisational language.

You can mass import synonyms using the “Import Synonyms” button. This takes a simple text file of the following format:

- 1: word 1, word 2, word 3, word-4, 12th word
- 2: Greece, surgeon, pokemon, War and Peace

An ID number (starting at “1”) to uniquely identify this synonym, followed by a colon, and a list of words separated by commas. One set per line.

Synonyms in this list this **CANNOT INCLUDE** commas, nor colons.

Import Synonyms

Text

```
1: word 1, word 2
2: some other word, word 3
|
```

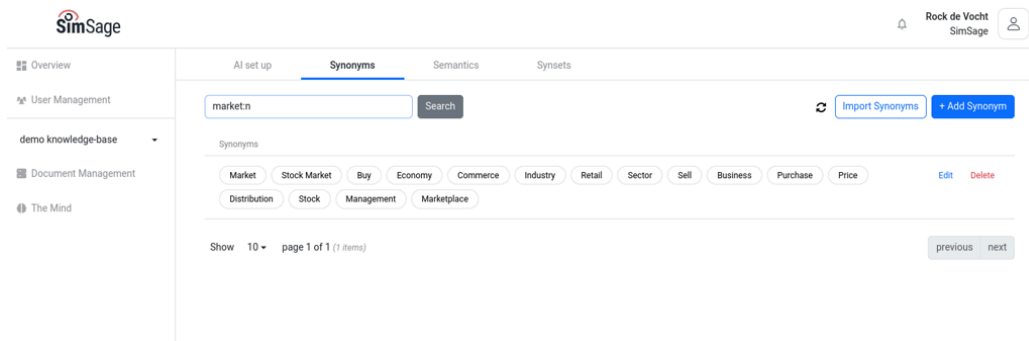
- Synonyms are comma separated
- One set of synonyms per line

Cancel

Import

Example Import Dialog for synonyms

User synonyms are all considered nouns (things), as opposed to verbs or adjectives. The system already has a rich set of adjectives and verbs for relationships. You can see what the system has defined for a synonym by typing it in the “Search Synonym...” Search section. Suppose you want to see what is defined for the noun “market”, search for “market:n”. The verb “market”, search for “market:v”, the adjective/adverb “market”, search for “market:a”. Or all combined without any part of speech type filtering, search for “market” just by itself.



“market” the noun, as defined inside SimSage

Semantics

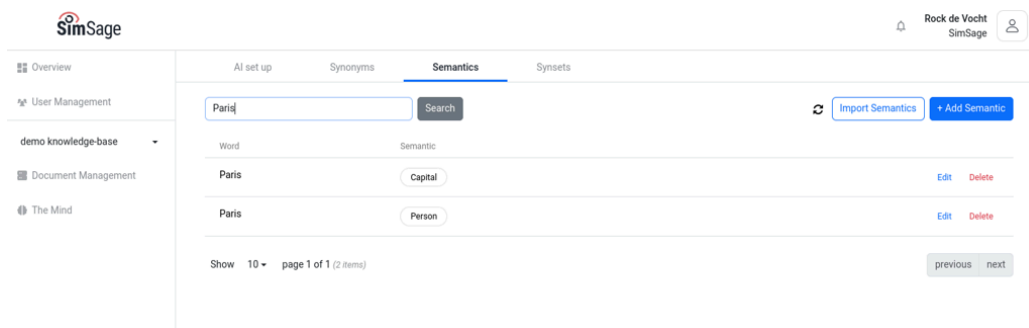
Semantics are generalisations. A semantic is an “is a” relationship. For instance “Toyota” is a car, “Jeep” is a car. This is completely arbitrary and debatable but still useful. Defining semantics enables you to use the general (“car” in our example) to look for all the specifics in one go.

Given the following imported semantics

toyota: car
jeep: car
mercedes: car

You can use “entity: car” in the SimSage search system to find instances of “Toyota, Jeep, and Mercedes”. Unlike synonyms, semantics ARE case sensitive. Leaving items lower case gives the best match results, however, if case sensitivity is important, you can mix case. In this case only matching case will work for semantics.

There is a large set of predefined semantics in SimSage. This includes a large set of English names, large cities of the world, capitals of the world, countries of the world, continents, a few brands and law-firms. All in all around 4.5 Million semantics are predefined.



“Paris” as both a person and a Capital

You can mass import your own semantics using the “Import Semantics” dialog. The semantic itself is its own ID, so you don’t need ID numbers for semantics.

Import Synonyms

Text

```
jeep: car  
toyota: car  
mercedes: car
```

- Semantics consist of a word colon semantic (e.g., Mark: person)
- One semantic per line

Cancel

Import

Language Changes

As mentioned above, any change to synonyms, or semantics requires reprocessing of all data. It is therefore a good idea to have all this information set up, up front. We realize however that this isn't always possible. So we've provided the tools to do both.