Structured Analytics – Language identification Setup Basics

This guide contains a workflow with basic information for setting up language identification to find foreign language content in your documents.

For more detailed information, refer to the documentation site.

Language identification setup

The setup for running language identification is comprised of three components:

1. Saved search
2. Structured analytics set
3. View

1. Saved Search Setup

Use the following conditions and fields to create the saved search used for language identification. You don't need to set a sort order on this search.

Search Name

There is no recommendation for the saved search name. Follow your team's normal protocol for naming searches.

Conditions

Extracted text size is greater than 0KB.

Fields

Any fields are acceptable.

2.Structured Analytics Set

Here are the steps and choices for creating a structured analytics set.

Structured Analytics Set Information

- **Name** - enter a name for the structured analytics set.
- **Prefix** - keep the default prefix or add your own prefix. Shorter prefixes (even just two characters, such as "LI") take up less space in your views.
- **Operations to run** - select **Language identification**.
- **Data source** - select the saved search you created above.

3. Language ID view

Once you run the structured analytics set, create the following view to see the results of your language identification operation.

Fields

- Control Number
- [SAS Prefix]::Primary Language
- [SAS Prefix]::Docs_Languages
- [SAS Prefix]::Docs_Languages:Language

Conditions

[SAS Prefix]::Primary Language is set

Sort

You don't need a specific sort order for this view. However, you can create separate searches and views for each document set as well as widgets and dashboards.

#				Control Number	Primary language	Docs_Languages	Docs_Languages::Language
				Filter	Filter	Filter	Filter
15	☐	Edit	📄	REL000013	Arabic	Arabic - 96%	Arabic
						English - 1%	English
						Other - 1%	Other
						Portuguese - 2%	Portuguese
16	☐	Edit	📄	REL000014	Hindi	Hindi - 100%	Hindi
17	☐	Edit	🖼	REL000014-0001	Other	Other - 100%	Other
18	☐	Edit	📄	REL000015	Chinese	Chinese - 97%	Chinese
						Chinese Traditional - 1%	Chinese Traditional
						Other - 2%	Other

Languages · + No Related Items · Add Widget · Document List · Items 1 - 100 (of 277) 100 ▼ per page