Analytics Index Quick Reference Guide

This guide provides step-by-step instructions for creating an Analytics conceptual index in order to run conceptual analytics tools (concept searching, find similar documents, categorization, clustering, and keyword expansion).

In addition, this guide explains the different stages of index creation.

Creating an Analytics conceptual index

1. Create a saved search to serve as your data source. The data source is the collection of documents on which you want to perform any conceptual analytics operation. This search should only pull back authored content fields (ex. extracted text) and typically needs no additional conditions.
2. Click the **Indexing & Analytics** tab and select **Analytics Indexes**.
3. Click **New Analytics Index**. The Analytics Index Information form appears.
4. Complete the following fields on the Analytics Index Information form.
    a. **Name** – enter a name for the index.
    b. **Index type** – select **Conceptual**.
    c. **Data source** – select the saved search you created in Step 1. The Training data source field automatically populates with this same saved search.
    d. **Order** – enter an order for the index. The order determines the relative position of the index in the search drop-down along with other search providers such as dtSearch and keyword search.

    Leave all other fields under Advanced Settings as default.

Note: For a complete description of the Analytics index fields, see the Relativity documentation site.

---

5. Click **Save**. The index console will appear.
6. Click **Run**.
7. If this is your first time running the index, it will automatically build the full index. Otherwise, select **Full**.

Index stages

Index creation consists of three stages: Population, Building, and Activating. The following steps occur during each stage:

Population

- All documents from the data source and training data source are staged and flagged for pre-processing.
- Document pre-processing occurs to clean up text in the following ways:
    - Numbers and symbols are ignored.
    - All words are made lowercase.
    - Filters found under Advanced Settings are applied (ex. email header filter).
    - Repeated content filters are applied.

Building

- Training data source documents and Latent Semantic Indexing (LSI) are used to build the concept space based on the relationships between words and documents.
- Data source documents are mapped into the concept space.
- Concept stop words (very common words) are filtered from the index to improve quality.

Activating

- Makes the index active and available in the search indexes drop-down.
- Saves the index to RAM which loads the index into memory. If you find yourself running out of free RAM on the Analytics server, deactivate your index.
    - Analytics indexes are automatically deactivated after 15 days of inactivity. You can reactivate the index from the index console.

Common workflows

There may be times when you need to update your index. Depending on the update you're making, you can save time by running an incremental population. The following table outlines workflows for different index updates.

Workflow	Index update
Adding new documents that: • Introduce new concepts • Make up **more than** 10% - 30% of your document population	1. Add documents to both the data source and training data source. 2. Click **Run**, then select **Incremental**.
Adding new documents that: • **Don't** introduce new concepts • Make up **less than** 10% - 30% of your document population	1. Add documents to the data source only. 2. Click **Run**, then select **Incremental**.
Removing documents from the data source or training data source	1. Remove documents from the data source or training data source. 2. Click **Run**, then select **Incremental**.
Updating stop words	1. Update stop words. 2. Click **Run**, then select **Full**.
Updating extracted text (ex. Updating poor quality OCR text)	1. Update extracted text. 2. Click **Run**, then select **Full**.
Updating filters (email header, repeated content)	1. Update filters. 2. Click **Run**, then select **Full**.