

Developing a specialized thesaurus for a niche domain with Synaptica

Create a KM for a niche domain — Nanotechnology



The key drivers for building the Knowledge Model (KM)

A global STM (Science, Technology & Medical) publisher built a new database product in the nanotechnology domain. To enhance the information retrieval process and enrich the user experience, the publisher decided to build and maintain a KM (in the form of a thesaurus).

Developing a thesaurus in a niche domain

The task was to refine and expand a two-level term list of basic hierarchies. The challenge was to build a nanomaterials thesaurus with its own set of unique challenges.

The huge list, having several thousands of terms and term relationships, created unique data-level needs for archiving and retention such as:

- Archiving original terms to ensure the traceability of deletions or modifications
- Retention of case-sensitive terms as hidden labels in the KM

Non-standard KM validation rules required special handling of polyhierarchy rules within the KM as poly hierarchy is not allowed in some branches.

Additionally, there was a high degree of domain complexity due to the prevalence of ambiguous and generic terms and redundant words in the form of grammatical variants-containing function words, verbs and their forms.

Since, a reference-able corpus of content was unavailable, the context of the ambiguous concepts had to be derived by researching external sources. Similarly, duplicate concepts had to be identified based on the context of the ambiguous concepts.

Standardize over 60,000 machine generated raw and poor quality terms



Ensure compliance with ANSI/NISO standards



Deployed Synaptica KMS platform



A hybrid, technology + SME delivery model

Scope (Scope eKnowledge Center) leveraged its unique delivery model — blending technology and expertise of Subject Matter Experts (SMEs).

The heavy lifting is done by technology, but subject experts from the industry and academia curate machine output to improve quality beyond automated systems.

Ensured high quality output through SME curation



Building the KM with Synaptica

Scope's partner Synaptica® developed customized features in the Synaptica KMS platform to handle the large vocabulary and the challenges relating to:

- Polyhierarchies
- Lexical variants
- Data standardization

The custom features built on the Synaptica KMS platform was used by Scope's SMEs to generate custom reports to meet the unique data level requirements such as:

- Label uniqueness
- Preservation of original labels
- Traceability of deletions

The KMS platform was used to validate the thesaurus based on standard validation rules and identify redundant hierarchical relations.

Delivered high precision search results



Project Snapshot

Term Standardization Phase

- Volume: 62,067 terms
- TAT: 4 months
- Accuracy 99%

Thesaurus Hierarchy Creation Phase

- Volume: 79,000 terms
- TAT: 8 months
- Accuracy: 98%

Output

- A comprehensive thesaurus of 79,000 terms and 150,000 relationships, delivered to the client as a SKOS (Simple Knowledge Organization System).

Richard Kobel

Vice President,
Business Development, USA
Tel: +1 631 375 1555
rkobel@scopeknowledge.com

Deborah Harman

Associate Vice President,
Business Development, UK & Europe
Tel: +44 770 203 6704
deborah@scopeknowledge.com

Scope eKnowledge Center Pvt. Ltd.

Corporate Office: No 41 Jayant Tech Park, 2nd Floor, Mount Poonamallee Road, Nandambakkam, Chennai – 600089

Tel: 091-44-4091 8900

www.scopeknowledge.com



An ISO 9001:2008 Company