

M+E

JOURNAL

Today's localization challenges are enormous. The opportunities are unprecedented. Is the industry ready for the mayhem?

GIVING VOICE TO CHAOS

SECURITY SOLUTIONS

The threats to our most valuable assets are many. M+E vendors are on top of it.

WORKFLOWS AND THE CLOUD

Much has changed in the way we track, access, move and store everything we deal with.

SMART CONTENT

The many ways the industry adopts new technologies to make content smarter.

22.02

THE WAY TO MULTI-LINGUAL CONTENT LOCALIZATION



Multi-lingual taxonomies can be used to support localized concepts

ABSTRACT: Content localization depends on language and imagery created specifically for regionalized markets. Localizing content text requires translation into one or more languages and possibly regional dialects. Managing concepts and textual blocks to serve to specific markets can be improved by using multi-lingual taxonomies to support localized concepts.

By Ahren E. Lehnert, Senior Manager, Graph Solutions, Synptica

Marketing materials, quite simply, are a combination of words and images. As simple as this is, there is nothing simple about creating, managing, and retrieving unstructured and semi-structured content.

One way to add structure and information to unstructured content is to add administrative and descriptive metadata. Administrative metadata is typically applied to content to describe things like the creation date, the creator (author, photographer, etc.), and copyright and licensing information, just to name a few. Descriptive metadata tells us what the content is about, including things like topics, people, and things in an image, and other internal and external information providing additional information to text and images.

TO HELP INFORM CONTENT STRATEGIES AND TACTICS, enterprises throughout the media and entertainment ecosystem are leveraging a wealth of data from different sources to achieve success in the face of competition.

One of the fundamental challenges in managing content, including its metadata, is the human factor. Even a single content creator and manager is inconsistent, writing text using different terms and applying various metadata values from day to day. This challenge is compounded when content is multi-regional and multi-lingual. Because human content curators are inconsistent, one answer may be to automate content creation and metadata application. While automated techniques are improving, there is often no substitute for human-made content, or, at the very least, human-in-the-loop content creation and maintenance.

What can we do to add structure to content in a more consistent and manageable way?

CONTROLLED VOCABULARIES AS METADATA

Controlled vocabularies have a long history in information management as sources for consistent terminology values. We see this in library catalogs, such as the Library of Congress Subject Headings (LCSH), and in science and educational fields, such as the NASA Thesaurus and the ERIC Thesaurus. Some publicly available and useful multi-lingual vocabularies include the General Multilingual Environmental Thesaurus (GEMET) and the UNESCO Thesaurus.

Controlled vocabularies can include flat lists, taxonomies, and more complex thesauri and ontologies. Controlled vocabularies are not just for academic and scientific publication, however. Hierarchical taxonomies have a place in the enterprise for content navigation and as sources for consistent metadata values. More pertinent for marketing content creation and localization are multi-lingual taxonomies supporting structures in one or more languages as a single source of truth.

Much like automating content creation and metadata tagging, automatic language translation has improved greatly. Despite these improvements, creating a one-to-one relationship between concepts in different languages isn't always possible, especially when trying to fit these concepts into hierarchical nestings which are equally sensible across languages and cultures. Because of these differences, taxonomy management systems do not typically support automatic language

translation. What they do support is the ability to create a taxonomy in a base language and allow the addition of labels in other languages.

For example, a single concept in the GEMET Thesaurus includes multiple concept labels (also called a preferred label) in several languages as indicated by the two-letter abbreviations. We can see the concept label

preferred label (SKOS) +	
وسائط سمعية بصرية	ar
audiovizual vasitələr	az
Аудио-визуална медия	bg
mitjans audiovisuaes	ca
prostředek audiovizuální	cs
audiovisuelle medier	da
Audiovisuelle Medien	de
οπτικοακουστικός τομέας	el
audiovisual media	en
medios audiovisuales	es
audiovisuaalmeedia	et
ikus-entzunezkoak	eu
audiovisuaalised viestintävälineet	fi
audiovisuel	fr

Two-letter abbreviation labels for languages in the GEMET Thesaurus.



Abren E. Lehnert is senior manager of graph solutions with Synaptica and an information and knowledge management professional with more than 20 years' experience in taxonomy, text analytics, search, and content and records management. He has developed enterprise taxonomies and knowledge development and retrieval strategies as a consultant and full-time employee in a broad range of industries. abren.lehnert@synaptica.com @abrenlehnert

in English (en) and supported language characters, such as Arabic (ar) and Bulgarian (bg).

An organization may create a taxonomy (or multiple taxonomies used together) in a base language such as English and then invite taxonomy editors to add labels in other languages. Each English language concept is mapped to one or more concept labels in other languages and the taxonomy can be displayed to content creators in a language of their choice.

TAXONOMIES WITHOUT BORDERS

Taxonomy management systems usually include the option for on-premise or cloud deployments. In today's remote work world with teams which may be situated in centralized and home offices across the globe, access to cloud-based content management systems is becoming the norm.

Applications used for content creation, management, and retrieval can easily pull multi-lingual taxonomy values from a taxonomy management system, allowing users to use consistent terminology values within the body of text or as typeahead values for metadata tagging. Because these values are consistent, content is then easier to gather by topic, author, or any other applied metadata value.


Behind the scenes, the mapping between language values also allows content creators to search for content tagged across languages. That way, a search for a subject in one language can bring back content tagged with the same subject in many languages. Grouping content in this way can reduce the amount of content recreation when appropriate content can't be found (even though it probably already exists) and facilitates content discovery across languages which the searcher may or may not speak.

SINGLE TAXONOMY BENEFITS

The overhead of a good deal of translation work is shifted to a single point in the content creation stream in the single source of truth taxonomy. To be clear, taxonomies won't include every possible sentence in every language to be used in marketing content. What they can include are the main marketing terms and phrases determined to be important by the business. Since these concepts change over time, they can be added, updated, or deleted quickly in a single location and made available for use globally across users, systems, and countries.

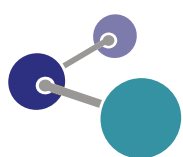
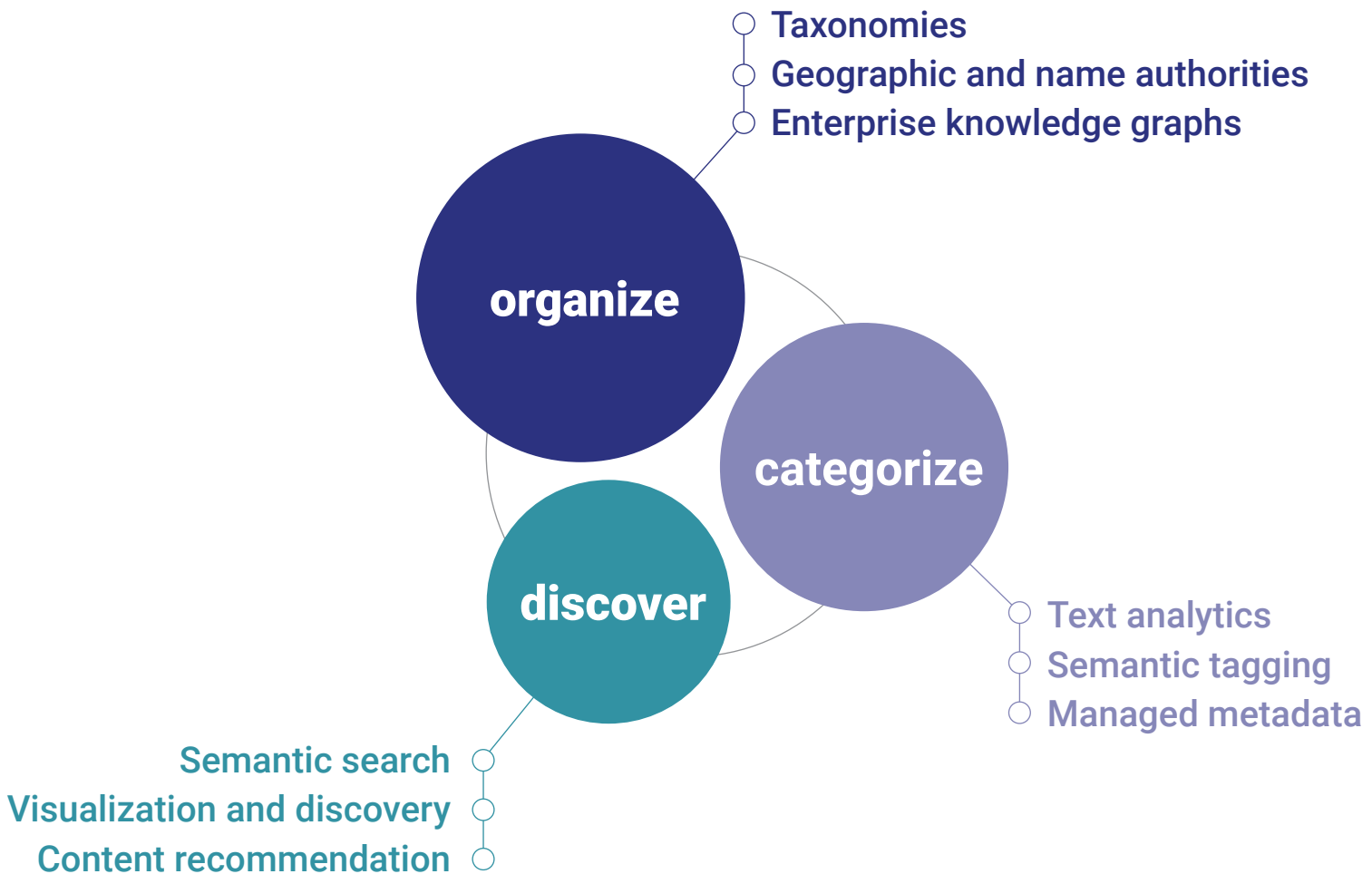
Using one or a few taxonomies supporting multi-lingual labels for content authoring and tagging reduces the complexity of developing many taxonomies in a single language for use in each regionalized content-producing location. Mapping separate taxonomies every time content is created (or not at all) creates a fragmented and chaotic information landscape in which it is difficult to find content and leads to time spent recreating content which already exists.

Mapping separate taxonomies in different languages once is time-consuming but can then be used with fewer changes over time. Creating a single taxonomy from existing vocabularies in one or more languages and using multi-lingual labels may require up-front effort, but the creation, maintenance, and governance reduces greatly over time.

The delivery of multi-lingual, localized marketing content supported by a single source of truth taxonomy on the back end streamlines the content creation and management process for better delivery to end users. 



Enterprise taxonomies and metadata that drive content discovery



synaptica