

What is Fodina TermCatch?

Fodina TermCatch is a web-based solution that supports creation and maintenance of standardized source language terminologies.

At the moment, English and Swedish are supported as source languages, and support for German is under development. The service is delivered to customers as a hosted solution.

Workflow

Create term sources

Fodina TermCatch harvests terms from source files like Word, txt, TMX, HTML and XML, using an AI-based extraction engine. The harvested terms are presented in their base forms even if they appear in inflected forms in the documents.

With the harvested terms, loads of metadata is provided. Some examples are:

- Number of times a term appears in a source
- Part of speech
- Type of term (acronym, name, unit, ...)
- Contexts where the term appears
- Importance score (weighted value of count and degree of specificity)

Term	Root	Root POS	Count	Score	
center display	display	NOUN	402	51.62	★
voice control	control	NOUN	143	17.94	★
Adaptive Cruise Control	Control	PROP	104	13.37	★
gear selector	selector	NOUN	86	10.88	★
windshield	windshield	NOUN	105	10.11	★
head-up display	display	NOUN	78	10.03	★

Figure 1: List of harvested terms

gear selector	selector	NOUN	86	10.88	★
IDs Term-ID: 685, Term-ID (Harvest): 1576					
Variants Gear selector, gear selector, gear selectors					
Labels NR, NP_POS					
POS Tags NOUN, NOUN					
Contexts (75 unique) <ul style="list-style-type: none"> 3: The gear selector is moved to N. 3: Gear selector positions for automatic transmissions (p. 392) 2: Gear selector positions 2: To move the gear selector from N to another gear, the brake pedal must be depressed and the ignition in mode II. 2: For vehicles with the small gear selector, the engine must be run-ning. 					
Context Tags Body Text: 42, List Paragraph: 27, Normal: 15, Heading 4: 1, Table Normal: 1 (6 unique)					

Figure 2: Metadata of a harvested term

TermCatch also allows import of terms from existing term lists and exported files from term databases like Acrolinx, or material names from ERP and PIM systems. The terms from imported term lists or term harvests are kept separate with full control of which term exists in which source.

A third type of term source is external term databases from which TermCatch can fetch terms and concepts automatically using API connections. Today, TermCatch has API connectors to Acrolinx and Quickterm term databases.

Create a Term View

Once all necessary term sources are available in TermCatch, you select a number of term sources (harvested terms, imported term lists or external term databases) and bring them into a Term View. TermCatch groups the terms from the selected sources into term clusters. Each cluster contains synonyms and spelling variants.

Term	Root	Root POS	Count	Score	
Cluster 2 - (2 Entries)					
Adaptive Cruise Control	Control	PROPN	104	13.37	☆
ACC	ACC	PROPN	13	1.88	☆
Cluster 3 - (2 Entries)					
windshield	windshield	NOUN	105	10.11	☆
windscreen	windscreen	NOUN	3	0.08	☆
Cluster 4 - (2 Entries)					
head-up display	display	NOUN	78	10.03	☆
head up display	display	NOUN	1	0.15	☆

Figure 3: List of term clusters

The term clusters expose term variants and possible inconsistencies. With this information at hand, the terminologists can make fact-based decisions about what terms to use – and what terms not to use – in the future.

Cluster 3 - (10 Entries)					
Concept 844f2a16-5819-42fc-947c-815b81164bd7					
cooling liquid	deprecated	en		1	0.03
coolant	preferred	en	4015	1	0.02
No Concept					
coolant				3622	13.71
cooling medium				6	0.05
cooling fluid				4	0.04
cooling liquid				1	0.03
medium				1	0.03
secondary refrigerant				2	0.03

Figure 4: One cluster where terms from the Acrolinx term database (top two rows) are clustered and compared with terms from four other sources including harvested terms from a TMX file and three term lists. The symbols to the left indicate in which sources the term is found.

The clustering mechanism in TermCatch sometimes overgenerates and adds terms to a cluster that aren't really synonyms in the realm of the given customer. (You can actually decide to what level TermCatch should overgenerate in each Term view.)

When reviewing a cluster, the terminologist can select terms from the cluster and “freeze” them to a concept. The concept then contains terms that are actually synonyms in the given company.

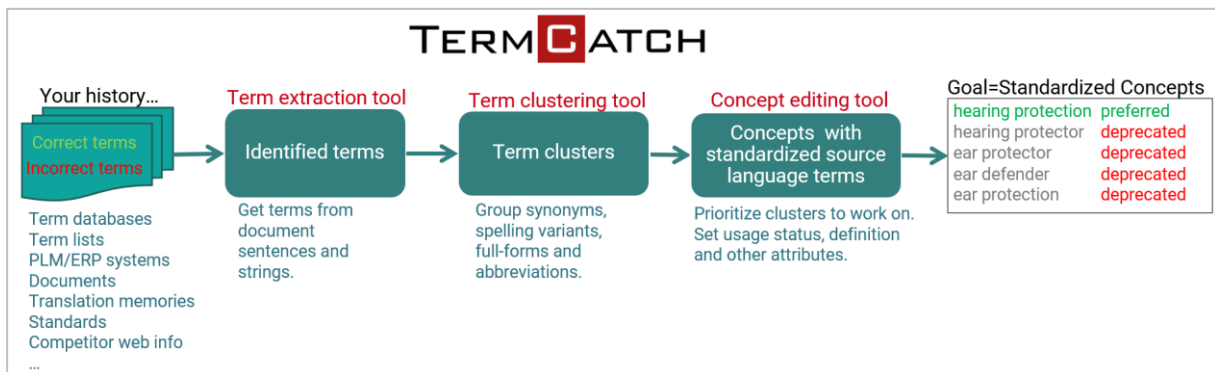
Make term decisions in the concepts

Once concepts have been formed, the terminologist can add attributes to the concept and each term, for example assign usage statuses like deprecated, admitted or preferred to the terms.

Push new and updated concepts

When one or several concepts are ready for publication, the concepts can be exported to different file formats. But since TermCatch has API connections to external term databases like Acrolinx and Quickterm, you could just push a button and have the concepts and terms transferred to an external term database.

In summary, TermCatch automates the two most time-consuming and hard-to-achieve activities in terminology buildup and maintenance: harvesting terms from many sources and understanding synonymy between the harvested terms.



Workflow in summary

1. Create term sources:
 - a. Import term lists
 - b. Harvest terms from document-style files
 - c. Harvest terms from web pages
 - d. Import concepts and terms from external term databases like Acrolinx and Quickterm
2. Create a Term View:
 - a. Select term sources to view together
 - b. Cluster the terms in the selected term sources
 - c. View terms and clusters
 - d. Sort and filter terms and clusters
 - e. Remove terms from clusters that they don't fit into
 - f. Create or edit concepts
3. Make term decisions in the concepts:
 - a. Assign statuses to the terms
 - b. Add attribute information to terms and concepts
4. Push new and updated concepts from TermCatch to an external term database like Acrolinx or Quickterm, or export terms to different file formats.

Use Cases

Build a standardized, value-adding terminology based on your content

If you don't have a standardized terminology, or very few terms available, you would benefit greatly from using TermCatch to get a working terminology quickly.

TermCatch helps you to work on the most needed terms and add them to your term database. You would do the following steps to build the initial version of your terminology:

- Harvest terms from large volumes of content, translation memories, web pages, standards, competitor web sites (!) and more
- Cluster terms automatically into synonym clusters
- Filter clusters based on different criteria (score, frequency, term sources...)
- Review clusters and create term concepts with status and other metadata
- Push your new concepts and terms to your master term database like Acrolinx or Quickterm

Maintain your terminology continuously

When your terminology process is up and running, new terms will be discovered and proposed:

- Writers, product owners and marketing people propose new terms
- Translators might propose terms for standardization
- Acrolinx identifies new high-frequent terms automatically when checking content

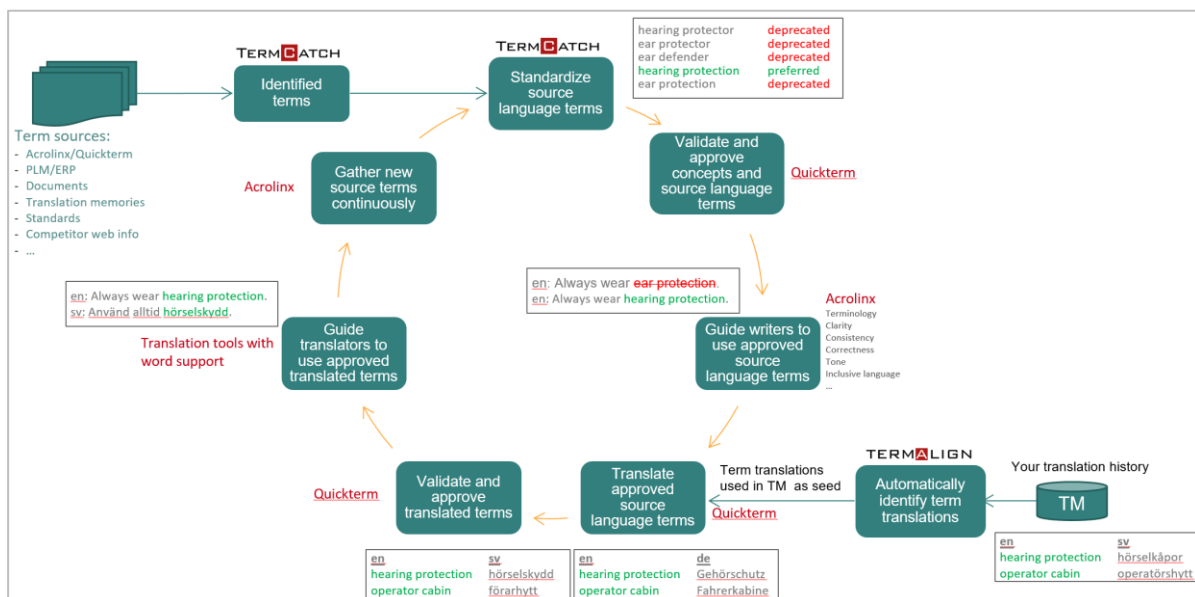
Before adding newly identified terms to your terminology, the terms should be reviewed and compared to your existing terms and concepts:

- You don't want to create terms that are synonyms to existing terms, and therefore would make your terminology inconsistent and of poor quality
- You may find new evidence for making new decisions on, for example, status or domains
- Or, quite simply, you want to add a newly found term as a deprecated synonym to an existing concept in your terminology

TermCatch compares the content of your existing term database with terms from any number of term lists. When terms from term lists cluster with terms in concepts in your term database, you can easily add terms to existing concepts or form new concepts. Thus, you are creating new enhanced concepts that replace concepts in your term database.

Process Alignment

TermCatch fits perfectly into the terminology process that Fodina advocates.



The TermCatch Service

- TermCatch is hosted in the Google Cloud
- Each customer gets their own instance of TermCatch, and no customer data is shared with other customers
- TermCatch is available as a web service where users authenticate with user id and password, or using SSO
- All data is encrypted, both at rest and in transition
- Fodina backs up all customer data on a nightly basis

TermCatch is a server-based application that doesn't require more from the client computer than a web browser and an open network connection to the TermCatch server.

Features

Term harvesting	<p>TermCatch harvests terms from many different types of source files. Today, we support these file formats: Word, txt, TMX, HTML and XML. For HTML and XML files.</p> <p>You can create content profiles that make TermCatch sensitive to your specific tags, for instance avoiding term harvesting from metadata tags.</p>
Term import from term lists	<p>Terms from different term lists can be imported into separate sources in TermCatch. In that way, we know and can use the source information in filtering and comparison operations.</p> <p>Today, TermCatch supports import from these file formats: txt, xlsx and json.</p>
Term import from or export to external term databases	<p>TermCatch can use APIs in Acrolinx and Quickterm to:</p> <ul style="list-style-type: none"> ▪ Pull concepts and terms from the external term database to TermCatch. ▪ Push new and updated concepts and terms from TermCatch to the external term database.
Term clustering	<p>As new terms are fed into the system by term extraction or term import, TermCatch automatically creates clusters of synonymous terms and spelling variants. This enables easy detection of erroneous term use.</p> <p>Each cluster can be modified by cluster operations, including the removal, editing or addition of terms. It's also possible to create new connections between terms or clusters.</p>
Concept creation	<p>A concept is the abstract object to which the terms refer. In the TermCatch Term Views, users confirm term synonymy by creating concepts out of automatically prepared term clusters.</p>
Filtering	<p>TermCatch can filter terms, or subsets of terms, with a sophisticated filter function.</p> <p>Today, TermCatch filters in many ways including regular expressions, part of speech, term type, term lists and more.</p>
Export	<p>Terms and clusters can be exported to several file formats. Today, TermCatch supports these formats: actif, xlsx and tsv.</p>

More information?

Please, do not hesitate to contact us!

- **info@fodina.se**
- **013 – 342 01 87**