
Confirmit
Confirmit ModelBuilder
User Guide

This is revision 4 of the Confirmit Model Builder User Guide published in November 2018. The information herein describes Confirmit Model Builder and its features as of Build nr. 3.15. New features may be introduced into the product after this date. Go to www.confirmit.com or check “News” on the Customer Extranet for the latest updates.

Table of Contents

What's New in this Revision?

The following changes have been made in revision 4 of the Confirmit Model Builder User Guide:

- Concept Miner information is added to the What is Model Builder chapter (see What is Model Builder? on page 1 for more information).
- The download URLs are added to the Logging In section (see Logging In on page 1 for more information).
- The text in the Search section is edited (see Search on page 5 for more information).
- A note in the Category List section is edited (see The Category List on page 13 for more information).
- The text in the Asking the Right Questions section is edited (see Asking the Right Questions on page 16 for more information).
- The Logical Operators section is renamed to Boolean Operators, and is extensively edited (see Boolean Operators on page 20 for more information).
- The text in the Importing your File section is extensively edited (see Importing your File on page 25 for more information).
- The Concept Miner chapter is added (see Concept Miner on page 28 for more information).
- The text in the How to Create a New Model section is edited (see How to Create a New Model on page 36 for more information).
- The Reordering Categories... section is added to the Creating a New Model chapter (see Reordering Categories in the Category Tree on page 41 for more information).
- The Excluding Categories... section is added to the Creating a New Model chapter (see Excluding Categories from the Report on page 41 for more information).
- The Survey Variables... section is added to the Creating a New Model chapter (see Survey Variables - Adding Survey Context into Models on page 42 for more information).
- The text and illustrations in the How to Test and Expression section are edited (see How to Test an Expression on page 49 for more information).

Note: The general layout and language in this document is continually being corrected, adjusted and improved to ensure the user has the best possible source of information. Only NEW information and details of functionality that has changed since the previous issue are listed here - minor corrections to the text and document layout are not listed.

Important
We need your feedback so we can improve this document and provide you with the information you require. If you have any comments or constructive criticism concerning the content or layout of this documentation, please send an email to documentation@confirmit.com. Please include in your email the section number and/or heading text of the section to which your comment applies.

1. What is Model Builder?

Confirmit Model Builder is an integral part of the Confirmit Genius text analytics application.

Model Builder is the application where you build the categorization model, which is used to analyze your text in Text Analytics. Model Builder houses Concept Miner, Confirmit's AI topic extraction tool, which helps you to understand your verbatim data, create topic tags to use in your categorization model, and speeds up the model building process.

The categorization model is built using a hierarchical tree structure and the nodes of the tree contain expressions, which are built up of keywords, Boolean operators and Concept Miner tags. These expressions determine the resulting categorization of your text.

For example, a shop could ask customers what they thought of the service they received while making their most recent purchase. A respondent's reply could read: "Your employees were very knowledgeable and helpful, but the floor was dirty and the shop was rather untidy." The meaning of this sentence and its clauses is obvious to a person who knows the language, and this sentence could result in a "staff knowledge" category (with keyword "knowledgeable") and a "staff helpfulness" category (with keyword "helpful"), the "shop cleanliness" attribute (with keyword "dirty") and "tidiness" attribute (with keyword "untidy"). However if you have several thousand such replies, all of them slightly different, extracting the categories by manual analysis becomes a laborious and inaccurate job. Analysis by computer then becomes the only feasible option. After automatic processing, the several thousand replies will provide the shop with a set of data they can use as a basis for changes such that the customer experience, and hopefully thereby customer retention, is improved.

Model Builder addresses the need to categorize and quantify topics and other information found in text. There is also, of course, the sentiment process, which adds the positive, neutral/mixed or negative sentiment analysis to the overall verbatim and also each of the categories, which further enriches your analysis. Note that Model Builder does not deal with the sentiment process – this happens in the Genius Text Analytics processing task from Horizons.

During the analysis part of the process, each word in each response is compared against a database (the model) of words, groups of words and Concept Miner tags that could be used in responses to the question. The model must therefore contain all the terms and tags that are to be searched for in the responses. Model Builder is the application that assists you with creating and compiling the model.

As the vocabulary used in different situations can differ considerably, the model must be tuned for the specific customer domain. For example, the retail branch, call centers, consumer electronics branch, hotel branch and vehicle workshop branch etc. will all need different models. And even within branches, different specializations will need to be fine-tuned; for example a hardware store will need a different model than a chemist. Template building blocks are available from Confirmit which can be used to speed-up the process.

1.1. Logging In

Model Builder is located on three separate URLs, for our three different SaaS sites. You will need to use the URL that corresponds to the SaaS site that you use for accessing Confirmit Horizons:

- US site: https://genius.us.confirmit.com
- EURO site: https://genius.euro.confirmit.com
- AUS site: https://genius.confirmit.com.au

When you go to the Model Builder URL, you are presented with the login page.

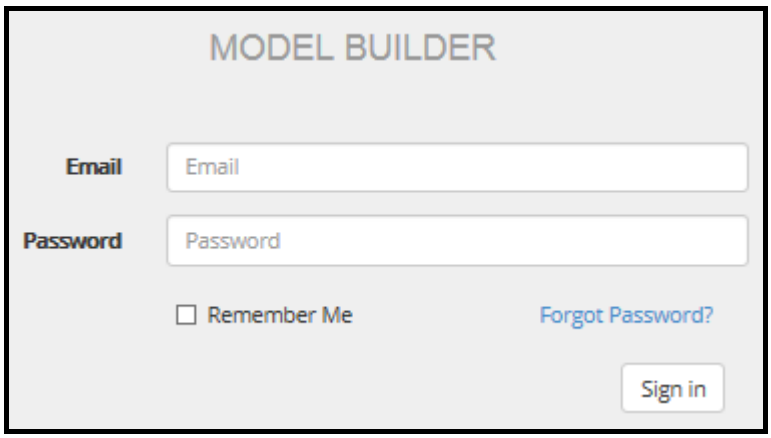


Figure 1 The Model Builder login page

Type your email address and password into the fields and click **Sign in**. In the event you have forgotten your password or you are a new user and do not yet have one, click **Forgot Password**. You are taken through the new password routine so you can reset your password (see Forgotten Password on page 2 for more information).

The Model List page opens. This shows all models that have been built within your customer account.

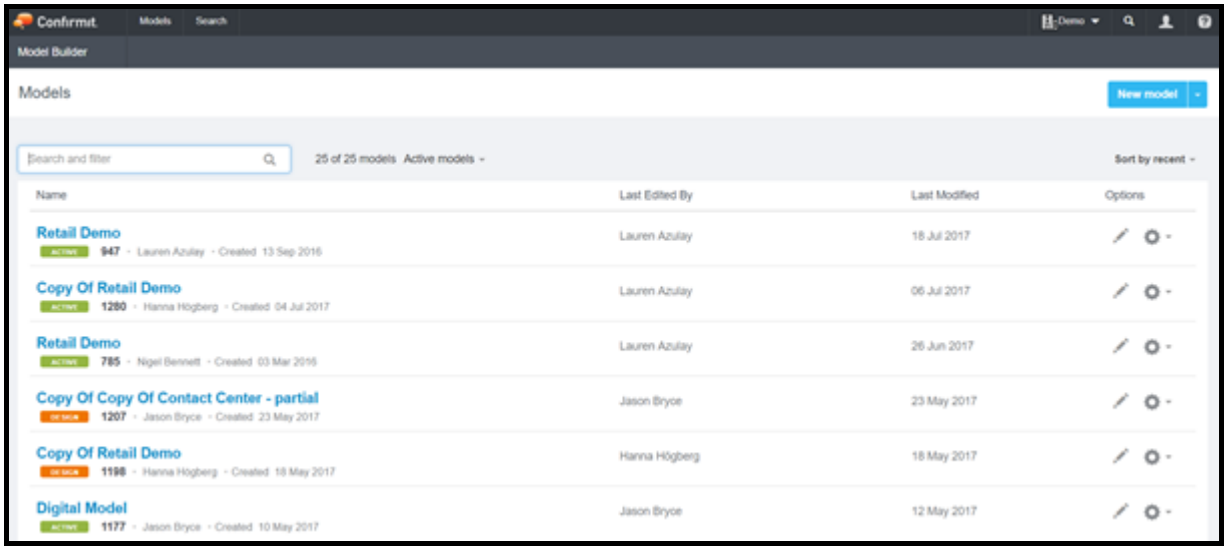


Figure 2 Example of the model list for a customer

Here you can select an existing model to work with, or create a new model.

Later when you have completed your work with Model Builder, close and log out of the application by going to the **User** menu and selecting **Log Out** (see The User Menu on page 11 for more information). This closes Model Builder correctly and returns you to the Log In page.

1.1.1. Forgotten Password

Your Model Builder password prevents unauthorized people from using your account and interfering with your models. You must therefore ensure your password is difficult to guess and is known only to you. This of course can create problems - if you forget your password then you will not be able to use Model Builder. The login page therefore includes the possibility for you to reset your password should you forget it, and also create a password the first time you log in to Model Builder. To do so:

1. Click the **Forgot Password?** link below the login page.

   The dialog shown below opens.

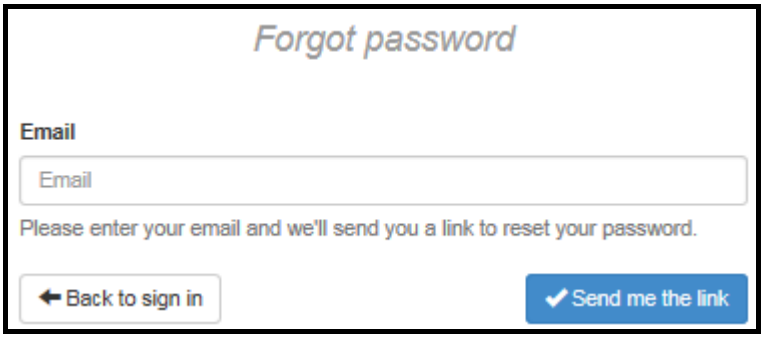


Figure 3 The "Forgot password" dialog

2. Type your email address into the field and click **Send me the link**.

   A message is sent to Confirmit. Assuming your email address is registered, an email will be sent to you containing a link and instructions for how you can reset your password.

Note: The reset password link is only valid for one hour. If you do not reset your password within the hour then the link will expire and you must repeat the procedure to be sent a new link.

1.1.2. Inactive-Screen Lock

For security reasons, when the Model Builder application has been inactive for 30 minutes it will be locked and you will have to log back in to continue. If you have not logged back in within three hours, Model Builder will close and any work that has not been saved will be lost. If you log back in within the three-hour period, Model Builder will reopen at the place you were last working.

Warning
Save your work at regular intervals, and ensure you save before you leave Model Builder unattended.

1.2. The Models List Details

When you open the Models page, each model that you have access to is displayed as one row in the list. To create a new model, click **New model** (see How to Create a New Model on page 36 for more information). To create a new model from an import, click on the drop-down arrow next to **New Model** then select **Import model** (see Importing a Model on page 24 for more information).

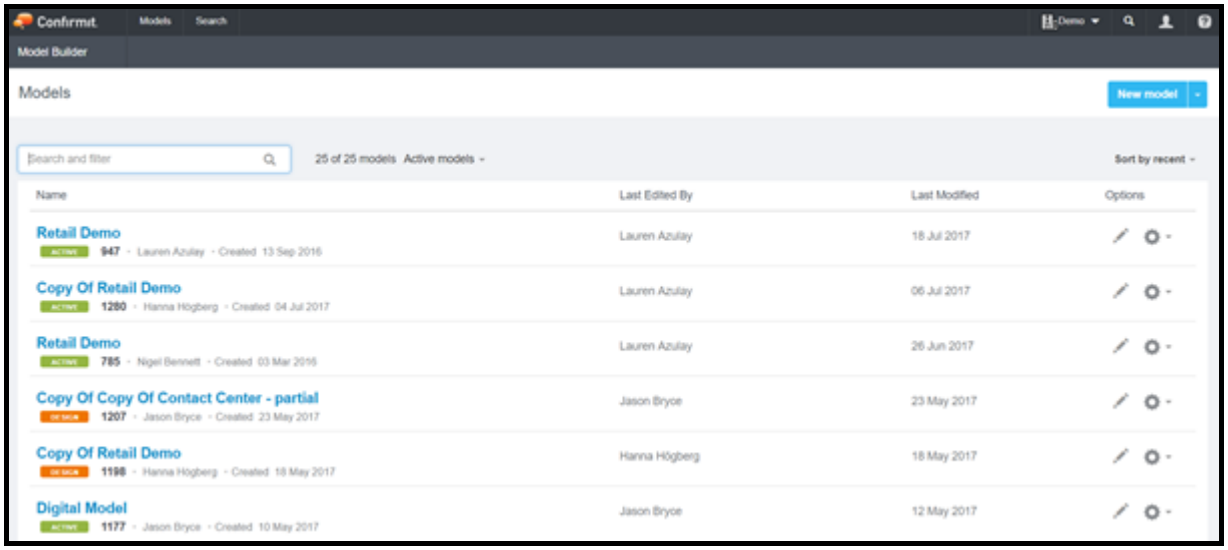


Figure 4 Example of the Models list

In the event you have many models available, you can use the Search field to find the model you wish to work with. The search looks at both the model names and numbers. You can type a text string into the Search field to filter the model list and display only those models with names that include the entered characters anywhere in the model name. Or you can search by model number. If for example you search for 15, Model Builder will return all models with a 15 anywhere in the name or in the number. Note that the search field is not case-sensitive. To re-display the full list, clear any characters from the field.

You can sort the model list by the most recent accessed, the newest (most recently created), and alphabetically by the model name. The button then indicated the sort selected. Click **Sort by...** to open the drop-down.

Click on the blue model name link to open the Design page for that model. The details displayed for each model include:

- **Model name -** this is the access link to the model. Click on the link to open the Design page for that model.
- **Model state –** the model is either in 'Design' mode or 'Active' mode. When it is in Design mode, it has not yet been made Active and cannot be referenced in Horizons. Once the model has been set to Active, it is in Active mode and can therefore be used in Horizons. You must also select which version of the model is to be the Active version (see Save as a Version on page 53 for more information).
- **Model ID -** when the model is created it is automatically given an identification number. This model ID is used when setting up Genius in Horizons.
- **Created date -** the date the model was created.
- **Last Edited By -** the user name of the person who last edited the model.
- **Last Modified -** the date the model was last modified.
- **View/Edit icon -** has the same effect as clicking the blue Model name link - see above.
- **Cog-wheel icon -** opens the **More Options** menu, enabling you to duplicate the model, archive the model, and open the list of categories.

1.3. The Menus

Model Builder holds a number of menus in a bar across the top of the screen and on the various pages in the application. These menus and tools provide access to the functionality.

1.3.1. Models

Click to return to the Models page, in which all the models to which you have access are listed.

1.3.2. Search

This opens the Search page.

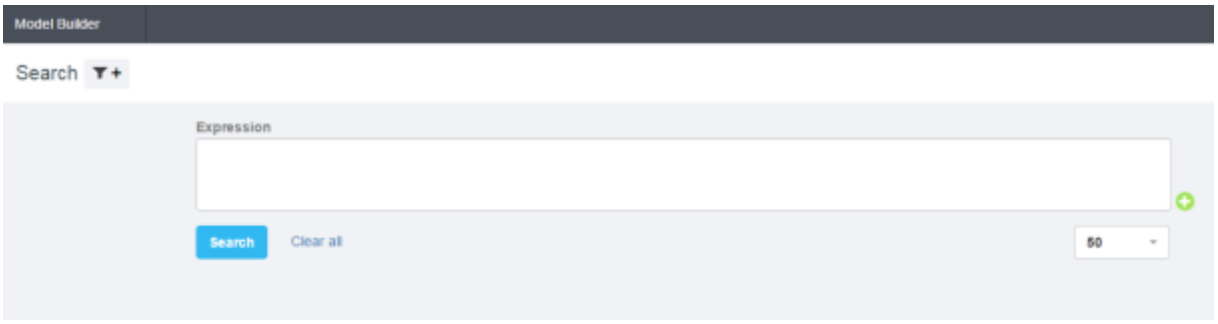


Figure 5 The Search page

Use this page to search the text content of any text that is in Model Builder. This area is great for testing proposed keywords or tags in your verbatim before you add them to your model expressions. The expression is your entire query that makes up what you are looking for in a category (more on this throughout the document). Adding keywords can sometimes be quite complex, in that you need to test which words and combinations of words achieve the most correct hits. There will be occasions where keywords merely return 'noise', but you will need to assess whether or not the hits-to-noise ratio is acceptable. The Search page can help you to decide this for individual words without having the results cluttered by the hits for all the other words in the expression. Once you are happy with what you have tested in the search boxes, you can add the expression into the design area of your model.

Tip!
If you have a large expression, you can expand the search box by dragging the lower right-hand corner.

Click on the green **+** icon at the side of the expression box to open up other options for testing. You have the choice of adding an AND, OR or AND NOT box. Click on **Search** to get your results (use **Ctrl-Enter** as a shortcut to clicking the **Search** button). .

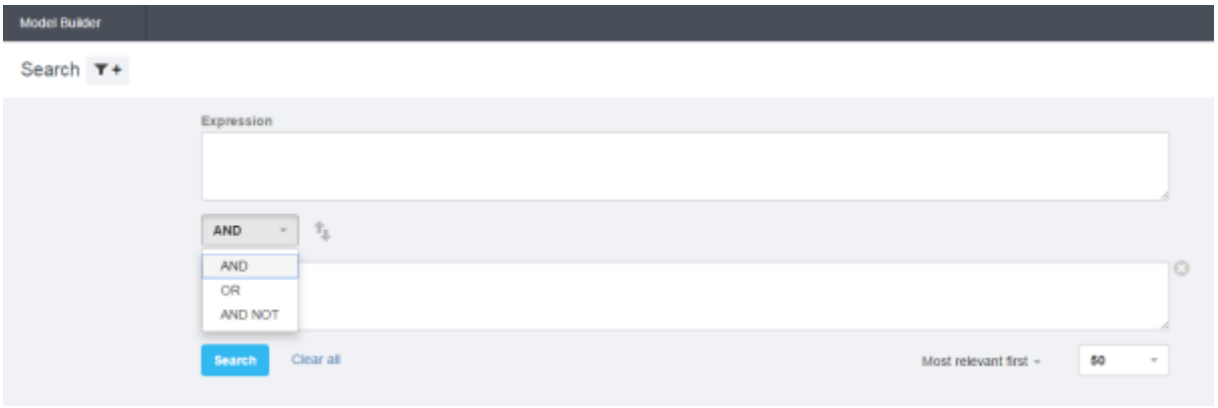


Figure 6 Adding an AND, OR or AND NOT box

Once you have a set of search results, you can choose how they are sorted. This allows you to look at different samples so you can test the accuracy of your results. Choose between 'Most relevant first', 'Least relevant first', 'Newest first', 'Oldest first' or Random order.

Figure 7 Sorting the search results

You can also flip the 2 search boxes around by clicking on the up and down arrow icons next to the drop-down list. This will only affect the AND NOT functionality, and it gives a good opportunity to test your expressions in the opposite order without having to copy the expressions into the reverse boxes.

You can open the filter options so that you can filter by a particular survey or by time period.

1. Click the filter icon beside the Search page title to open the filter.

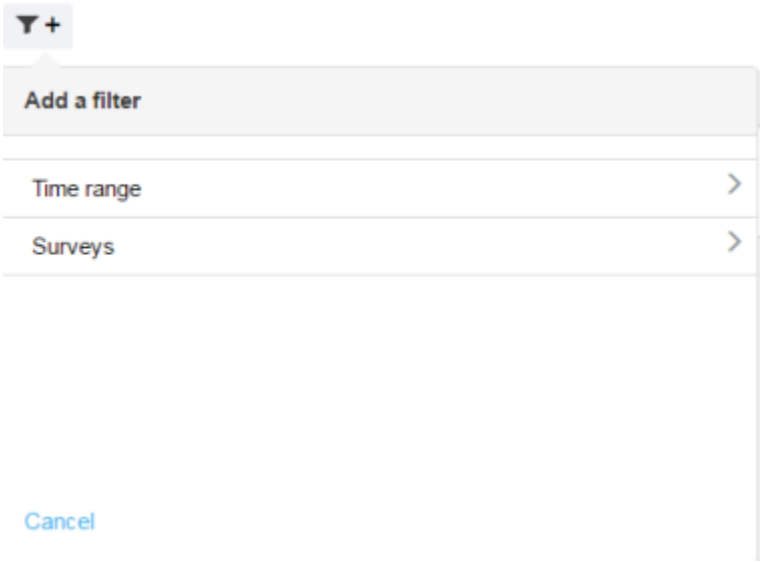


Figure 8 The filter overlay

2. Select one or more surveys to filter your results. You can also choose the question within the survey if that question has had the Text Analytics folder applied to it.

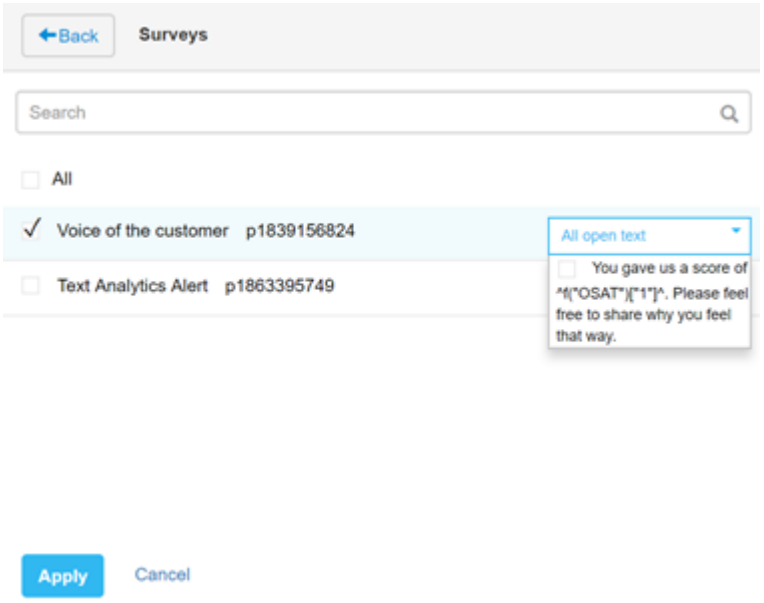


Figure 9 Selecting the surveys

3. Click **Apply** or select **Time Range**.

   Under **Time range** you can choose from a list of options (last 30 days, last 60 days, last 90 days or last 120 days) or select a custom range.

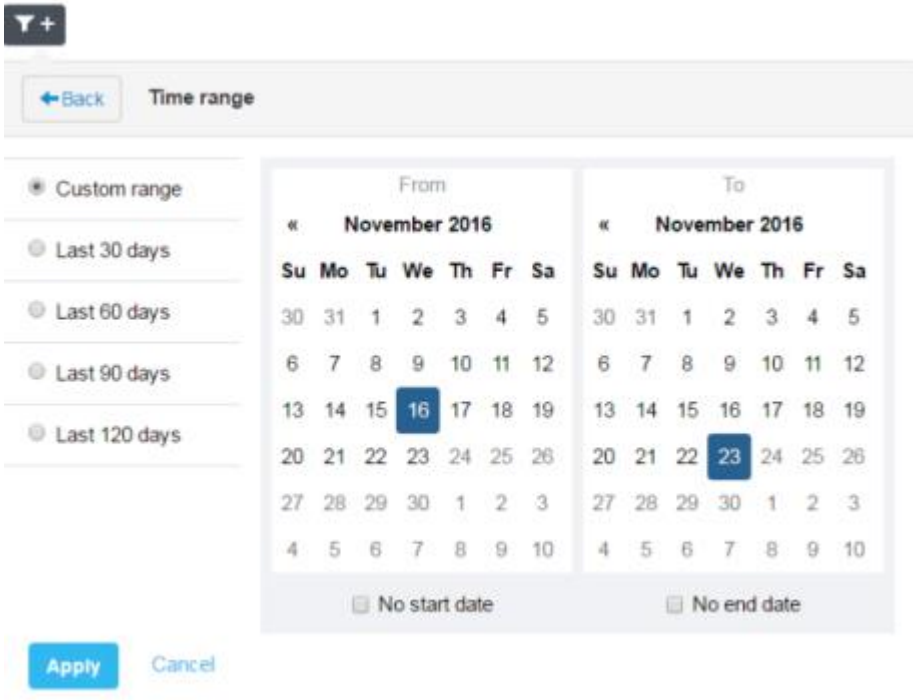


Figure 10 The Time range overlay

4. On completion, click **Apply**.

Once you have applied the filter you can perform your search and you will only see results from the selected surveys, questions and / or date range.

Your filter for the search page will be remembered during your session. If you navigate to another screen and then back to the Search page, the same filter will be applied. You will see this identified at the top of the screen. To remove the filter, click on the 'remove' icon to the right of the filter.

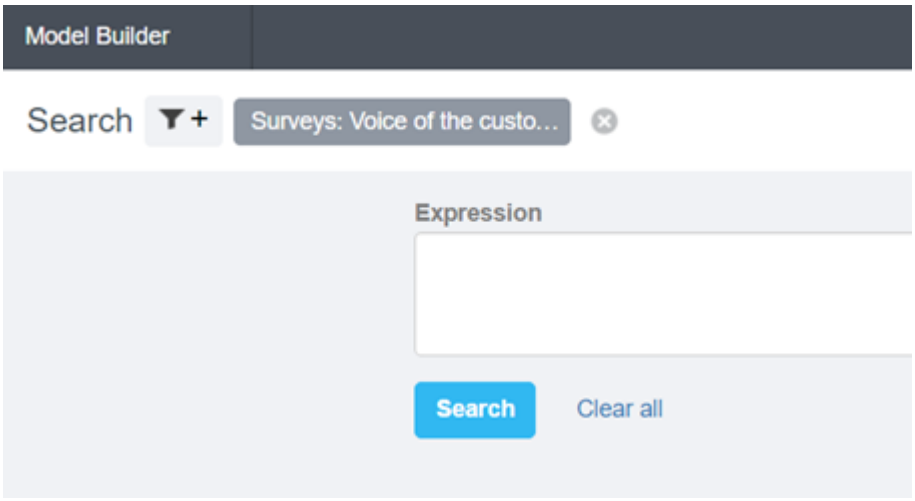


Figure 11 The search filter

1.3.3. Concept Miner

Concept Miner is Confirmit's AI topic extraction tool. Concept Miner will sort through data and provide the top topics for the given verbatim, and these topics can then be used as tags in your categorization model, helping you to speed up the categorization process (see Concept Miner on page 28 for more information).

1.3.4. Company Name

Your company account name is displayed next to the **User** menu icon. In the event you have administrator access to more than one customer, click the account name to open a list of the customers to which you have access, and select the customer you wish to work with. The list of models associated with that customer and to which you have access then opens.

1.3.5. Keyword Search

The keyword search (available from the search icon on the top navigation bar) can be used to find expressions where you may have used a particular keyword or set of keywords. Click the search icon and type your keyword into the search criteria field. The search will start automatically after a few seconds. The results are presented in the window under the search bar.

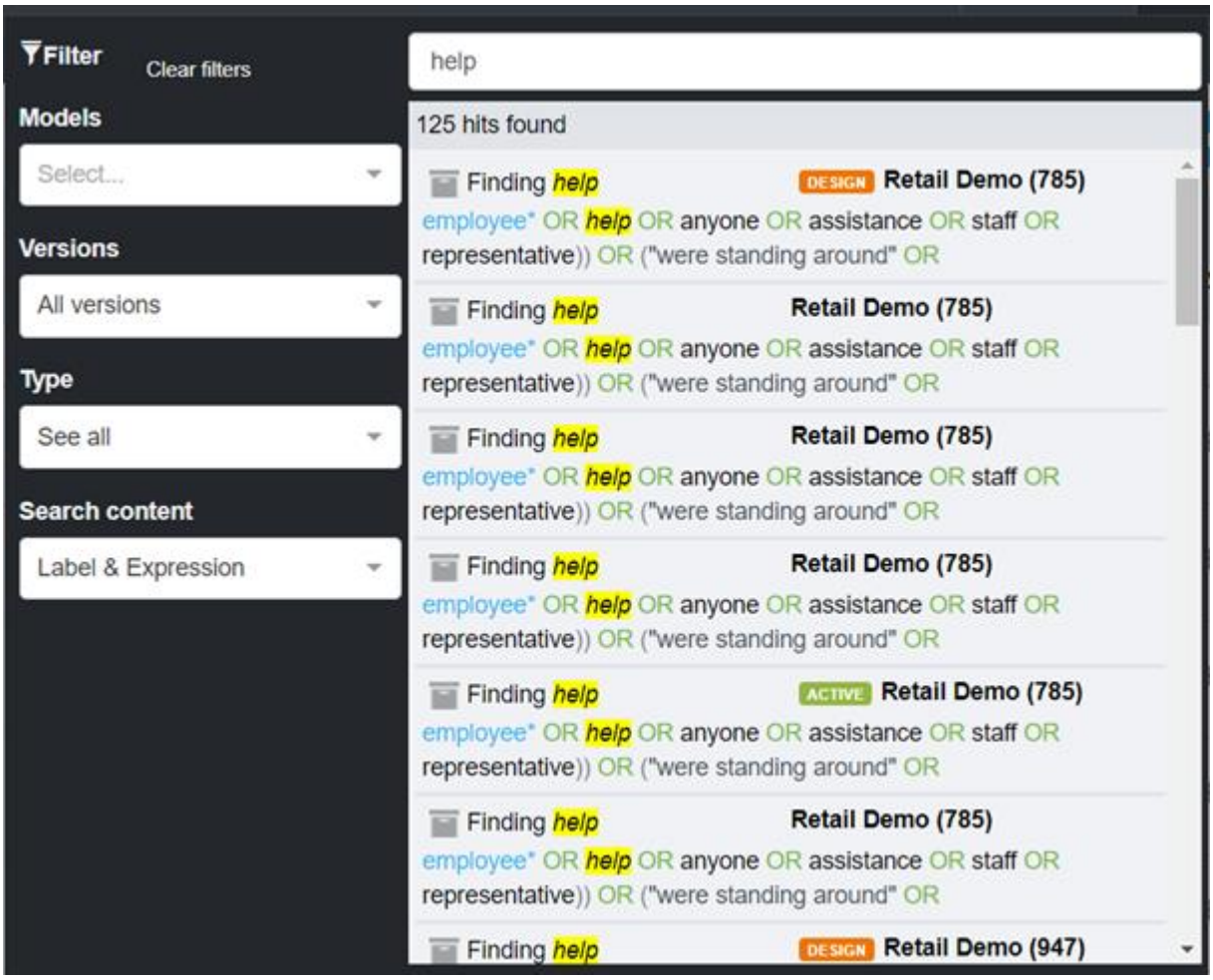


Figure 12 Using the keyword search functionality

Click on one of the expressions to go into the full expression pane.

Figure 13 The expression

Click **Go to expression** go to the model and the expression in the model.

The Filters

The filter panel on the left of the keyword search helps to refine your search, so that you're only looking through relevant expressions.

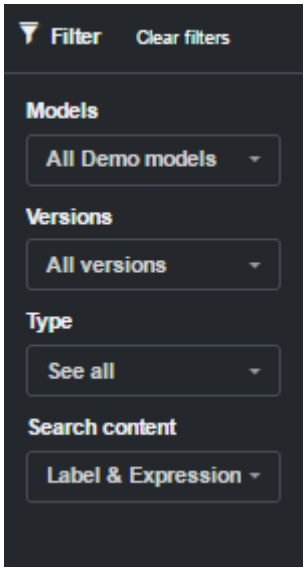


Figure 14 The filter panel

The options are:

- **Models -** you can choose to filter by a single model or by all that are in the account.
- **Versions -** you can choose between All versions, Active, Design or Other saved versions.
- **Type -** you can choose between All, Category or Attribute.
- **Search content -** you can choose between Label & Expression, Label or Expression.

1.3.6. The User Menu

The User menu  , located towards the right end of the main toolbar, provides access to commands applicable to you personally as the user.

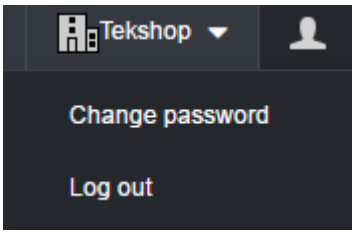


Figure 15 The User menu

- **Change password -** you can change your password from the **User** menu. To do this, click on **Change password**. You will be asked for your current password, your new password, and you will then have to repeat your new password to ensure that it is as you intended it to be.

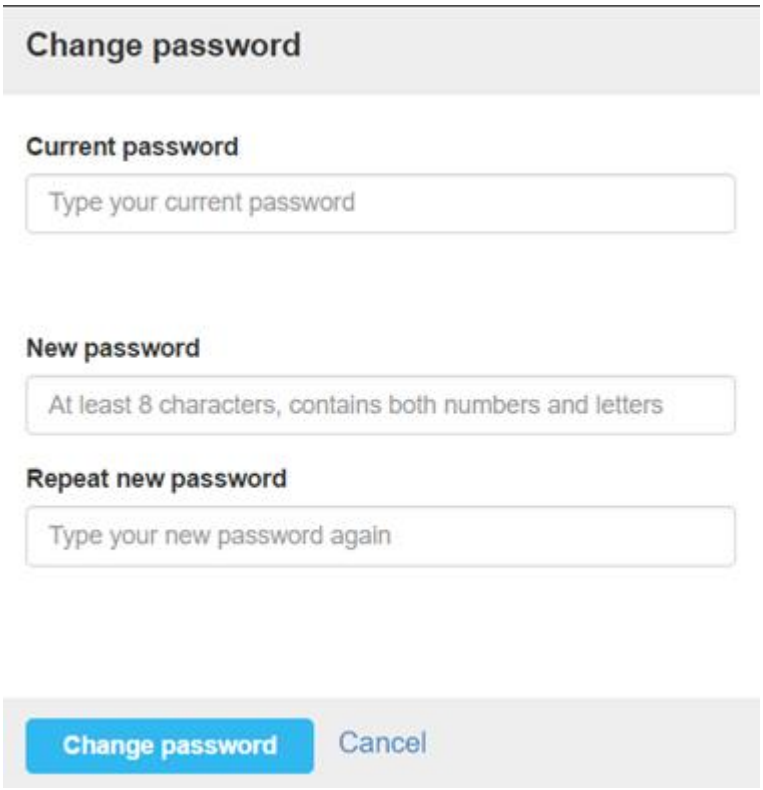


Figure 16 The Change password overlay

- **Log out -** click to log out of Model Builder and return to the login page (see Logging In on page 1 for more information). The next time you log in you will be taken to the Models list. In the event you have access to more than one company, you will be taken to the Models list for the company you were last working with when you logged out.

1.3.7. The More Options Menu

The **More options** icon , located towards the right end of each model row in the Model List, opens the More Options menu.

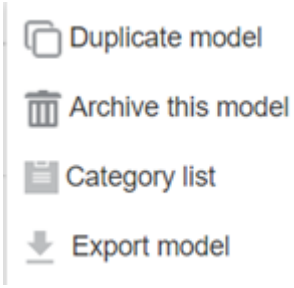


Figure 17 The More Options menu

The options in this menu are:

- **Duplicate model** - if you need to create a new model that you know will be similar to an existing model, you can create a copy of the existing model and edit the copy, thus saving much time.
- **Archive** - when you no longer need to use a model and you wish to remove it from the Model List, you can archive it (see Archiving Models on page 12 for more information).
- **Category list** - opens a dialog box that gives you access to the category list (with sub-categories and attributes) and the associated category and attribute IDs (see The Category List on page 13 for more information).
- **Export Model –** exports the model into an Excel spreadsheet. It is exported in the same format needed to import the model, so it can be used to import into another account (see Export Model on page 15 for more information)

1.4. Archiving Models

To remove a model from your model list, click in the cogwheel on the Model List page and select **Archive this model**; the model will be archived.

To view all archived models, click on the **Active models** link next to the search box on the Model List page, then choose **Archived models**.

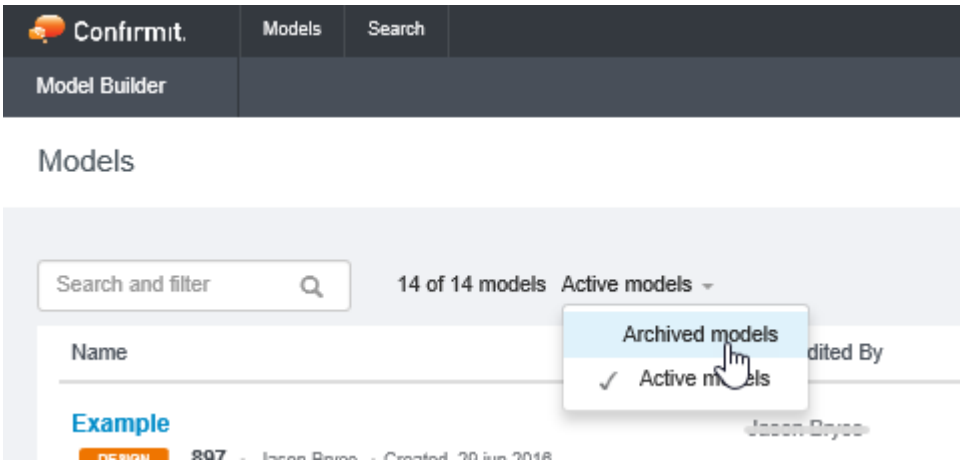


Figure 18 Selecting the Archived Models list

From the Archived Models page you can **View** the archived model or **Restore** it. Restoring an archived model puts it back into the Active Model list page, where you can continue to use and edit it.

1.5. The Category List

Open the **Category List** from the **More options** menu on the Model List page, or from the Overview page (**Get list of categories and attributes**). This is a dialog box that lists the categories in the model (with sub-categories and attributes) and the associated category and attribute IDs. You can copy this list from the dialog box into an Excel® spreadsheet and save it as a tab delimited text file for use in the hierarchy table in Database Designer.

Category and sub-category IDs start with a 'c' and attribute IDs start with an 'a'. Uncategorized comments can also be transferred into the report, and these are given the prefix **uc**. Uncaptured attributes can be transferred into the report and are given the prefix **ua**. Note that there are 3 tabs in the Category list pop up. Depending on how your project is set up and how you are reporting on your data, you will need to choose the correct list:

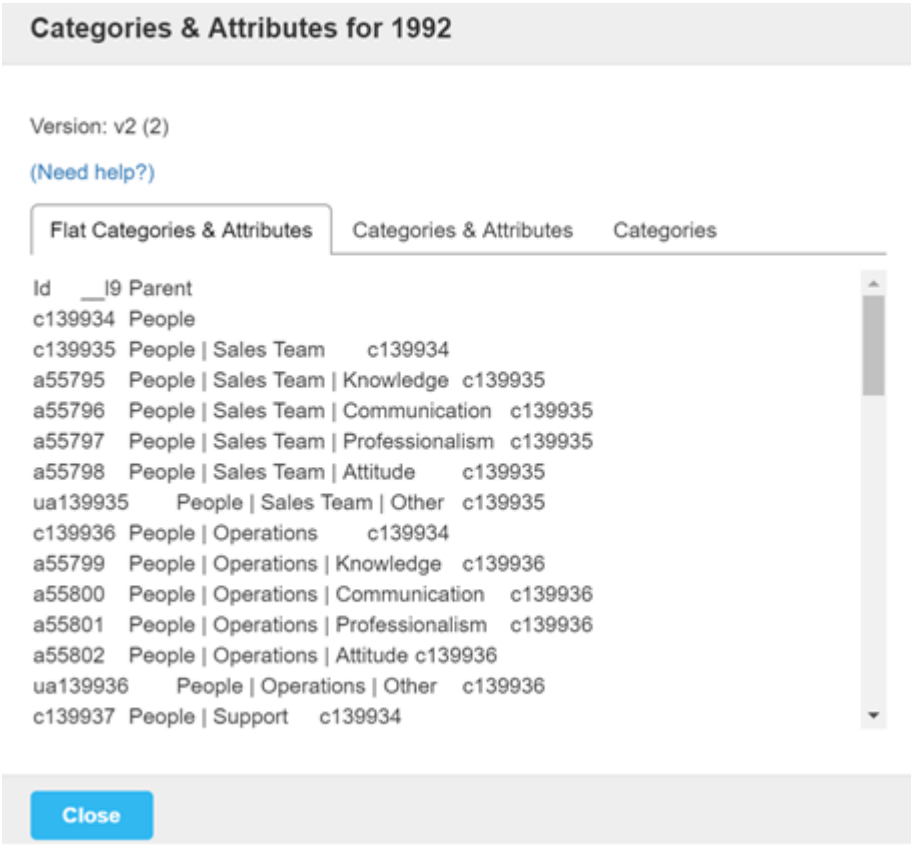


Figure 19 Example of categories and attributes for a model

- **Flat Categories and Attributes -** use this category list when you have a hierarchical model and you are going to be using the Text Analytics standard report template for your report. It has the names of the categories, sub-categories and attributes flattened out, with pipe character (|) in between. The report uses the pipe character to parse the information so that you still have the hierarchical structure where needed, but you can also have the full category / sub-category / attribute name in the charts (you need this for the context).
- **Categories and Attributes -** use this category list if you have a report is not the standard template (i.e. a bespoke report) or if your report was set up before the template was developed. Please note that this is the category list that must be used with the Discovery Analytics Text Analytics widget, if you want to view the true hierarchy.
- **Categories -** this category list is a legacy option and only to be used if you have an older, flat model.

If you are at all unsure of which list to use, contact your Program Manager or Account Manager for advice.

Note: If you want to use both the Reportal template and the Discovery Analytics widget for reporting, you will need to have both tables in your schema in Database Designer. In the set up in Horizons, the loopreference must reference the category list with the | (pipe) character. The HierarchyReference variable must reference the category list without the pipe character.

Note also that you are only given the category list for the Active version. This is the version that is in use in Horizons. If you change your active version, you will either need to manually update your Database Designer table in Horizons (see the Genius chapter of the Authoring manual to see how this is done) or you can choose to 'Synchrionize model to database designer tables' (currently on feature toggle), which will sync the table automatically. Don't forget to also check 'with full path in label' if you need the pipe (|) character to be included for your reports..

Important
Ensure you use the 'Categories and Attributes' tab to get your category and attribute IDs. The 'Categories' tab is for use in older projects that use a flat category structure and that have not yet been transferred to the hierarchical model.

1. Select the full list, then use the **Ctrl+C** keys on your keyboard to copy the list.
2. Paste your list into an Excel spreadsheet.

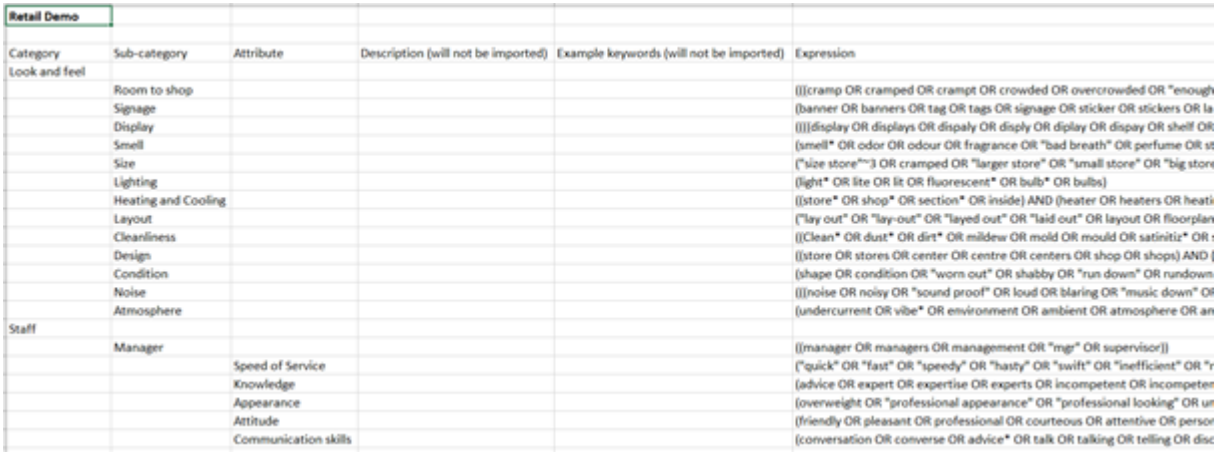
   The copy/paste routine will put the content into the appropriate columns, ready for loading into your Database Designer table in Horizons.
3. Save the Excel spreadsheet in tab delimited text format.

   This file will be used to reference your hierarchy in Horizons in a Database Designer table. Refer to the Genius chapter of the Authoring User Guide for further information.

1.6. Export Model

The **Export Model** menu item is located in the **More Options** menu on the Model List page (see The More Options Menu on page 12 for more information).

Click on the option in the menu to download the model as an Excel spreadsheet. This will be in the same format as that used to import a model.



Retail Demo					
Category	Sub-category	Attribute	Description (will not be imported)	Example keywords (will not be imported)	Expression
Look and feel					
	Room to shop				((cramp OR cramped OR crampt OR crowded OR overcrowded OR "enough
	Signage				(banner OR banners OR tag OR tags OR signage OR sticker OR stickers OR la
	Display				(((display OR displays OR dispaly OR disply OR diplay OR dispay OR shelf OR
	Smell				(smell* OR odor OR odour OR fragrance OR "bad breath" OR perfume OR st
	Size				("size store"~3 OR cramped OR "larger store" OR "small store" OR "big store
	Lighting				(light* OR lite OR lit OR fluorescent* OR bulb* OR bulbs)
	Heating and Cooling				((store* OR shop* OR section* OR inside) AND (heater OR heaters OR heati
	Layout				("lay out" OR "lay-out" OR "layed out" OR "laid out" OR layout OR floorplan
	Cleanliness				((Clean* OR dust* OR dirt* OR mildew OR mold OR mould OR satinitiz* OR s
	Design				((store OR stores OR center OR centre OR centers OR shop OR shops) AND (
	Condition				(shape OR condition OR "worn out" OR shabby OR "run down" OR rundown
	Noise				(((noise OR noisy OR "sound proof" OR loud OR blaring OR "music down" OR
	Atmosphere				(undercurrent OR vibe* OR environment OR ambient OR atmosphere OR an
Staff					
	Manager				((manager OR managers OR management OR "mgr" OR supervisor))
		Speed of Service			("quick" OR "fast" OR "speedy" OR "hasty" OR "swift" OR "inefficient" OR "r
		Knowledge			(advice OR expert OR expertise OR experts OR incompetent OR incompeter
		Appearance			(overweight OR "professional appearance" OR "professional looking" OR un
		Attitude			(friendly OR pleasant OR professional OR courteous OR attentive OR perso
		Communication skills			(conversation OR converse OR advice* OR talk OR talking OR telling OR disc

Figure 20 Example of a model spreadsheet

2. General Information

This chapter explains the basic principles of the process and defines the terminology.

2.1. Where do you Start

In the knowledge that we are seeking, the subjects are typically your customers or consumers, and the objects are typically elements of your products and services. The relationships are the customers' views and opinions; their sentiments. The first step in creating a model is to determine which objects you are interested in. Objects can be grouped into categories, and the objects within each category are arranged as a hierarchy. The depth of the hierarchy for a category depends on the level of granularity needed to determine the knowledge that you can act upon, and also on the volume of texts that will match the category.

For example, the ordering channel for a product could be categorized as "Online", "Call Center" or "Store". "Staff" could be divided into "Managers", "Associates", "Cashiers" etc. Remember that you will need to obtain these keywords from your customers' verbatims, so you will need to create sub-categories based on the terminology that your customers use. The temptation might be to refine the hierarchy to many levels, but often the consumer does not easily distinguish things to the same detail so it may not be worthwhile spending time defining a complex model.

On the other hand, not refining the model sufficiently will mean losing out on valuable intelligence. For example, having a single category of "Flight" does not allow you to separate out analysis of different aspects of the flight, such as legroom, cabin service, menu choice, entertainment, etc. These are items that customers are likely to express opinions about, so it makes sense to build these hierarchy levels into the categorization model.

At the same time as you are thinking about the objects that are being discussed in the texts, and how they can be classified, you should also think about the way opinions about the objects can be categorized. For example, what opinions could be expressed about the staff in a store, or about the flight experience? How helpful were the staff in the store, or how knowledgeable were they? How comfortable were the seats, how extensive was the choice of in-flight entertainment etc? These are Attributes and they need to be carefully thought through such that they can be segmented into useful groupings that can typically be determined from analysis of textual content. While we might be interested in consumers' opinions of the change in color of our logo, they are unlikely to express an unsolicited opinion about it. If we were to ask them for an explicit opinion using an open-ended question, then we already have the context and we do not need to look for it in the text.

2.2. Asking the Right Questions

One way to approach building your categorization model and associated attributes is to think about how you would ask structured questions. With text analytics, you are effectively asking questions of the open-ended text, and the answers are in the structured output you get in the categorization model and the sentiment analysis of the opinions in the text.

Using an open-ended question instead of a series of closed questions means you are not explicitly asking about a particular aspect that you are interested in, such as the friendliness of the cabin staff. However you can imagine asking that question of the open-ended text responses and using that thought to help build your categorization model. This can also help to segment which questions you want to ask explicitly in a closed-question format, and which questions you can leave to be asked by the analysis of open-ended text responses.

In a survey, the way an open-ended question is worded impacts the answers. If you say to a customer "Tell us why you like the product", then you will get answers about aspects of the product but not necessarily opinions, such as "The price", "The design", "The range of add-on modules available", etc. This is fine as long as you have previously determined from a closed question that the customer actually likes the product, as you can then derive actionable insight from this. Furthermore, an open-ended question that is non-neutrally worded (such as "What are we doing well?" or "What do we need to improve?" will likely have unpredictable effects upon any sentiment engine's results.

If you are considering analyzing unsolicited comments, for example from social media or call-center notes, then ideally you should analyze these comments alongside your survey comments. In this case your survey question should be less loaded, for example "Please provide any further comments you have about the product". Your text analysis will then ask the questions of the solicited response text in the same way as it is asking questions of the unsolicited text from support or social media interactions

2.3. The Elements of a Categorization Model

The Genius categorization model is built in a tree hierarchy structure. Some basic definitions:

- **Category -** an object that you have defined within the business (see Categories and Sub-Categories on page 17 for more information).
- **Sub-category -** a further refinement of the category.
- **Attributes** – The elements of opinion. These are applied to categories and sub-categories, and have their own expressions (see Attributes on page 18 for more information).

2.4. Categories and Sub-Categories

The categories and sub-categories are the main parts of the categorization tree. The categories describe the main touch points of the business, such as 'Staff', 'Store' and 'Product', and these can be broken down into sub-categories that further define these categories. For example “Staff” could be broken down into 'Management', 'Associates', 'Trainees' and 'Other Staff', while 'Product' could be broken down into the different products that you offer, or sub-categories based on groups of products.

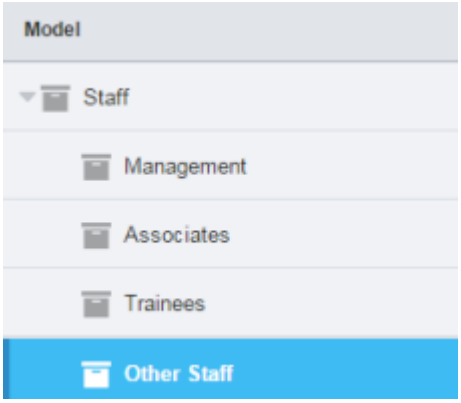


Figure 21 Example of categories and sub-categories

Expressions are added onto the child nodes, that is the sub-categories that have no further children (in this case Management, Associates, Trainees and Other Staff). This is where you add all the keywords that will add verbatims into that category. Parent nodes, that is the categories that contain leaf or child nodes (Staff in the above example), do not have their own expressions. Their expressions are made up of the expressions of the child nodes joined with an OR Boolean operator.

You can easily see whether your categories or sub-categories contain an expression by looking at the icon in front of the name. If the icon is colored, there is an expression associated with it. If the icon is clear, there is no expression.

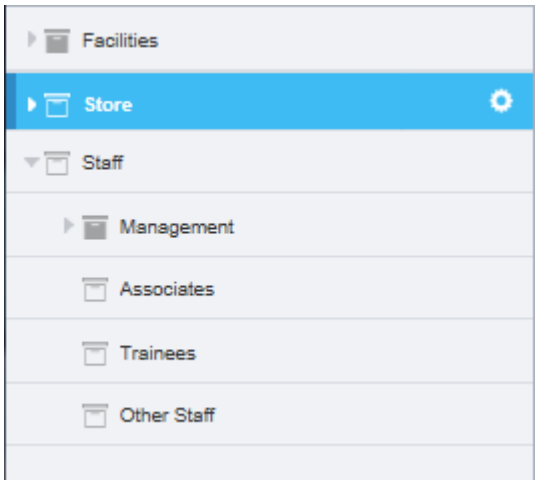


Figure 22 Example of complete and incomplete categories

In the example above, Management has an expression, but Associates, Trainees and Other Staff do not. Staff is therefore also shown as incomplete because of the incomplete nodes beneath it.

Note: If you add a sub-category to a category that already has an expression, then you will be asked if you want to move the expression to the sub-category you are creating. If you decline this invitation then the expression will be deleted from the parent category.

Warning
To avoid disappointment, save your work at regular intervals, and ensure you save before you leave Model Builder unattended (see Inactive-Screen Lock on page 3 for more information).

2.5. Attributes

Attributes are the elements of the categorization model that you wish to measure; your customers' opinions. These could be for example Product knowledge, Friendliness, Speed of service of the Staff, Location, Opening hours, Cleanliness of the Store, Availability, Reliability, Variety of the Product etc.

These attributes can only be applied to the lowest node of the tree. Attributes are joined to their category / sub-category nodes with an AND Boolean operator. So the expression that Genius will use for determining if text is referencing the friendliness of management could be:

```
(manager OR managers OR management OR "mgr" OR supervisor OR "in charge")
AND (friendly OR friendliness OR "not friendly"~3)
```

By analyzing only those verbatim texts that match this expression, Genius can then use its sentiment algorithm to determine the opinion expressed by the customers regarding the friendliness of the manager.

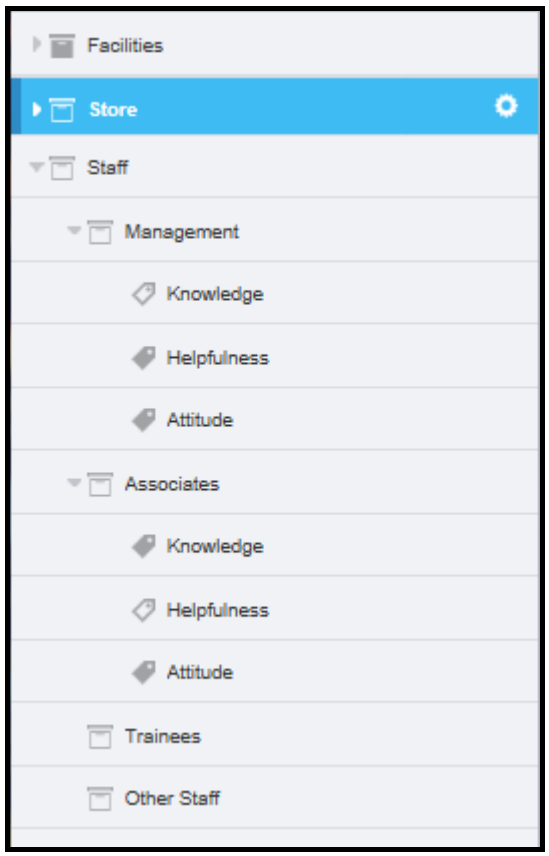


Figure 23 Example of categories, sub-categories and attributes

The example above shows the attributes for Management and Associates. Some of the attributes do not yet have expressions, so the icons are shown as empty. The “incomplete” status for the icons is passed back up to the parent node (Staff) so that you can see that there is an empty attribute even when the hierarchy is closed.

2.6. The Uncaptured Attribute

When setting up the model, you have the option to automatically generate attributes for uncaptured verbatims. This option is also available on the Model Overview page (see The Model Overview Tab on page 38 for more information). If you select this option, once you have at least one attribute added to the category tree an additional attribute **Uncaptured** appears in the tree. This automatically shows you all of the verbatims that have been captured by the sub-category expression (“Sales” in the example illustration) and excludes the attributes that have already been set up (“Helpfulness” and “Knowledge” in the example illustration). Topic Discovery shows a list of words that are used when the respondent talks about the ‘Sales’ function. You can drill deeper into these words (see Topic Discovery on page 52 for more information) in order to find keywords that can be added to existing attribute expressions or new attributes that should be added.

The Uncaptured Attribute is useful for helping to build out your model, so it’s worth switching it on when you’re in model building mode. You can also transfer the results for Uncaptured Attributes through to your reporting (use the ‘ua’ prefix in the Database Designer table). If you do not want these results in your reporting, ensure the toggle is switched off and save your active version so that they are not present in your category list.

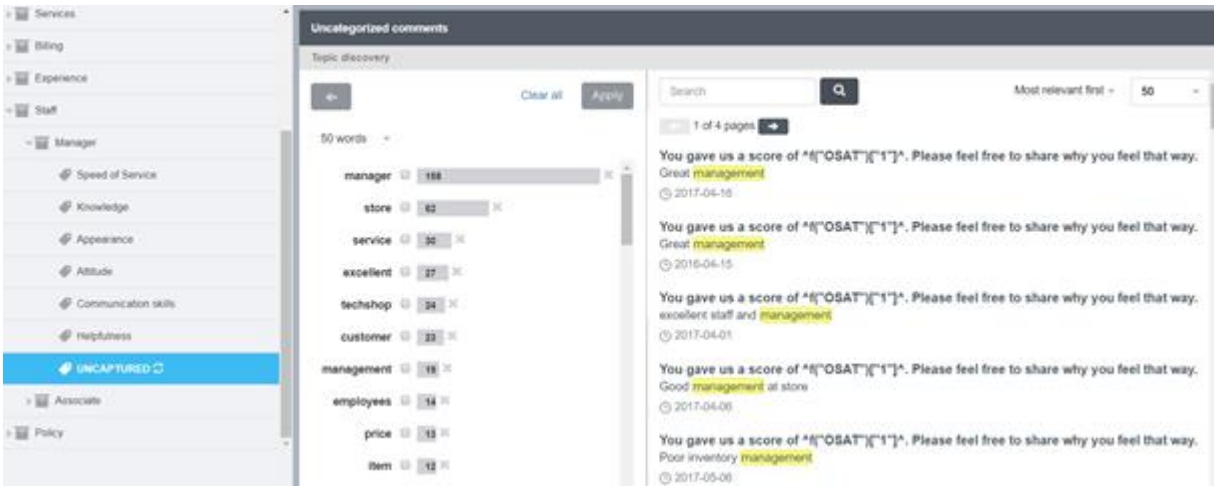


Figure 24 The Uncaptured attribute and Topic Discovery

The Topic Discovery words are useful in helping to understand whether you are missing keywords for your current attributes or whether you need to define some more attributes based on the frequency of the topics in the Topic Discovery word list (see Topic Discovery on page 52 for more information).

2.7. Boolean Operators

You need to use Boolean operators in the search field when you write search expressions. The Boolean operators combine search words in your expression.

Logical AND

The AND operator will combine two (or more) search words. For example if you want to search for posts containing both of the car manufacturers Audi and Mercedes, your search expression will be *audi AND mercedes*. Only posts where both words occur will be returned in the search result. Note that the order in which the words appear, and their distance apart, will not be taken into consideration.

Logical OR

The OR operator will return posts where any of the given search words are represented. For example if you want to search for posts that mention either of the phone companies Orange or Vodafone, your search expression will be *orange OR vodafone*. Any posts containing either Orange or Vodafone will then be returned.

Logical NOT

The NOT operator is used to exclude posts that contain the specified search words. For example if you want to search for all posts that mention iPhone, but exclude posts that mention Samsung, your search expression would be *iphone AND NOT samsung*.

Note: Logical operators are always written in UPPER CASE. To make your search expressions easier to read we suggest that you always write your search words in lower case. For example both "AUDI AND MERCEDES" and "audi AND mercedes" will return the same results, but the latter is easier to read.

Grouping the expressions using parenthesis

The search expression can be grouped by parenthesis as one logical entity, which can be further connected with other logical entities by Boolean operators. For example, *(hotel AND lodging) AND (clean OR cleaned)*. This expression will provide results which contains both hotel and lodging with clean or cleaned.

Parenthesis are very useful and can be used when nesting logic blocks together. In the example above we have *a = (hotel AND lodging)* and *b = (clean OR cleaned)*, but if we add *((hotel AND lodging) AND(clean OR cleaned)) AND NOT (dirt OR dirty)* we are creating *c = (dirt OR dirty)* to exclude dirty hotels and lodgings.

To make this easier we can think of this as *(a OR b) AND NOT c*.

Quoted search

When we have keywords that occur together in an exact order, using quoted search makes it possible to capture this. For example, say we are looking for *check in* at a hotel. You can search for *((check OR checking) AND in)*, but this would yield too many results. The more elegant solution would be to use *"check in" OR "checking in"* to preserve both recall and precision.

Proximity search

You can use the proximity operator (~) to specify a maximum number of words between your search words. If you want to find information about clean rooms you would use the proximity search. Your search expression could then be *"room clean"~3* . In this example the expression will return only those posts where there are a maximum of three words between *clean* and *room*.

For example, *"The room was clean"*, *"The room was very clean and nice"* and *"The room was bright and clean"* are all valid results.

One important thing to be aware of is that the proximity search will also find the reverse order of your query, but the reversal will count as two words. This means that *"room clean"~3* also finds *"clean room"~1*. This means that *"We got a clean bright room"* is also a valid result, but not *"We got a clean and nice room"*.

You will need to test the maximum number to use with your proximity indicator, to see which number produces clean results and which starts to bring in unwanted results. Proximity operator can also be used when there are more than two keywords in the double quotation mark.

Wild card search

The wildcard operators (* and ?) ensure that all words that start or end with the preceding or following text characters will be captured. For example, *categor** will find words such as category, categories, categorisation and categorization, and **common* will find words such as common and uncommon. Wild cards do not work within quotation marks.

The ? can be used to wildcard a single character in a word. For example both Categorization in US English and Categorisation in UK English could be captured by using the keyword *categori?ation*.

Fuzzy search

Fuzzy search returns more search results by allowing some "errors" in the search term. For example, *communication~* will find words like communication, communications, comunication, comuunication etc. When doing fuzzy search, the tilde sign (~) must follow immediately after a term. Optionally you can specify how many characters are allow to be wrong in the term. For example, *communication~1* allows one character to be wrong, while *communication~2* allows two mistakes. If there is no parameter specified, the fuzzy search will allow two characters to be wrong by default.

Tip!
Fuzzy search is more useful for long keywords.

Regular expressions

Regular expressions can be used as part of the expressions. Model Builder supports regular expressions through Lucene. Go to the link below for help on writing these expressions.

```
https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-regexp-query.html#regexp-syntax
```

Contact your Account Manager if you want to learn more from our Analytics team.

Exact search

The *content_exact* search term allows us to search for something exactly as it is written, with punctuation, hyphens or other symbols, which the normal search does not provide. For example searching for *"off-white"* will give you results that contains off-white, but also off and white together. Using *content_exact:"off-white"* will provide only results that contain off-white. The table below lists some query and result examples.

Query	Example results from standard search	Example results from content_exact search
W210	W210, W-201, DSC-W210, W - 210	W210
wi-fi	WiFi, Wi, Fi, Wi-Fi	Wi-Fi, wi-fi
SX	SX210, 180SX, 40S-X	SX
oleary	OLeary, O'Leary	oleary
SD500	SD-500, SD500, SD 500, SD 500s	SD500
42	42, 6.42, 42kg	42

Table 1 Results from standard search and content_exact search

Special characters

Characters that are not alphanumeric cannot be searched for like normal characters. However they can be used when prefixed with content exact.

Example: *content_exact:$*, *content_exact:50%*

Wild cards can also be used with the exact search like this: *content_exact:$**

Some characters are special and have been reserved by the search system. They can still be searched for, but they have to be escaped. They are escaped if they have backslash (\) before them or double quotation marks ("") around them.

Examples: *content_exact:\+, content_exact:"+"*

	Special characters	Special characters
+	)	“
-	{	~
&&	}	*
||	[	?
!	]	:
(	^	/

Table 2 List of special characters that must be escaped

Term counter

You can search for the number of terms that occur in a response using *content_termcount*. It is also possible to do a range lookup. This type of filtering can be useful when you have lots of responses with only a few words, typically one to four that you want to filter out. For example: *content_termcount:1* will find any response that has exactly one term.

Range search example: *content_termcount:[1 TO 4*] will find any response that has from one to four words in it.

Length

Searching for the number of characters used in a response can be done with *content_length*. As with *termcount*, this can also be range lookup. For example: *content_length:3* will find all responses that have exactly three characters, and *content_length:[1 TO 5]* will find all responses that have one to five characters.

References and Keywords in Model Builder search expressions

Model Builder offers a range of fields that you can use to better refine your search queries. The following lists all these fields and gives examples of how they will impact your search results.

Field name	Explanation	Examples of usage
subject	Performs a search that looks for matching terms in the subject of a comment. The subject stores the survey question label from Horizons.	subject:why subject:"Why did you give us that score?"
date	Performs a search with date restriction.	date:"2010-01-01" (will only give results from January 2010). date:[2010-01-01 TO 2010-06-01] (will only give results between January 1st 2010 and June 1st 2010)
time	Limits your search to a specific timestamp.	android AND time:[00:00:00 TO 01:00:00] (will give you all talk about Android between midnight and 01:00)
content_exact	The exact version of the content field. This will give stricter results.	content_exact:"iPhone7+" content_exact:"$500"
language	Limit your search to a specific language. Uses standard language unicodes.	language:en language:se
projectid	Enables you to use the Horizons projectid (survey p-number) in your query.	projectid:p189794685
content_termcount	Searches for a specific number of terms or a range of terms.	content_termcount:1 (will only give results that has one word in the response). content_termcount:[1 TO 5] (will give results that has from one to five words in the response).
content_length	Searches for a response with a specific number of characters or a range of characters.	content_length:1 (will only give results that has one character in the response). content_length:[1 TO 5] (will give results that has from one to five characters in the response).

Table 3 Searchable fields quick reference

3. Importing a Model

You can import a new model into Confirmit from a specially formatted Excel file. To find the Import functionality, click on the drop-down on the **Create model** button. You will need to follow the wizard in order to be able to import your model structure into Model Builder.

3.1. Download the Excel Template

To download an Excel template:

1. In Model Builder, click on the drop-down on the **Create model** button.

   The Create New - import screen opens.

Figure 25 The Create New - import overlay

2. Click **Download**.

   The Excel template downloads.

3. Open the template to see an example of how the file must be formatted for your categorization model.

   You can use the downloaded template as the basis of your file.

Techshop					
Category	Sub-category	Attribute	Description (will not be imported)	Example keywords (will not be imported)	Expression
Staff					
	Sales				sales OR "account manager"
		Knowledge			knowledge OR detail
		Professionalism			professional*
		Communication			email OR communication OR language
	Services				services OR tam OR tss OR "technical services"
		Speed of service			speed
		Knowledge			knowledge OR detail
		Professionalism			professional*
	Technical Support				"technical support" OR "tech support" OR support
		Speed of service			speed
		Knowledge			knowledge OR detail
		Professionalism			professional*

Figure 26 Example of an Excel template

Add your model definition to the spreadsheet in the required format (don't forget to add a name in the first column and row). Ensure you follow the format, with category, sub-category and attributes on different rows. You can also add in an expression, which will be imported into the model. Do not leave any line spaces between the Category header and where you start adding category names.

4. Save your file as an .xlsx file.

3.2. Importing your File

1. On the Create new - import screen, click **Choose file**.
2. Find and select the .xlsx file that you have previously saved

   The selected file is indicated on the screen.

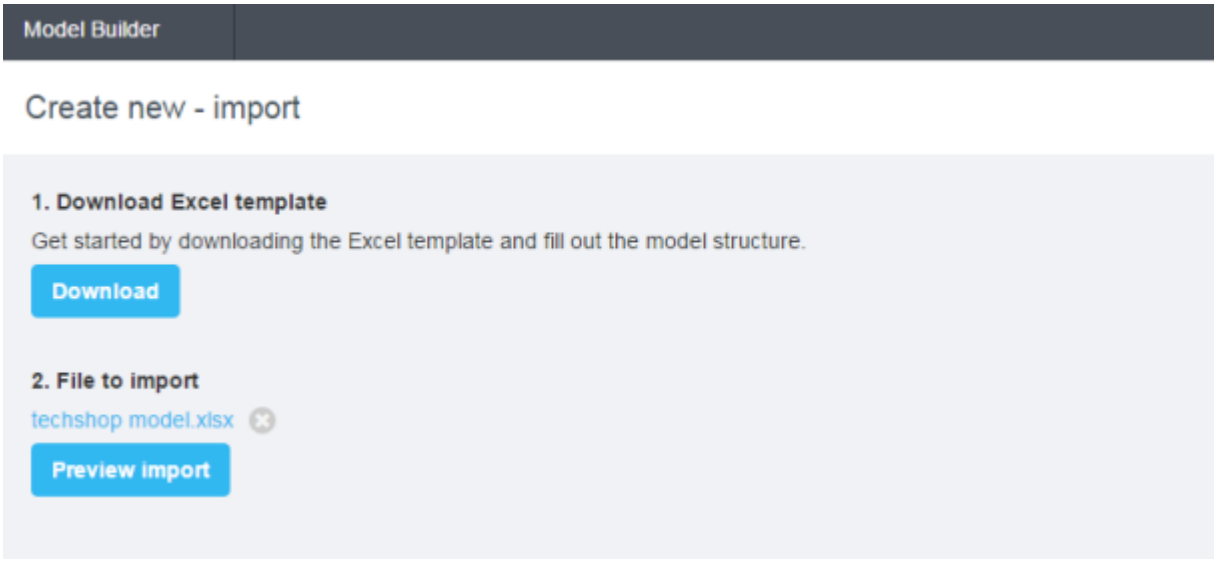


Figure 27 Selecting the file to import

3. Click **Preview import**.

   The preview page opens.

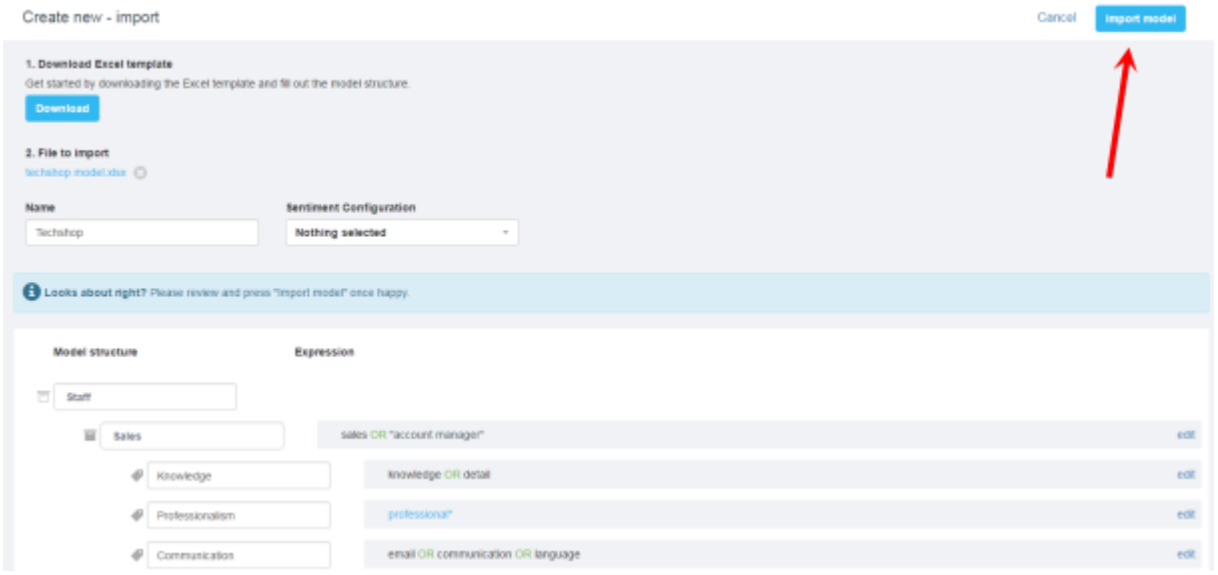


Figure 28 Previewing the import file

4. Review the file. If it is correct, click **Import model** (arrowed).

   The model is added to Model Builder. Don't forget to 'Save as version' and set to 'Active' once you are ready to use it in Horizons.

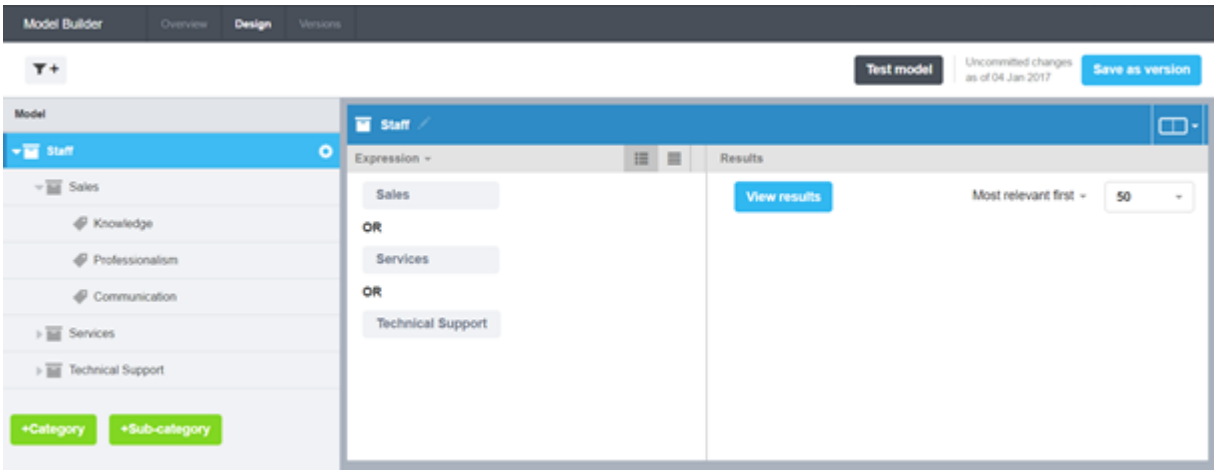


Figure 29 The file in Model Builder

If the file has some errors, you will need to fix them before you can import the file. You will see the errors, as there will be an error caution in front of the category name and the expression box will be colored red.

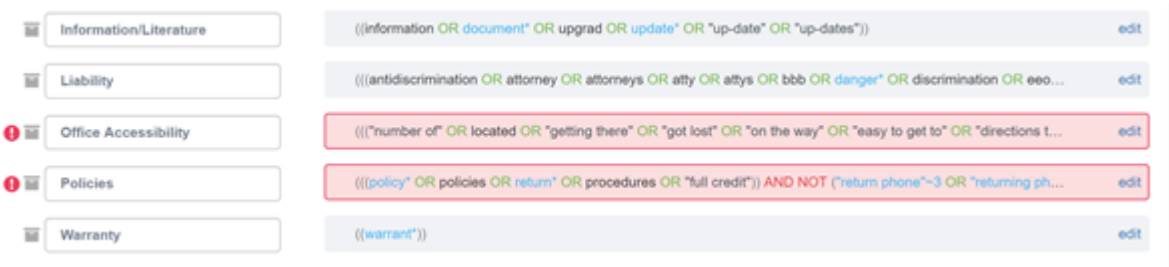


Figure 30 Example of a file with some errors

To fix the errors, click on the **edit** link and a pop up box will appear which will show the expression that needs to be fixed.

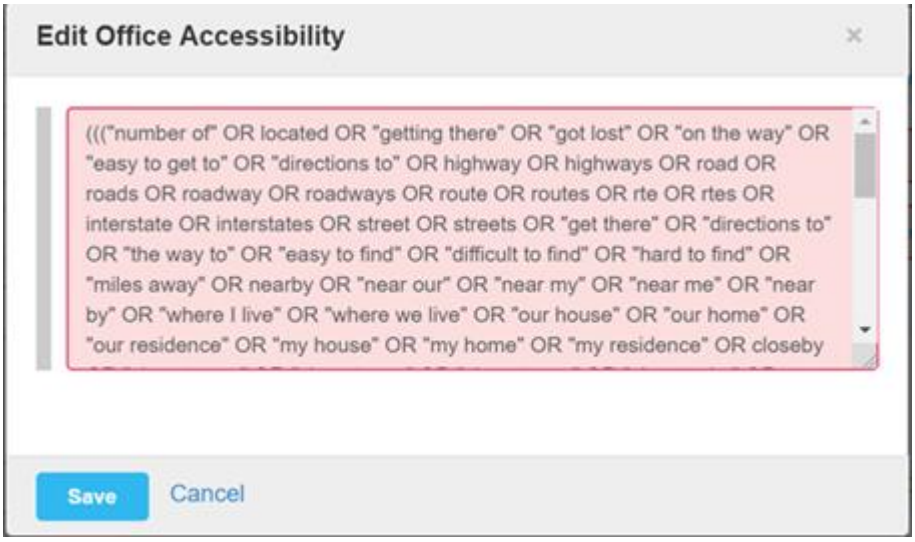


Figure 31 Example of a file with some errors

You will need to go through the expression and correct any mistakes. Things to look out for include:

- Ensuring you have the correct number of parentheses (using Notepad++ helps to show that you have the correct number, and helps to align where these should be put).
- Ensuring you are using Boolean operators in the correct places and that you have not got repeated operators. Ensure your Boolean operators are all written in full upper-case.
- If you are importing a model from another vendor, you will need to remove any parts of the expression that are specific to that vendor. If you are having difficulty working this out, please speak to Support, who will be able to help you.

Once your model has no more red boxes, you will be able to import your model.

4. Concept Miner

Concept Miner takes your project verbatim and clusters them into the most talked about topics. It uses machine learning to analyze the themes according to how the respondents talk about them. The topics are not linear, as they would be using Boolean, so you might find that the topic talks about staff friendliness and efficiency in the same topic. You can select the topics that are appropriate and make them into tags, which can be used in your categorization model, alongside your normal Boolean queries. Tags understand a lot more than keywords, so you will be able to cover more of your verbatim with less work.

4.1. Getting Started

Go to the Concept Miner link in the main navigation bar. If you cannot see this link, speak to your Account Manager. The functionality is currently in "Limited Availability" so only stakeholder customers who have agreed to provide feedback have access to it.

Concept Miner opens at the Core list page. A core is the list of topics that the Concept Miner finds, based on the verbatim analyzed when it is created. It contains the knowledge of your data.

If you have any existing cores these will be listed in the core list page. The first time you go here you will not have any cores so the list will be empty.

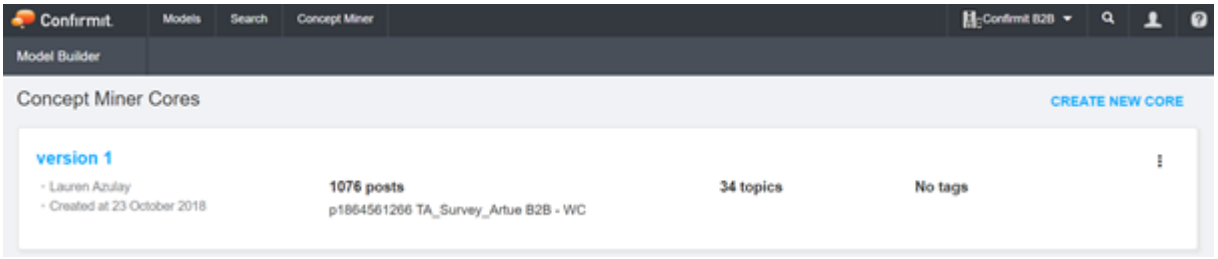


Figure 32 Example of the Core List

4.2. Creating a Core

First you need to create a core:

1. Click **Create new core** towards the right side of the page.

   The Create core page opens.

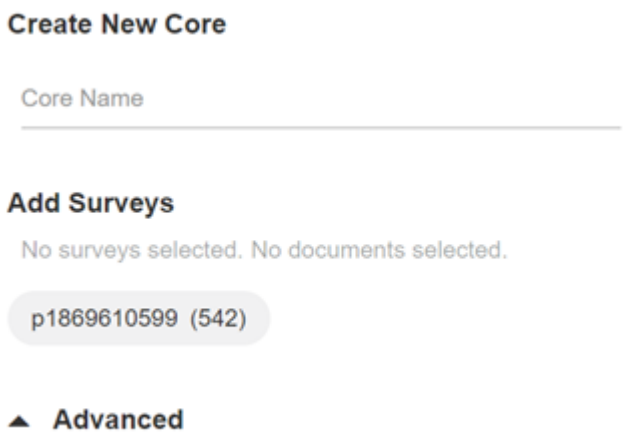


Figure 33 The Create core page

2. Provide a name for the core.

   It is recommended to give your core a descriptive name, such as the data you have chosen to run it on, so it is easier to find later.

3. Choose the data set that you want to run through Concept Miner.

   You will see all the projects that have previously been run for Text Analytics – you can choose one or more of these. Note that you must have more than 100 documents in the project. If you have more than 50,000 documents, the Concept Miner will take a random sample of 50,000 and run the core on this. It is worthwhile running the core on multiple projects if they have the same kind of data.

4. Decide on the number of topics you want to have in your core (35 is the default number - you should keep it at 35 to start).

   Advanced settings: You may skip this for the first run of data so that you can see what Concept Miner comes back with. You should however use this on subsequent runs to improve your data accuracy.

5. Click the **Create core** button at the bottom of the page.

   Your core is now being trained - return to the Core list page to view progress. Note that this can take some time.

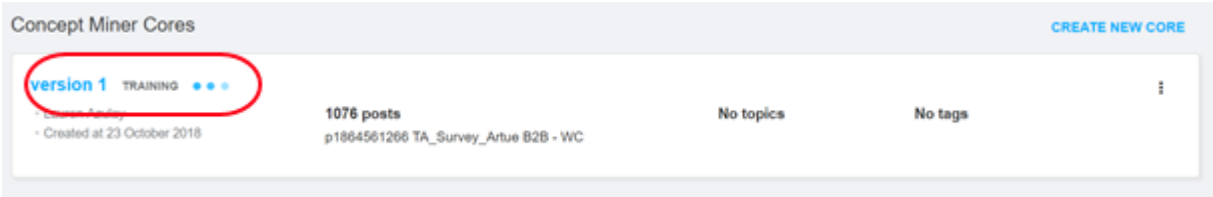


Figure 34 The Core List page while the core is training

On completion, the Topics found list is displayed.

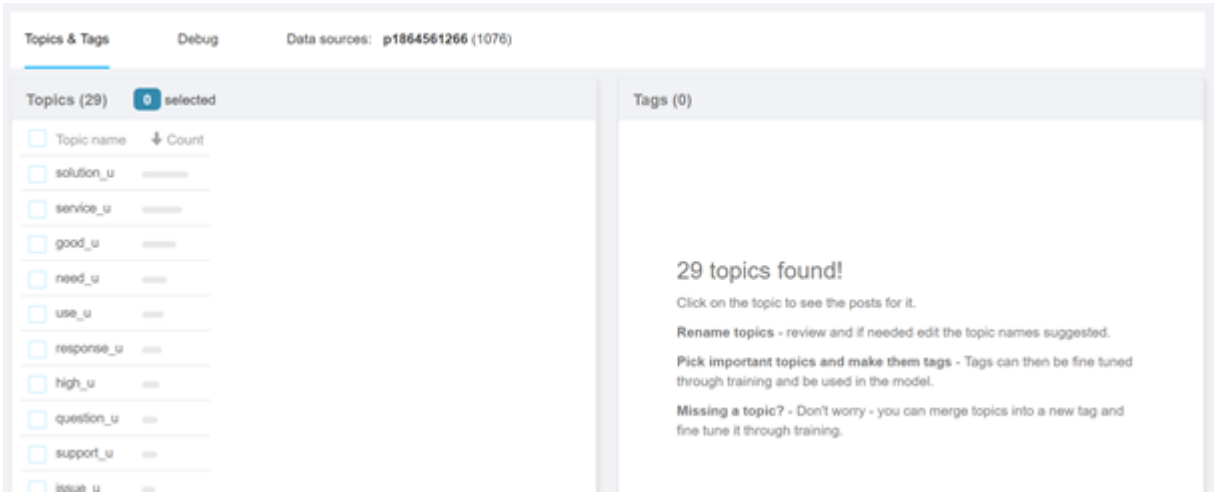


Figure 35 The topics found list

Initially, the topics will all be named with a **_u** suffix. This indicates that the topic is "unsupervised"; Concept Miner has suggested them based on its unsupervised analysis. You can run the core again using guided topics, which will give you a different core with supervised topics.

4.3. Analyzing the Topics

You now need to analyze the topics to assess what has been found and decide its relevance.

Click on one of the topics to view the detail contained in the topic. You will see the top words that make up the topic, and the verbatim that are captured within it. The bigger and darker blue the word, the more relevant it is to the topic. Less relevant words will be smaller and lighter blue.

The verbatim are sorted by confidence level, so the posts that it considers to be the best match for the topic are shown first (highest confidence). The green Confidence level slider indicates the distribution of the documents. As you move the slider, the confidence level is displayed and this confidence selection will change the verbatim viewed below.

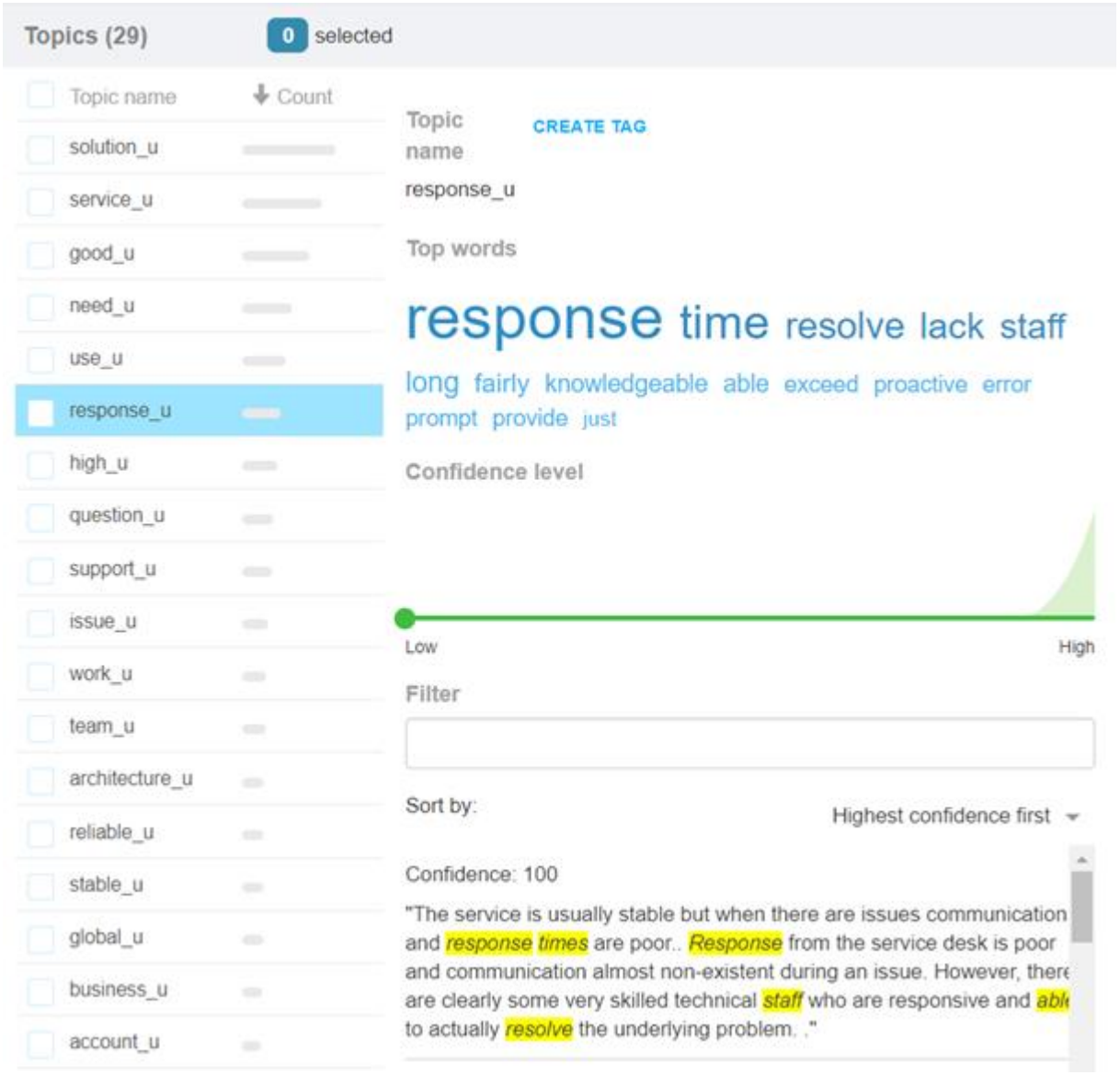


Figure 36 Viewing the topics and verbatim

When Concept Miner analyzes the verbatim, it finds topics based on the associations between words that have commonly been used together. It develops a knowledge of a theme, which will allow it to bypass bad grammar and spelling mistakes.

When viewing the topic, look at the top words and then go through the verbatim with the highest confidence to gain an understanding of what the topic is about. Make sure that 'Highest confidence first' is selected and go through these. Some topics will be very evident, some will seem like they are almost a topic but could do with some work, and some will not be good topics. Go through the key topics and note the ones that make sense, along with their top key words.

If desired you can then start to work on a new core based on what you now know from Concept Miner. You can add some guided topics and seed words (find this in the 'Advanced' drop down), which will give Concept Miner some guidance on what to look for in the text. If you already have a categorization model or a code frame, you might use that knowledge to seed Concept Miner here, or use the key topics that you've noted from your previous Concept Miner run(s).

Add the topic name that you want to use into the topic name field on the left, and add a comma-separated list of the words that Concept Miner needs to look out for in the field on the right. If you try to add a topic name without adding a list of seed words, an error will be indicated and you will not be able to continue creating the core, and the same will happen if you add seed words with no topic name. Click the blue **+ADD TOPIC** link to add more rows for further topics.

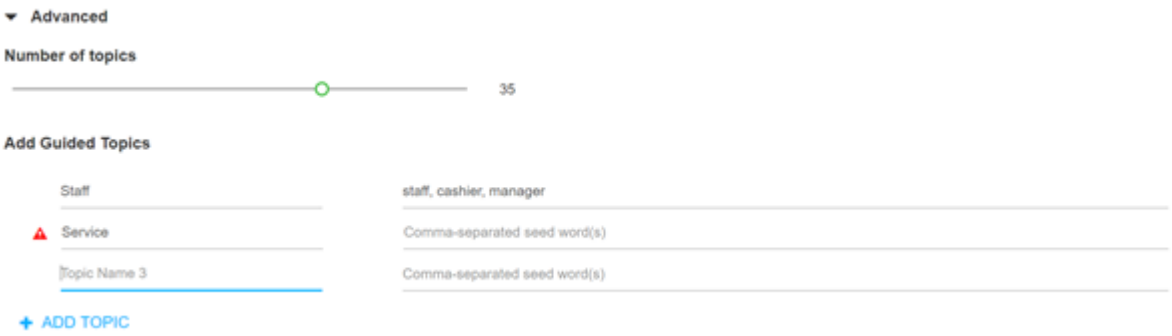


Figure 37 Adding topics

Tip:
When the Concept Miner runs, it first prepares the data. Part of that process is lemmatization, which breaks the word down into its root. This means that you do not have to add all forms of the word, but just the basic word, for example, Service instead of service, serviced, services, servicing, etc.

4.4. Re-running the Core

The second time you run the core, you should see results that are more relevant to your needs. You will notice that any guided topics do not have a **_u** suffix as these are not unsupervised. Go through the list of topics again and see if there's anything that you might want to add to the guided topics in a further run. If you're happy with what you now have you can start to review your topics to make tags to be used when building your category model.

Note: A tag is a label you give your Concept Miner topic to make it available for building your categorization model.

When reviewing the topics a second time, first check your top words and the verbatim that have been captured with the highest confidence. Then click on the arrow next to 'Highest confidence first' and select 'Lowest confidence first' from the drop down. You will then see the verbatim with the lowest confidence rating ordered from the top. You can adjust the slider to the level that looks right (that is, all the verbatim that are displayed are now definitely part of the topic), and then click **Create tag** at the top of the page. You'll see that the tag will be created towards the right side of the page. You will need to give the tag a name, and the confidence slider should be in the same position as where you left it on the left side. You can still adjust it here before saving your tag.

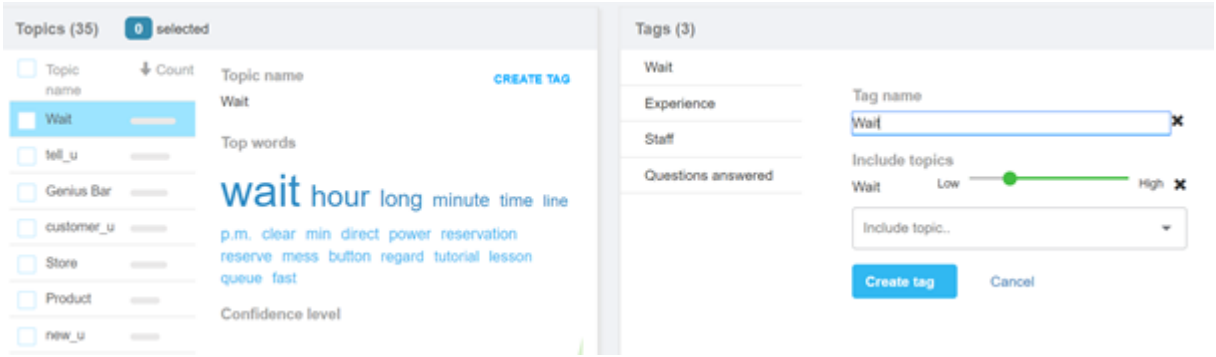


Figure 38 Reviewing the topics again, and adding tags

You can include other topics in your tag before clicking the **Create tag** button. These topics will then be given the **_u** suffix - see above.

Note: Once you have created the tag, you cannot change it. So if you have to change a tag then you will have to delete it and recreate it.

Once created, the tag shows which topics have been included, and the confidence level that has been set.

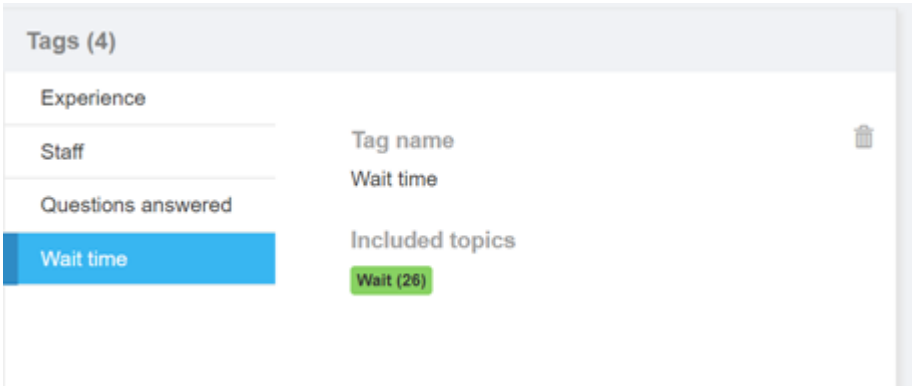


Figure 39 The list of tags

4.5. Quick Create Tags and Merge Topics as Tag

There are two other ways to create tags: “Quick create tags” and “Merge topics as tags”. These options appear once you have selected more than one topic from the list on the left.

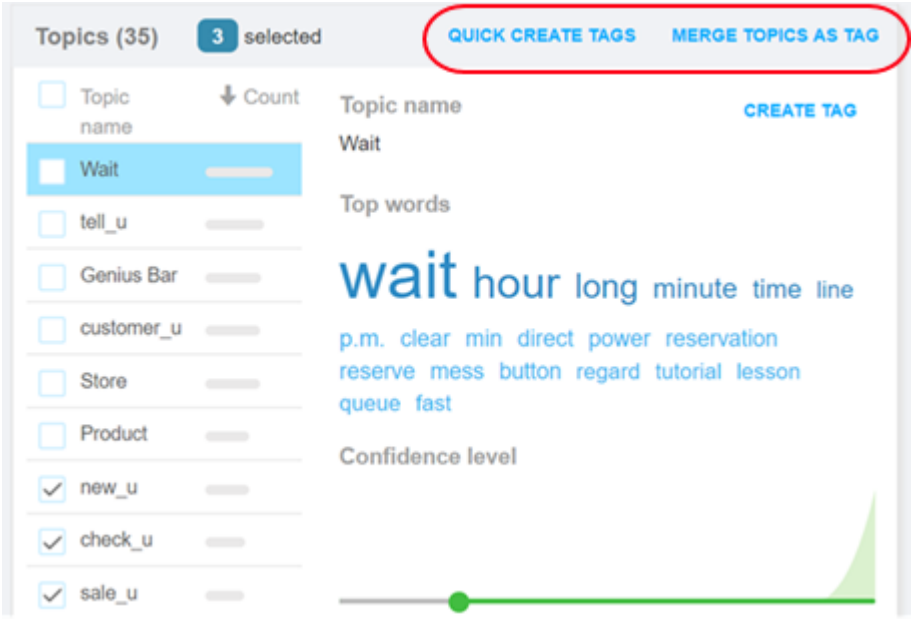


Figure 40 The additional options

If you want to use **Quick create tags**, it is recommended that you first do the work to ensure your topics are in good shape (that is, rename them to useful names and ensure that the confidence level is correct on each of them) before clicking it. The option will take the list of topics that you have selected and create them as tags, automatically, as is, on the right. Note that you will not be able to change their names or their confidence levels once you have "quick created" them.

Merge topics as tag allows you to take the selected topics and create one tag out of them.

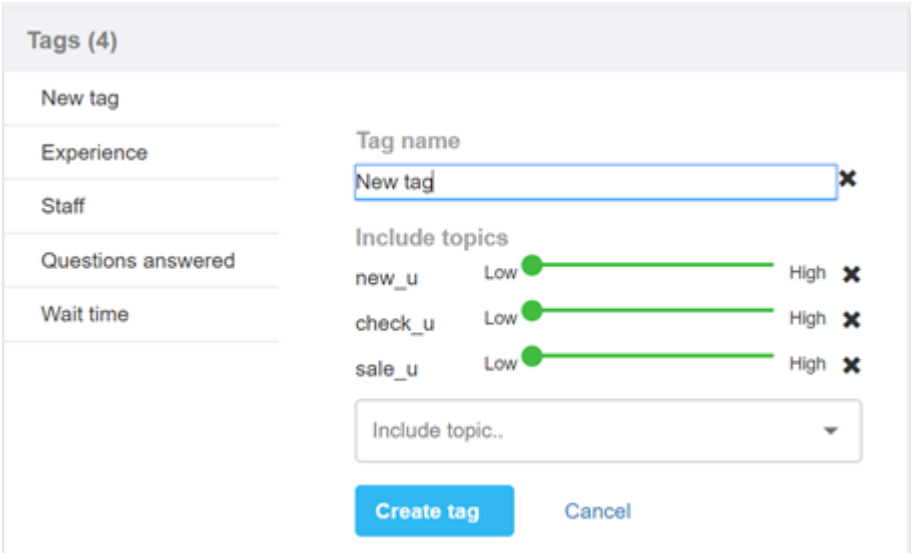


Figure 41 Merging topics as one tag

Give the merged tag an appropriate name and adjust the confidence levels on each of the added topics. Once you're happy with the adjustments, click **Create tag** to create the new tag. Note that you will not be able to change the name or the confidence levels once you have created the tag, you need to change anything then you will have to delete it, make the changes and recreate it.

4.6. Activating the Core

Once you have been through all of your topics and created the tags that you want to use, it is time to activate your core. To do this:

1. Go back into the Core list page.
2. Go to the **Options** menu.
3. Select **Activate**.

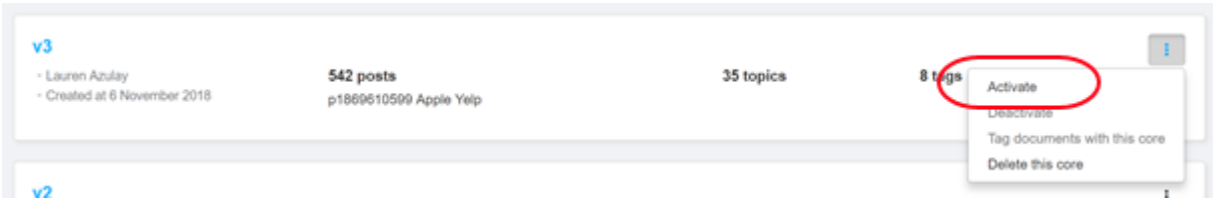


Figure 42 Activating a core

Once you have selected **Activate**, you will see that this core will become your 'Active Core'.

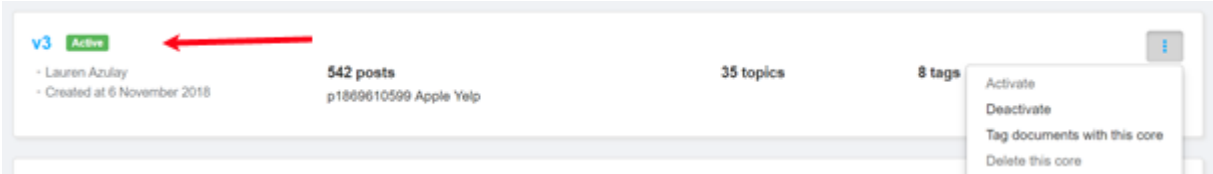


Figure 43 The active core

Now you need to make Concept Miner tag the documents with your active core. Select **Tag documents with this core** to start the process. It may take some time to process all the documents that you have in your Model Builder, and you can see the progress at the bottom of the core detail in the core list.

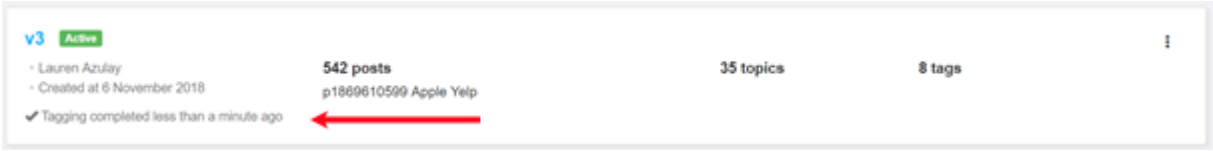


Figure 44 Tagging completed

You are now ready to use the tags that you have created in Model Builder. Go into the Model Builder tab and either choose to set up a new model or open an already created model. Set up a category and, in the expression box, start typing #. You will see the list of tags from your active Concept Miner core appear in the list. Choose whichever is relevant from the list. You can add Boolean as well as Concept Miner tags to your expression syntax.

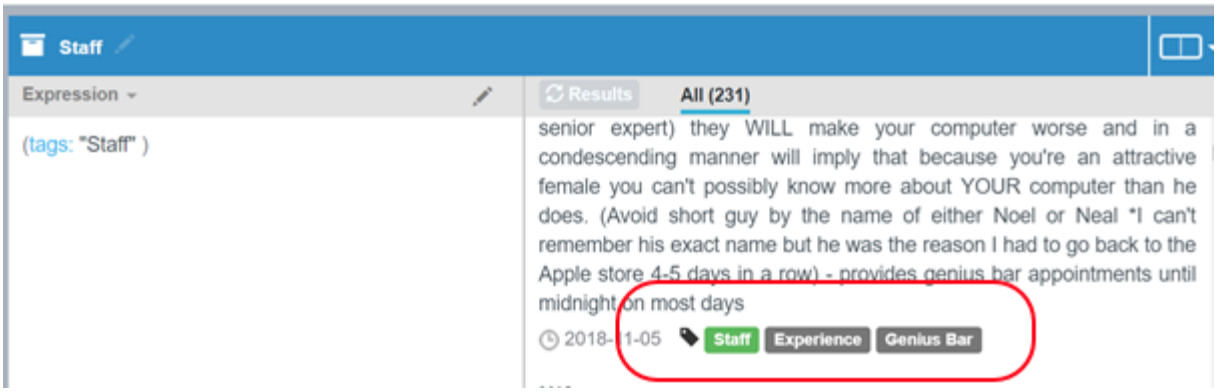


Figure 45 Adding the tags

When viewing results, now that the core has tagged all your documents, you will see tags as extra meta information for all of your verbatim or documents. The tag will be highlighted green if it is a tag that has been used in the expression that you're currently viewing results for. The tags will be grey if they are tags that are contained in the document but are not part of the current expression that you are viewing.

5. Creating and Maintaining a New Model

A categorization model must be built around how your customers are speaking about your business; borrowing a model from another business will not work. So before you start to build the categorization structure you will need to have a fairly detailed understanding of what your customers are saying about you. Based on this you can then identify the major themes or categories that define your business structure, and decide where Text Analytics will provide value by reporting sentiment about each category.

When creating a new model it is best to start simple and to build up more complex categories as you explore your verbatim data. Having identified the main categories, you must then identify the sub-categories. Here you should start with simple sub-categories; you can always create further sub-categories as you find them.

You can then create the expression for each sub-category.

Identify what customers are saying about these categories that you want to be able to measure; these will be your Attributes that the analysis will look for.

Create the expressions for the Attributes.

Warning
To avoid disappointment, save your work at regular intervals and ensure you save before you leave Model Builder unattended (see Inactive-Screen Lock on page 3 for more information).

5.1. How to Create a New Model

1. In the Models page, click **New Model**.

   The Create Model dialog opens.

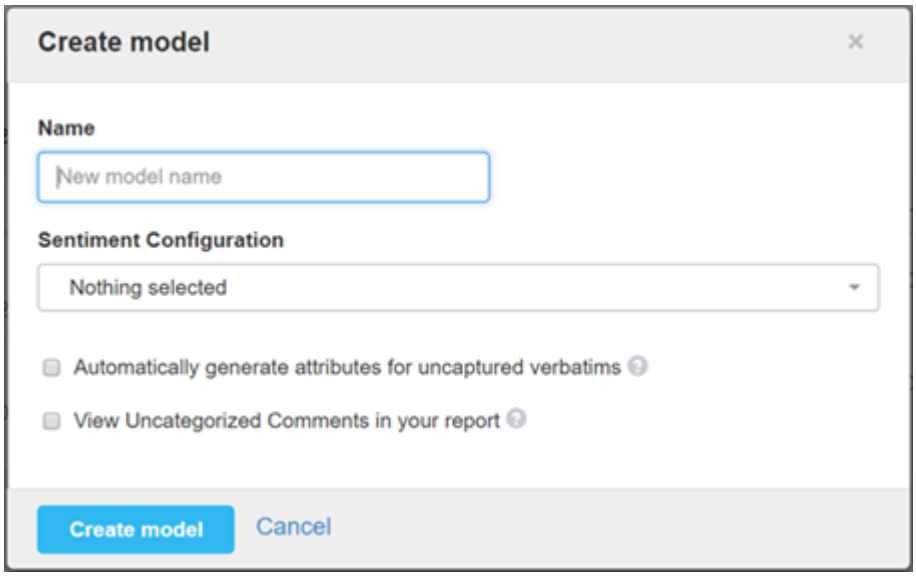


Figure 46 The Create Model overlay

2. Type in a name for your new model.
3. Select the sentiment configuration you want to use.

   The options available to you are in the drop-down list. You will either have the choice of the Confirmit English DL v1 sentiment configuration or perhaps a configuration that has been adjusted for your business. If in doubt, use the Confirmit English DL v1 configuration and speak to your Account Manager if you feel the sentiment needs some adjustment.

4. Decide whether you want to automatically generate attributes for uncaptured verbatims for your model.

   This helps to find other attributes mentioned in the comments for a category or sub-category (see The Uncaptured Attribute on page 19 for more information).

5. Decide whether you want to see Uncategorized Comments as a category in your report.

   This shows you all the comments that have not been captured by your categorization model, so could help you to understand any emerging topics or whether you need to perform some maintenance on your model.

6. Click **Create model**.

   The model is created and given a model ID number, and the Add Category overlay opens.

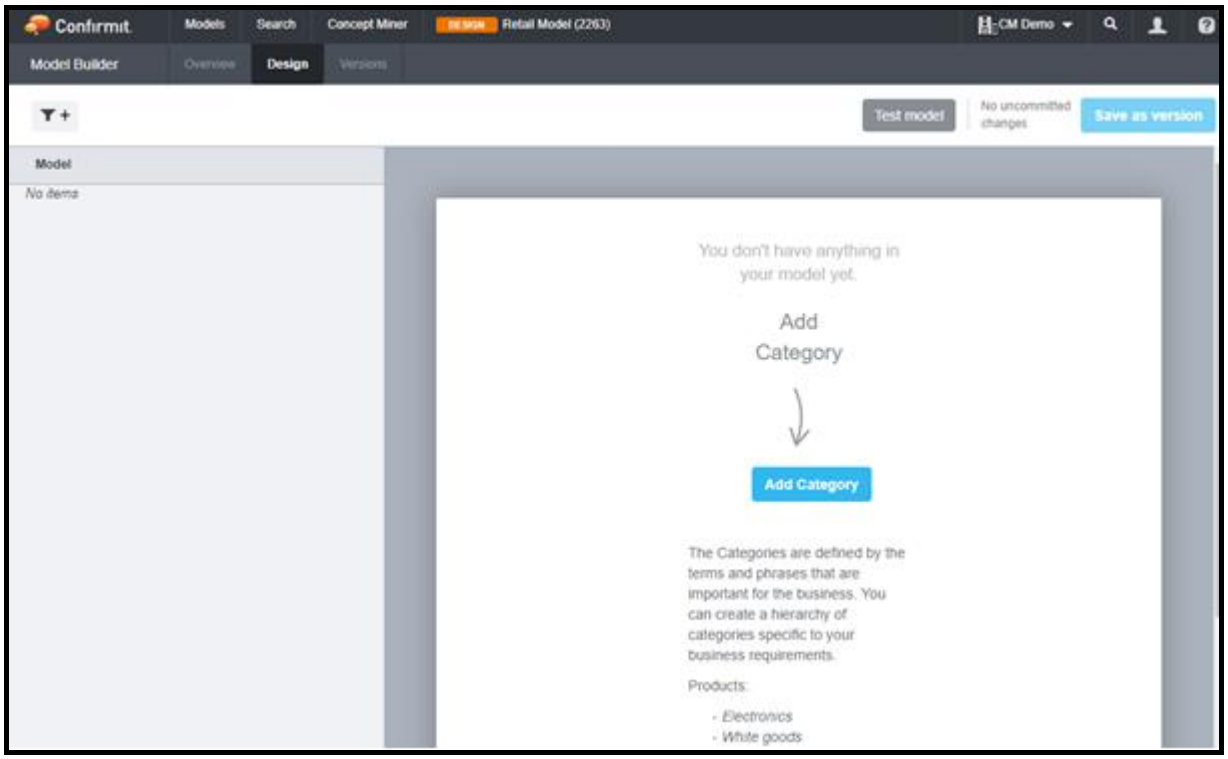


Figure 47 The Add Category overlay

7. Click **Add Category**.

   The Create a new Category overlay appears.

Figure 48 The Create a new Category overlay

8. Type the category name into the field, then press **Enter** or click **Create**.

   The category pane opens. You can now create your expression for this category or choose to create a sub-category if you want to refine your model further. If this category is to be a leaf node - the last level in the category hierarchy before you add attributes - then enter the required expression in this overlay.

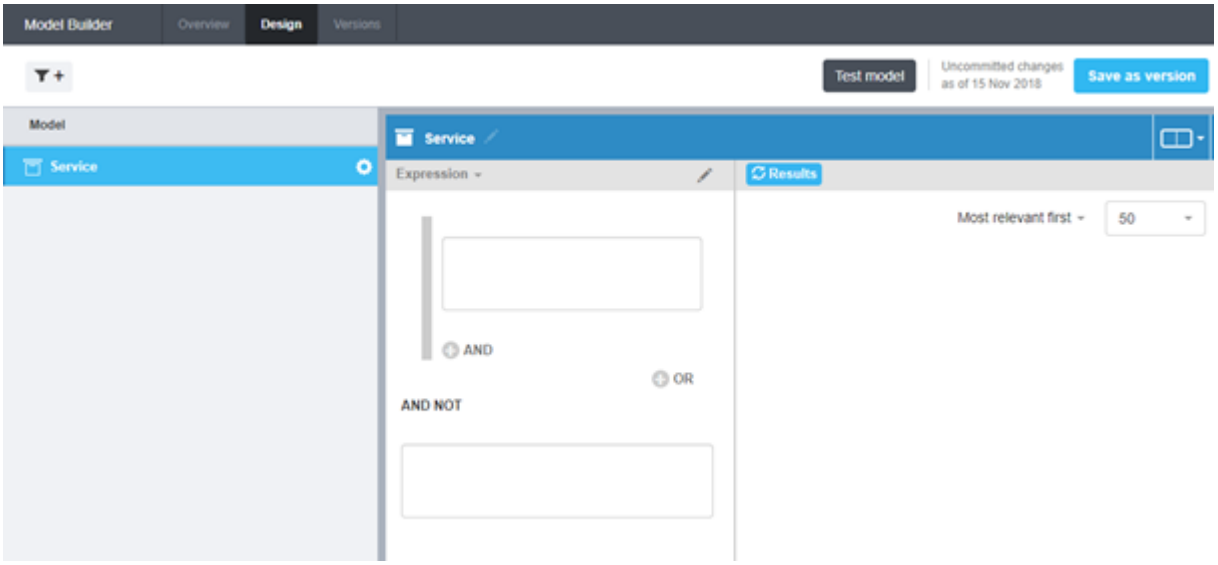


Figure 49 The Category pane

Warning
To avoid disappointment, save your work at regular intervals and ensure you save before you leave Model Builder unattended (see Inactive-Screen Lock on page 3 for more information).

5.2. The Model Overview Tab

The Overview tab contains all the background information about your model.

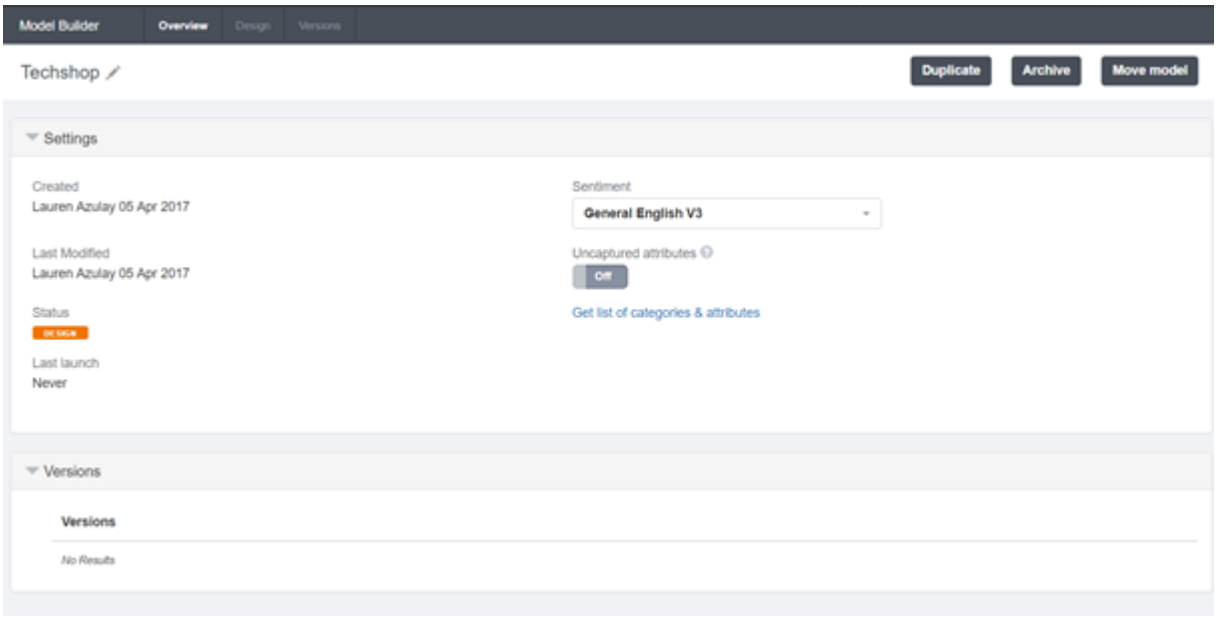


Figure 50 Example of the Model Overview tab

Use this tab to:

- Edit your model name (click on the edit icon next to the model name).
- See who created and who last modified your model, with the dates.
- **Status –** shows an orange **Design** button if the model has not yet been saved and marked as Active (this also means that it cannot be used in Horizons) and a green **Active** button if the model has been saved as a version. The first time it is saved as a version, it is set to Active. Note that you can only use a model in Horizons if it has the green **Active** button.
- **Last saved –** the date that this model was last saved as a version.
- **Sentiment –** you can select the sentiment configuration from the drop down menu. If you do not know which configuration to use, speak to your Account Manager.
- **Uncaptured attributes –** an on/off switch. If set to **On**, uncaptured attributes are generated automatically in your model. If set to off, the Uncaptured attributes folder is not generated.
- **Get list of categories and attributes list –** displays the list of categories, sub-categories, attributes and their associated IDs, along with the parent hierarchy reference, which you need to copy and paste into a Horizons database designer table.
- **Survey variables** (feature toggled) – if you have brought any survey variables through in your Text Analytics task, you will see the name, codes and answers in this section (see section xxxxx)
- **Versions -** a list of saved versions of the model, along with any notes that you have written when saving the version. Click on the name and it will take you into that saved version, so that you can test the model and view the results of the expressions. From this screen you can duplicate this version of the model, which copies that version and gives it a new model ID. The version that is currently active will have a green **Active** button next to it.

The buttons towards the upper-right corner of the screen are:

- **Duplicate –** makes a copy of the model and gives the duplicate a new model ID. You are redirected to the new model immediately, and the new model is called ‘Copy of ...’

- **Archive –** archiving a model takes it out of your Model list and puts it into your Archived models list. To view the Archived models list, click on the Active models drop-down on the Model List page (see Archiving Models on page 12 for more information). From the Archived models list you can restore the model back into your Active models list.
- **Move Model –** if you have access to more than one account you can move a model from one account to the other. Moving the model removes it from the originating account and puts it into the destination account.

5.3. The Design Tab

The Design tab is where you make any changes to the hierarchy and expressions in the model.

5.3.1. Creating the Categorization Hierarchy

You can now start to create your categorization tree. From your first category, you have 3 options:

1. You can click on the **+Category** button at the bottom of the categorization tree to add another category.

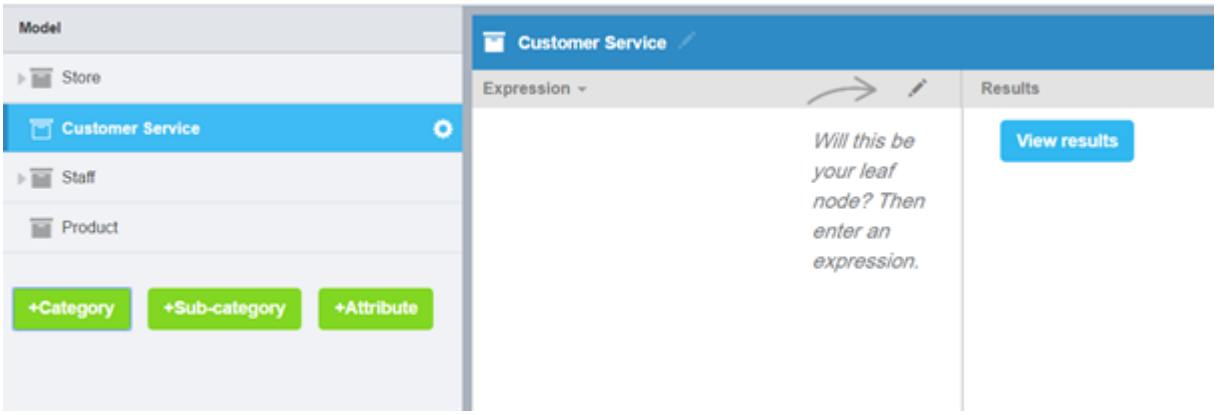


Figure 51 To add a Category

2. Click **+Sub-category** and **+Attribute** to add to add a Sub-category and an Attribute respectively.

Build up your categorization tree based on how your model needs to look, with Categories and Sub-categories describing the touch points in the business and Attributes describing the emotions or measures of those touch points.

5.3.2. How to Edit a Category Name

To edit the names of the categories, sub-categories or attributes:

1. Click on the **Edit** icon in the blue bar.

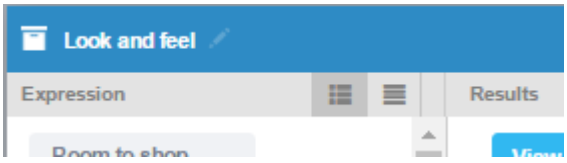


Figure 52 Click the Edit icon in the blue bar

The text field opens.

Figure 53 The text field opens

2. Edit the name as required, then press **Return** or click out of the field to close it.

5.3.3. Reordering Categories in the Category Tree

The order that the categories appear in the category tree is the same order that you will see in your report. You can drag and drop categories within the category tree so that they appear in the required order.

1. Move the pointer into the category tree - the pointer becomes a hand.
2. Use the hand to drag a category to the required location.

Note: You cannot drag sub-categories or attributes out of their current category – everything moves together.

Once you have completed re-ordering your categories, save the model version to Active to push the changes to the report. This will happen automatically if you have used the **Auto-sync categories list** function when setting up your Text Analytics task in Horizons. Refer to the Genius chapter in the Professional Authoring User Guide for further details.

5.3.4. Excluding Categories from the Report

You can exclude categories, sub-categories and attributes from your report. This is useful for generic catch-all categories or for excluding uncaptured attributes from your reports. To do this:

1. Click the cogwheel icon next to the category name to open the drop-down menu.
2. Select **Exclude from report**.

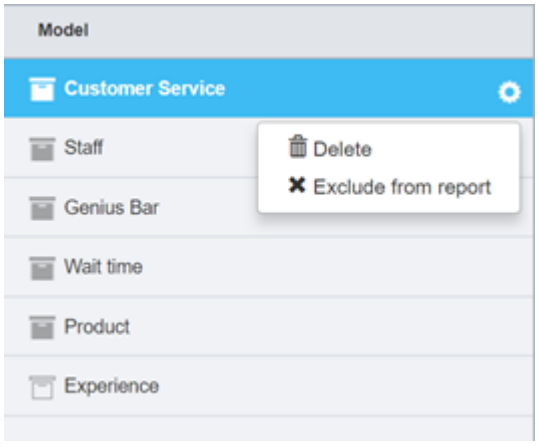
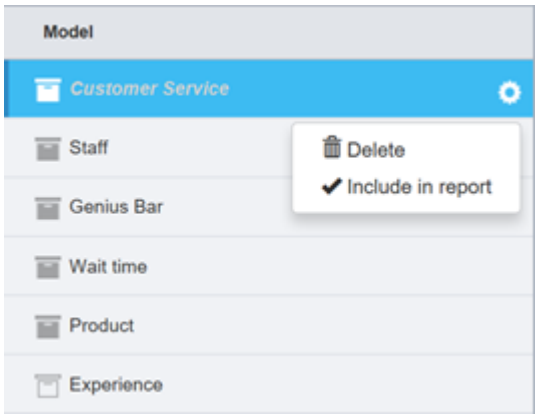


Figure 54 Excluding a category from a report

Once you have selected to exclude the category, it becomes grayed out and italicized. To include a previously excluded category back into the report:

1. Click the cogwheel icon next to the category name to open the drop-down menu.
2. Select **Include in report**.



Once you have completed excluding or including your categories, save the model version to Active. The changes will appear in your report the next time the Text Analytics task is run. This will happen automatically if you have used the **Auto-sync categories list** function when setting up your Text Analytics task in Horizons. Refer to the Genius chapter in the Professional Authoring User Guide for further details.

5.3.5. Survey Variables - Adding Survey Context into Models

You can add background variables into your model expressions for the purpose of:

- Giving additional context to the category.
- Improving the accuracy of your expressions.
- Limiting a category, sub-category or attribute to a value in your background data.

To be able to bring these variables into the model, you must first set this up in the Text Analytics task in Horizons. Once you have brought this information through to Model Builder, you can start to include the syntax in your expressions.

Note: This feature is not visible to all users – at present you must request access to it as it is behind a feature toggle.

Both the code and the label are transferred in the task and both are searchable. Both the code and label will appear in the Model Overview page.

To reference a survey variable that is a string, you must use this syntax:

```
SurveyType_s:”touchpoint”
```

where SurveyType is the exact name of your background variable, _s indicates the variable's value is a string, and "touchpoint" is the value you want to reference (in quotation marks).

To reference a survey variable containing values that are integers, you must use this syntax:

```
SurveyType_i:3
```

where SurveyType is the exact name of your background variable, _i indicates the variable's value is numeric, and 3 is value you want to reference.

Be sure to use the correct name of the survey variable that you have brought through from your survey, so that the system can identify it correctly and include it in your expression.

You can also search the code, not just the answer lists. To do this, use:

```
SourceProject_code_i:10
```

These are currently configured to be exact matches. The search is case sensitive, and it must match the entire string; partial matches are not accepted.

Refer to the Genius chapter in the Professional Authoring User Guide for further details on bringing the survey variables through from Horizons.

5.4. Writing an Expression

Your expressions are made up of the keywords or tags that describe your category, sub-category or attribute. Click on the **Edit** icon to start writing your expression.

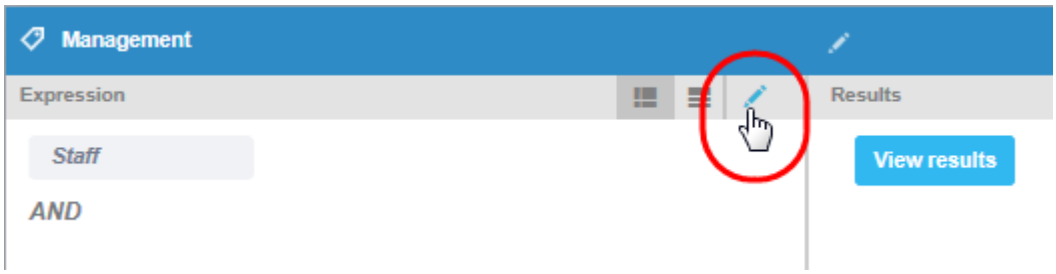


Figure 55 Starting the expression

You build up the expression by adding Concept Miner tags and combinations of keywords, and by testing the expression against your verbatims from the survey to ensure that they are collecting the correct verbatims for that category. When you start out, you can test keywords in the expression pane by clicking on **View results**. If you find you are getting verbatim that shouldn't be in that category, then you will need to refine your expression by using more targeted keywords or proximity indicators and AND NOT operators. Bear in mind that excluding keywords using AND NOT will ensure the entire verbatim is excluded from that category, even if both relevant and irrelevant information is in the verbatim.

You will also need to add in misspellings or use the fuzzy logic operator to ensure that misspelled words and words that are written in text speak are included in the analysis, although some of these may be captured by the Concept Miner tags.

If you have a longer expression, the process can be simplified by using the Search page (see Search on page 5 for more information). Here you can test tags and keywords, test the proximity of the words to one-another, and test the effect of using an AND NOT on your expression. It is recommended to review a good set of random results to validate that the expression brings through a good percentage of accurate hits. You should aim to reduce noise (unwanted hits) as much as possible, but this may not always be possible due to the unlimited number of ways people express themselves in text. You will need to decide how much noise is acceptable for each category. If you have an active core, you can also see other tags that are related to the verbatim, which you can add to another point in your model, to ensure that you're capturing everything. A simple goal is to improve the results until the point of diminishing returns — where the effort you invest starts to outweigh the improvements in your results.

Important
When writing expressions in Model Builder, you should save the expression at regular intervals and between adding new keywords or tags, as any changes will be lost if you leave Model Builder for 30 minutes with no interaction. This is especially valid if using a separate tab with your Search page opened.

You can use the different AND and OR boxes to separate parts of your expression and to format it for easier reading. You might like to keep certain concepts or words and misspellings together so that it is easier to read and edit afterwards.

AND boxes will always be associated with an OR box. This means that at least one of the keywords in the AND box will need to be present in the verbatim as well as at least one of the keywords from the OR box. You can have a separate OR box that will not have the AND box association, if you click on the button for a different OR box. The gray line on the left side shows you the association between the boxes.

Figure 56 Example of the association indicator line

The global AND NOT box towards the bottom of the pane is used when you have some keywords that need to be excluded from the entire expression. You can also add local AND NOTs in the other OR and AND boxes.

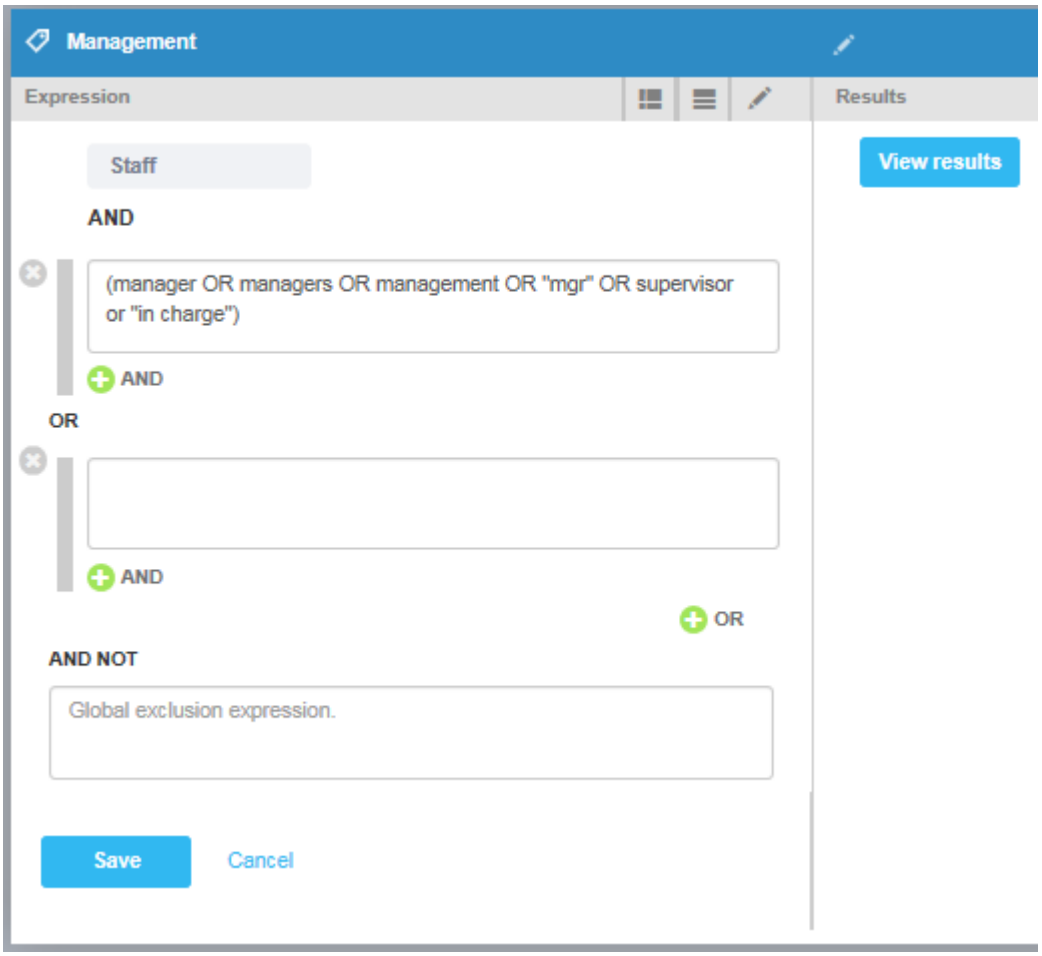


Figure 57 The AND, OR and AND NOT fields

Warning
Save your work at regular intervals, and ensure you save before you leave Model Builder unattended.

5.4.1. Adding Comments to Expressions

You can add comments into an expression to provide information for future reference. These comments will not be included in the analysis, yet can be very helpful for keeping notes, documenting your work in progress, or describing the expression you're writing.

To add a comment:

Type the characters **/*** at the beginning of the comment text, and the characters ***/** to close the comment.

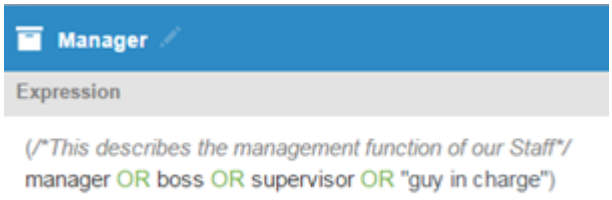


Figure 58 Example of a expression comment

5.4.2. Error Messages

As you type your expression, it is validated in the background. In the event you make a mistake and your expression becomes invalid, if you are still in the expression box the sides will go red. If you leave the expression box and the expression is not valid, an error message will be displayed and you will not be able to save your expression until the errors have been rectified. You will need to ensure that the parentheses match up, that all the Boolean operators are in capitals and that the required formatting is correct. Checking your expression in an application such as Notepad++ can help to verify the parentheses are matched.

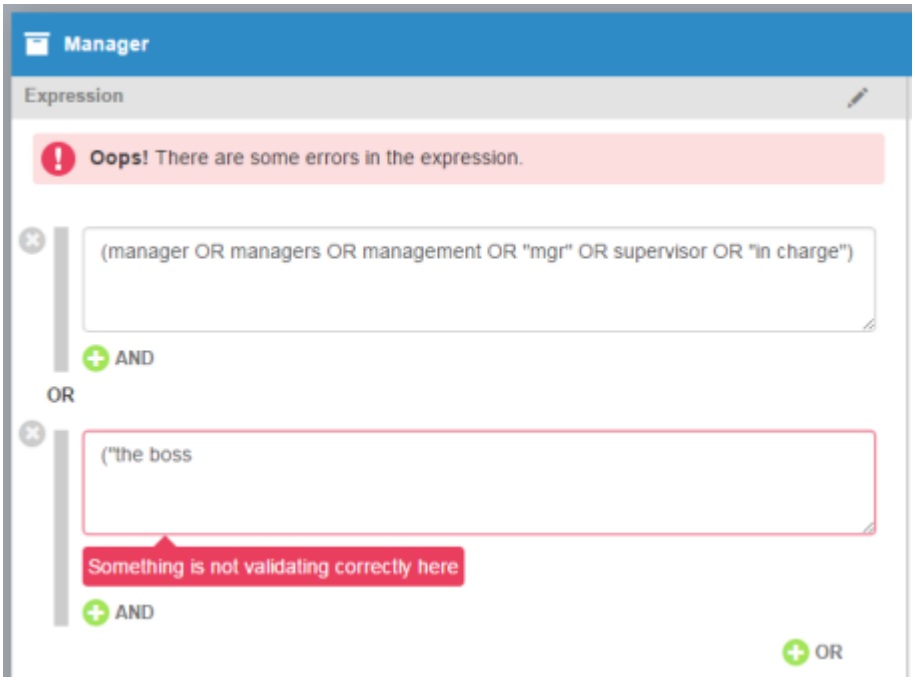


Figure 59 Example of an error message

5.4.3. The Survey Filter

If you have Genius Text Analytics on more than one survey, the results that you see when clicking on 'View results' are from all of the open text verbatim that have the Genius folder attached for analysis. To view results from a reduced set of surveys, use the survey filter.

1. Click the **Filter** button to open a list of the survey names and related Project IDs that have had Genius Text Analytics added to the open text variables. You can choose which open end you would like to filter on by selecting this on the right hand side of the survey name..

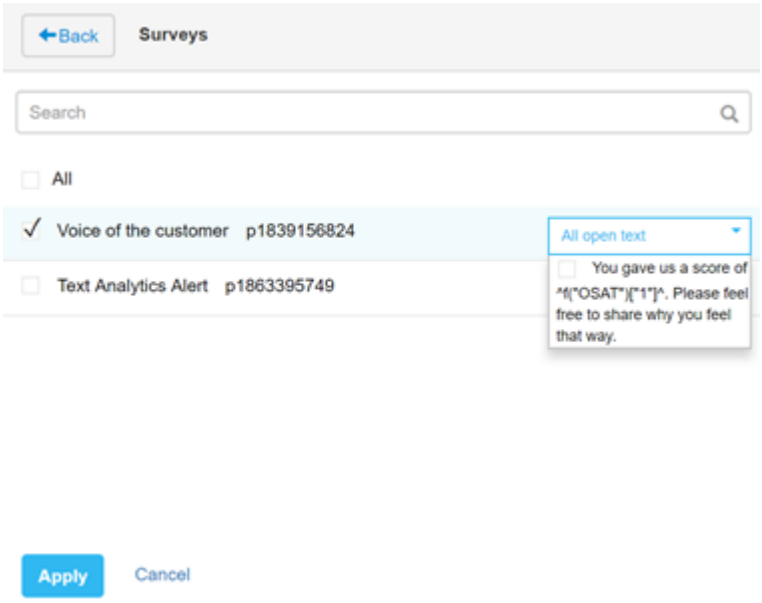


Figure 60 Example of the survey filter

2. Select the survey(s) you wish to use and click **Apply** or select **Time Range**.

   Under **Time range** you can choose from a list of options (last 30 days, last 60 days, last 90 days or last 120 days) or select a custom range.

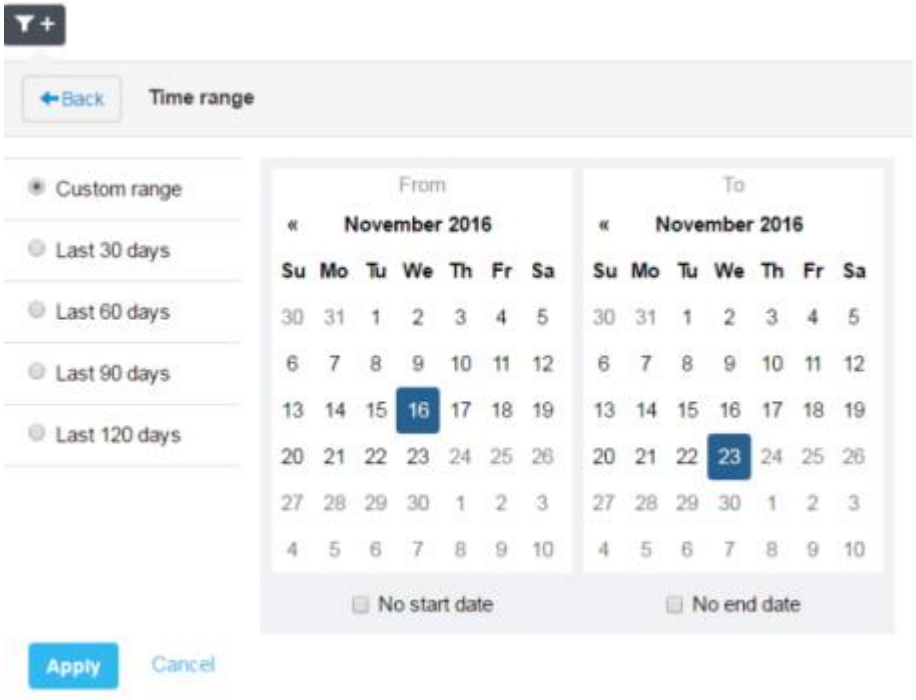


Figure 61 The Time range overlay

3. On completion, click **Apply**.

Once you have applied the filter you can perform your search and you will only see results from the selected surveys, questions and / or date range.

The filter that has been applied is indicated next to the filter button. This will remain visible for the duration of the session or until you cancel the filter by clicking on the (**X**) to remove it.

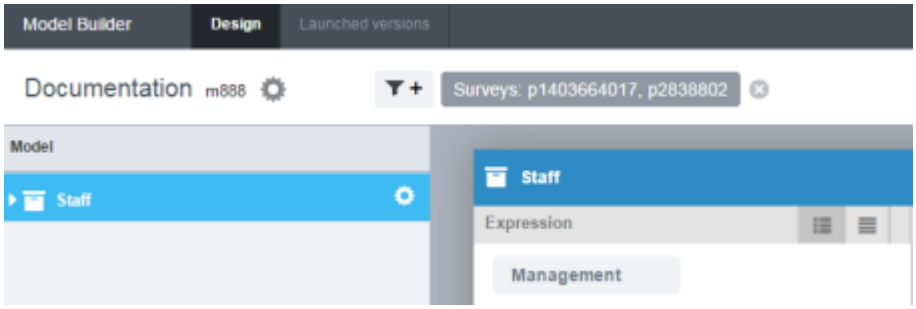


Figure 62 A filter applied

If you apply a filter to a survey, it will be remembered for that session. If you switch screens during your session, the survey filter will stay activated. To remove the filter, click the delete icon to the right of the filter name.

Warning
Save your work at regular intervals and ensure you save before you leave Model Builder unattended (see Inactive-Screen Lock on page 3 for more information).

5.5. How to Edit an Existing Expression

To edit an existing expression:

1. Click on the **Edit** icon in the expression pane.

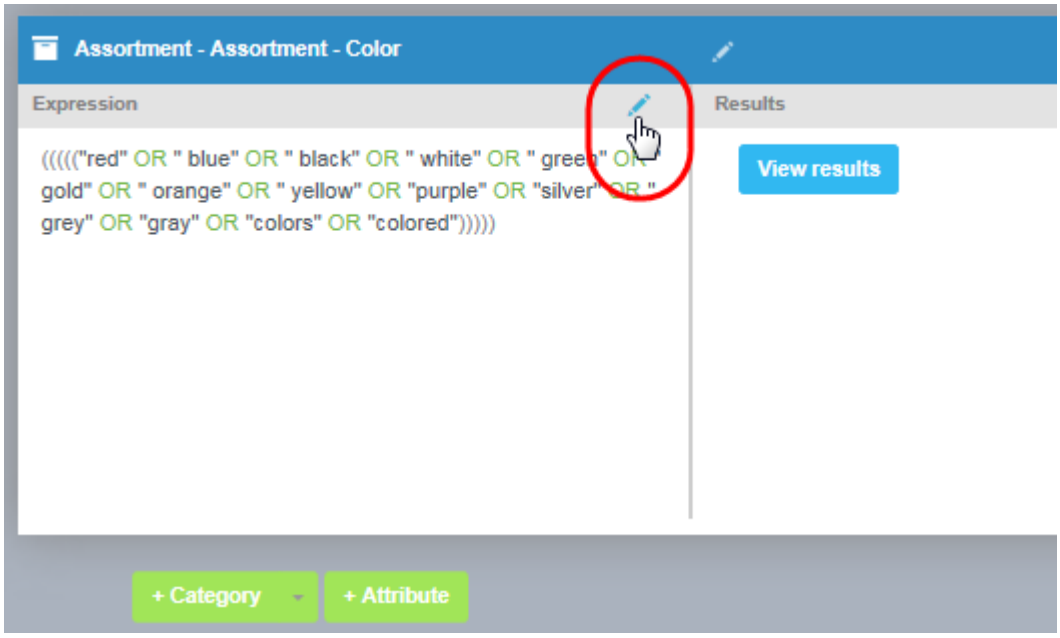


Figure 63 Click the Edit icon

This opens the Edit Expression pane for editing.

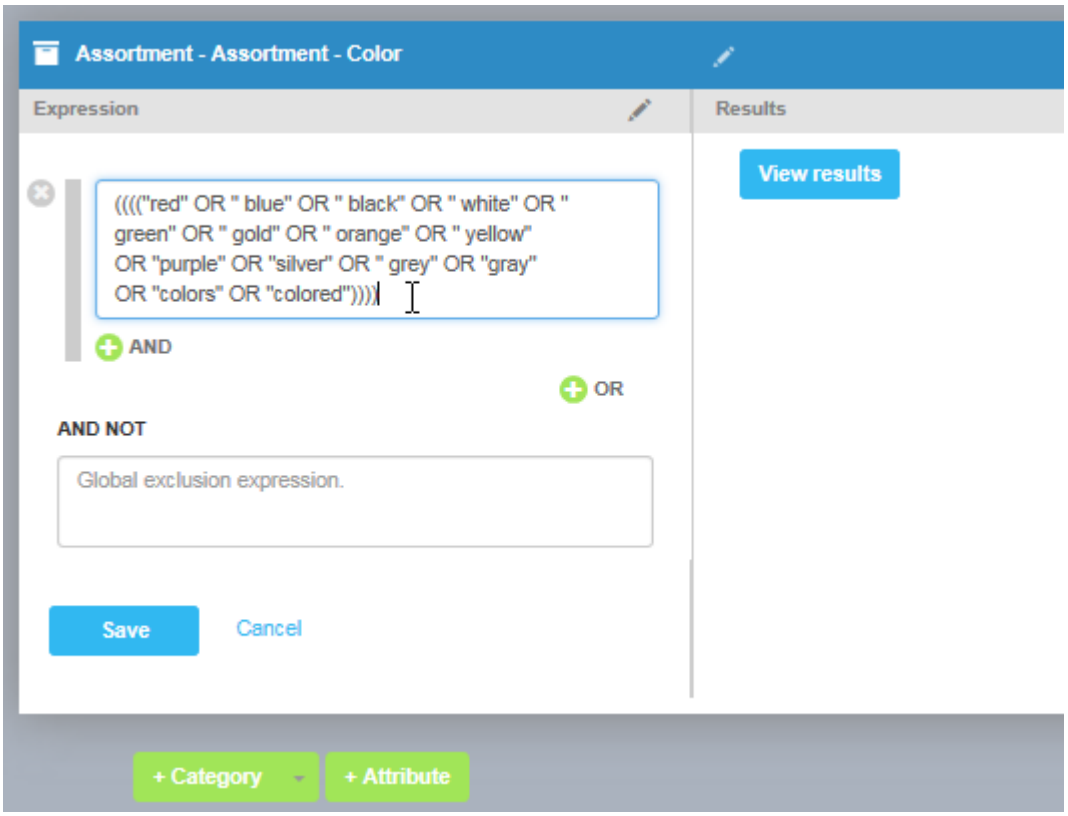


Figure 64 The Edit Expression pane opens

You can now click into the expression and add and take away keywords and add additional operators as required.

5.6. Making the Best Use of the Screen Space

You can switch between views by clicking on the icon on the right hand side of the expression pane:

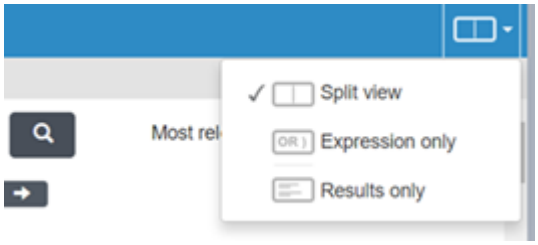


Figure 65 The split screen options

- **Split view** shows you the expression, with the results shown next to it.
- **Expression only** shows you just the expression, if you want to be able to look through the expression without looking at the results
- **Results only** shows you just the results, without the expression next to it.

5.7. How to Test an Expression

To test whether your expression is working well, click **Results** in the right pane opposite your expression.

You will see that the keywords from the expression have been highlighted in yellow. This will help you to assess why the verbatim was captured and whether or not it is correct for the category, sub-category or attribute that you defined.

Once you have clicked **Results** you can filter your result set by using the search bar at the top of the right pane.

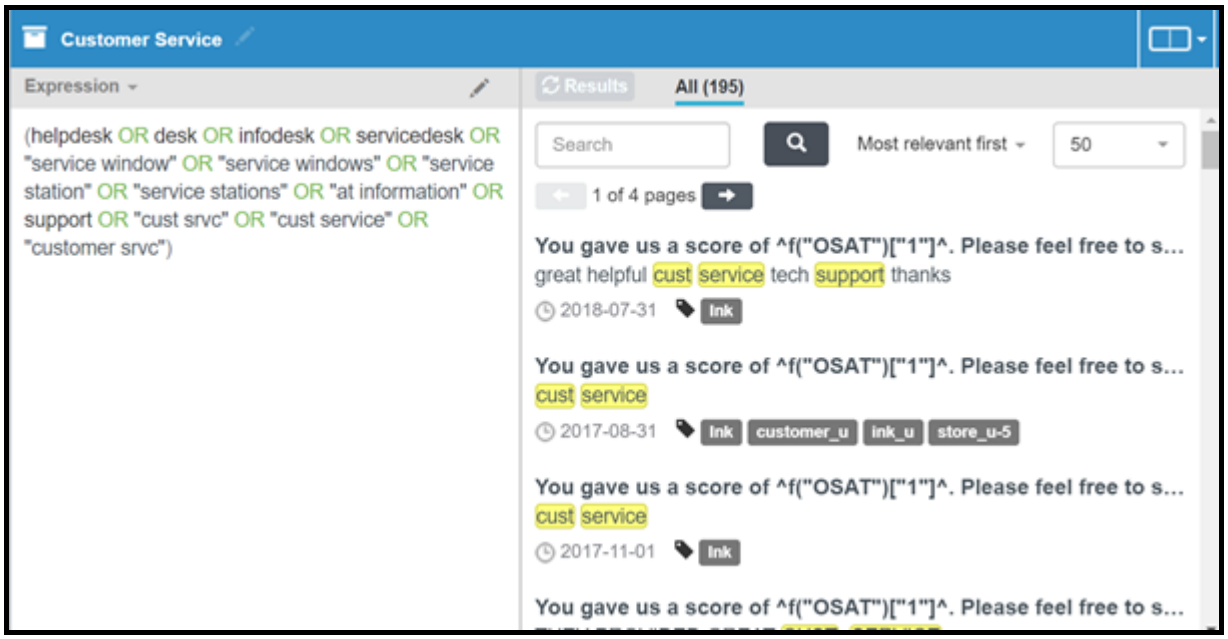


Figure 66 Testing an expression - 1

If you make a change to the expression (either remove or add keywords), the **Results** button will go blue, showing that you can refresh your results. If you do this before saving the expression, you will be able to see 'All' results as well as 'will add' and 'will remove'. This allows you to see what effect your changes will have on your model before making the save.

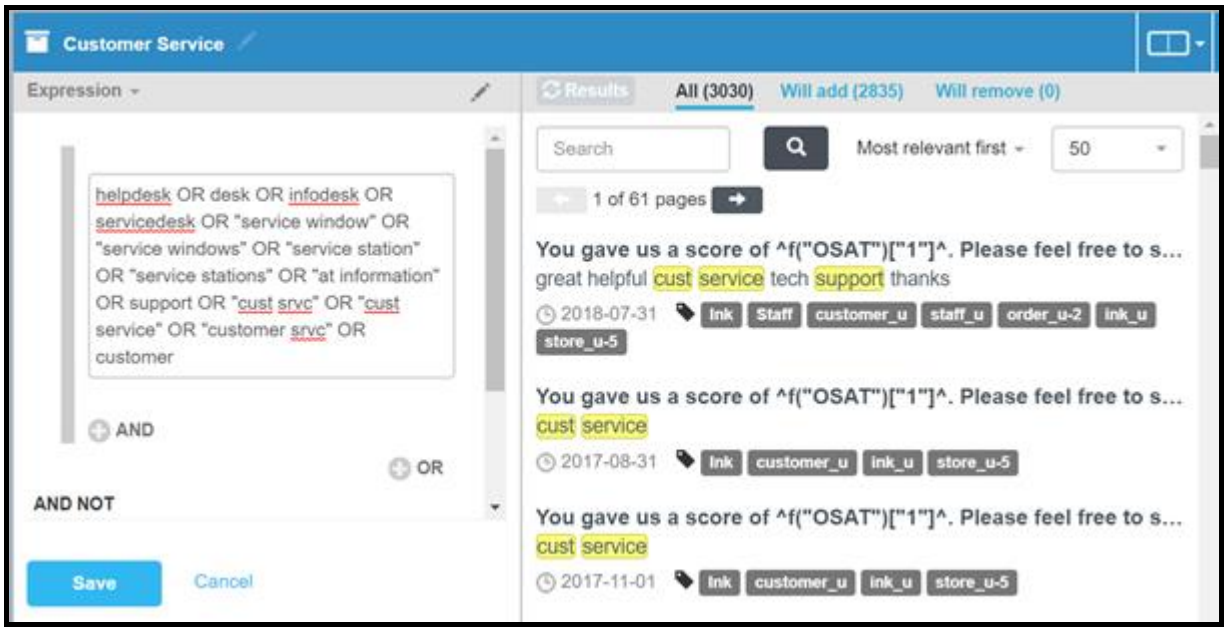


Figure 67 Testing an expression - 2

You can sort the results by ‘Most relevant first’, ‘Least relevant first’, ‘Newest first’, ‘Oldest first’ and 'Random order' so that you can test a good sample of the results.

To view only the results, click on the **Split view** icon and select **Results only**.

5.8. Testing the Model

You can test the model against your verbatim set at any time. To run a test, click the **Test Model** button towards the upper right corner of the Model Builder screen.

This will run all the available verbatims through the entire model, and once it has completed running it will show as the number of hits against each category.

A new folder will also be placed at the bottom of your categorization tree, called ‘Uncategorized comments’. This folder contains all the verbatims that have not been categorized by the rest of the model. You can go into this folder to look through the results to see whether there are any misspellings or keywords that need to be added to the existing model or whether you need to add any further categories, sub-categories or attributes.

If you make changes to the model while these hits are still showing, the expressions that have been affected (and therefore are no longer showing accurate hits) will now be greyed out, indicating that the hits are no longer accurate.

5.9. Uncategorized Comments

The Uncategorized Comments folder appears after you have tested your model against all the verbatims from the surveys that are currently saved in Model Builder, or the filtered set of verbatims if you’ve selected to filter by a specific survey or question.

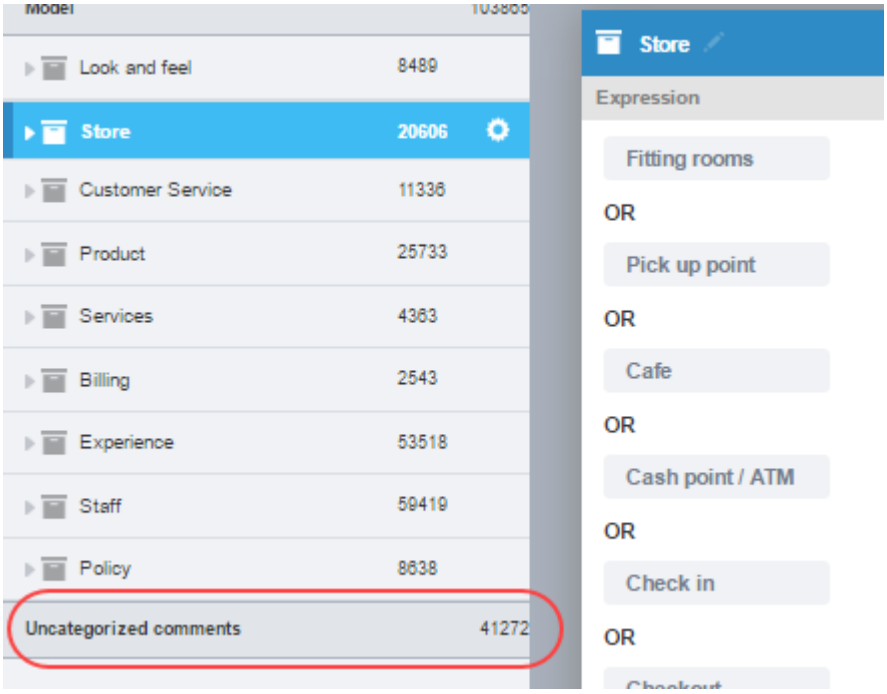


Figure 68 The Uncategorized Comments folder

The folder contains all comments that have not been captured by the existing model. You can explore these comments to find new categories, sub-categories and attributes or add to existing categories, sub-categories and attributes to improve your model. If you have an active Concept Miner core, you might get some quick and easy wins from looking at the tags that have been assigned to any verbatim in the Uncategorized Comments. We've seen that Concept Miner is good at working out categories in cases where it hasn't been easy to find a keyword to place that verbatim into the category .

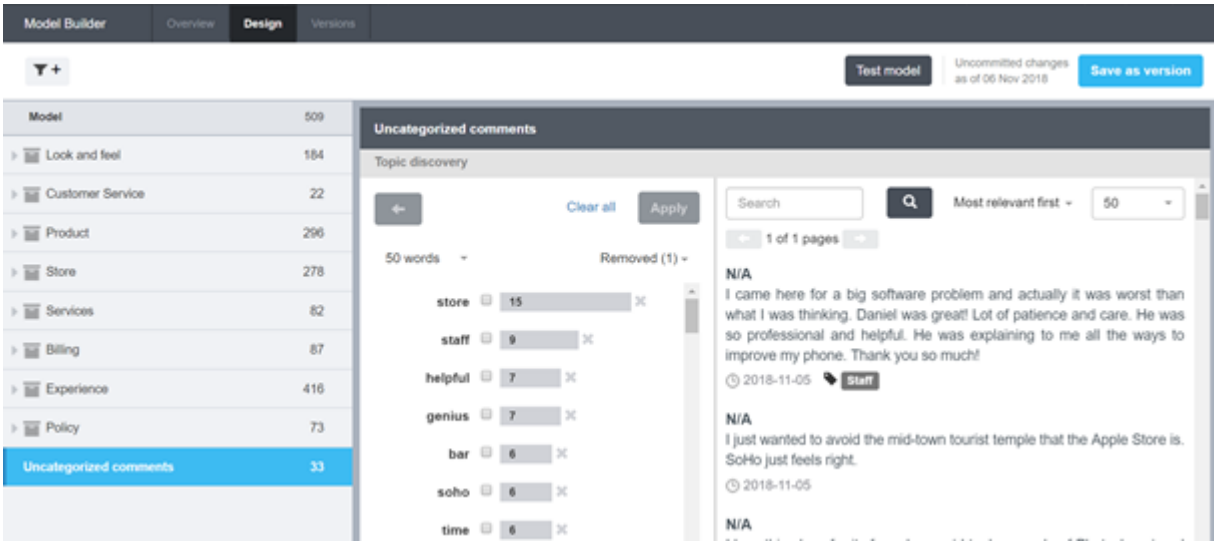


Figure 69 The Uncategorized comments

You can choose to report on your Uncategorized Comments in your report. To do this, ensure that you have the Uncategorized Comments code (starts with a prefix of **uc**) in your category list in the Database Designer table. If you do not want to report on Uncategorized Comments, do not add this code to your category list.

5.9.1. Topic Discovery

Topic Discovery is a feature that allows you to see the most frequently used words in your Uncategorized Comments. By default the top 50 words are shown, along with their frequency of appearance, in the left pane. Use the drop-down to select between the top 10, 50, 100 or 200 words. All comments related to the listed words are shown in the results pane on the right. Again, use the drop-down to select between the top 10, 50, 100 or 200 comments.

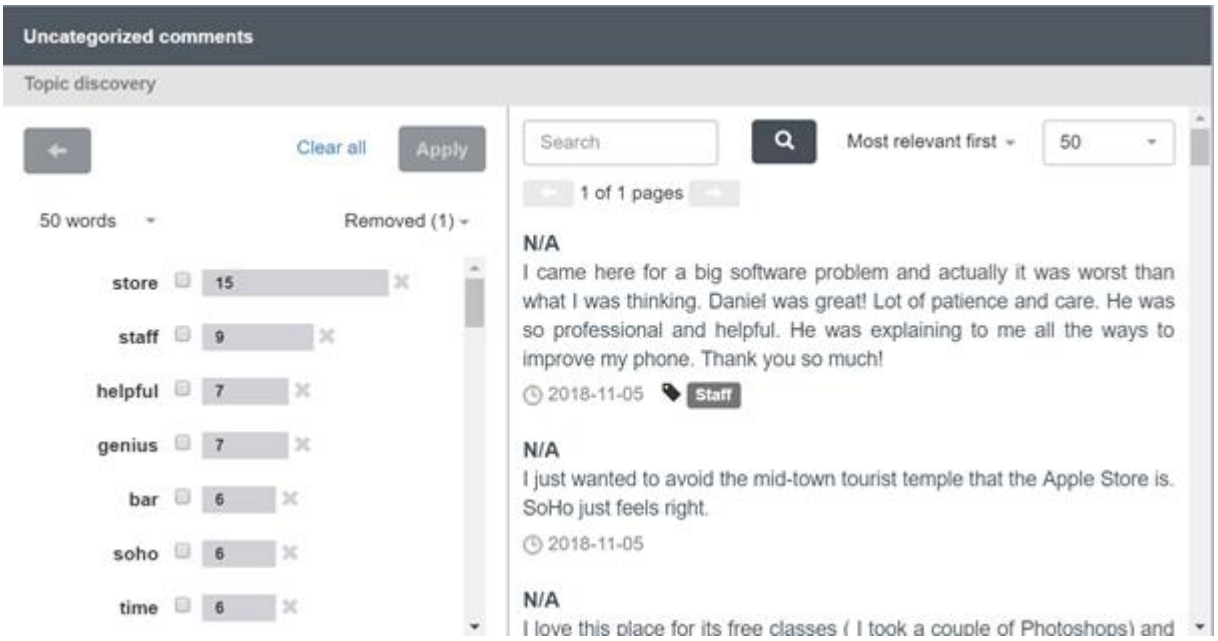


Figure 70 The Topic discovery pane

To explore and filter the results further:

- If there are words that you do not want to be included in the top words list, for example brand names or company names, remove them from the list by clicking on the **x** next to the frequency count. If you later decide you need to re-include them in the list, click the **Removed** button to display a list of removed words, then click **Add back** for the appropriate word(s).
- To display all the verbatims that include a particular keyword, check the checkbox for the word or words and click **Apply**. The verbatims that include the selected keyword(s) are then listed in the pane on the right, with the keywords highlighted. A new list of the most used words associated with the now-displayed verbatims is then presented in the left pane. You can repeat this process as many times as you wish to focus in on particular combinations of keywords.

  You can use these words to help you to update existing category expressions or to create new categories.

Click **Clear all** to take one step back in the sequence of filters.

5.10. Save as a Version

Model Builder supports versioning of the categorization models. This is to keep models consistent so that you have a benchmark for comparisons with previous periods. Maintenance of a model is important, and by using versioning you can save up all your changes and launch the complete version of the model on a particular date, enabling you to keep track of when the changes were made.

For a model to be used in Horizons, you must save it as a version and set it as Active. The first time that you save a model as a version, you will see that the checkbox that asks you whether you want to set the version as Active is ticked by default. You will see that the label changes from Design to Active. Once launched the model is given a version number, and any edits that are made to the model after it has been launched are made in the 'working version'. The next time you 'Save as version', the 'set this version as the active version' checkbox will not be checked and you will need to tick the checkbox to make it the active version that is used in your Horizons analysis. Any changes to the structure of the model (for example if you have added any new categories, sub-categories and attributes) will need to be updated in the Horizons Database Designer table. Refer to the Genius User Guide for more details.

You can set any version as the Active version. To set a version as Active, go to the cogwheel icon in the Versions list and select **Set to Active**.

The **Save as version** button is located towards the right side of the Design page for the Model. Update text next to the **Save as version** button tells you whether you have uncommitted changes and the date that the first set of changes were made. Once you save the next version of the model, this update will reflect whether you have made any further changes.

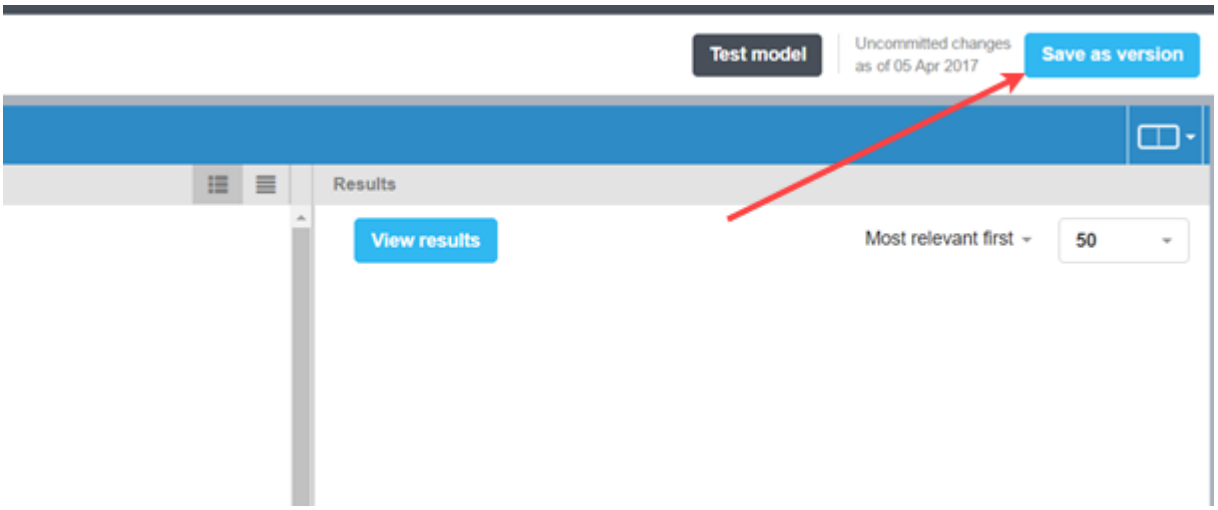


Figure 71 The Save as version button on the Design page

To save your model:

1. Go to the Design page for the model you wish to launch.
2. Select the **Save as version** button in the upper-right corner of the page.

   The Save model overlay appears.

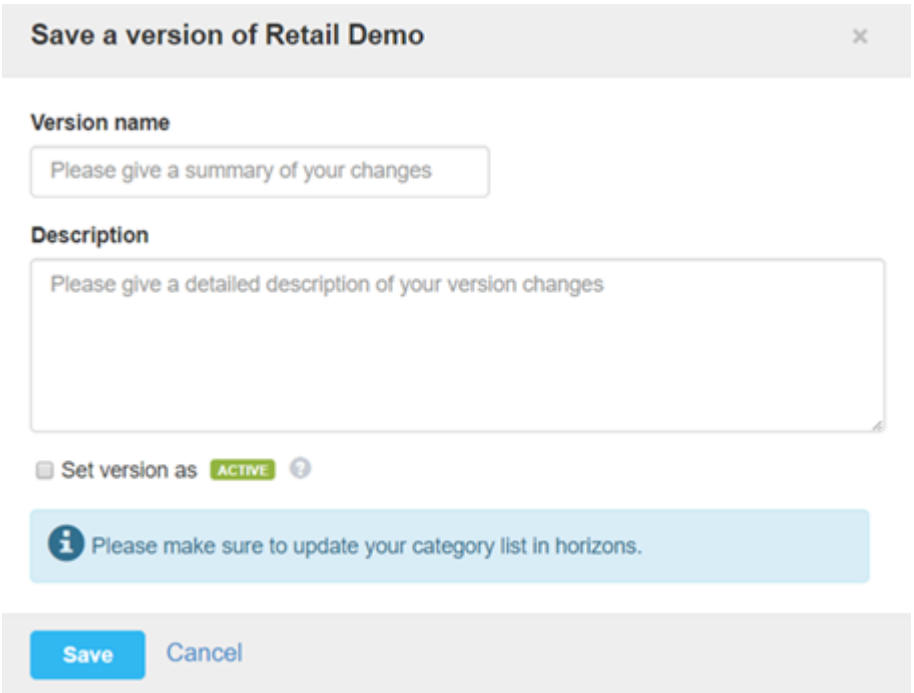


Figure 72 The Save model overlay

3. Add a name for the version as appropriate (we recommend using the model name and a date), and a description for future reference.

   If this is the first time you have saved the model, the 'Set version as active' checkbox will be checked by default. This sets the model to Active so that it can be seen by Horizons. If it not the first time, you will need to check the checkbox to make the version 'Active'. You might not want to set every save of the model to active as this will affect the results that you get in Horizons.

4. Click the **Save** button.

   The model is "locked" and is saved.

5. If you have made changes to the structure of the model (that is, you have added, moved, renamed or deleted any categories, sub-categories or attributes), you will need to update these in the model hierarchy (in Database Designer) in Horizons.

Note: In the event one or more categories or attributes are empty then a message will be displayed informing you of this.

The next time changes are made to the model, a new "working version" will be created automatically and the changes will be made to that version. That version will then remain the working version until it is saved, whereupon it will be locked and a new working version will subsequently be created.

Making changes such as changing the sentiment configuration, switching on Uncaptured Attributes and changing category names will all need to be saved as a version in order to see these changes reflected in your project in Horizons. The new version will need to be set as Active.

If you have made structural changes to the model, that is you have added new categories or attributes, you will need to update your Database Designer table in Horizons to reflect this change. The instructions for this are given in the Genius chapter of the Professional Authoring manual.

5.11. Viewing your Versions

Once you have saved your model, you can view that model in the 'Versions' tab (next to the Design tab on the secondary navigation tab). All your saved versions are available here.

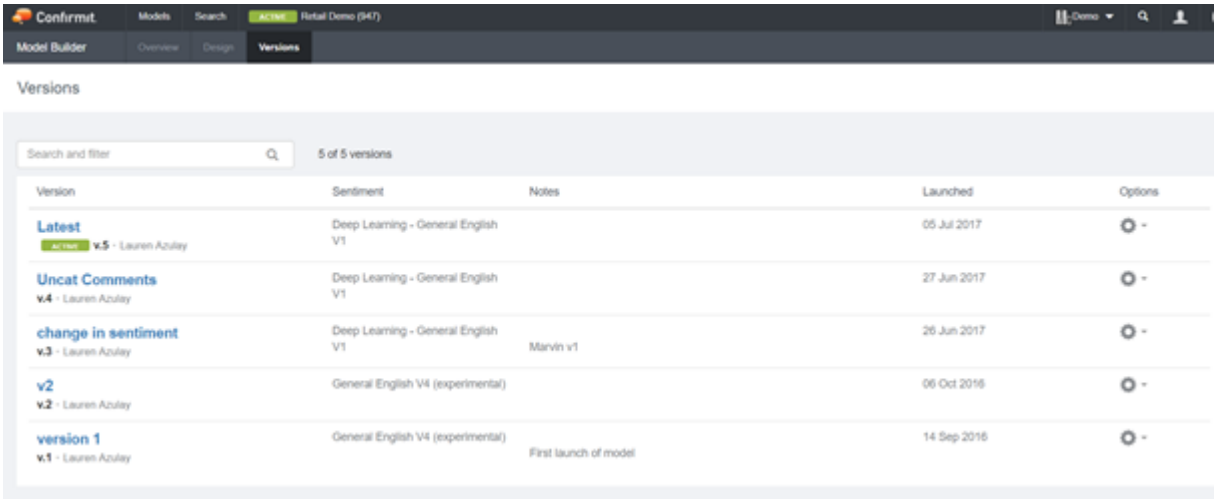


Figure 73 The Versions tab

A number of options are available from the cog-wheel icons on this page.

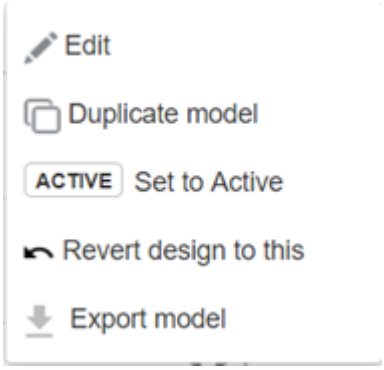


Figure 74 The cog-wheel options

- **Edit -** you can edit the description and name of the version from the cogwheel on the main launched versions page.
- **Duplicate model -** duplicates the version that you've selected and creates a new model from that version.
- **Set to Active -** sets the selected version as the active version for the model
- **Revert design to this -** takes the selected version and sets it as the current working version.
- **Export model -** exports the model to an Excel® spreadsheet.

6. Putting it All Together

Now that you have a categorization model that is active (i.e. you have launched it), you can set up the Genius folders in Horizons. Refer to the Genius chapter in the Professional Authoring User Guide for further details on how to achieve this.

Once you have set this up and run the Text Analytics task the first time, you will see your verbatims in Model Builder and the survey name and PID will be referenced in the survey filters.

Note: If you are running Text Analytics on a lot of projects, you will not see all your projects in the survey filter. Once there are more than 100 projects that have been run on the same account, the system will only show you the 100 project ID's with the highest number of verbatims.

Index

M

N

O

P

Q

R

S

T

U

V

W