Define rules to categorize responses in your survey data

Bilal Siddiqui
Freelance consultant
XML4Java

Working with type patterns

Part 1 describes concepts, types, and patterns. Recall that types are similar concepts that are put together in one placeholder. Concept patterns are combinations of a concept with a type. Each combination contains the survey responses that mention the concepts that are combined. Figure 36 and Figure 37 of Part 1 to refresh how combinations of concepts form a concept pattern.

IBM® SPSS® Text Analytics for Surveys organizes survey data as patterns of concepts, but also as patterns of types, which are referred to as type patterns. A type pattern combines one type with another type, thus combining all concepts of one type, one-by-one, with all concepts of another type.

With type patterns, you can see which of the responses fall into which combination. The following details show how to show combinations of type patterns.

Switch the results window to display type patterns by using the same drop-down list that you used in Figure 27 of Part 1, as shown in Figure 1.
Figure 1. Switching results window to display type patterns

![Image of results window]

Figure 2 shows a list of type patterns that SPSS Text Analytics for Surveys generates from the sample touchscreen survey data by using its built-in linguistic resources.

Figure 2. The list of type patterns that SPSS Text Analytics for Surveys generated by using its built-in linguistic resources

![List of type patterns]

Type patterns are easy to understand by looking at what is contained in the second type pattern `<Unknown> + <Positive>` (with 105 responses) shown in Figure 3.
Figure 3. Concept patterns and responses that are contained in the \textit{<Unknown> + <Positive>} type pattern

You can see that the \textit{<Unknown> + <Positive>} type pattern is a collection of concept patterns. For example, Figure 3 contains touchscreens + <Positive>, touch + <Positive>, touch gestures + <Positive>, and fingers + <Positive> patterns that are also present in Figure 33 of Part 1.

Try to match the number of responses in each of these concept patterns in both Figure 3 and Figure 33 of Part 1. The concept patterns that are shown in Figure 33 of Part 1 are the same as the concept patterns that constitute \textit{<Unknown> + <Positive>} type pattern of Figure 3.

You can say that the \textit{<Unknown> + <Positive>} type pattern is a combination of all concept patterns between concepts of the \textit{unknown} type and the \textit{Positive} type.

The first type pattern in Figure 2 is \textit{<Unknown> + <>}. The \textit{unknown} type is a placeholder for all concepts that do not fit into any type. SPSS Text Analytics for Surveys uses angle brackets (<> to represent a type that is not distinguishable as a specific type. <> is a collection of unknown types.

Expand \textit{<Unknown> + <>} to see that it is a collection of unknown types with unknown concepts, as shown in Figure 4.
Consider the usefulness of type patterns. Recall from the description of **Figure 40** in Part 1 that the combinations in the touch gestures + <Positive> patterns are useful because this pattern reveals precisely those responses that say something positive about touch gestures.

Therefore, it is wise to combine features and applications of touchscreen devices into a touchscreen-specific type (which you call <TouchFeaturesNApps>). SPSS Text Analytics for Surveys combines the <TouchFeaturesNApps> type with the existing default <Positive> type in a <TouchFeaturesNApps> + <Positive> type pattern. Such a pattern shows all positive responses for touchscreen features and applications in one placeholder.

It is time to start configuring touchscreen-specific linguistic resources in SPSS Text Analytics for Surveys.

### Configuring linguistic resources specific to touchscreen terminology

This series has thoroughly explored concepts, types, concept patterns, and type patterns, which are important linguistic features of SPSS Text Analytics for Surveys. Start to build these resources according to domain-specific terminology suitable for a touchscreen survey with the following steps.

1. Configure a library that is named **Touchscreen** that contains touchscreen-specific linguistic resources.
2. Recall the fingers (Figure 14 in Part 1) and touch gestures (Figure 18 of Part 1) concepts. Because these two concepts are used interchangeably in this touchscreen survey, combine the fingers and touch gestures concepts as one concept named gestures in the Touchscreen library.

3. Similarly, configure different concepts that are related to staying connected by using touchscreen devices as one concept named social media.

4. Configure a type that is named TouchFeaturesNApps, which includes gestures and social media concepts as features and applications of touchscreen devices.

5. Explore the effect of touchscreen-specific resources on results of extraction process in SPSS Text Analytics for Surveys.

6. Learn about categories, another important and powerful feature of SPSS Text Analytics for Surveys. Use the instructions in this article to define criteria or rules to categorize survey responses.

Configuring a new library

SPSS Text Analytics for Surveys comes with a resource editor to configure linguistic resources such as concepts and types. Therefore, the first step toward a domain-specific library is to switch to the editor view.

Click View > Resource Editor as shown in Figure 5.

Figure 5. Use the View menu to access the Resource Editor

The Resource Editor opens, as shown in Figure 6.
The main **Resource Editor** window includes three smaller windows, as shown in Figure 6. In the upper-left, the **libraries** window shows default libraries such as **Local Library**, **Opinions Library**, and **Budget Library**.

The upper-right window is empty. You use the **resources** window to specify resources (or concepts) for the **TouchFeaturesNApps** type.

The lower **synonyms** window is used to specify synonyms for a concept.

In the libraries window, right-click **Opinions > New Library** as shown in Figure 7.

**Figure 7. Creating a new library**

You see a **Library Properties** dialog box, as shown in Figure 8.
Figure 8. The Library Properties dialog box

![Library Properties dialog box]

Enter the name of your new library in the Name field along with a simple annotation in the Annotation field as shown in Figure 9.

Figure 9. Entering the name of your new library

![Library Properties dialog box]

To publish and reuse your libraries across different projects, you can click Publish as shown in Figure 9. However, publishing is outside the scope of this article.

Click OK to see your new Touchscreen library in the list of available libraries in the libraries window as shown in Figure 10.

Figure 10. A new Touchscreen library added to the list of existing predefined libraries

![Libraries window]

The Touchscreen library is empty. The next step is to start adding touchscreen-specific concepts to the library. For this purpose, use the synonyms window that is shown in Figure 6.

The synonyms window is a table with four columns. The leftmost column counts concepts, each row shows one concept. The top row is shown as Row 0, which is empty and does not contain any concept definition. This row is used to configure new concepts.
The second column is **Target**, which contains name of the concept. The third column is **Synonyms**, which specifies all synonyms for the concept. The fourth column, **Library**, specifies the library that is to hold the concept you are configuring.

To configure a new concept, click the **Target** column in Row 0 and enter **gestures** as name of the new concept.

Then, click the fourth column, **Library**. It shows a list of available libraries, including the Touchscreen library that you just created. Select **Touchscreen**.

**Figure 11. Configuring synonyms**

As you enter data for your new gestures concept in Row 0, the data goes into Row 1. Row 0 becomes empty again so that you can enter another new concept.

Now type in the touchscreen-specific synonyms for your gestures concept in the **Synonyms** column. Synonyms are touch gestures, finger, swiping, and zoom.

As you type synonyms in the **Synonyms** column, the column expands to accommodate multiple lines of text, as shown in **Figure 12**.

**Figure 12. Entering synonyms for the gestures concept**

After you enter synonyms, click **Enter** and the configuration for your new touchscreen-specific concept is ready.

**Figure 13. Configuration for the gestures concept**
The new concept is configured in Row 1. If you scroll up, you can also see an empty Row 0, where you can enter another concept. Using Row 0, configure social media, as shown in Figure 14.

**Figure 14. gestures and social media concepts that are configured in SPSS Text Analytics for Surveys**

You have successfully configured two concepts for your Touchscreen library. Next, see the effect of these two concepts on extraction results. Click Analysis > Text Analysis, as shown in Figure 15.

**Figure 15. Switching back to the Analysis view**

SPSS Text Analytics for Surveys knows that old extraction results are no longer valid because you enhanced the linguistic resources by adding concepts. Therefore, SPSS Text Analytics for Surveys asks you to repeat the process of extraction to get updated results, as shown in Figure 16.
In Figure 16, the results window has a yellow background, which shows that the results are outdated. Above the yellow results window, SPSS Text Analysis for Surveys prompts you to Extract to update concept list. Click Extract. SPSS Text Analytics for Surveys compiles and shows a new list of concepts as shown in Figure 17. (The extraction can take some time.)

Figure 17. New list of concepts that are based on updated linguistic libraries

Compare the new list of concepts in Figure 17 with initial results of extraction in Figure 8 of Part 1. In Figure 17, the gestures concept has 37 responses. The responses included in the gestures
concept in Figure 17 include the touch gestures and fingers concepts that are mentioned in the responses because you added those words as touchscreen-specific synonyms of gestures.

The fingers concept in Figure 8 of Part 1 has 15 responses. Similarly, the touch gestures concept has 14 responses in Figure 8 of Part 1. These two concepts (along with other concepts swiping and zoom with fewer responses) are all merged into the gestures concept in Figure 17, giving a total of 37 responses in the gestures concept. For this reason, you don't find any fingers and touch gestures as independent concepts in Figure 17.

Similarly, the responses included in the social media concept in Figure 18 include the concepts that are merged into social media.

**Figure 18. Responses included in the social media concept**

Switch the results window to display concept patterns. The gestures and social media concepts are reflected in concept patterns. Three concept patterns, social media + , (shown in Figure 19), gestures + <Positive> (shown in Figure 20), and gestures + , (shown in Figure 21).
Figure 19. The social media + . concept pattern

Figure 20. The gestures + <Positive> concept pattern
From the results of the concepts that are shown in Figure 17 and the concept patterns of Figure 19, Figure 20, and Figure 21, the most useful results are the ones in the gestures + <Positive> concept pattern of Figure 20. The responses in Figure 20 indicate positive sentiment about gestures are collected in one place. The advantage of defining a touchscreen-specific gestures concept is that regardless of whether a response uses the word "gesture," "touch gesture," "finger," "swiping," or "zoom," the gestures concept includes the response within the gestures + <Positive> concept pattern.

The gestures + <Positive> concept pattern of Figure 20 demonstrates how SPSS Text Analytics for Surveys combines user-defined and built-in concepts in a useful manner.

You configured touchscreen-specific concepts and you looked at the effect of new concepts on extraction results. The next section describes how to configure a touchscreen-specific type named TouchFeaturesNApps.

**Configuring a domain-specific type**

Open the Resource editor view and expand the Touchscreen library. Expanding the library reveals the types that are contained in this library. Currently, the only entry is a NewType(0) child of the Touchscreen library.
To configure a new type, right-click \texttt{NewType(0)}.

Click \texttt{New Type}. The \texttt{Type Properties} dialog box opens.

To configure a new type, right-click \texttt{NewType(0)}.

Click \texttt{New Type}. The \texttt{Type Properties} dialog box opens.
Specify `TouchFeaturesNApps` in the **Name** field for the new type.

Leave the default value **Entire Term** in the second field **Default match** to indicate that responses included in this type are to contain entire terms that match the synonyms of concepts that you include in this type.

Similarly, leave the default value `Touchscreen` in the **Add to** field to indicate that the new type is added to the Touchscreen library.

Because features related to inflection are helpful in linguistic analysis, select the **Generate inflected forms by default** check box. For example, *connect* and *connected* are inflected forms of the same word and therefore can be treated as the same word in this analysis.

SPSS Text Analytics for Surveys also enables you to apply colors to types. In this example, the default option is selected, but you can play with the use of colors.

I also provided a simple annotation for the `TouchFeaturesNApps` type. Figure 25 shows a completed form of the **Type Properties** dialog box.

**Figure 25. Completed form of the Type Properties dialog box**

Click **OK** to see the new `TouchFeaturesNApps` type added to the Touchscreen library.

**Figure 26. TouchFeaturesNApps type added to the Touchscreen library**
Now you can add concepts to the **TouchFeaturesNApps** type. For this purpose, use the empty row of the **Resources** table in the Resources window at the right side of the libraries window in **Figure 26**.

Select the **TouchFeaturesNApps** type in the libraries window and start entering concepts in the **Resources** table. Configure one concept in each a row.

You can enter just the name of the concept **social media** in the **Term** column. The rest of the data comes from the **Type Properties** dialog box that is shown in **Figure 25**.

**Figure 27. The social media concept that is configured as a term in the TouchFeaturesNApps type**

After you configure **social media** as a term in the **TouchFeaturesNApps** type, responses that contain this term or any of its synonyms are included in **TouchFeaturesNApps**.

Similarly, configure the **gestures** concept as another term in the resources table, as shown in **Figure 28**.

**Figure 28. The gestures concept that is configured in the TouchFeaturesNApps type**

**TouchFeaturesNApps** in extraction results

To see how the **TouchFeaturesNApps** type affects extraction results and analysis, switch back to the **Text analysis** view, extract fresh results, and then switch the results window to display types.
In Figure 29, the TouchFeaturesNApps type contains 37 responses from the gestures concept and 25 responses from the social media concept.

In addition, SPSS Text Analytics for Surveys included three other concepts (social media with touch, gestures with touchscreens, and social media on touchscreens) with one response each in the TouchFeaturesNApps type. These concepts contain the social media and gestures terms; therefore, bringing them together into the TouchFeaturesNApps type helps in compiling analysis results.

Switch the results window to show type patterns to see how the TouchFeaturesNApps type combines with other types. Figure 30 shows TouchFeaturesNApps in several type patterns, such as:

- `<TouchFeaturesNApps> + <>`
- `<TouchFeaturesNApps> + <Positive>`
- `<TouchFeaturesNApps> + <Contextual>`
- `<TouchFeaturesNApps> + <Negative>`
Figure 30. Type patterns shown after the new TouchFeaturesNApps type is configured

The <TouchFeaturesNApps> + <> type pattern includes two concept patterns: social media + . in Figure 19 and gestures + . in Figure 21.

Look at the responses and concept patterns included in the <TouchFeaturesNApps> + <Positive> type pattern.
Figure 31 shows the <TouchFeaturesNApps> + <Positive> type pattern, which brings together responses that include positive mention of features and applications of touchscreen devices, such as gestures. Recall from Figure 20 that the gestures + <Positive> concept pattern collects responses that include positive mention of gestures. Therefore, you can say that the <TouchFeaturesNApps> + <Positive> type pattern is a higher level (and maybe more useful) compilation of gestures + <Positive> and similar concept patterns.

Categorizing based on your own criteria

In addition to extracting concepts and recognizing types and patterns, SPSS Text Analytics for Surveys includes another powerful feature that is called categorization of responses. You can create categories so that SPSS Text Analytics for Surveys can place responses into categories.

SPSS Text Analytics for Surveys supports several forms and ways of categorization. This article demonstrates one important form of categorization that is based on category rules.

Categorizing based on rules

Define rules when you want to categorize responses that are based on questions such as:

- "Bring me all responses that mention any of the features or applications of touchscreens, but don't mention anything negative"
- "Make a collection of responses that mention features but don't mention anything random in addition."

Such questions become obvious and important after you analyze concepts and patterns and want to move further, based on initial findings.
To configure categories for the touchscreen survey data, create an empty category. Click **Categories > Create Empty Category** as shown in Figure 32.

**Figure 32. Using the Categories menu to create an empty category**

![Diagram showing the Categories menu with Create Empty Category highlighted.](image)

The **Category Properties** dialog box opens. Specify `TouchFeaturesNApps` as the name of the new category as shown in Figure 33.
Figure 33. Using the Category Properties dialog box to create a new category

![Category Properties dialog box](image)

Click **OK**. The new category `TouchFeaturesNApps` is listed under **All Records** as shown in **Figure 34**.

Figure 34. A new empty category is created

![Category list](image)

The new `TouchFeaturesNApps` category is empty; therefore, the number of responses that are included is 0 in this category, which is shown in the rightmost column. The category is empty because you did not configure any rules that SPSS Text Analytics for Surveys can use to find responses relevant to this category.

To configure a rule for categorization, right-click the `TouchFeaturesNApps` category and select **Create Category Rule** as shown in **Figure 35**.
Figure 35. Creating a category rule

SPSS Text Analytics for Surveys has a category rule editor that you can use to create category rules. Select **Create Category Rule** to open the rule editor.

Figure 36. The category rule editor
Notice that the rule editor has a toolbar. Specify `<TouchFeaturesNApps>` in the editor to add all responses included in the TouchFeaturesNApps type to the TouchFeaturesNApps category.

**Note:** To specify a type when you create a category rule, you must enclose the name of the type in angle brackets (`<` and `>`) as in this example: `<TouchFeaturesNApps>`.

**Figure 37. Authoring a simple category rule**

Click **Save & Close**. As shown in **Figure 38**, your simple new rule is listed in the TouchFeaturesNApps category.

**Figure 38. A simple category rule added to the TouchFeaturesNApps category**

The TouchFeaturesNApps category now contains 64 responses, all of which come from the TouchFeaturesNApps type that is shown in **Figure 29**.

This simple category rule is not helpful because it collects responses from only one type. Enhance the simple rule to make it more useful.

Right-click the TouchFeaturesNApps category and select **Create Category Rule** (as in **Figure 35**). A second category rule is shown in the rule editor.

This time, enter `<TouchFeaturesNApps> & !(<Negative>)` in the rule editor. The ampersand (`&`) and exclamation point (`!`) characters are used in the way they are used in normal programming logic. The ampersand (`&`) character represents logical AND operator, meaning responses that fulfill both `<TouchFeaturesNApps>` and `!(<Negative>)` criteria. The exclamation point (`!`) character represents
logical NOT, which means !(<Negative>) collects responses that do not belong to the <Negative> type.

Therefore, <TouchFeaturesNApps> & !(<Negative>) collect responses that mention features and applications of touchscreen devices and do not contain anything negative.

You can use the toolbar on top of the editor to insert & and !( ) code. Figure 39 shows the rule editor.

**Figure 39. Authoring the second rule**

![Rule editor screenshot]

Click **Save & Close** to add a second rule to the TouchFeaturesNApps category.

**Figure 40. Second rule added to the TouchFeaturesNApps category**

<table>
<thead>
<tr>
<th>Category</th>
<th>#Desc.</th>
<th>#Resp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Records</td>
<td></td>
<td>294</td>
</tr>
<tr>
<td>- Uncategorized</td>
<td></td>
<td>230</td>
</tr>
<tr>
<td>- No concepts extracted</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>TouchFeaturesNApps</td>
<td>2</td>
<td>64</td>
</tr>
<tr>
<td>Check &lt;TouchFeaturesNApps&gt;</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>f &lt;TouchFeaturesNApps&gt; &amp; !(&lt;Negative&gt;)</td>
<td>62</td>
<td></td>
</tr>
</tbody>
</table>

The second <TouchFeaturesNApps> & !(<Negative>) rule has 62 responses that are associated with it. These results indicate a total of 64 responses that mention features and applications of touchscreen devices and 62 of those do not mention anything negative.

You can author a third rule <TouchFeaturesNApps> & (<Negative>), which collects those two responses that mention something negative along with touch features and applications as shown in Figure 41.
Figure 41. More categories

![Table showing category rules and responses](image)

**Figure 41** shows the total of the five rules. These rules demonstrate what you can achieve by using category rules. Explore the additional categories that you can configure in SPSS Text Analytics for Surveys.

**Conclusion**

This article explores text analytics with SPSS Text Analytics for Surveys. **Part 1** describes how to use built-in resources of SPSS Text Analytics for Surveys and notes salient points from sample survey data. Part 2 shows how to develop domain-specific resources to fine-tune analytic results and how to configure rules to categorize survey data.
## Downloads

<table>
<thead>
<tr>
<th>Description</th>
<th>Name</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample data for touchscreen survey</td>
<td>SampleSurveyData.zip</td>
<td>16KB</td>
</tr>
</tbody>
</table>
Resources

- Learn about all features of SPSS Text Analytics for Surveys by reading its official user guide on the IBM website.
- See this demonstration about IBM SPSS Text Analytics for Surveys.
- The Lousy Linguist posted about his findings after he used a trial version of IBM SPSS Text Analytics for Surveys.
- Ashley Richards wrote a review of IBM SPSS Text Analytics for Surveys.
- Download IBM SPSS Text Analytics for Surveys.
- Read Using Text Mining Techniques to Analyze Students’ Written Responses to a Teacher Leadership Dilemma (Yuejin Xu and Noah Reynolds). The paper describes a research project that used IBM SPSS Text Analytics for Surveys to analyze survey responses.
- Check out this workshop on text analytics.
- Read Analytics Basics: Understanding Survey Data (Neil Mason, ClickZ, June 2010) about understanding of survey data.
- Try Natural Language Toolkit, which is an open source platform to work with projects that involve linguistics.
- Follow developerWorks on Twitter.
- Stay current with developerWorks technical events and webcasts that focus on various IBM products and IT industry topics.
About the author

Bilal Siddiqui

Bilal Siddiqui is an electronics engineer, XML consultant, technology evangelist, and frequently published technical author. He is the founder of XML4Java.com, a company that is focused on simplifying e-business. After he graduated in 1995 from the University of Engineering and Technology, Lahore, Bilal began designing software solutions for industrial control systems. Later, he turned to XML and built web- and WAP-based XML processing tools, server-side parsing solutions, and service applications. Since 2006, he has focused exclusively on open source tools and solutions based on Java and XML technologies. A strong advocate of open source tools, he not only designs solutions that are based on them but also trains software and IT personnel at Lahore universities in using open source technologies. Bilal is the author of *JasperReports 3.6 Development Cookbook* (Packt Publishing, 2010).

© Copyright IBM Corporation 2014

Trademarks
(www.ibm.com/developerworks/ibm/trademarks/)