

# Learning Module - Content-based analysis

We use graphs to represent many kinds of networks. Nodes represent entities, and edges represent relationships. One type of behaviour that typically happens in many different types of networks is nodes passing information from one to another along these edges. This information can be represented as Content.

In this module we will go through how to look at Content in Constellation and look at answering some Content-based analytical questions. We will cover:

- Content Attributes in Constellation
- Using the Conversation View to view Content
- Using the Conversation View to add translations
- Using the Histogram View to find Content
- Extracting content as features

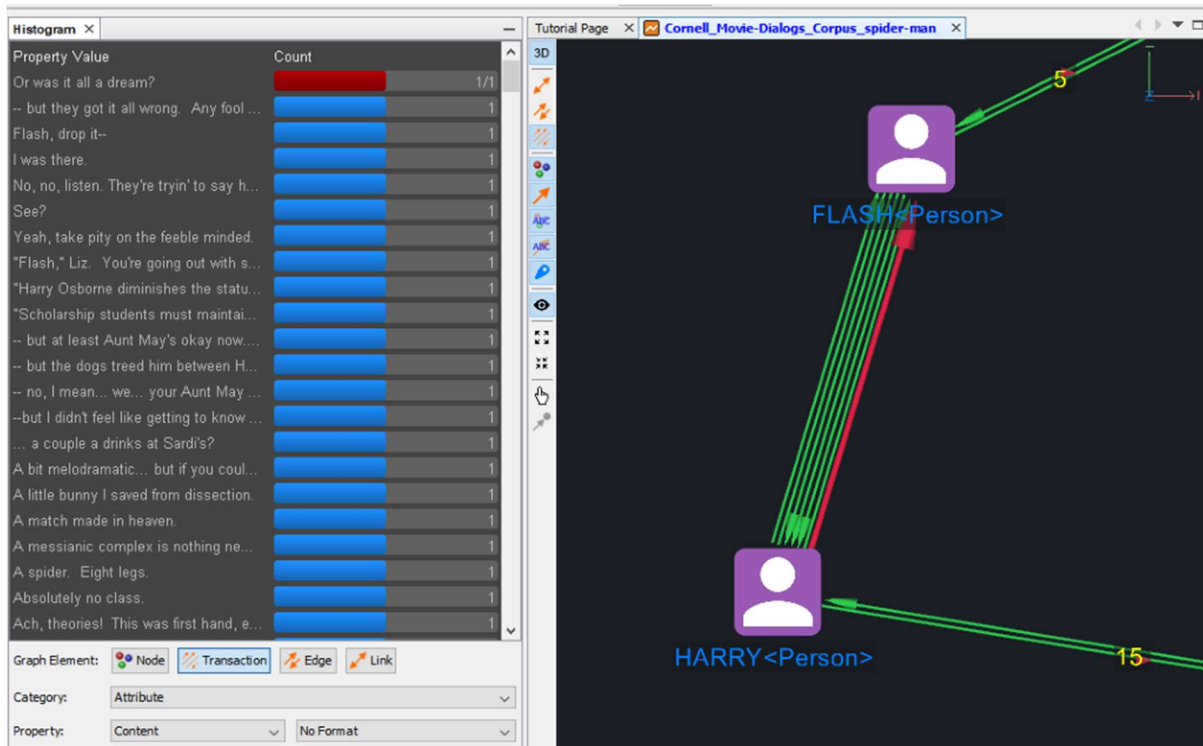
---

## Content Attributes in Constellation

Content is typically stored as an Attribute within Constellation; content passed between entities usually being found as a Transaction Attribute, which can be viewed and interacted with via the Attribute Editor and the Histogram View as any other attribute.

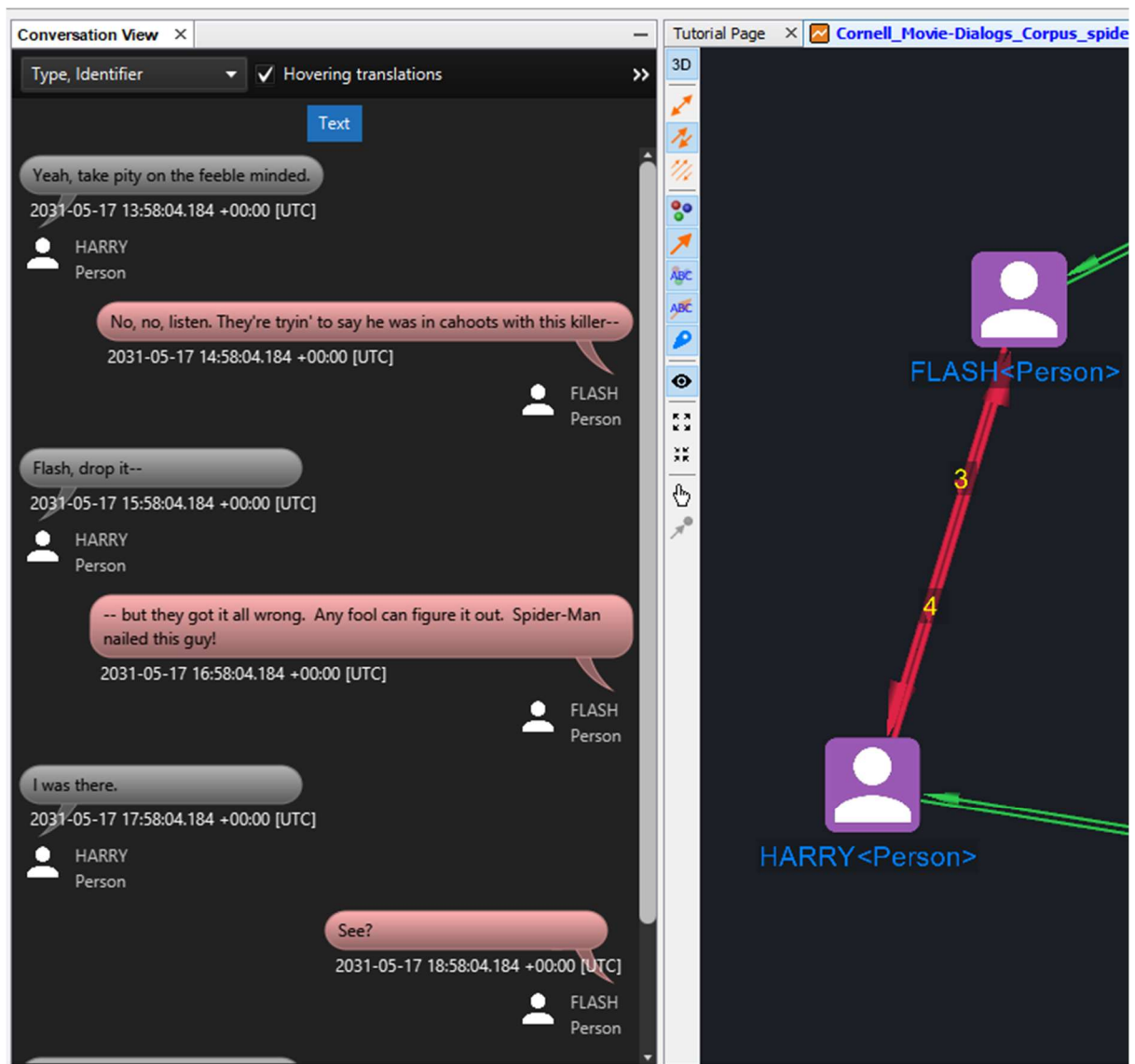
The screenshot displays the Constellation software interface. On the left, a graph shows two nodes, FLASH and HARRY, both labeled as <Person>. They are connected by a transaction edge, represented by a bundle of green lines. The edge has a weight of 15. The Attribute Editor panel on the right shows the details of the selected transaction. The transaction is labeled 'Communication' and has a content of 'Or was it all a dream?'. The date and time is 2031-05-17 19:58:04.184 +00:00 [UTC]. The identifier is 97013. The transaction is directed, has a solid line style, and is selected. The color is Emerald, and the width is 1.0.

Attribute	Value	Edit
Activity	<No Value>	Edit
Content	Or was it all a dream?	Edit
DateTime	2031-05-17 19:58:04.184 +00:00 [UTC]	Edit
Identifier	97013	Edit
Label	Communication	Edit
Source	<No Value>	Edit
Type	Communication	Edit
color	Emerald	Edit
dim	False	Edit
directed	True	Edit
line_style	SOLID	Edit
selected	True	Edit
visibility	1.0	Edit
width	1.0	Edit

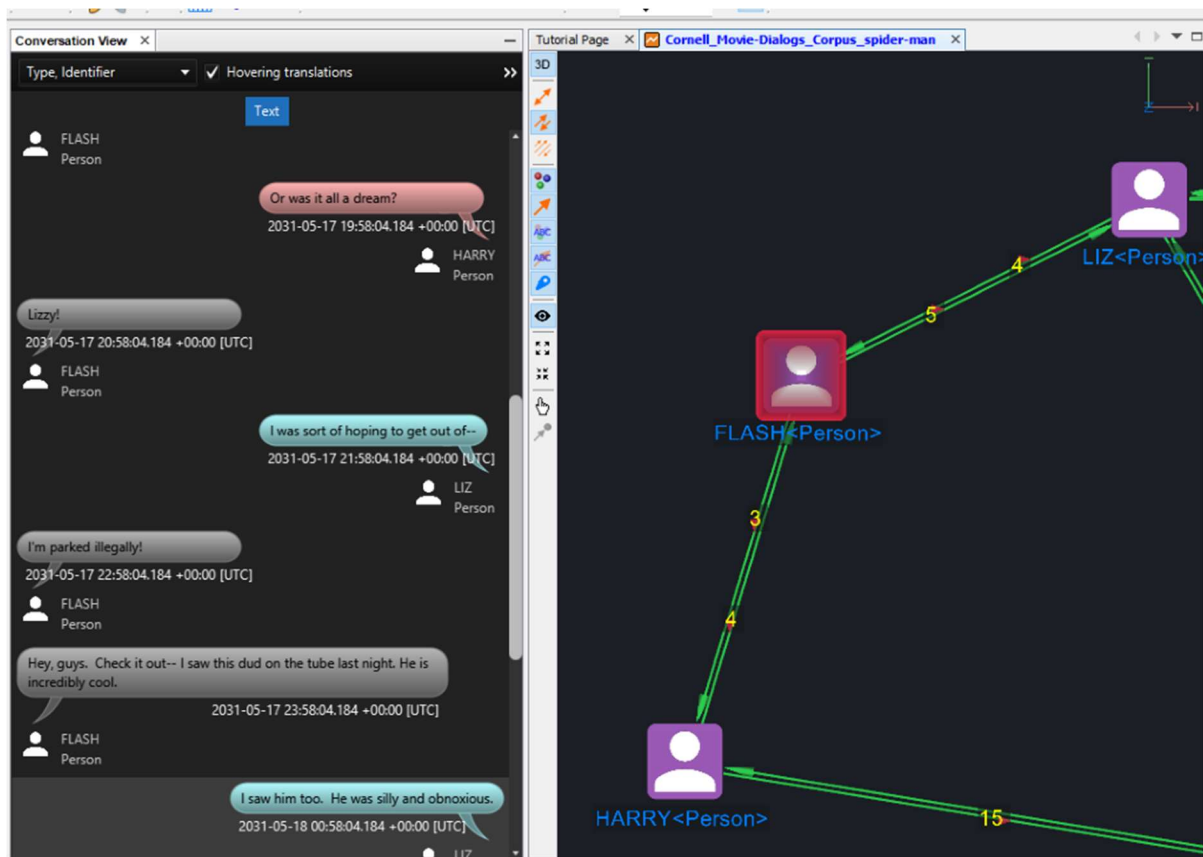


## Using the Conversation View to view Content

The Conversation View provides a handy interface for viewing Content in a natural, human readable way. It presents information as Conversations, as you might view them in a mobile phone messenger app. By selecting transactions with content, they will be displayed with the timestamp information, as well as node Identifier and Type information.

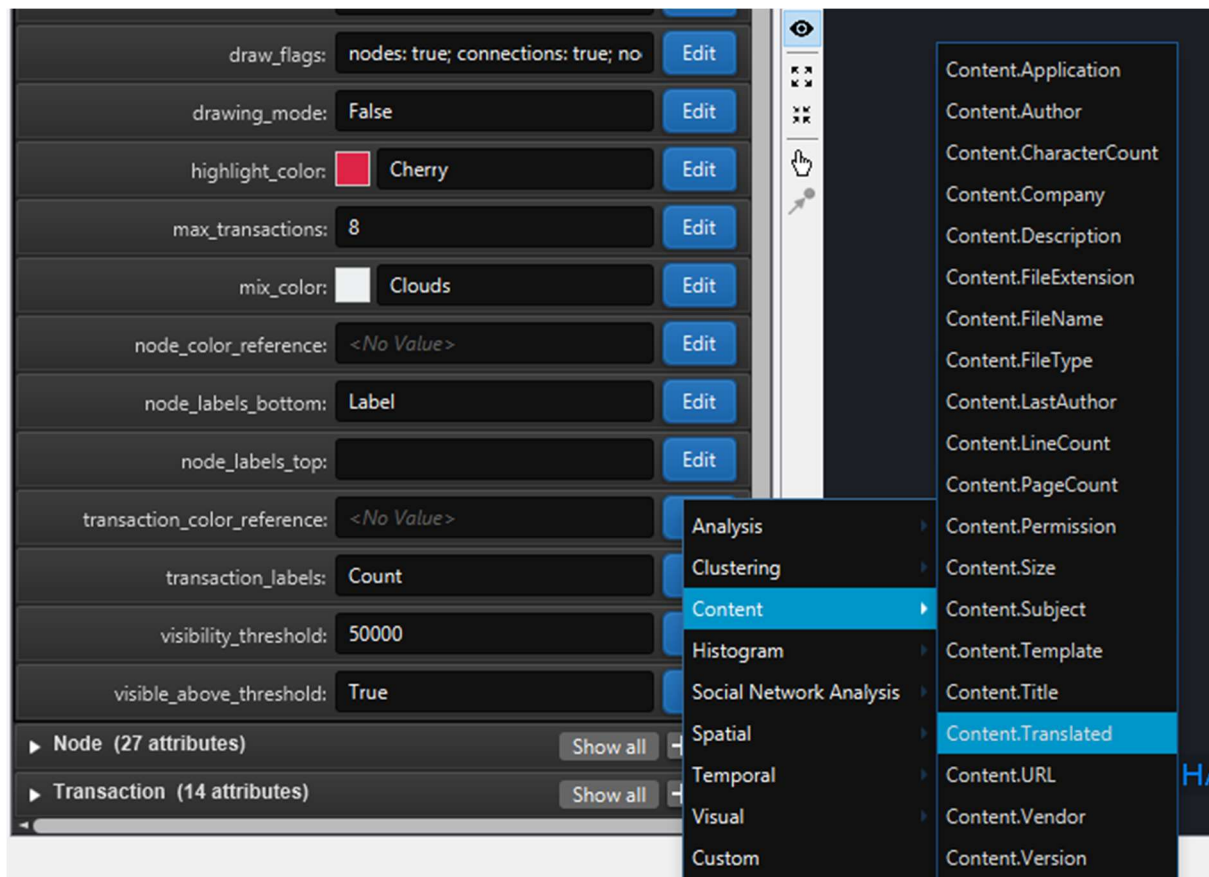


If you select a Node, the Conversation View will display all the outgoing and incoming information being sent between that Node and its neighbours.

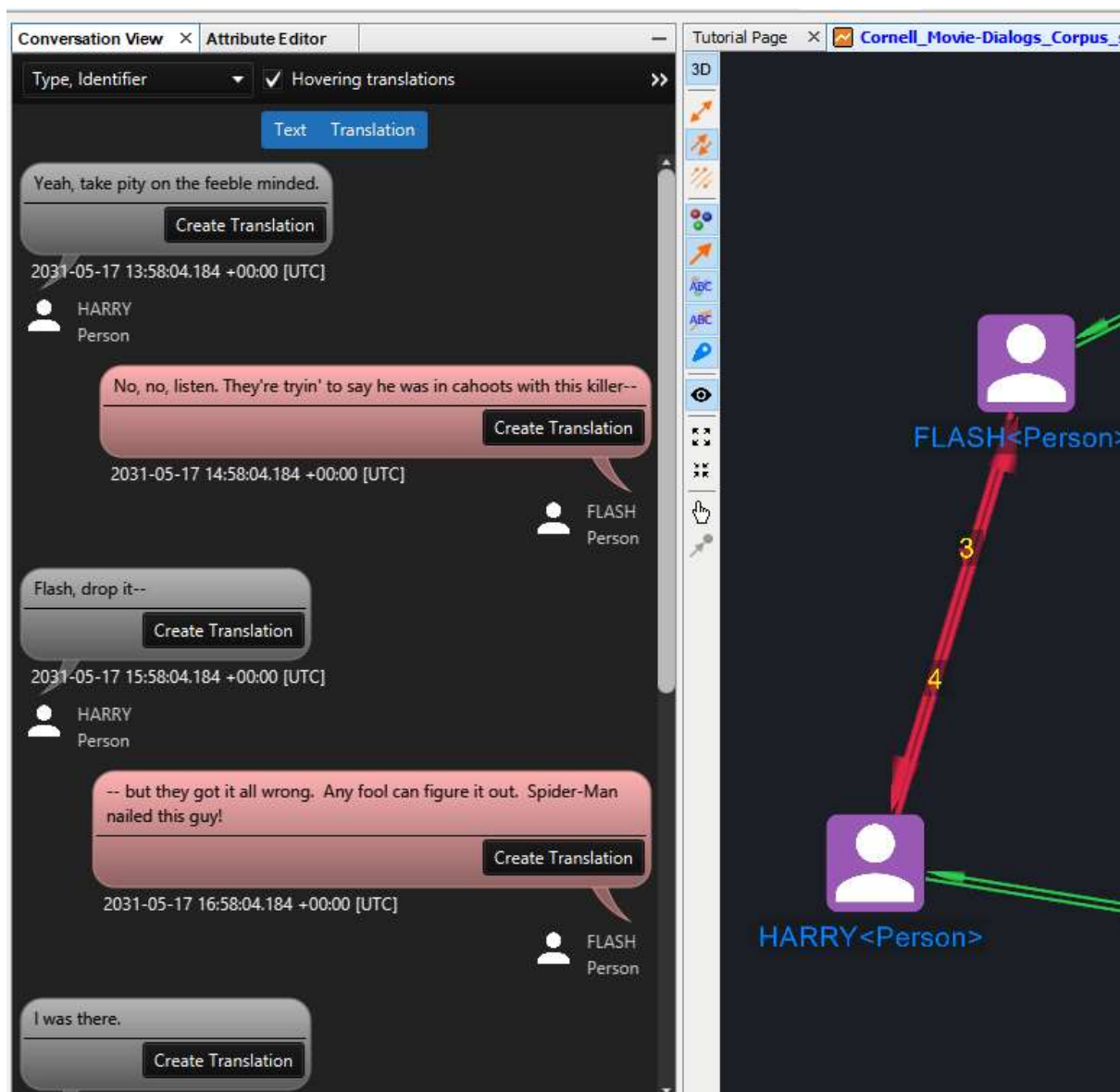


## Using the Conversation View to add translations

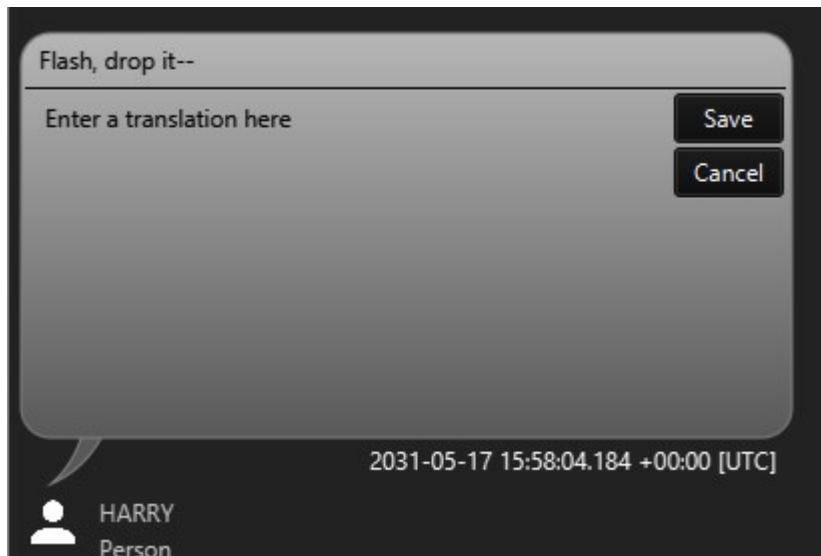
If assistance is required to help interpret the content on the graph, it is possible to add translations that can be configured in the Conversation View. First, add the Content.Translated Attribute using the Attribute Editor



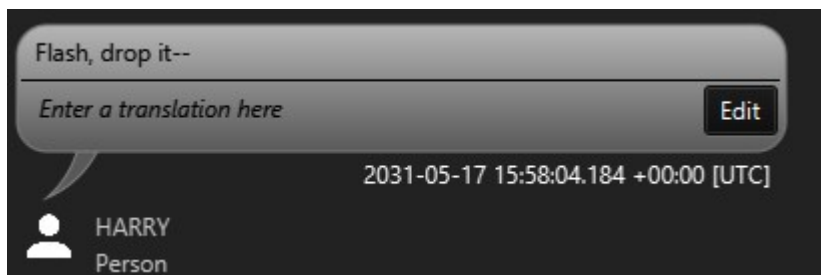
The Conversation View will then provide the option of adding translations, which will be saved to this new Attribute.



Clicking on Create Translation will allow you to enter text. Click Save to store your translation or Cancel to exit.

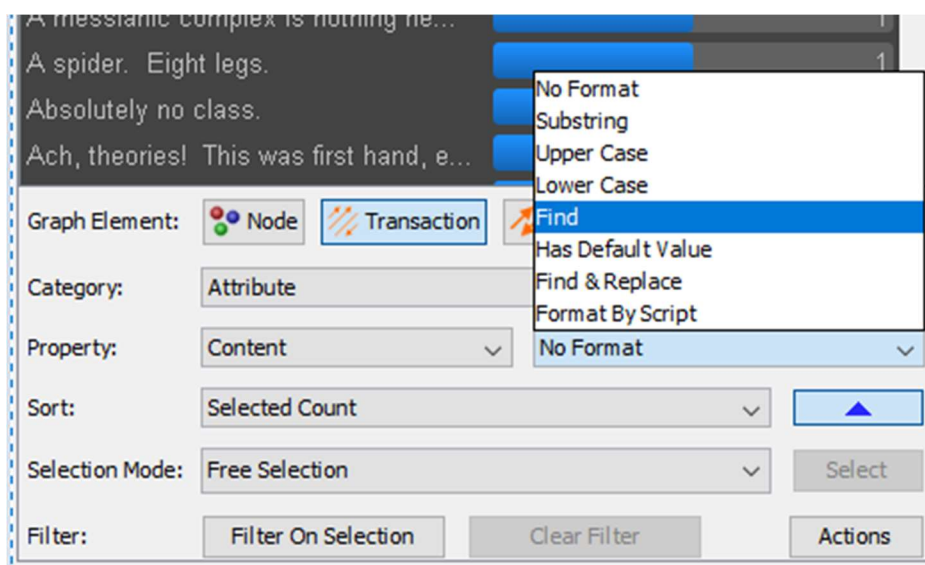


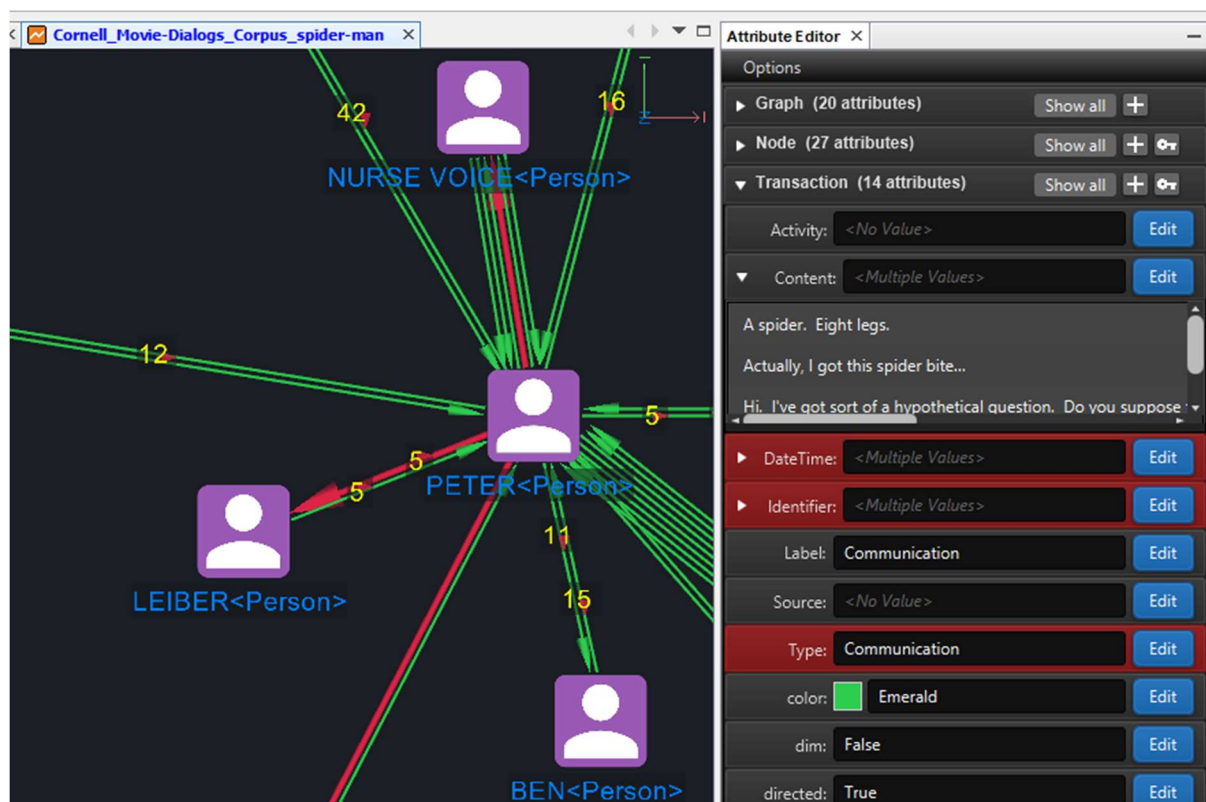
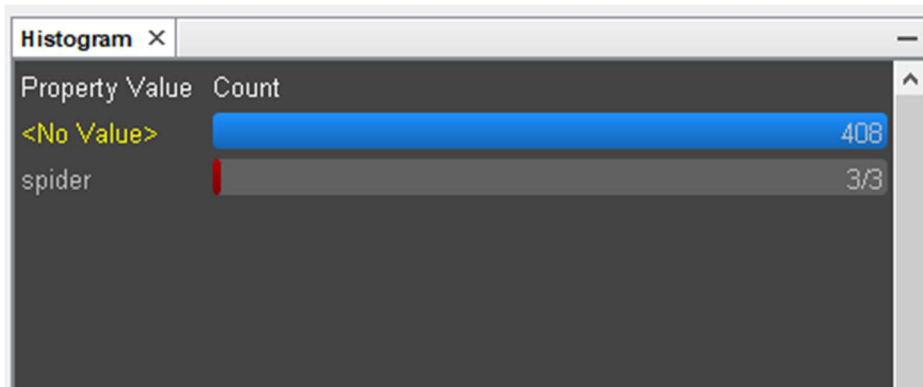
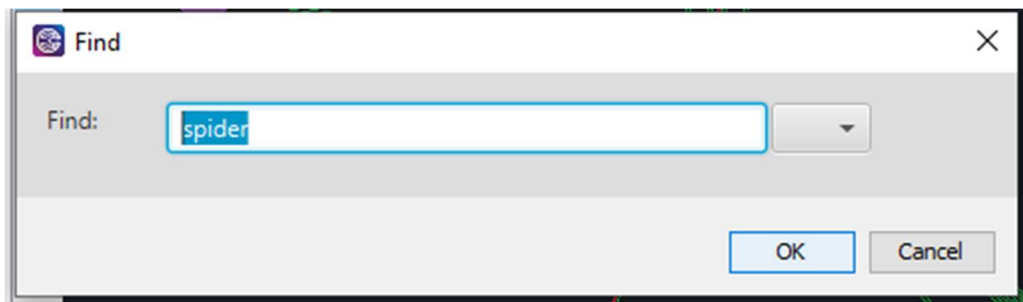
Once saved, the translation will appear below the Content in italicised text and can be edited.



## Using the Histogram View to find Content

Formatting the Histogram View is a useful way to find Content of interest. Use the Find format to search for transactions that use either specific words, or patterns of text via Regular Expressions (<https://blog.usejournal.com/regular-expressions-a-complete-beginners-tutorial-c7327b9fd8eb>)

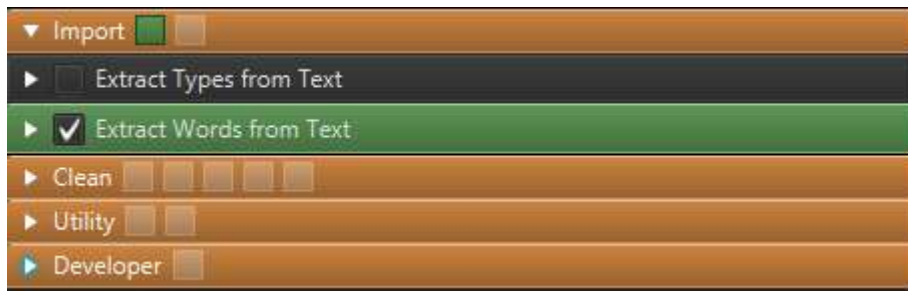




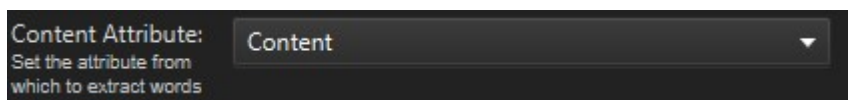
### Extracting content as features

Finally, we can use the Extract Words from Text Data Access Plugin to add Content to the graph as Nodes, which will allow us look at content usage in a graph context.

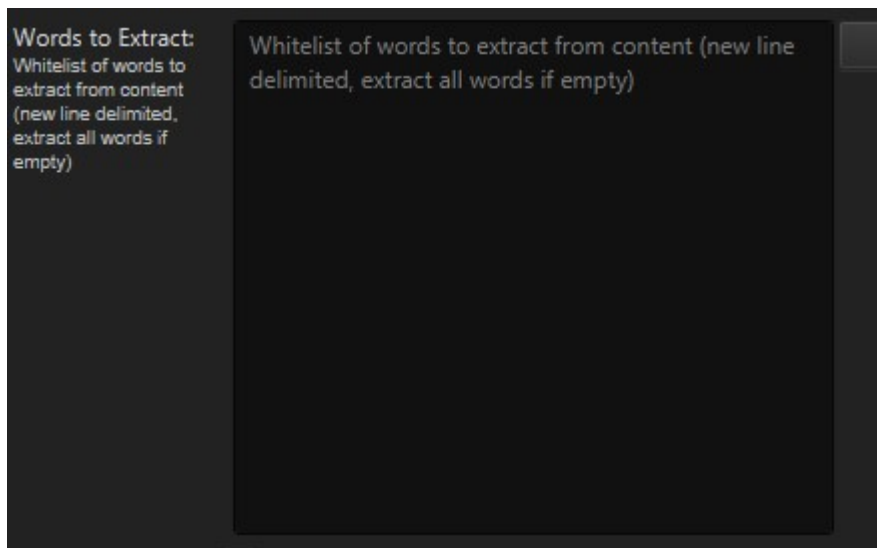




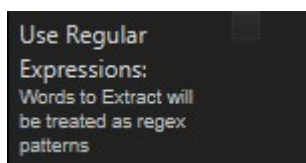
Firstly, set the Content Attribute to Content – this will tell the plugin which Transaction Attribute to extract words from.



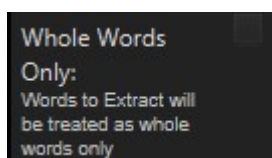
Next, if you want to find specific words, enter them in the Words to Extract box as a line delimited list. If you leave this blank, the plugin will extract all words.



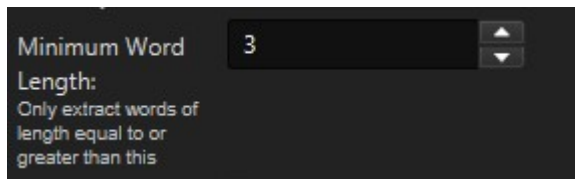
You can toggle whether this will extract either exact matches or regular expressions with the next parameter.



The next parameter will toggle whether to search for whole words only, or whether you want to extract partial words

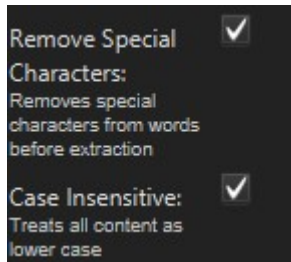


If extracting all words from text, the next parameter will determine the minimum word length to extract.



Minimum Word Length: 3  
Only extract words of length equal to or greater than this

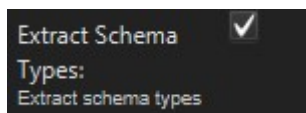
The next two parameters will toggle how content will be cleaned up before the extraction.



Remove Special Characters: ☒  
Removes special characters from words before extraction

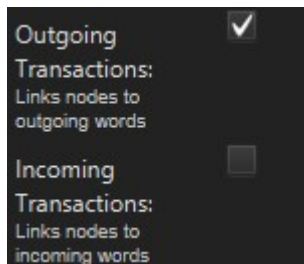
Case Insensitive: ☒  
Treats all content as lower case

The next parameter will make the plugin behave similarly to the **Extract Types From Text** plugin, only it will be extracting Types from Content.



Extract Schema Types: ☒  
Extract schema types

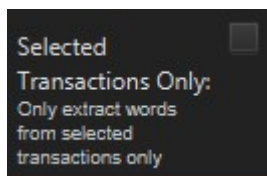
The next two parameters will determine how results will be linked in Constellation. You can link existing nodes to words if they appear in Outgoing Transactions, Incoming Transactions, or both.



Outgoing Transactions: ☒  
Links nodes to outgoing words

Incoming Transactions: ☐  
Links nodes to incoming words

And finally, you can specify whether to only extract words from Content of selected transactions only.



Selected Transactions Only: ☐  
Only extract words from selected transactions only

Now for some examples. Below, the word “spider” is being extracted from content.

Data Access View

Conversation View

Attribute Editor

Options

?

+

★

Quality Score: N/A

Go

Type to search for a plugin

Step 1

name for the query

Range:

The date and time range to query

Days: 1 2 3 4 7 14 Months: 1 3 6 12 2

Absolute range

Import

Extract Types from Text

Extract Words from Text

Content Attribute:

Set the attribute from which to extract words

Content

Words to Extract:

Whitelist of words to extract from content (new line delimited, extract all words if empty)

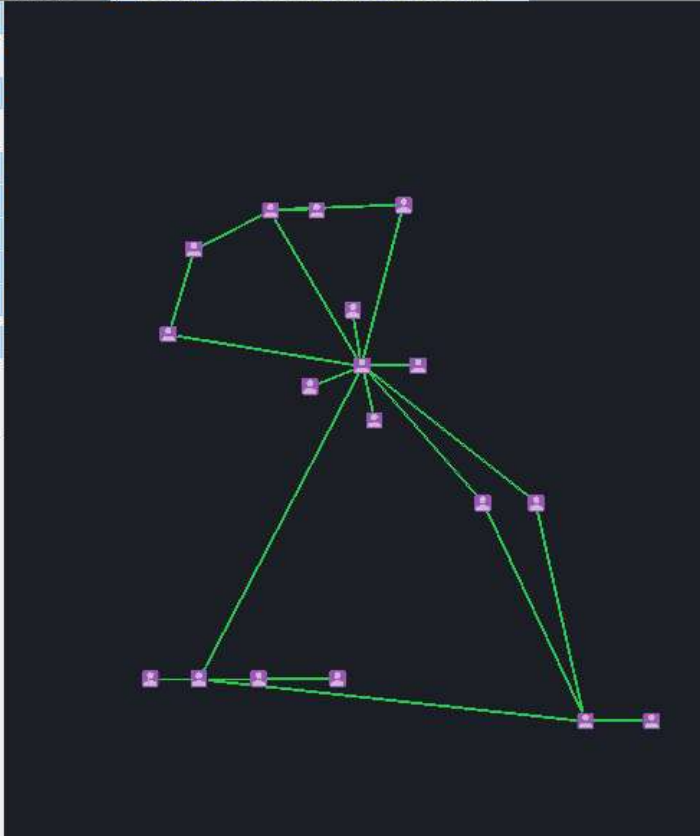
spider

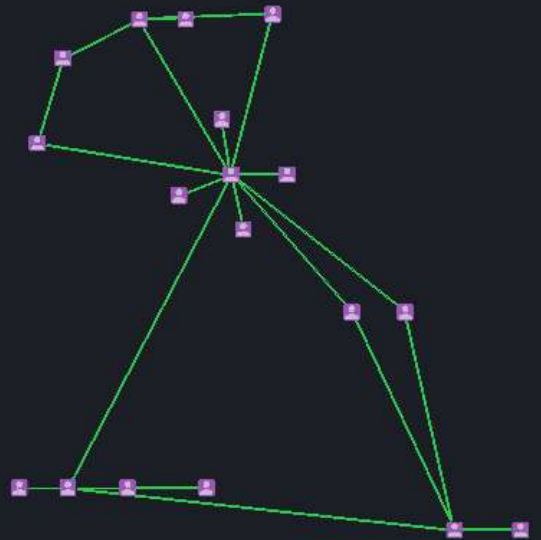
Use Regular

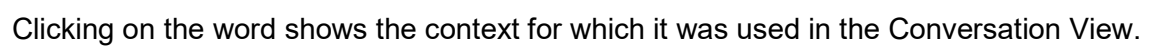
Tutorial Page

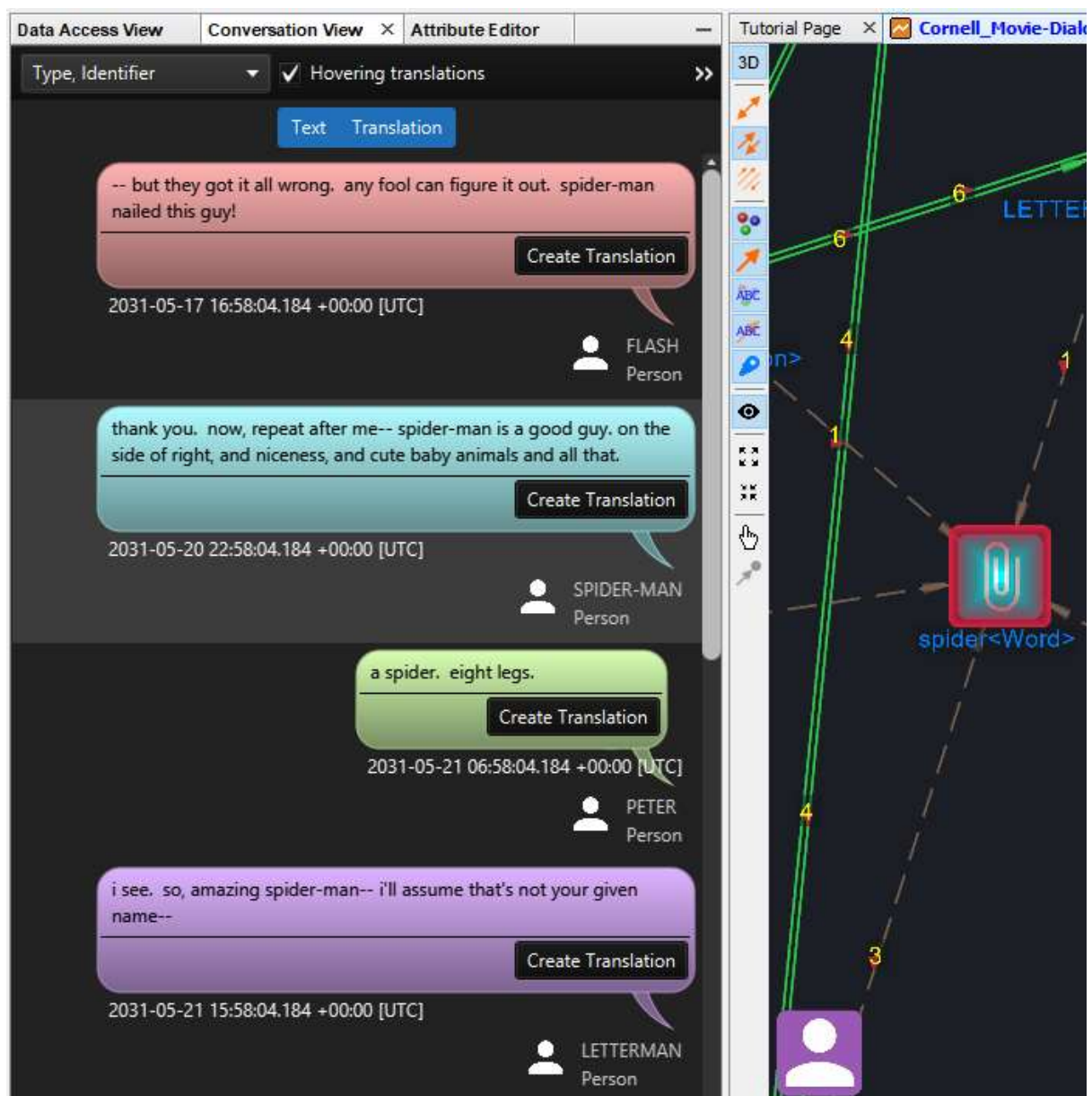
Cornell\_Movie-Dialogs\_Corpus\_spider-man

3D





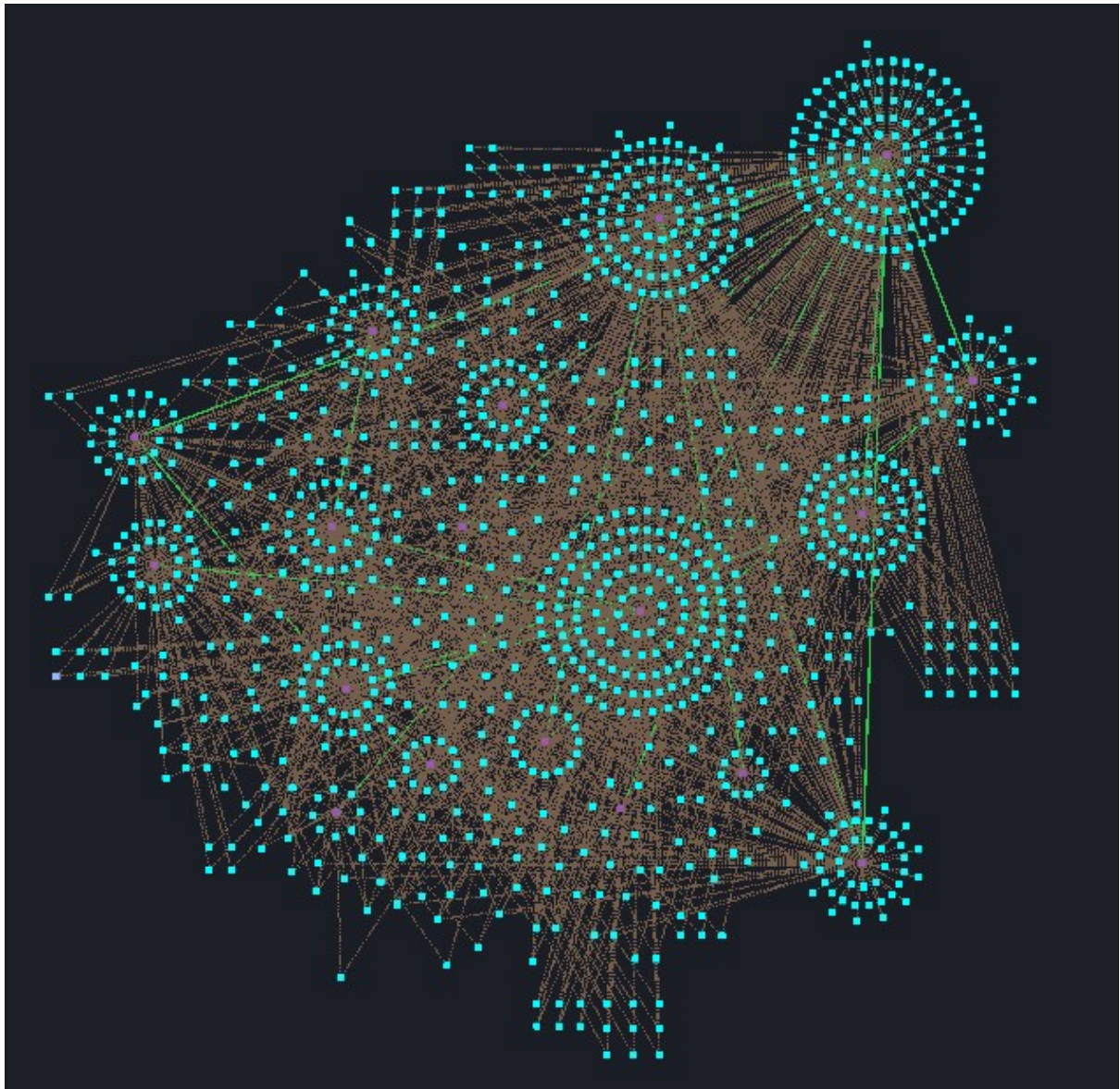




The neighbour count of the word tells us how many entities used the word, and the transaction count tells us the number of transactions in which the word was used.

Extracting all words looks something like this:





We can filter on Word type in the Histogram View to see which words are used the most, or we can select entities and use a One Hop Induced Subgraph to find which words they use in common.