The world of Business Intelligence (BI) has long supported visualization. With BI you can do many things – 
- drill down 
- drill across 
- summarize, and so forth.

There is great value in being able to see things in a visual fashion. While lists are one form of displaying results, lists do not allow the person – at a glance – to survey what is being presented. In order to use lists, you have to dig into the information. And when there are a large amount of entries into the list, digging into the list is not easy. All to often important results become hidden or hard to find.

But with visualization, what was hard to see with reports and lists become easy to see. Sorted and aggregated, visuals tell a story a mile away. It is for these reasons that visualization of results is a popular approach to looking at BI.

But there is a problem.

Visualization in the conventional sense does not apply to the world of unstructured data.

There are some good reasons why visualization does not apply to conventional technology. The biggest reason is that visualization is primarily for numeric data and unstructured data is primarily textual. Trying to feed unstructured textual data into traditional BI technology is like trying to make AC current electricity work in a DC current system. It is simply a gross mismatch.

Thus it is no surprise that traditional BI technology simply is not appropriate to the visualization of unstructured data.

There is a new solution for unstructured visualization. The new solution recognizes that display of numeric data is not appropriate for unstructured data. So the new solution takes an entirely innovative approach. The new solution takes the approach that textual data itself can and should be displayed.
The first step is to use the unstructured data collection component in order to find and prepare the textual data for visualization.

The preparation is done by reading the unstructured data, looking for the data that needs to go into the visualization. The data can be read from a wide variety of sources, such as .txt files, .doc files, .xls files, .pdf files, email, email attachments and so forth. Once the unstructured data is read, it is placed into a work area where there are words and indexes. The words are what will go into the visualization. The indexes are where the sources of the words are.

As the words and indexes are created, there is all sorts of processing that occurs.

Depending on what the analyst wants the visualization to look like, the analyst can –

- eliminate words
- edit words
- stem words back to their root
- count words
- rank words, and so forth.

2 UNSTRUCTURED VISUALIZATION
All this back end work is done before the words are ready to be presented to the analytical engine for visualization.

Once the words have been gathered and processed, they are ready to go into the visualization engine.

At this point the unstructured text component passes off the words and indexes that have been captured to a SOM (self organizing map) creation component. The SOM creation component takes the words and indexes and processes them. Once processed by the SOM creation component a mapping of the words is created. The mapping contains the words that have been processed ranked according to their number of occurrences and their importance. A SOM, or self organized map, is created.

Once the mapping is created, the mapping can be searched and visualized in many ways. The data is organized according to its contents inside documents. Many documents can be handled, and the relationship between documents is created by the SOM creation component.

The contents of a document can be searched. The theme of the documents can be searched collectively. The individual occurrences of words can be examined. The words can be searched in a drill down manner. The information found in the documents can be clustered. And these are but a few of the capabilities of the SOM structuring of unstructured data.

In many ways a SOM is like business intelligence for numeric data, such as that provided by Business Objects, Cognos, and others.

For more information about textual visualization, contact Carol Renne at 303-973-3788 or at crenne@inmondatasystems.com.