

# Generating Annotated Graphs using the NLG Pipeline Architecture

Saad Mahamood, William Bradshaw and Ehud Reiter

Arria NLG plc

Aberdeen, Scotland, United Kingdom

{saad.mahamood, william.bradshaw, ehud.reiter}@arria.com

## Abstract

The Arria NLG Engine has been extended to generate annotated graphs: data graphs that contain computer-generated textual annotations to explain phenomena in those graphs. These graphs are generated alongside text-only data summaries.

## 1 Introduction

Arria NLG<sup>1</sup> develops NLG solutions, primarily in the data-to-text area. These solutions are NLG systems, which generate textual summaries of large numeric data sets. Arria's core product is the Arria NLG Engine,<sup>2</sup> which is configured and customised for the needs of different clients.

Recently Arria has extended this core engine so that it can automatically produce annotated graphs, that is, data graphs that have textual annotations explaining phenomena in those graphs (see example in Figure 1). This was developed after listening to one of our customers, whose staff manually created annotated graphs and found this process to be very time-consuming. The annotated graph generation process is integrated into the NLG pipeline, and is carried out in conjunction with the generation of a textual summary of a data set.

In this short paper we summarise the relevant research background, and briefly describe what we have achieved in this area.

## 2 Background: Multimodality and NLG

Rich media such as web pages and electronic documents typically include several modalities in a given document. A web page, for example, can contain images, graphs, and interactive elements. Because of this there has been an interest within

the NLG community in generating multimodal documents. However, basic questions remain as how best to combine and integrate multiple modalities within a given NLG application.

### 2.1 Annotated Graphics

Sripada and Gao (2007) conducted a small study where they showed scuba divers different possible outputs from their *ScubaText* system, including text-only, graph-only and annotated graphs. They found that divers preferred the annotated graph presentation. The *ScubaText* software could not in practice produce annotated graphs for arbitrary input data sets and automatically set the scale based on detected events, so this was primarily a study of user preferences.

McCoy and colleagues have been developing techniques to automatically generate textual summaries of data graphics for visually impaired users (Demier et al., 2008). This differs from our work because their goal is to replace the graph, whereas our goal is to generate an integrated text/graphics presentation.

There were several early systems in the 1990s (Wahlster et al., 1993; Feiner and McKeown, 1990, for example), which generated integrated presentations of figures and texts, but these systems focused on annotating static pictures and diagrams, not data graphics. The WIP system, which combined static diagram images and text, used a deep planning approach to produce tightly integrated multimodal documents; it is not clear how robustly this approach handled new data sets and contexts.

### 2.2 Embodied Conversational Agents

In recent years the challenge of combining multiple modalities such as text, speech, and/or animation has been addressed in the context of Embodied Conversational Agents or ECAs. One example is the NECA system (Krenn et al., 2002).

<sup>1</sup>Arria NLG plc (<https://www.arria.com>)

<sup>2</sup>For more information see: <https://www.arria.com/technology-A300.php>

It allowed two embodied agents to converse with each other via spoken dialogue while being able to make gestures as well. From an architectural perspective, NECA used a pipeline architecture in some ways similar to the standard NLG data-to-text pipeline (Reiter and Dale, 2000). Document Planning is handled by the Scene Generator, which selects the dialogue content. The ‘Multi-modal NLG’ (M-NLG) component handles Microplanning and Surface Realisation, and also deals with specifying gestures, mood, and information structure. Thus the textual output generated by the surface realiser in the NECA M-NLG component is annotated with metadata for other modalities. In particular, information on gestures, emotions, information structure, syntactic structure and dialogue structure (Krenn et al., 2002) are also included to help inform the speech synthesis and gesture assignment modules.

### 2.3 Background: Psychology

The question of whether information is best presented in text or graphics is in principle largely one for cognitive psychology. Which type of presentation is most effective, and in which context? The answer of course depends on the communicative goal, the type of data being presented, the type of user, the communication medium and other contextual factors.

In particular, a number of researchers (Petre, 1995, for example) have pointed out that graphical presentations require expertise to interpret them and hence may be more appropriate for experienced users than for novices. Tufte (1983) points out that statistical graphs can be very misleading for people who are not used to interpreting them.

Alberdi et al (2001) report on a number of psychological studies on effectiveness of data visualisations which were performed with clinicians in a Neonatal Intensive Care Unit (NICU). At a high level, these studies found that visualisations were less effective and less used than had been hoped. Detailed findings include the following:

- Although consultants, junior doctors, and nurses all claimed in interviews to make heavy use of the computer system which displayed visualisations, when observed onward only senior consultants actually did so; junior doctors and nurses rarely looked at the computer screen.

- Senior consultants were much better than junior staff at distinguishing real events from noise (sensor artefacts).
- Even senior consultants could only identify 70% of key events purely from the visualisations.

Law et al (2005) followed up the above work by explicitly comparing the effectiveness of visualisations and textual summaries. The textual summaries in the experiment were manually written, but did not contain any diagnoses and instead focused on describing the data. Law et al found that clinicians at all levels made better decisions when showed the textual summaries; however at all levels they preferred the visualisations.

A similar study with computer generated summary texts produced by the Babytalk BT45 system (Portet et al., 2009), conducted by van der Meulen et al (2008), found that decision quality was best when clinicians were shown manually written summaries; computer generated texts were of similar effectiveness to visualisations. An error analysis of this study (Reiter et al., 2008) concluded that computer generated texts were much more effective in some contexts than in others.

An implication of the above studies is that in many cases the ideal strategy is to produce both text and graphics. This increases decision effectiveness (since the modalities work best in different situations), and also increases user satisfaction, since users see the modality they like as well as the one which is most effective for decision support.

### 2.4 Annotated Graphs in NLG Engine

We have extended our NLG Engine to generate annotated graphs as well as texts; an example output, generated by a demonstration system, is shown in figure 1. This example shows a very simple textual output; examples of more complex textual output are on the Arria website<sup>3</sup>.

This example output shows a comparison of performance between the FTSE 100 and a given stock portfolio. The value of the FTSE 100 is used as a performance benchmark to see if a given stock portfolio is performing better or worse than compared to the stock market in general.

As can be seen in the graph in figure 1, the annotations are small text fragments, which are placed

<sup>3</sup>A more detailed example is given here: <https://www.arria.com/case-study-oilgas-A231.php>

### Monthly Report for October, 2011

This month the market was positive with FTSE gaining 9.23 percent. Your portfolio rose by 22.84 per cent to 176777.5 GBP. Your simple return stands at 16.736 per cent.

Looking at your holdings, Mining stocks rose the most with a 0.42 percent advance. During this period Antofagasta advanced 26.71 per cent. London Stock Exchange Group enjoyed a gain over the last 5 days. Overall eight of your stocks rose and one fell. Your exposure to the Mining sector stands at 52.5 making your portfolio highly skewed. With recent purchase of Barclays shares your exposure to the Banking sector grew to 15.80 per cent. The most volatile security during this month was BAE Systems.

#### SUMMARY

Current Value 1.76777455E7 Simple Return 16.736%

Security Name	Last Price (GBX)	Quantity	Today's % Value Change	Current Value
Aberdeen Asset Management	174.77	200	0.113%	349.54
Antofagasta	1137.6	5200	0.267%	59155.2
BAE Systems	2.77	5350	0.069%	148.2
Barclays	185.09	7350	0.25%	13604.12
BHP Billiton	1770.23	15350	0.151%	271730.3
HSBC Holdings	490.36	16350	0.118%	80173.86
Lloyds Banking Group	32.49	17550	-0.029%	5702
London Stock Exchange Group	847.31	17570	0.107%	148872.37
Sage Group	283.0	17670	0.082%	50006.1

### Portfolio Value vs FTSE100

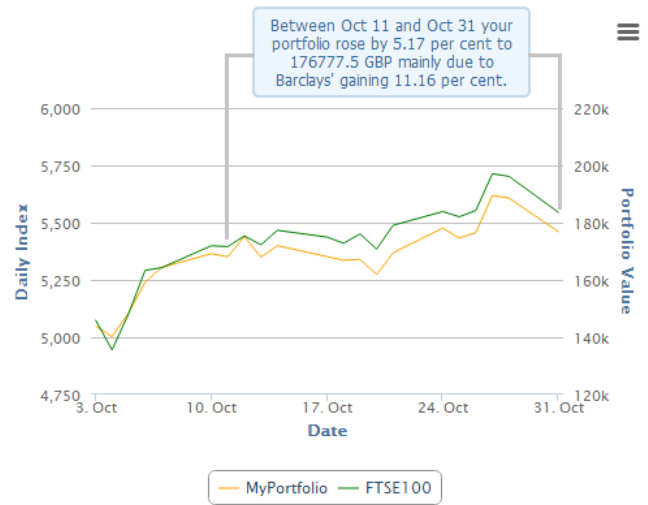


Figure 1: Combined text and annotated graph detailing the stock portfolio performance

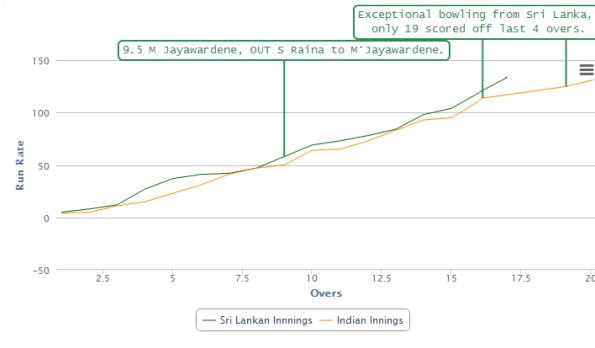


Figure 2: Graph illustrating stacking capabilities when two annotations intersect each other

on top of the graph, and are linked to the relevant events in that occur in the graph. Annotations can also be placed at the bottom of graphs and at the sides and can rearrange themselves depending on the space available. In figure 1 the range annotation used indicates the reason for the increase in the value of a given stock portfolio over a particular time period. If one or more annotations collide or intersect a stacking algorithm is used prior to presentation to rearrange the placement of colliding annotations as shown in figure 2.

Figure 3 illustrates the architecture that is used by our NLG engine. The data analysis and data interpretation modules analyse the input data and produce a set of messages which can be communicated to the user in the generated report. The document planner decides on which messages should be communicated overall, and where messages should appear (for example, situational analysis text, diagnosis text, impact text, graph annotation, or a combination of these). The document planner also decides on the type of graph used, and which data channels it displays; these data channels must include any channels which are annotated, but in some cases other channels are displayed as well.

Once document planning is complete, the visualisation planning module generates the graph design, including X and Y scale and the position of the annotations on the graph. The time range shown in the graph is largely determined by the annotation messages. In other words, the decision about what data to show on the graph is partially driven by the annotations.

The annotation microplanner and realiser generate the actual annotation texts, using special rules which are optimised for annotations (which need

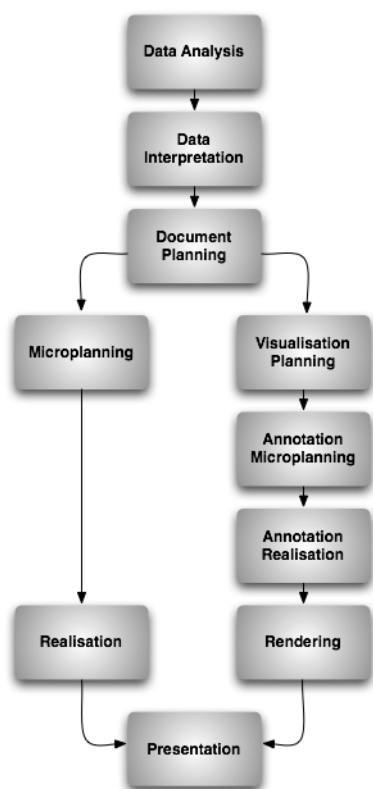


Figure 3: Pipeline architecture of the Arria NLG Engine

to be short and have different referring expressions). After this has been completed, a renderer produces the actual annotated graph. The final task lies with the presenter module, which recombines the separately generated summary text (generated by the NLG Microplanning and Realisation modules) with the annotated graphs.

### 3 Conclusion

Annotated graphs are a very appealing mechanism for combining text and data graphics into a single multimodal information presentation; this is shown both by the findings of Sripada and Gao (2007) and by the experiences of our customers. Amongst other benefits, we believe that annotated graphs will address some of the deficiencies in data graphics which were pointed out by Alberdi et al (2001), by helping users (especially inexperienced ones) to more easily identify key events in a graph and also to distinguish real events from sensor artefacts and other noise.

We have developed software to create annotated graphs, by modifying the standard NLG data-to-text pipeline as described above. Our clients have

reacted very positively so far, and we are now exploring extensions, for example by making annotated graphs interactive.

### References

- E. Alberdi, J. C. Becher, K. J. Gilhooly, J. R.W. Hunter, R. H. Logie, A. Lyon, N. McIntosh, and J. Reiss. 2001. Expertise and the interpretation of computerised physiological data: Implications for the design of computerised physiological monitoring in neonatal intensive care. *International Journal of Human Computer Studies*, 55(3):191–216.
- S. Demier, S. Carberry, and K. F. McCoy. 2008. Generating textual summaries of bar charts. In *Fifth International Natural Language Generation Conference (INLG 2008)*, pages 7–15. Association for Computational Linguistics.
- S. Feiner and K. R. McKeown. 1990. Generating Coordinated Multimedia Explanations. In *Sixth Conference on Artificial Intelligence Applications*, volume 290-296.
- B. Krenn, H. Pirker, M. Grice, S. Baumann, P. Piwek, K. van Deemter, M. Schroeder, M. Klesen, and E. Gstrein. 2002. Generation of multi-modal dialogue for a net environment. In *Proceedings of KONVENS-02*, Saarbruecken, Germany.
- A. S. Law, Y. Freer, J. Hunter, R. H. Logie, N. McIntosh, and J. Quinn. 2005. A comparison of graphical and textual presentations of time series data to support medical decision making in the neonatal intensive care unit. *Journal of Clinical Monitoring and Computing*, 19(3):183–194.
- M. Petre. 1995. Why Looking isn't always Seeing: Readership Skills and Graphical Programming. *Communications of the ACM*, 38:33–44.
- F. Portet, E. Reiter, A. Gatt, J. Hunter, S. Sripada, Y. Freer, and C. Sykes. 2009. Automatic generation of textual summaries from neonatal intensive care data. *Artificial Intelligence*, 173(7-8):789–816.
- E. Reiter and R. Dale. 2000. *Building Natural Language Generation Systems*. Cambridge University Press.
- E. Reiter, A. Gatt, F. Portet, and M. van der Meulen. 2008. The Importance of Narrative and Other Lessons from an Evaluation of an NLG System that Summarises Clinical Data. *Fifth International Natural Language Generation Conference (INLG 2008)*, pages 147–155.
- S. G. Sripada and F. Gao. 2007. Summarizing dive computer data: A case study in integrating textual and graphical presentations of numerical data. In *MOG 2007 Workshop on Multimodal Output Generation*, pages 149–157.

- E. Tufte. 1983. *The Visual Display of Quantitative Information*. Graphics Press.
- M. van der Meulen, R. Logie, Y. Freer, C. Sykes, N. McIntosh, and J. Hunter. 2008. When a graph is poorer than 100 words: A comparison of computerised natural language generation, human generated descriptions and graphical displays in neonatal intensive care. *Applied Cognitive Psychology*, 24:77–89.
- W. Wahlster, E. André, W. Finkle, HJ. Profitlich, and T. Rist. 1993. Plan-based integration of natural language and graphics generation. *Artificial Intelligence*, 63:387–427.