

# Towards Generating Textual Summaries of Graphs

Kathleen F. McCoy, M. Sandra Carberry, Tom Roper  
Dept. of Computer and Information Sciences  
University of Delaware  
Newark, DE 19716, USA  
mccoy@cis.udel.edu, carberry@cis.udel.edu, roper@cis.udel.edu

Nancy Green  
Dept. of Mathematical Sciences  
University of North Carolina at Greensboro  
Greensboro, NC 27402, USA  
green@uncg.edu

## Abstract

For people who use text-based web browsers, graphs, diagrams, and pictures are inaccessible. Yet, such diagrams are quite prominent in documents commonly found on the web. In this work we describe a method for rendering important aspects of a line graph as plain text. Our method relies heavily on research in natural language generation and is motivated by an analysis of human-written descriptions of line graphs. In this paper we concentrate on determining which aspects of the line graph to include in a summary.

## 1 Introduction

The amount of information available electronically has increased dramatically over the past decade. The challenge is to develop techniques for providing effective access and utilization of such information so that all individuals can benefit from these resources and so that the information is readily available when needed. Unfortunately, many knowledge sources are provided in a single format; thus they are not accessible to individuals who cannot assimilate information in that format. For example, individuals with visual impairments have limited access to data stored as graphs; this prevents them from fully utilizing available information resources. In addition, even for individuals without disabilities, such graphical information is readily accessible only in environments with high-bandwidth transmission and viewing facilities. For example, recent developments have led to cellular telephones for accessing the Web. The use of such devices highlights the need for tools that convey information in other modalities.

In this work we attempt to make the major information content of some kinds of graphs available in textual form. We are not interested in creating a description of the graph itself, rather we wish to produce a summary of the major points the graphic is intended to convey in an understandable and well-structured text. To illustrate our approach, consider the case of a line graph. In choosing what is important to include in a summary, we attempt to take into account the presentational goals that led the author to produce a line graph (as opposed to another form of representation) in the first place. In addition, we consider what aspects of line graphs are “typically” included in descriptions of such graphs, and look at organizational methods found in naturally occurring line-graph summaries.

In this paper we first outline some related research. This includes work on both providing alternative modes of delivering graphical information, and on natural language generation research that is relevant to this problem. Next we lay out our proposed architecture for summarizing a graph. We give some detail of two components of this architecture: the intention recognition component and the component responsible for choosing aspects of the graph to include in the summary.

## 2 Related Research

There has been some work on providing alternative modes of presentation of graphical/diagrammatic information for people who have visual impairments. For example, [Zhang and Krishnamoorthy, 1994] describe a package designed for people with visual impairments to study graph theory. Input to their system is an adjacency matrix for a graph. The system generates a description of the graph tailored to the particular purpose of graph theory applications. Because of this, items of importance can be predefined. In our work we wish to take in graphs that are part of larger texts where the graph has been included to make a particular point or points. Thus, one of the major research efforts involves determining the appropriate description given a graph. We conjecture that the important things to say will be dependent on the intentions that led to the selection of that particular graphical device, as well as (visual) features of the data displayed in the graph itself.

Several systems exist that attempt to allow a user to “visualize” a graphic. For instance, [Meijer, 1992] describes a system that converts a video image into a “soundscape”. Others describe a transformation of information possible by taking advantage of the underlying structure of the information [Barry et al., 1994]. Notable projects include ASTER [Raman, 1997] and several projects from the Science Access Project at Oregon State [Bargi-Rangin et al., 1996], [Gardner, 1998], [Walsh and Gardner, 2001], [Walsh et al., 2001]. These projects rely on meta-document

information to help interpret the graphical components in the intended way. A great deal of work also comes from the Multimodal Interaction Group (part of the Glasgow Interactive Systems Group [Brewster, 1994]). Much of this work involves the application of earcons [Blattner et al., 1989] and sonically-enhanced widgets. This work and others (e.g., [Kurze et al., 1995], [Blenkhorn and Evans, 1995] [Kennel, 1996]) relies on a human to “translate” the graph into a form where the earcons and widgets can be used.

While these systems are clearly extremely valuable, their basic focus is on graphical elements and their rendering in an alternative medium. In our work, rather than enabling the user to reason on a “reproduced” graphical image, we concentrate on describing the important content the graph is intended to convey. We take the view that the graph was produced in order to get across particular information. Thus one can view the summary we wish to generate as an alternative mode of getting across the important informational content of the graph.

Most previous work in the area of Natural Language Generation surrounding graphical presentation has centered on generating multi-media explanation (e.g., [McKeown et al., 1992], [Wahlster et al., 1993], [Arens and Hovy, 1990]). In particular, that work attempts to generate a presentation that contains both text and graphics. In contrast, our work is aimed at providing a textual explanation of a human-generated graphic.

Two projects that do deal with the kind of data of interest here are (1) the caption generation system associated with the SAGE project [Mittal et al., 1998] and (2) the PostGraphe system [Fasciano and Lapalme, 2000], [Fasciano and Lapalme, 1996], [Lapalme, 1999]. The caption generation system is concerned with generating a caption that enables the user to understand how to read the data in the graphic (i.e., how the graphic expresses the data). The system is thus driven by an identified set of “graph complexities” that the caption explains to the user. Rather than graph complexities, our proposed work must identify and convey aspects of the data depicted in the graph that are important to the user. Thus, a large portion of our work involves identifying features that convey importance.

The PostGraphe system [Fasciano and Lapalme, 2000], [Fasciano and Lapalme, 1996], [Lapalme, 1999] generates statistical reports that integrate both graphics and text (accompanying caption). The primary input is data in a tabular form, an indication of the types of values in the columns of the table, and the intentions that the user wishes to convey. Examples of intentions handled by PostGraphe include: present a variable, compare variables or sets of variables, and show the evolution of a variable with respect to another variable. A user may have multiple intentions to be achieved in a single report. PostGraphe uses a schema-based planning mechanism [McKeown, 1985] to plan both the graphics and the text. The graphical schema is chosen primarily on the basis of the input intentions using information about the kinds of graphics that are most effective in achieving those intentions. While the kind of information included in the PostGraphe captions are of the type we would hope to generate, there are significant differences between the input and goals of the two projects. A significant portion of our work concerns determining what these goals should be given a previously generated graphic. For example, since we don't have the input intentions, we must deduce these given the graph itself.

Note that the caption associated with the line graph we are attempting to summarize is also potentially available to our system. However, we have not yet considered how the caption may be useful. In this paper we concentrate on deducing the goals of the graphic through the properties of the graphic itself. Future work will consider how the caption might affect what is included in the explanation.

### 3 System Components

A well-known adage is that a picture is worth a thousand words. This is because a picture captures a great deal of information, some of it only tangential to the main purpose of the picture. Our goal is to communicate only the most significant and relevant information, while retaining the ability to convey additional detail if requested by the user. In addition, the designer of a visual display has some underlying purpose, which can affect how the display is created. For example, a pie chart emphasizes comparisons while a line graph brings out trends.

Our overall methodology for summarizing a graph consists of the following components:

1. Extract the basic components of the graph, identify their relationship to one another, and classify the screen image into one of the classes of graphs being handled using techniques as described in [St. Amant and Riedl, 2000], [Futrelle et al., 1995], and [Futrelle, 1999].
2. Hypothesize possible intentions of the graph designer, based on characteristics of the type of graphical display selected. For example, both a scatter plot and a pie chart can be used to graph how an entity (such as government income) is divided up among several categories (such as social welfare, military spending, etc.); however, a graphical designer will choose a pie chart if the intent is to convey the relative distributions as opposed to their absolute amounts.
3. Once we have determined the goals of the graphic designer, we still must decide what propositions should be expressed in the summary. There are likely to be many interesting things within a graphic, but we wish to select the few most emphasized ones for summarization. These are identified in our architecture through the

use of a set of features where the features that we look for are dependent on the overall characteristics of the graphic (e.g., the features looked for in a line graph will be quite different from those in a pie chart) and the intentions determined in the previous phase. The feature use and importance values for features was motivated by a study in which human subjects were asked to describe line graphs. The use of these features will produce a set of propositions along with a rating of each indicating how important they are.

4. Construct a coherent message that conveys the most significant propositions conveyed by the graphical display and translate the message into English sentences. For generating the actual sentences we turn to FUF/SURGE [Elhadad and Robin, 1996]. This is a unification-based natural language sentence generation system that is able to translate between formal specifications of English sentences into English sentences.

The following sections concentrate on the subproblems of identifying the intentions of the graph and selecting relevant features to include in the summary.

## 4 Recognizing Intention

A great deal of research is being conducted focusing on the automatic generation of multimedia presentations, such as that of the SAGE visualization group (e.g., [Kerpedjiev and Roth, 2000], [Kerpedjiev et al., 1998], [Green et al., 1998]). They propose that speech act theory can be extended to cover graphical presentations. A graphic can bring about an intended change in the reader's cognitive state in the same way that text can.

If we accept this claim, then we have a basis for trying to summarize a given graphic. There are well-researched and well-documented results concerning the construction of effective graphical presentations, and the means by which humans decode them. If we assume:

1. the author is being cooperative and efficient: the author is including information pertinent to the presentation, and avoiding extraneous information that might confuse or hide the intended information
2. the author is competent: the author is aware of the conventions for effective graphic design

then we are able to extract from the graphic a set of intentions the graphic is likely intended to convey.

The goal of our intention recognizer is the inverse of the graph design process: namely, we wish to use the graphic display as evidence to hypothesize the communicative intentions of the graph's author. This is a plan recognition problem [Carberry, 1990]. Typically, both planning and plan recognition systems are provided with operators that describe how an agent could accomplish different goals; these operators can include constraints, preconditions, subgoals and subactions, and effects of actions. Planners start with a high-level goal, select an operator that achieves that goal, and then use other operators to flesh out the details that describe how to satisfy preconditions and subgoals. On the other hand, a plan recognition system is provided with evidence (here an action performed by an agent such as the use of a particular graphical device), and chains backwards on the operators to deduce one or more high-level goals that might have led the agent to perform the observed action as part of an overall plan.

To cast intention recognition from graphic displays as a plan recognition problem, we must represent the knowledge that might have led the graphic designer to construct the observed graphic. We will construct two types of operators, goal operators and task operators. Goal operators will specify the tasks that must be supported in order to achieve various high-level intentions. Task operators will be of two types: cognitive and perceptual. Cognitive operators will specify the tasks that must be performed in order to accomplish cognitive tasks such as performing comparisons or summary computations. Perceptual operators will specify the tasks that must be performed in order to accomplish perceptual tasks such as looking up the value of a particular attribute.

The task of this module is one of plan recognition. This module has as input the graphical display and any accompanying discourse context, and must reason backwards to identify the author's intentions. In our architecture, input to the plan recognizer will be the graph type (pie chart, line graph, etc.) identified by the visual extraction module. The plan recognizer will also have access to an examination module that can answer questions about the graph and aid the plan recognizer in determining whether operator constraints are satisfied. For example, the plan recognizer might need to know whether an attribute could be a primary key. From this evidence, the system will chain backwards to identify perceptual and cognitive tasks that are supported by the graph and from these hypothesize the plausible intentions of the graph designer. In most cases, multiple intentions will be hypothesized. These will be ordered according to how well the requisite cognitive and perceptual tasks are enabled.

## 5 Identifying Important Features

Once the overall intention is recognized (e.g., get the reader to recognize a trend in the data), the system must determine what should be said in order to convey the trend to the user. To determine how this might be done, we conducted an experiment where we asked human subjects to produce summaries of some graphs that we provided. The graphs in the experiment were line graphs whose intention was arguably to get the reader to recognize a trend.

An analysis of the resulting summaries led us to conjecture the use of features of graphical objects to help select important propositions to include in the summary.

We define a graph feature as an attribute of some subset of the components of the graph. Such features can vary from local features that apply to a single point in the graph to more global features that apply to a segment of the graph or even the entire graph. In the case of line graphs, a local feature might be a maximum value at a point and a global feature might be resemblance of a portion of a graph to a parabola.

The relevant features will be different for different kinds of graphs (e.g., line graph versus pie chart). In studying our generated summaries and a number of collected line graphs from popular media such as USA Today and Newsweek, our analysis suggests several relevant features that we hypothesize will be useful in generating summaries:

- **Extrema:** the extreme values of a line graph are often significant. These are point values, and include the initial value, global minimum, global maximum, and final value
- **Slope:** slope refers to the rate of change of a line segment (or sequence of line segments viewed as a single straight line from the starting point of the initial segment to the terminal point of the last segment). Slope is relevant because it indicates how quickly the data is changing.
- **Variance:** variance refers to the raggedness, or jumpiness, of a sequence of line segments. A sequence of line segments that form a straight line would have zero variance. Variance may be significant since it indicates the stability or consistency of changes in the data.
- **Analogy:** analogy refers to a concept that a portion of the data resembles. For example, a sequence of line segments may overall resemble a parabola or sine wave. If the user is familiar with the selected concept, comparison to it provides a compact method of enabling the user to easily visualize the data in his mind.
- **Cycles:** a cycle is a repeated pattern in the line graph. As with analogies, cycles provide compact methods of characterizing graphical information.

Further experimentation needs to be done to identify a complete set of features. However, this set is a starting point that can be added to as further analysis is done.

We define significance of a feature to be the strength of the feature in the graph. For example, the more ragged a sequence of line segments, the more significance will be assigned to the variance feature. We hypothesize that significance of a feature cannot be determined in isolation. For example, the significance of a feature will be affected by the domain of the graph, with some features of greater importance in some domains than in others.

In addition, significance will also be affected by the portion of the graph that the feature spans. For example, large variance over a small segment of a graph will have less significance than large variance over a longer segment of the graph. We propose to evaluate our corpora of ranked features and explanations produced by human subjects to identify what aspects of the graph and its domain should be taken into account in estimating significance.

The result of reasoning on these features will be a set of propositions -- one for each feature whose significance is great enough to include in the summary.

These propositions will be organized and eventually rendered as text by the remaining system components.

## 6 Conclusion

We have set forth a strategy for creating a summary for a graph. What we have presented is a method that has the potential for paralleling human generated summaries. Future work might consider how to combine this basic methodology with world knowledge, user preferences, and consideration of the context of the presentation. Such knowledge will ultimately be necessary for completely natural sounding, natural language summarization of any given graphic.

## 7 References

- [Arens and Hovy, 1990] Arens, Y. and Hovy, E. H. (1990). How to describe what? Towards a theory of modality utilization. In *Proc. of the 12 Conference of the Cognitive Science Society*. Cognitive Science Society.
- [Bargi-Rangin et al., 1996] Bargi-Rangin, H., Barry, W. A., Gardner, J. A., Lundquist, R., Preddy, M., and Salinas, N. (1996). Scientific reading and writing by blind people - technologies of the future. In *Proceedings of the 1996 CSUN Conference on Technology and Persons with Disabilities*, Los Angeles, CA.
- [Barry et al., 1994] Barry, W. A., Gardner, J. A., and Raman, T. V. (1994). Accessibility to scientific information by the blind: Dotsplus and ASTER could make it easy. In *Proceedings of the 1994 CSUN Conference on Technology and Persons with Disabilities*, Los Angeles, CA.
- [Blattner et al., 1989] Blattner, M., Sumikawa, D., and Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human Computer Interaction*, 4(1): 11--44.

- [Blenkhorn and Evans, 1995] Blenkhorn, P. and Evans, D. G. (1995). Access to graphical information for blind people through speech and touch. In *Proc. of the Second TIDE Congress*, pages 298--301, Paris.
- [Brewster, 1994] Brewster, S. A. (1994). Providing a structured method for integrating non-speech audio into human-computer interfaces. PhD thesis, University of York, UK.
- [Carberry, 1990] Carberry, S. (1990). *Plan Recognition in Natural Language Dialogue*. ACL-MIT Press Series on Natural Language Processing. MIT Press, Cambridge, Massachusetts.
- [Elhadad and Robin, 1996] Elhadad, M. and Robin, J. (1996). An overview of surge: a re-usable comprehensive syntactic realization component. In *Proceedings of the Eighth International Generation Workshop (INLG96) (demonstration session)*, Herstmonceux Castle, Sussex, UK.
- [Fasciano and Lapalme, 1996] Fasciano, M. and Lapalme, G. (1996). PostGraphe: A system for the generation of statistical graphics and text. In *Proceedings of the Eighth International Generation Workshop (INLG96)*, pages 51-60, Herstmonceux Castle, Sussex, UK.
- [Fasciano and Lapalme, 2000] Fasciano, M. and Lapalme, G. (2000). Intentions in the coordinated generation of graphics and text from tabular data. *Knowledge and Information Systems*.
- [Futrelle, 1999] Futrelle, R. (1999). Ambiguity in visual language theory and its role in diagram parsing. In *Proceedings of the IEEE Symposium on Visual Languages*, pages 172-175.
- [Futrelle et al., 1995] Futrelle, R. P., Zhang, X., and Sekiya, Y. (1995). Corpus linguistics for establishing the natural language content of digital library documents. In Adam, N., Bhargava, B., and Yesha, Y., editors, *Digital Libraries. Current Issues*, pages 165-180. Springer-Verlag.
- [Gardner, 1998] Gardner, J. A. (1998). The quest for access to science by people with print impairments. *Computer Mediated Communication*, 5(1).
- [Green et al., 1998] Green, N., Carenini, G., Kerpedjiev, S., Roth, S., and Moore, J. (1998). A media-independent content language for integrated text and graphics generation. In *Proc. of the Workshop on Content Visualization and Intermedia Representations (CVIR'98)* of COLING '98 and ACL'98, Montreal, Canada.
- [Kennel, 1996] Kennel, A. R. (1996). Audiograf: A diagram-reader for the blind. In *Proc. of the Second Annual ACM Conference on Assistive Technologies*, pages 51-56, Vancouver, British Columbia, Canada. ASSETS '96.
- [Kerpedjiev et al., 1998] Kerpedjiev, S., Carenini, G., Green, N., Moore, J., and Roth, S. (1998). Saying it in graphics: from intentions to visualizations. In *Proc. of Information Visualization*, Research Triangle Park, NC. IEEE.
- [Kerpedjiev and Roth, 2000] Kerpedjiev, S. and Roth, S. (2000). Mapping communicative goals into conceptual tasks to generate graphics in discourse. In *Proc. of Intelligent User Interfaces 2000*, New Orleans, Louisiana.
- [Kurze et al., 1995] Kurze, M., Perie, H., Morley, S., Deconinck, F., and Strothotte, T. (1995). New approaches for accessing different classes of graphics by blind people. In *Proc. of the Second TIDE Congress*, pp. 268-271, Paris.
- [Lapalme, 1999] Lapalme, G. (1999). Comparing PostGraphe with the rags architecture. In *Proc. RAGS Workshop -- Reference Architectures for Generation Systems*, Edinburgh, Scotland, UK.
- [McKeown, 1985] McKeown, K. (1985). *Text generation: Using discourse strategies and focus constraints to generate natural language text*. Cambridge University Press, Cambridge UK.
- [McKeown et al., 1992] McKeown, K., Feiner, S., Robin, J., Seligmann, D., and Tanenblatt, M. (1992). Generating cross-references for multimedia explanation. In *Proc. of the 1992 National Conference on Artificial Intelligence*, pp. 9-16, San Jose, CA. AAAI.
- [Meijer, 1992] Meijer, P. B. L. (1992). An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, 39(2): 291-300.
- [Mittal et al., 1998] Mittal, V. O., Moore, J. D., Carenini, G., and Roth, S. (1998). Describing complex charts in natural language; a caption generation system. *Computational Linguistics*, 34(3): 431--468.
- [Raman, 1997] Raman, T. (1997). *Auditory User Interfaces*. Kluwer Academic Publishers, Boston.
- [St. Amant and Riedl, 2000] St. Amant, R. and Riedl, M. (2000). A perception/action substrate for cognitive modeling. In *HCI. International Journal of Human-Computer Studies*. To appear.
- [Wahlster et al., 1993] Wahlster, W., Andre, E., Finkler, W., Profitlich, H. J., and Rist, T. (1993). Plan-based integration of natural language and graphics generation. *Artificial Intelligence*, 63:387-427.
- [Walsh and Gardner, 2001] Walsh, P. and Gardner, J. A. (2001). Tiger, a new age of tactile text and graphics. In *Proceedings of the 2001 CSUN Conference on Technology and Persons with Disabilities*, Los Angeles, CA.
- [Walsh et al., 2001] Walsh, P., Lundquist, R., Gardner, J. A., and Sahyun, S. (2001). The audio-accessible graphing calculator. In *Proc. of the 2001 CSUN Conference on Technology and Persons with Disabilities*, Los Angeles.
- [Zhang and Krishnamoorthy, 1994] Zhang, S. H. and Krishnamoorthy, M. (1994). Audio formatting of a graph. In *Proc. of the 1st Annual ACM Conference on Assistive Technologies*, pages 148-156, Los Angeles, CA. ASSETS.