Understanding Information Graphics: A Discourse-Level Problem *

∗Sandra Carberry, ∗Stephanie Elzer, ∗∗Nancy Green, ∗Kathleen McCoy, and ∗Daniel Chester
∗Dept. of Computer Science, University of Delaware, Newark, DE 19716
(carberry, elzer, mccoy, chester@cis.udel.edu)
∗∗Dept. of Math. Sciences, Univ. of North Carolina at Greensboro, Greensboro, NC 27402
(nlgreen@uncg.edu)

Abstract

Keywords: graphics, understanding, discourse, plan-based models

Information graphics that appear in newspapers and magazines generally have a message that the viewer is intended to recognize. This paper argues that understanding such information graphics is a discourse-level problem. In particular, it requires assimilating information from multiple knowledge sources to recognize the intended message of the graphic, just as recognizing intention in text does. Moreover, when an article is composed of text and graphics, the intended message of the information graphic (its discourse intention) must be integrated into the discourse structure of the surrounding text and contributes to the overall discourse intention of the article. This paper describes how we extend plan-based techniques that have been used for understanding traditional discourse to the understanding of information graphics. This work is part of a project to develop an interactive natural language system that provides sight-impaired users with access to information graphics.

1 Introduction

Information graphics (non-pictorial graphics such as bar charts and line graphs) are a variant of language with many similarities to other forms of communication. Information graphics are prevalent in information resources since they enable complex information to be assimilated perceptually with ease. Unfortunately, knowledge sources such as information graphics are not accessible to some users. For example, individuals with impaired eyesight have limited access to information graphics, thus preventing them from fully utilizing information resources.

Some information graphics are only intended to display data values; (Yu et al., 2002) developed a pattern recognition algorithm for summarizing interesting features of automatically generated graphics of time-series data from a gas turbine engine. However, the overwhelming majority of the graphics that we have examined (taken from newspaper, magazine, and web articles) appear to have some underlying goal, such as getting the viewer to believe that interest rates have fallen substantially and that this would therefore be a good time to refinance a mortgage. We have found that understanding information graphics is a discourse-level problem. In particular, it requires assimilating information from multiple knowledge sources to recognize the intended message of the graphic, just as recognizing intention in text does. Moreover, the communicative intention of the information graphic must be integrated into the discourse intentions of the surrounding text.

We are developing an interactive natural language system that infers the intended message underlying an information graphic, augments it with related interesting features of the graphic, provides an initial summary of the graphic, and then responds to followup questions from the user. This paper presents the system architecture, shows why interpreting information graphics is a discourse-level problem, and outlines how we extend techniques that have been used for understanding traditional discourse to the understanding of information graphics.

[0]The work of the third author was supported by the National Science Foundation under Grant No. 0132821.

2 A Natural Language Modality

Information is the key to knowledge and effective decision-making. But information is useful only if it is accessible in a form that can be easily assimilated. For sighted users, information graphics capture complex information and enable it to be assimilated perceptually with ease. For individuals who have serious sight-impairments, documents that contain information graphics pose challenging problems. Although devices have been developed for conveying information graphics in alternative mediums such as musical tones or tactile images, these approaches have serious limitations. For example, systems that attempt to convey graphics via a soundscape(Meijer, 1992) do not facilitate easy comparison of two line graphs linked in a single graphical display. Moreover, these approaches require the user to construct a "mental map" of the graphic, which is difficult for congenitally blind users who do not have the personal knowledge to assist them in the interpretation of the image(Kennel, 1996). The underlying hypothesis of our work is that alternative access to what the graphic looks like is not enough — the user should be provided with the message and knowledge that one would gain from viewing the graphic in order to enable effective and efficient use of this information resource. To accomplish this objective, we are developing an interactive natural language system for communicating the content of an information graphic. Our methodology offers promise as a means of providing access to information graphics without expensive equipment, with few limitations on the complexity of the graphic that can be handled, and with relatively little cognitive load on the user.

3 Architecture and Overview

Our current work is concerned with bar charts, line graphs, and pie charts, although eventually we will handle other kinds of graphics. Figure 1 shows the architecture of our system for conveying information graphics. The visual extraction component (VEC) analyzes the graphic and provides an XML representation of the graphic to the intention recognition component (IRC). The IRC is responsible for recognizing the intended message of the information graphic and sending it to the content planning component (CPC), which will augment the intended message of the graphic with related interesting features. The message organization component (MOC) then organizes the most salient propositions into a coherent summary, which will be rendered in natural language and conveyed to the user via speech synthesis. The followup question component (FQC) will allow the user to interactively seek additional information about the graphic.

Our work thus far (Section 4) has focused on understanding an information graphic so that its intended message can be conveyed to the user. Section 4.1 discusses the extension of speech act theory to the generation and understanding of information graphics. Section 4.2 argues that understanding information graphics is a discourse-level problem in which the system must recognize the intended message of the graphic and integrate it into the intentions of any surrounding text; it further argues that understanding information graphics requires similar kinds of knowledge and processing as does the understanding of traditional textual discourse. Section 4.3 provides a brief overview of the visual extraction component that analyzes the graphical image and constructs an XML representation of the graphic for use by the graphic understanding system. Section 4.4 then describes how we have extended techniques used for understanding traditional discourse and dialogue to the understanding of information graphics. Section 5 gives a brief overview of future work on the rest of the system. The Appendix contains information graphics that are part of the corpus on which our work is based.

4 Understanding Information Graphics

4.1 Intention in Information Graphics

Information graphics are a variant of language. As noted by Clark(Clark, 1996), language is more than just words. It is any "signal" (or lack of signal when one is expected), where a signal is a deliberate action that is intended to convey a message. According to speech act theory, a speaker or writer executes a speech act whose intended meaning he expects the listener or reader to be able to deduce(Searle, 1970; Grice, 1969; Clark, 1996). In their work on multimedia generation,

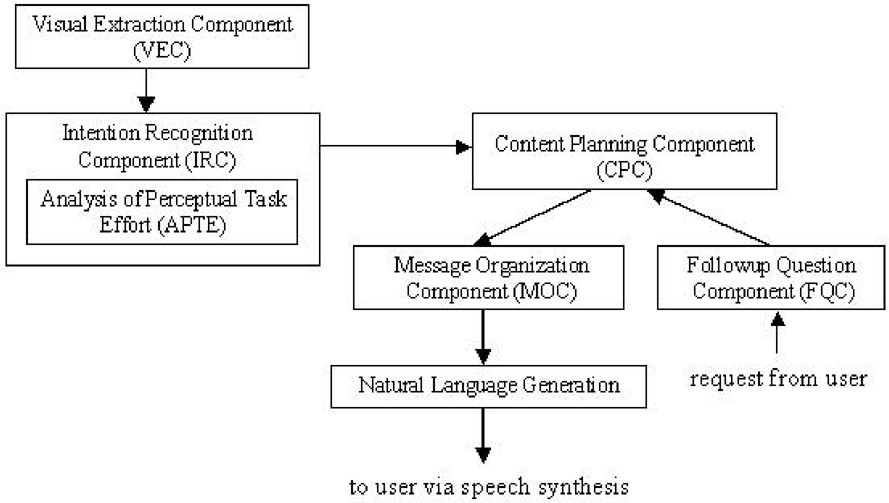


Figure 1: System Architecture

the AutoBrief group proposed that speech act theory could be extended to cover the generation of graphical representations(Kerpedjiev and Roth, 2000). They developed a multimedia presentation system that generated text and information graphics. It included 1) an algorithm that could map communicative goals to a set of perceptual and cognitive tasks that must be enabled for a viewer to recognize the goals and 2) an automatic graph designer that used constraint satisfaction to construct an information graphic that best facilitated those tasks, subject to competing constraints among the tasks.

The overwhelming majority of information graphics accompanying newspaper and magazine articles appear to carry a message that the designer intends to convey to the viewer by virtue of the graphic's design and the data presented in the graphic. Consider the graphic in Figure 9. It conveys the message that the salary of women in science, mathematics, and engineering fields is consistently less than that of men in the same fields. Other messages could have been conveyed by a different graphic design. For example, by grouping the bars for men together, grouping the bars for women together, and ordering the bars for each group by height, the graphic would have conveyed the message that both men and women earn the least in the social sciences and the most in engineering. Or if the bars for Computer/Mathematical Sciences were highlighted in Figure 9 by coloring them significantly differently from the other bars in the graphic, the graphic would have invoked a comparison of the discrepancies between male and female salaries in Computer/Mathematical Sciences and the salary discrepancies between men and women in other fields. Although a graphic's caption can be helpful in identifying its intended message (as in Figure 8), Corio performed a large corpus study(Corio and Lapalme, 1999) in which he found that captions are often missing or fail to provide any indication of what the information graphic conveys (as in Figures 6 and 10). Thus we cannot rely entirely on the presence of useful captions to identify the intended message of an information graphic.

Language research has posited that the listener or reader who is interpreting a speech act identifies its intended meaning by reasoning about the observed signals and the mutual beliefs of author and interpreter(Grice, 1969; Clark, 1996). Applying this to graphical displays, it is reasonable to presume that the author of a graphic similarly expects the viewer to use perceptual skills along with other knowledge sources to deduce from the graphic the message that he intended to convey. Thus we are applying speech act theory in the reverse direction of the AutoBrief project, namely to the recognition of the intended message underlying an information graphic.

4.2 A Discourse Level Problem

This section argues that interpreting information graphics is a discourse-level problem — not only is it necessary to recognize the intention of the graphic as noted in Section 4.1, but understanding an information graphic requires similar kinds of knowledge and processing as does understanding traditional discourse.

Grosz and Sidner contended that discourse has a structure comprised of discourse segments. Each discourse segment has a discourse segment purpose that contributes to the discourse purpose or intention underlying the overall discourse(Grosz and Sidner, 1986). When an article is comprised of text and graphics, the graphic generally expands on the text and contributes to the discourse purpose of the article. Consider the graphic and partial surrounding text reproduced in Figure 6. Nowhere in the text is it stated that the income of black women has risen dramatically over the last decade and has reached the level of white women. Yet this message is clearly conveyed by the graphic and contributes to the overall communicative intention of this portion of the article — namely, that there has been a "monumental shifting of the sands" with regard to the achievements of black women. Not only does the intended message of the graphic (its discourse segment purpose) contribute to this overall intention, but in fact the discourse intention of the graphic helps to recognize the overall intention.

Even when the graphic stands in isolation as in Figures 7 and 8, understanding the graphic is a discourse-level problem. Grosz and Sidner(Grosz and Sidner, 1986) claim that a robust model of discourse understanding must use multiple knowledge sources in order to recognize the complex relationships that utterances have to one another. Information graphics have similar complex relationships among their component elements. Not only might the graphic include multiple elements that must be related to one another (such as multiple lines in a line graph, or individual bars in a bar chart), but information graphics often include highlighting of certain elements to make them particularly salient (as in Figure 10) or include captions that might contribute to recognizing the graphic's intention. The graphic in Figure 8 includes such a helpful caption, although many graphics, such as the ones in Figures 7 and 9, do not.

Furthermore, identifying the intended message of a composite graphic (one comprised of multiple individual graphics) requires relating the individual graphics to one another to identify the intended message of the composite. Figure 11 illustrates a composite information graphic. The discourse purpose of the composite graphic is that audits of affluent taxpayers are declining with respect to audits of all taxpayers. This message can only be deduced by relating the two individual graphics and their underlying messages.

Moreover, understanding information graphics requires the use of multiple knowledge sources. In earlier work on recognizing expressions of doubt, we developed an algorithm that combined linguistic, contextual, and world knowledge and applied it to the recognition of complex discourse acts(Carberry and Lambert, 1999). In the case of information graphics, the corollary to linguistic knowledge is perceptual knowledge, by which one recognizes the individual elements of the graphic (for example, the bars in a bar chart), the relation of the individual elements in the graphic to one another, the type of graphic (line graph, bar chart, pie chart, etc.), and what the different graphic types can be used to convey. For example, both a scatter plot and a pie chart can be used to portray how an entity (such as government income) is divided up among several categories (such as social welfare, military spending, etc.); however, a graphic designer will choose a pie chart if the intent is to convey the relative distributions as opposed to their absolute amounts. Furthermore, a particular type of graphic (such as a line graph) might be appropriate for conveying several different intentions (maximum data point, data trend, data variation, etc.).

Contextual and world knowledge are also essential for understanding information graphics. Contextual knowledge includes the caption associated with the graphic, any highlighting of graphic elements that affects the focus of attention in the graphic, and the discourse structure and focus of attention in any surrounding text. World knowledge consists of mutual beliefs between designer and viewer about entities of interest to the intended viewing audience. For example, if an information graphic appears in a document targeted at residents of New York City, then both the designer and the viewer will mutually believe that entities such as New York City, its football and baseball teams, etc. will be particularly salient to the viewer. Our methodology for understanding information graphics takes these knowledge sources into account.

4.3 The Visual Extraction Component

The visual extraction component (VEC) captures much of the perceptual knowledge discussed in Section 4.2. It is responsible for recognizing the individual components comprising the graphic, identifying the relationship of the different components to one another and to the graphic as a whole, and classifying the graphic as to type. Extracted components include not only the bars, lines, or wedges of a graphic but also the titles of the axes, the legend, and the graphic's title or caption. The present implementation deals only with gray scale images (in pgm format) of bar charts, pie charts, and line graphs, though eventually it will be extended to handle color and other kinds of information graphics. Words and numbers that appear in the chart are associated with particular bars, wedges and lines by their proximity to the chart component in question. The output of the visual extraction component is an XML file that describes the chart and all of its components.

4.4 Applying Discourse Understanding Strategies

Many researchers have cast the understanding of discourse and dialogue as a plan recognition problem — that is, the writer or speaker (or characters in the case of a story) has an underlying goal and a plan for accomplishing that goal, and understanding requires that the reader or listener infer the plan and in turn the goal that the plan is intended to achieve. (Perrault and Allen, 1980; Wilensky, 1983; Litman and Allen, 1987; Carberry, 1990; Charniak and Goldman, 1993; Ardissonno and Sestero, 1996) are just a few examples of such systems.

Since understanding information graphics is a discourse-level problem, we are extending plan inference techniques to recognizing the intended message of an information graphic(Elzer et al., 2003) and to identifying its contribution to an extended discourse that includes both text and graphics. Planning and plan inference systems require knowledge about goals and how they can be achieved. Typically, this is provided by a library of operators. Each operator encodes a goal in its header; the body of the operator encodes the subgoals that must be accomplished in order to achieve the operator's goal. A planning system starts with a high-level goal, and uses operators to decompose the goal into a set of simpler subgoals, which eventually decompose into primitive subgoals that can be accomplished by primitive actions in the domain. On the other hand, a plan inference system starts with the primitive goals associated with observed actions, and uses the operators to chain backwards to higher-level goals which the lower-level subgoals contribute to achieving. In the case of traditional discourse and dialogue, the subgoals in the plan operators are either communicative or domain goals, and the observed actions that start the plan inference process are the speech acts represented by the utterances in a story or a dialogue.

To extend plan inference to information graphics, the plan operators must include goals that can be accomplished by viewing an information graphic, as opposed to being the recipient of an utterance. As discussed in Section 4.1, the AutoBrief project(Kerpedjiev and Roth, 2000) developed an algorithm to map communicative goals to a sequence of perceptual and cognitive tasks that the graphic should support. *Perceptual tasks* are tasks that can be performed by simply viewing the graphic, such as finding the top of a bar in a bar chart; *cognitive tasks* are tasks that are performed via mental computations, such as computing the difference between two numbers. We draw on the AutoBrief notion of perceptual and cognitive tasks enabled by an information graphic. Our plan operators not only encode knowledge about how to achieve domain and communicative goals (the latter of which may require that the viewer perform perceptual and cognitive tasks) but they also encode knowledge about how information-access tasks, such as finding the value of an entity in a graphic, can be decomposed into simpler subgoals. Figures 2 and 3 present two plan operators for achieving the goal of finding the value <v> of an attribute <att> for a graphical element <e> (for example, the value associated with the top of a bar in a bar chart). The body of the operator in

Goal: Find-value(<viewer>, <g>, <e>, <ds>, <att>, <v>)
Gloss: Given graphical element <e> in graphic <g>, <viewer> can find the value <v> in dataset <ds> of attribute <att> for <e>
Data-req: Dependent-variable(<att>, <ds>)
Body: 1. Perceive-dependent-value(<viewer>, <g>, <att>, <e>, <v>)

Figure 2: Operator for achieving a goal perceptually

Goal: Find-value(<viewer>, <g>, <e>, <ds>, <att>, <v>)
Gloss: Given graphical element <e> in graphic <g>, <viewer> can find the value <v> in dataset <ds> of attribute <att> for <e>
Data-req: Natural-quantitative-ordering(<att>)
Display-const: Ordered-values-on-axis(<g>, <axis>, <att>)
Body: 1. Perceive-info-to-interpolate(<viewer>,<g>,<axis>,<e>,$<l_1>$,$<l_2>$,<f>)
2. Interpolate(<viewer>, $<l_1>$, $<l_2>$, <f>, <v>)

Figure 3: Operator that employs both perceptual and cognitive subgoals

Figure 2 specifies that the goal can be achieved by a primitive perceptual task in which the viewer just perceives the value; this could be done, for example, if the element in the graphic is annotated with its value, as are the bars in the bar chart in Figure 8 of the Appendix. On the other hand, the body of the operator in Figure 3 captures a different way of finding the value, one that presumably requires more effort. It specifies the perceptual task of finding the values $<l_1>$ and $<l_2>$ surrounding the desired value on the axis along with the fraction <f> of the distance that the desired value lies between $<l_1>$ and $<l_2>$, followed by the cognitive task of interpolating between the retrieved values $<l_1>$ and $<l_2>$.

Our operators contain *data requirements* (labelled Data-req) which the data must satisfy in order for the operator to be applicable in a graphic planning paradigm; they may also contain *display constraints* (labelled Display-const) which constrain how the information graphic is constructed if this operator is part of a final plan. In the case of plan recognition, these constraints are used in reverse. The display constraints are used to eliminate operators from consideration, since if a graphic does not satisfy the operator's display constraints, then the operator could not be part of a plan that led to the graphic. If a graphic meets the display constraints of an operator, then the data requirements are used to limit how the operator's parameters might be instantiated.

4.4.1 Beginning the Plan Inference Process

Traditional plan inference systems used for language understanding start with the primitive goal achieved by the speech act in the dialogue or discourse. In the case of information graphics, the role of the speech act is played by the primitive perceptual tasks that the viewer performs on the graphic. To limit the set of perceptual tasks that are considered, we make two observations:

- The graphic designer has many alternative ways of designing a graphic, and the design choices facilitate some perceptual tasks more than others. Following the AutoBrief work(Kerpedjiev and Roth, 2000) on generating graphics that fulfill communicative goals, we hypothesize that the designer chooses a design that best facilitates the tasks that are most important to conveying his intended message, subject to the constraints imposed by competing tasks.

- Entities may become particularly salient by virtue of highlighting in the graphic (for example, coloring certain elements different from the others, annotating an element with an asterisk, or exploding one piece of a pie chart[1]), by their mention in the caption or surrounding text, or via world knowledge

[1] (Mittal, 1997) discusses a variety of such design techniques in the context of distorting the message inferred from a graphic.

capturing mutual beliefs about entities of interest to the intended audience. We hypothesize that the designer relies on the viewer recognizing particularly salient entities, in order to make certain perceptual tasks more salient to the viewer.

As noted in Section 4.1, one cannot rely on a graphic's caption to provide the intended message of the graphic. Consequently, the plan inference process starts with both the set of tasks that are best enabled by the information graphic and the set of tasks (if any) that are particularly salient. These will be referred to as *candidate tasks*. The next two subsections describe how candidate tasks are identified.

Identifying the Best Enabled Tasks The APTE (Analysis of Perceptual Task Effort) submodule, shown in Figure 1 as part of the Intention Recognition Component, captures perceptual knowledge about performing primitive perceptual tasks[2], and it encapsulates the results of cognitive psychology research to estimate the relative effort required for different tasks. The output of APTE is the set of perceptual tasks that are best enabled by the graphic. These become candidate tasks.

Each APTE rule captures a primitive perceptual task that can be performed on a particular type of information graphic, the conditions (graphic design choices) that affect the difficulty of performing that task, and the estimated effort expended by a viewer if those conditions are satisfied in the graphic. The condition-computation pairs are ordered so that the ones producing the lowest effort estimates appear first in a rule.

To derive the effort estimates in the rules, we have followed the GOMS approach(Card et al., 1983) by breaking down the tasks that are regarded as primitive in our plan operators into even more basic component tasks, and then summing the effort estimates for these very basic tasks. Lohse's work(Lohse, 1993) is an example of the GOMS architecture applied to predicting performance on graph comprehension tasks, and many of our effort estimates are based on Lohse's research. For example, Figure 4 displays the APTE rule for the task of finding the value associated with the top of a bar in a bar chart. If the bar is annotated with its value, then condition-computation pair B1-1 estimates its effort as 150 units for discriminating the label (based on work by Lohse(Lohse, 1993)) and 300 units for recognizing a 6-letter word (John and Newell, 1990). If the bar is not annotated with its value but is aligned with a tick mark on the axis, then condition-computation pair B1-2 estimates the perceptual effort in terms of the distance to the dependent axis (in order to capture the degrees of visual arc scanned(Kosslyn, 1989)) plus the effort of discriminating and recognizing the label. Figure 5 displays the APTE rule associated with the first subgoal in Figure 3. It estimates the effort for the primitive task *Perceive-info-to-interpolate* as the effort of the scan to the dependent axis (based on (Kosslyn, 1989)), the effort of discriminating the intersection location on the axis (150 units based on (Lohse, 1993)), plus the effort of the saccade to each label (230 units each (Russo, 1978)) along with the effort involved in discriminating and recognizing the labels. Similarly, there is a cognitive rule (not discussed here) for estimating the effort associated with the cognitive task *Interpolate* (the second subgoal in the operator in Figure 3). (Elzer et al., 2003a) presents a more extensive discussion of the cognitive principles underlying the APTE rules.

Given the XML representation of an information graphic, each APTE rule that is applicable to the graphic produces an effort estimate for the task captured by the rule. When a task might be instantiated in multiple ways and still satisfy the conditions of a condition-computation pair (for example, the task of finding the value of the top of a bar could be instantiated for each bar in a bar chart), only the instantiation that produces the lowest effort estimate becomes a candidate task. (If the bars are not annotated with values, then the instantiation that will produce the lowest effort estimate for the task of finding the value of the top of a bar in a bar chart would be the bar with the shortest scan to the dependent axis.) This is consistent with the idea that the graphic designer will make the important tasks easy to perform. The set of perceptual tasks that require the least effort become candidate tasks.

[2] Primitive perceptual tasks are those that we do not decompose into a set of simpler subtasks; this is not to be confused with the notion of a psychological primitive.

Rule-1:Estimate effort for task Perceive-dependent-value(<viewer>, <g>, <att>, <e>, <v>)
Graphic-type: bar-chart
Gloss: Compute effort for finding the exact value <v> for attribute <att> represented by top <e> of a bar <b> in graph <g>
B1-1: IF the top of bar <b> is annotated with a value,
THEN effort=150 + 300
B1-2: IF the top <e> of bar <b> aligns with a labelled tick mark on the dependent axis,
THEN effort=scan + 150 + 300

Figure 4: A rule for estimating effort for the primitive perceptual task *Perceive-value*

Rule-2:Estimate effort for task
Perceive-info-to-interpolate(<viewer>,<g>,<axis>,<e>,<l_1>,<l_2>,<f>)
Graphic-type: bar-chart
Gloss: Compute effort for finding the information needed for interpolation, including the labels <l_1> and <l_2> on either side of entity <e> on axis <axis> in graph <g>, and the fraction <f> that is the distance between <l_1> and entity <e> on <axis> relative to the distance between <l_1> and <l_2>
B2-1: IF <axis> is labelled with values THEN effort=scan + 150 + ((230 + 150 + 300) x 2)

Figure 5: A rule for estimating effort for the primitive perceptual task *Perceive-info-to-interpolate*

Identifying Particularly Salient Tasks *Salient tasks* are those that the viewer might perform because they relate to entities that are in the viewer's current focus of attention, as determined by contextual knowledge provided by the caption, highlighting, and the surrounding text and by world knowledge in the form of mutual beliefs about items of particular interest to the viewing audience.

Ideally, a caption will provide clues about the message that an information graphic is intended to convey, and thus noun phrases in captions represent salient entities.[3] The graphic designer can also call into focus certain aspects of the graphic by using attention-getting devices such as coloring it differently from the rest of the graphic, annotating it with an arrow, etc. Our working hypothesis is that if the graphic designer goes to the effort of employing such attention-getting devices, then the highlighted items almost certainly contribute to the intended message. Thus the attributes of these highlighted items (for example, the attributes of a highlighted bar in a bar chart), which are captured in the XML representation of the graphic, are also regarded as salient entities. Salient entities also include those that world knowledge suggests are mutually believed to be of interest to the viewing audience. We envision in the future using the notion of lexical chains(Silber and McCoy, 2000) to identify entities that the accompanying text makes particularly salient. Perceptual tasks that are instantiated with a salient entity and that can be performed on the graphic are designated *salient tasks*.

[3] Verb phrases in captions also provide evidence, but they suggest particular operators of interest rather than instantiated perceptual tasks, and thus we associate verbs with operators in the plan library.

4.4.2 The Search Process

Candidate tasks consist of the set of perceptual tasks that require the least effort and the set of salient tasks. Once the set of candidate tasks has been identified, plan inference begins. Initial candidate plans are constructed from each operator in which a candidate task appears as a subgoal; the root of the candidate plan is the goal of the operator, and its children are the subgoals in the body of the operator. Chaining from the root goal to other operators whose body contains the root goal as a subgoal produces larger candidate plans with higher-level goals as the new root goal.

Plan inference systems have used a variety of heuristics to evaluate candidate plans and to select the candidate plan to expand further. These heuristics help to guide the search through the space of candidate plans in order to hypothesize the plan that best represents the user's intentions. These heuristics have included increasing the rating of partial plans as their arguments become instantiated(Perrault and Allen, 1980), preferring coherent discourse moves(Litman and Allen, 1987; Carberry, 1990), and biasing the plan inference process based on knowledge about the user group(Gertner and Webber, 1996). We have identified several kinds of evidence for guiding plan inference from information graphics, including the estimated effort required by a candidate plan, the basis for instantiating parameters in the plan, adherence to the proximity compatibility principle from cognitive science research, and the relation between a candidate plan and the established discourse context.

Since our working hypothesis is that the graphic designer tried to enable those tasks necessary to recognize his intended message, candidate plans that require substantially more effort than other candidate plans are less likely to represent the intentions of the designer. The effort associated with a candidate plan is measured as the sum of the effort of the tasks comprising it.

There are many ways that a parameter in a task or subgoal might become instantiated, and the basis for the instantiation provides evidence about the likelihood that a hypothesized candidate plan represents the graphic designer's intentions. If an instantiation is suggested by highlighting or a caption or entities that are particularly salient to the targeted audience, that partial plan should be evaluated more favorably since the designer of the graphic has provided reasons for the viewer to use these instantiations in recognizing his intentions. Similarly, if the instantiation is one of several possible alternatives with no reason for preferring one over the other, then the partial plan should be evaluated less favorably since the designer did not give the viewer any reason to prefer one over the other. This relates to Allen's forking heuristic(Perrault and Allen, 1980). The proximity compatibility principle(Wickens and Carswell, 1995) also suggests that candidate plans which use similarly encoded elements (for example, all red bars) in an integrated fashion should be evaluated more favorably than those that do not.

If there is a context established by the text preceding or surrounding the graphic, then candidate plans whose root goal contributes to the existing discourse context should be preferred. If the surrounding text has a reference to the graphic, then focusing heuristics(Carberry, 1990) will prefer candidate plans that relate most closely to the current focus of attention at that point in the surrounding text. However, the surrounding text often does not refer to accompanying graphics, as is the case in the Newsweek article whose excerpt is shown in Figure 6. Future work will investigate how we should handle instances such as this.

5 Response Generation and Followup

The intended message of the graphic must be augmented with additional propositions that convey interesting features that a viewer would glean from the graphic. For example, the intended message of the graphic in Figure 6 appears to be that the income of black women has risen dramatically over the last decade and reached the level of white women. But other interesting features of the graphic might include the trends over the past several decades, periods where they were closest, etc. In future work, we anticipate developing a methodology for identifying propositions that expand on the message of the graphic designer and for including the most salient of these in the summarization of the graphic. We also envision responding to followup requests for further information about the graphic by selecting the highest ranking propositions that were not included in the initial message, organizing them into a coherent response, and conveying it to the user.

6 Summary

This paper has argued that understanding information graphics is a discourse-level problem. Not only must the system recognize the intended message of the information graphic, but the recognition process requires similar kinds of knowledge sources and similar kinds of processing as does the understanding of traditional discourse and dialogue. Moreover, when an article is composed of text and graphics, the intended message of the information graphic must be integrated into the discourse structure of the surrounding text, and it contributes to the overall discourse intention of the article.

References

L. Ardisonno and D. Sestero. 1996. Using dynamic user models in the recognition of the plans of the user. *User Modeling and User-Adapted Interaction*, 5(2):157–190.

S. Carberry and L. Lambert. 1999. A process model for recognizing communicative acts and modeling negotiation subdialogues. *Computational Linguistics*, 25(1):1–53.

S. Carberry. 1990. *Plan Recognition in Natural Language Dialogue*. ACL-MIT Press Series on Natural Language Processing. MIT Press, Cambridge, MA.

S. Card, T. Moran, and A. Newell. 1983. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ.

E. Charniak and R. Goldman. 1993. A bayesian model of plan recognition. *Artificial Intelligence*, 64:53–79.

H. Clark. 1996. *Using Language*. Cambridge University Press.

M. Corio and G. Lapalme. 1999. Generation of texts for information graphics. In *Proceedings of the 7th European Workshop on Natural Language Generation EWNLG'99*, pages 49–58.

S. Elzer, N. Green, and S. Carberry. 2003a. Exploiting cognitive psychology research for recognizing intention in information graphics. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. To appear.

S. Elzer, N. Green, S. Carberry, and K. McCoy. 2003. Extending plan inference techniques to recognize intentions in information graphics. In *Proceedings of the Ninth International Conference on User Modeling*. To appear.

A. Gertner and B. Webber. 1996. A Bias Towards Relevance: Recognizing Plans Where Goal Minimization Fails. In *Proc. of the Thirteenth National Conference on Artificial Intelligence*, pages 1133–1138.

H. P. Grice. 1969. Utterer's Meaning and Intentions. *Philosophical Review*, 68:147–177.

B. Grosz and C. Sidner. 1986. Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12(3):175–204.

B. John and A. Newell. 1990. Toward an engineering model of stimulus response compatibility. In R. Gilmore and T. Reeve, editors, *Stimulus-response compatibility: An integrated approach*, pages 107–115. North-Holland, New York.

A. Kennel. 1996. Audiograf: A diagram-reader for the blind. In *Second Annual ACM Conference on Assistive Technologies*, pages 51–56.

S. Kerpedjiev and S. Roth. 2000. Mapping communicative goals into conceptual tasks to generate graphics in discourse. In *Proc. of the International Conference on Intelligent User Interfaces*, pages 60–67.

S. Kosslyn. 1989. Understanding charts and graphs. *Applied Cognitive Psychology*, 3:185–226.

D. Litman and J. Allen. 1987. A Plan Recognition Model for Subdialogues in Conversation. *Cognitive Science*, 11:163–200.

G. Lohse. 1993. A cognitive model for understanding graphical perception. *Human-Computer Interaction*, 8:353–388.

Peter B. Meijer. 1992. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, 39(2):291–300, February.

V. Mittal. 1997. Visual prompts and graphical design: A framework for exploring the design space of 2-D charts and graphs. In *Proc. of the Fourteenth National Conference on Artificial Intelligence*, pages 57–63.

R. Perrault and J. Allen. 1980. A Plan-Based Analysis of Indirect Speech Acts. *American Journal of Computational Linguistics*, 6(3-4):167–182.

J. Russo. 1978. Adaptation of cognitive processes to eye movement systems. In J. Senders, D. Fisher, and R. Monty, editors, *Eye movements and higher psychological functions*. Lawrence Erlbaum, Hillsdale, NJ.

J. Searle. 1970. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, London.

G. Silber and K. McCoy. 2000. Efficient text summarization using lexical chains. In *Proc. of the International Conference on Intelligent User Interfaces*, pages 252–255.

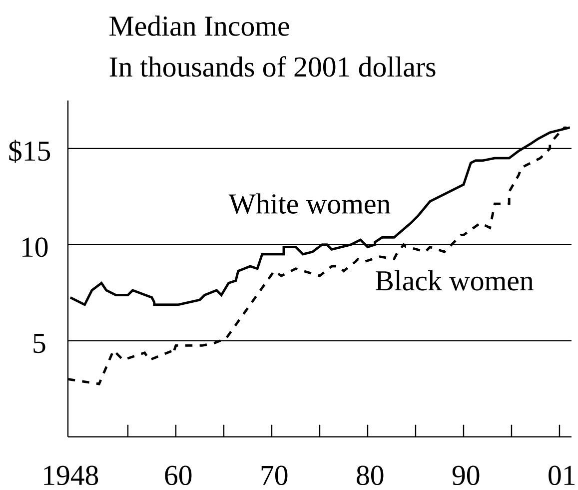
C. Wickens and M. Carswell. 1995. The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, 37(3):473–494.

R. Wilensky. 1983. *Planning and Understanding*. Addison-Wesley.

J. Yu, J. Hunter, E. Reiter, and S. Sripada. 2002. Recognising visual patterns to communicate gas turbine time-series data. In *ES2002*, pages 105–118.

Appendix of Graphics from our Corpus

Graphic from Newsweek Article



Relevant Text from Newsweek Article

This is not to say that black women have climbed the storied crystal stair. They remain "in the proving stage", observes Alabama executive Alice Gordon. Nearly 14 percent of working black women remain below the poverty level. And women don't yet out-earn black men. But the growing educational-achievement gap portends a monumental shifting of the sands. College-educated black women already earn more than the median for all black working men — or, for that matter, for all women. And as women in general move up the corporate pyramid, black women, increasingly, are part of the parade. In 1995 women held less than 9 percent of corporate-officer positions in Fortune 500 companies, according to Catalyst, a New York-based organization that promotes the interests of women in business. Last year they held close to 16 percent, a significant step up. Of those 2,140 women, 163 were black — a minuscule proportion, but one that is certain to grow.

Figure 6: Excerpt from Newsweek Magazine

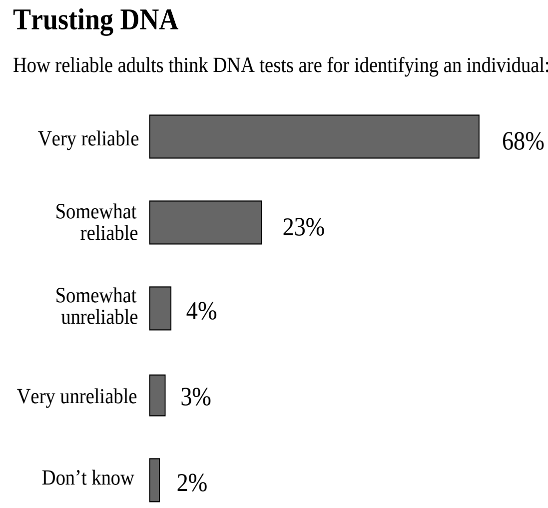


Figure 7: Standalone Graphic from USA Today

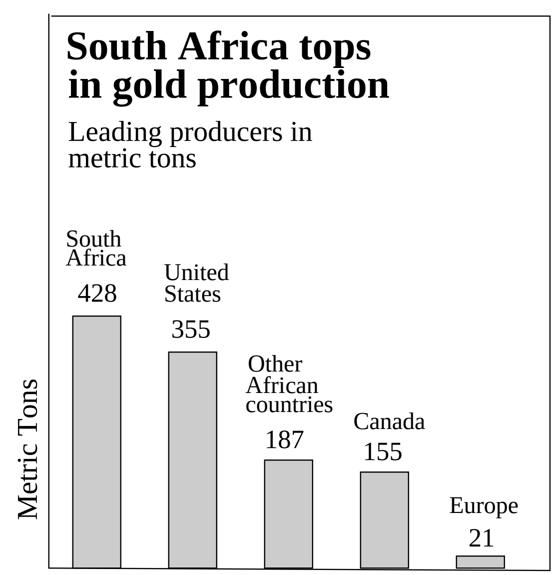


Figure 8: Standalone Graphic from USA Today

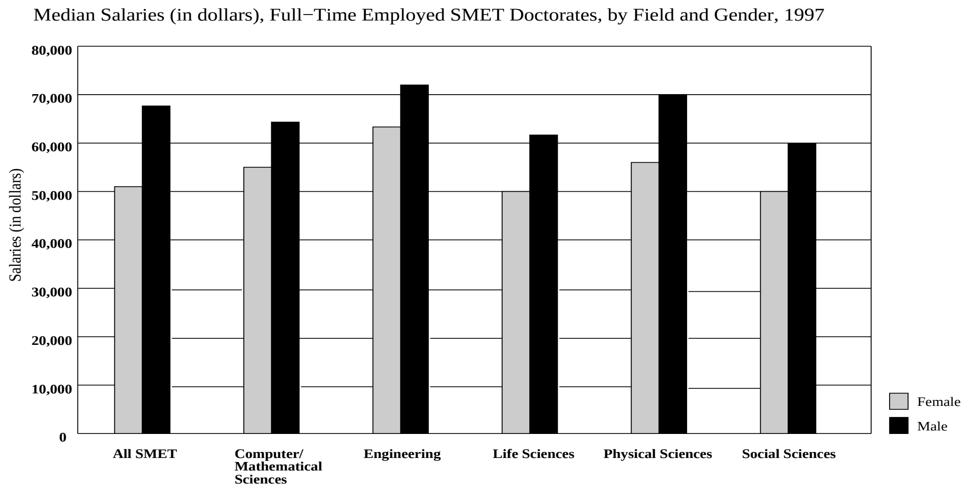


Figure 9: Graphic from Report of the NSF Committee on Equal Opportunities in Science & Engineering

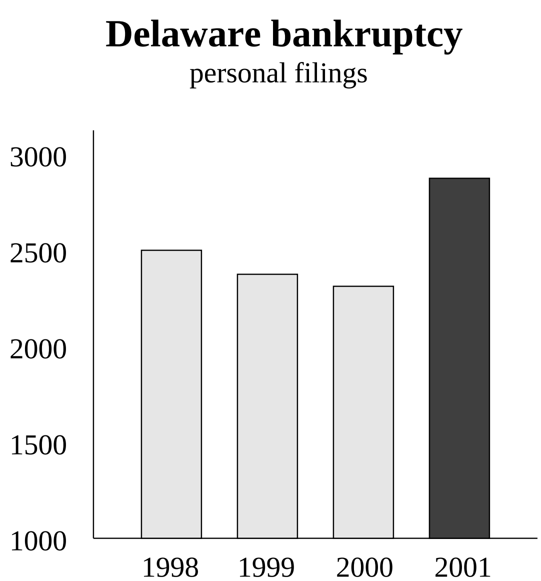


Figure 10: Graphic from Wilmington News Journal

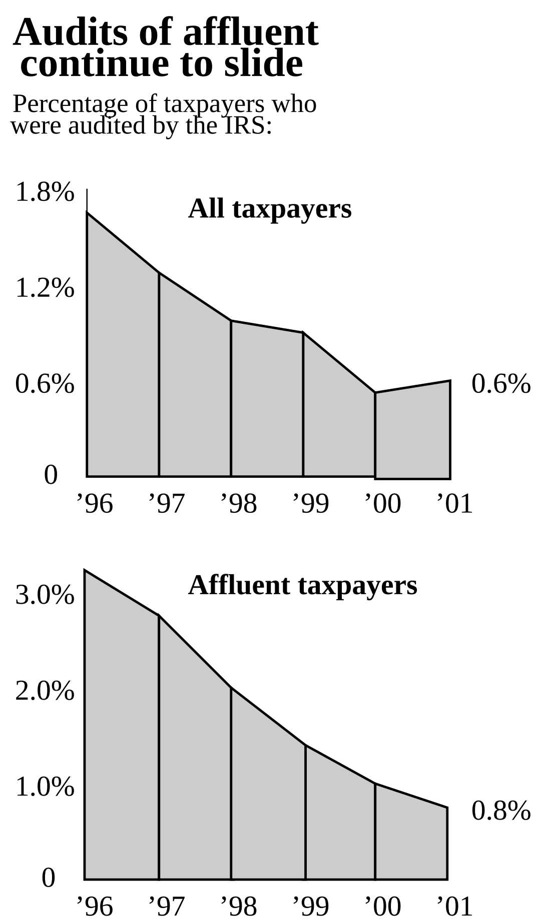


Figure 11: Graphic from USA Today