

Chapter 1

Bar Charts in Popular Media: Conveying Their Message to Visually Impaired Users via Speech*

Stephanie Elzer, Edward Schwartz, Sandra Carberry, Daniel Chester, Seniz Demir and Peng Wu

1.1 Introduction

Information graphics such as bar charts, line graphs, and pie charts play a vital role in many documents found on the Web. In contrast to graphics generated for the purposes of data visualization, which are intended to allow the viewer to visually explore the data, we posit that the majority of information graphics that appear in popular media are designed to convey a clear message to the viewer [5]. These visual constructs serve as a communication medium between the graphic designer and the viewer, since they enable the viewer to quickly and easily perform complex tasks such as comparing entities or identifying trends [26, 8, 2] in order to infer the message being conveyed by the graphic designer.

Unfortunately, while information graphics provide an invaluable tool for sighted individuals, they pose a serious obstacle to users with sight impairments. Many documents on the Web can be accessed by visually impaired users through screen reading software that reads the information on a computer screen using synthesized speech. If the developer of the web page has supplied alternative text (or “alt text”) for graphics in the HTML as specified in the guidelines of the web accessibility initiative (WAI) [27], most screen readers will read this text to the user. However, the vast majority of web pages are developed without broad accessibility in mind, and alt text is not supplied, thus making the content of the document’s graphics inaccessible to a visually impaired user. As one of our visually impaired users facetiously noted, “Of course, half the time they [web pages] don’t even label [the graphics] to

Stephanie Elzer and Edward Schwartz
Dept of Computer Science, Millersville University, 1 South George St, Millersville PA 17551,
e-mail: stephanie.elzer@millersville.edu

Sandra Carberry, Daniel Chester, Seniz Demir and Peng Wu
Dept of Computer & Info Sciences, University of Delaware, 103 Smith Hall, Newark, DE 19716

* This material is based upon work supported by the National Institute on Disability and Rehabilitation Research under Grant No. H133G080047

say this is a chart, or, it just says ‘Graphic207,’ it’s like, ‘Okay, wow. Thank you. That helped.’ ” Most of the visually impaired users whom we interviewed said that they simply skip any graphics that they encounter, and many expressed their frustration with the situation. For example, one user said that when they come to an information graphic on a web page, “I just get frustrated. And just go on because I know I won’t get any information from it.” Our research [11, 5] has shown that the message conveyed by an information graphic in a multimodal document is very often not captured by the graphic’s caption or the article’s text; thus it is essential that visually impaired viewers have alternative access to the content conveyed by information graphics. The challenge addressed by the SIGHT (Summarizing Information GraphHics Textually) system is to provide effective access to information graphics for sight impaired viewers who are disenfranchised by the growing use of visual information displays.

1.2 Approach and Related Work

Consider the graphic in Figure 1.1, where the graphic’s message is that of the train stations shown, Lancaster had the highest number of passengers in 2001. A visually impaired user who is presented with a listing of the 18 data points contained within the graphic is going to have to do a significant amount of work to determine that Lancaster had the highest number of passengers. Alternatively, if the visually impaired user is given a haptic rendering of the graphic, or another alternative representation of the graphic’s appearance, it will be very difficult for the user to distinguish the difference between the “heights” of the bars representing Lancaster and Harrisburg. In fact, without the visual cues indicating that Lancaster is salient (the bolding of the label and data value for Lancaster), the user may be completely unaware that the designer was attempting to draw attention to the entity with the maximum value in the graphic.

Our work is concerned with providing access to information graphics from popular media such as newspapers and magazines, such as the one shown in Figure 1.1. Our approach to this challenge is a novel one — rather than providing the visually impaired user with alternative access to what the graphic looks like or a listing of all of the data points contained in the graphic, we attempt to provide users with the message and knowledge that one would gain from viewing the graphic. Our longer term goal is an interactive system that will provide a richer textual summary, with the graphic’s inferred message as the core content, and that can respond to requests for further information about the graphic or requests for justification of the system’s inferred message when it appears to conflict with the article’s text.

A number of projects have attempted to make graphs accessible to visually impaired viewers by reproducing the image in an alternative medium, such as sound [22, 1], touch [15] or a combination of the two [25, 29, 16, 21]. One particularly

² This is based on a bar chart from a local newspaper (the Lancaster Intelligencer Journal on April 10, 2002).

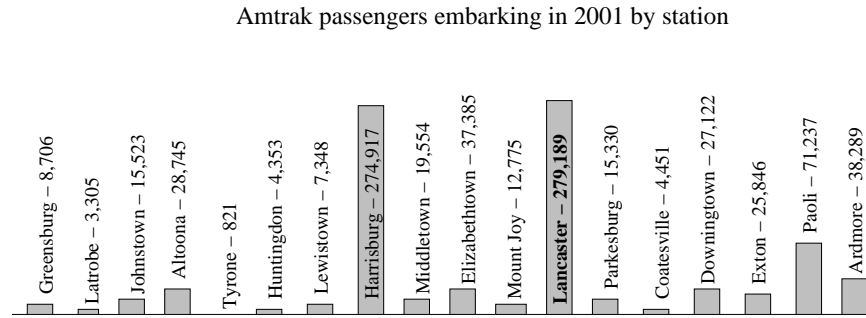


Fig. 1.1 Bar Chart Showing Yearly Ridership Data²

interesting project is that of Yu and Brewster [29]. In this project, they investigate the development and usefulness of web-based multimodal graphs which use haptic devices and auditory output to communicate the contents of the graphs to the users. The web pages containing the graphics must be properly formatted with the coordination of embedded haptic and audio features controlled by Cascading Style Sheets [29]. Although the evaluation of their system does demonstrate the usefulness of the approach when compared to traditional tactile diagrams, they note that the process of preparing the graphics is laborious [29].

Aside from the use of sound and touch, there is some research involving presenting graphics via text. Kurze [18] describes a system that generates text in a presentation tool used to convey the content of a graphic. In this system, a verbal description of the diagram's properties, such as the style of the diagram, the number of data sets, the labels of axes and the ranges of axes, is output through a text-to-speech device [18]. More recently, Ferres et al. [14], describe the inspectGraph system, which has the goal of providing accessibility for blind users to the graphs published in Statistics Canada's "The Daily." When the graph creator saves the graph as part of an Excel spreadsheet, a plug-in generates the inspectGraph files that are required for later processing. inspectGraph supports the communication of the contents of graphics in two ways: 1) it generates short, static textual descriptions of the content and appearance of the graphic that can be referred to by tags in HTML files, and 2) there is a navigation tool that allows users to query and navigate the content of the graph.

However, all of these approaches require the user to build a "mental map" of the graph – a task that is very difficult for the congenitally blind because they have no personal knowledge regarding the appearance of information graphics [16]. In addition, many of the other systems require 1) special equipment or 2) preparation work (such as model creation) by a sighted individual. For example, inspectGraph [14] requires that support files be generated by a plug-in when the graph is created.

Consequently, existing systems have not been successful in solving the graph accessibility issue for visually impaired individuals. Thus it is imperative that novel approaches be investigated.

1.3 The Importance of Making Information Graphics Accessible

When we asked our visually impaired users whether they felt that the lack of access to the content of information graphics hindered their ability to comprehend the full meaning of a document, the responses were quite varied. Some users felt that it severely impeded their understanding, while others expressed their belief that while an information graphic might add some detail, it generally repeated what was already in the text of the document. In order to assess the relative importance of being able to access the information contained within an information graphic, we conducted a corpus study [5] whose primary goal was to determine the extent to which the message conveyed by an information graphic in a multimodal document is also conveyed by the document's text. We analyzed 100 randomly selected graphics from our collected corpus of information graphics, along with the articles in which they appeared. The selected articles were taken from magazines (such as Newsweek, Business Week, Fortune, and Time) and local and national newspapers. In 39% of the instances, the text was judged to fully or mostly convey the message of the information graphic. However, in 26% of the instances, the text conveyed only a little of the graphic's message. Most alarming was the observation that in 35% of the instances in our analyzed corpus, the text failed to convey any of the message, yet the message conveyed by the graphic was important to full understanding of the article. These findings demonstrate that it is crucial that mechanisms be developed for providing visually impaired individuals with alternative access to the content of information graphics.

The usefulness of the text within information graphics is also important when considering the broader issue of the accessibility for visually impaired users of all images on the Web. For example, the WebInSight [3] project is aimed at providing useful alternative text for images found on the web, such as those found in navigation bars, as form buttons, and to display textual and visual content. WebInSight attempts to extract any text found in the images and supply it to the user via the alternative text attributes of the `img` tag in the HTML. Blind users access the web through a proxy that allows WebInSight to supply alternative text for the images being accessed. Alternative text is generated dynamically through OCR (optical character recognition) and web context labeling. Text that is generated is stored in a database for faster future retrieval. This method shows great promise for many critical navigational and functional images (such as those found in menus or on buttons), since the important text is often embedded in the image. However, in light of the corpus study presented in [11], the captions found in information graphics are unlikely to consistently supply helpful information about the contents of the graphic. In fact,

one could envision a system such as ours being used to generate the alternative text for information graphics within a framework such as WebInSight.

The remainder of this chapter presents our implemented system for enabling visually impaired users to gain access to the information provided by simple bar charts that appear on a web page. First we present the design goals and implementation of the browser extension that allows visually impaired users to interact with our system. Next we outline our overall methodology for inferring the intended message of an information graphic and conveying it to the user via spoken language. We then present the user evaluations of the system and discuss how they are influencing the future directions of our work. The focus of this chapter is on the interaction of the user with the system interface, the impact of the design decisions upon the user experience, and the user evaluations and feedback.

1.4 System Architecture

The architecture of our system is shown in Figure 1.2. The following subsections discuss the various components of our system, beginning with the interface. Note that while the browser extension will work for any type of information graphic, the scope of the work currently implemented for the image processing and message inference components is limited to the processing of simple bar charts. By simple bar charts, we mean bar charts that display the values of a single independent attribute and the corresponding values for a single dependent attribute (such as in the graphic in Figure 1.1). Although our system currently handles only simple bar charts, our methodology is currently being extended to other types of information graphics [4, 28].

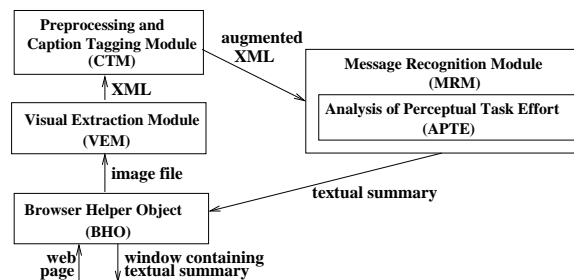


Fig. 1.2 System Architecture

1.5 Browser Extension

1.5.1 Design Goals

In designing the interface to the SIGHT system, we considered the following three goals: 1) the system should be usable by as many visually impaired individuals as possible, 2) the system should not require the use of special equipment or assistance from a sighted helper, and 3) the interface presented to users should be easy to use and compatible with their navigation preferences. Each of these design goals influenced the form and functionality of the browser extension that serves as our user interface.

In order to achieve the first goal, reaching the largest potential audience of users, we have implemented our browser extension specifically for Internet Explorer (preferably version 6.0 or later). Since Microsoft's Internet Explorer currently holds the majority web browser market share [23], most web related software designed for visually impaired users works well with Internet Explorer. Similarly, JAWS, produced by Freedom Scientific, holds 65% of the screen reader market share [20]. Thus we have implemented our browser extension for Internet Explorer and tested it using JAWS. However, the concepts applied here are extensible to other implementation platforms (see Section 1.5.4).

With respect to the second design goal, visually impaired computer users are already encumbered by having to buy relatively expensive screen reading or magnification software. Any solution which requires additional special equipment would likely limit potential users, as well as hinder the portability of the computer system it is attached to. Thus, being able to utilize our system with only a web browser and screen reading software is a strong advantage. The ability of our system to automatically infer the intended message of an information graphic ensures that intervention, in the form of assistance or preparation work, by a sighted individual (including the developer of the web page) is not necessary.

The final goal for our browser extension was that it should be as easy as possible for users to find and select an information graphic, signal for the system to infer the graphic's summary and convey it to the user, and then return to the original position in the web page. For this reason, our system is completely keystroke driven and utilizes keys that do not interfere with any of the current navigational settings for JAWS.

1.5.2 Interacting with the System

When navigating a web page, JAWS users have many options. When the web page is initially opened, JAWS begins reading the content of the web page, from top to bottom. The actual content that JAWS reads is highly configurable by the user, but typically includes any text on the page, the screen text pertaining to links and

buttons, and the alternative text associated with graphics. Additionally, users could choose to press the “tab” key to traverse the content of the page, use quick navigation keys (such as G and Shift+G), or use the cursor (arrow) keys to control their navigation through the content of a web page. In order to avoid conflicts with the existing navigation commands in JAWS, we chose CONTROL+Z as the key combination for launching our system. If the user comes across a bar chart during their navigation of a web page, they can hit CONTROL+Z to launch our application and receive a textual summary of the information conveyed by the bar chart. For example, if the user encountered the graphic shown in Figure 1.3, they could hit CONTROL+Z and a dialog box containing the summary of the graphic would appear. For this particular graphic, our system produces the summary *“This bar chart titled ‘The notebook spiral’ shows that the dollar value of average laptop prices fell from 2000 to 2003 and then falls more slowly until 2007.”* By default, JAWS will read the contents of the dialog box aloud as soon as it is displayed. A user can also have JAWS reread the graphic’s inferred message, or close the dialog by pressing the spacebar.

THE NOTEBOOK SPIRAL

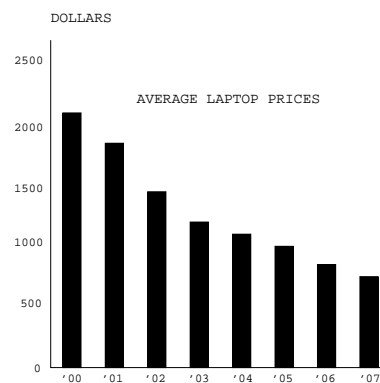


Fig. 1.3 Graphic with a Changing Trend Message³

This type of interaction requires a very tight coupling between our application and the web browser, because our application needs to be able to determine which graphic is currently in focus within the web browser and listen for user keystrokes. We achieved the proper level of integration by implementing our user interface as a Browser Helper Object for Internet Explorer.

Browser Helper Objects (BHOs) are special add-on components that enable the customization of Internet Explorer (version 4.0 or later). BHOs are tied to the main window of the browser and are created as soon as a new browser window is created. BHOs are implemented as in-process Component Object Model (COM) compo-

³ Graphic from BusinessWeek, September 5, 2005.

nents, and they run in the same process space as the browser; this means that they can perform virtually any action on the available windows and modules. Our BHO hooks onto Internet Explorer's message loop and captures all of the keyboard events within the browser, looking for the CONTROL+Z combination. Upon detecting the CONTROL+Z combination, the BHO queries the Document object of the Internet Explorer instance to determine which object is currently in focus within the browser. If the object in focus appears to be a graphic containing a bar chart, our system attempts to infer the intended message of the bar chart.

1.5.3 Identifying Bar Charts

When a user navigates to a graphic, JAWS attempts to identify the graphic by using the "alt text" (if present) or the file name. However, these sources often do not contain any indication that the image represents a bar chart. Not surprisingly, some users have stated in their user evaluations that they often do not know when an image represents an information graphic. Therefore, as soon as the web page has loaded, our browser extension scans all the images on the page for ones that appear to be bar charts. The scan uses simple tests, such as whether the graphic has fewer than 20 gray levels, and whether or not rectangles with aligned edges (as they would be aligned along an axis) are present in the image. If an image is considered likely to be a bar chart, the browser extension appends the message, "This image appears to be a bar chart. Press Control+Z to determine the summary." to any existing alt text for the image.⁴ JAWS then reads this text to the user when the user navigates to the image. Alternatively, if the user presses Control+Z and the object in focus does not appear to be a bar chart, the message "The selected graphic does not appear to be a bar chart," is read to the user by JAWS.

1.5.4 Extensibility of the Browser Extension

While the current version of the user interface has been designed specifically with JAWS and Internet Explorer in mind, we expect similar solutions to work for other applications. For example, extensions similar to BHOs can be developed for Mozilla's Firefox browser using the Cross Platform Component Object Model (XP-COM). Regarding the use of screen readers other than JAWS, our BHO in Internet Explorer will work with any screen reader; it is simply a matter of investigating how the focus of Internet Explorer and the screen reading software interact and of ensuring that the keystroke combination does not conflict with existing screen reader functionality. For visually impaired users who primarily use a screen magni-

⁴ Any images that pass the pre-processing checks for bar charts are also added to the tab order of the page, so that users tabbing through the content will not miss these graphics.

fier (such as ZoomText), the text produced by our BHO can be handled in the same manner as text in any other application.

1.6 Processing the Image

After our system is launched by the user through the browser extension, the image is processed by the Visual Extraction Module (VEM). The VEM is responsible for analyzing the graphic's image file and producing an XML representation containing information about the components of the information graphic including the graphic type (bar chart, pie chart, etc.) and the textual pieces of the graphic (such as its caption). For a bar chart, the representation includes the number of bars in the graph, the labels of the axes, and information for each bar such as its label, the height of the bar, the color of the bar, and so forth [7]. This module currently handles only electronic images produced with a given set of fonts and no overlapping characters. In addition, the VEM currently assumes standard placement of labels and axis headings. Work is underway to remove these restrictions. But even with these restrictions removed, the VEM can assume that it is dealing with a simple bar chart, and thus the problem of recognizing the entities is much more constrained than typical computer vision problems. An example of the XML representation of a bar chart produced by the VEM is given later in the chapter.

The XML representation is then passed to the Preprocessing and Caption Tagging Module (CTM). The preprocessor augments the XML with salience information such as a bar that is colored differently from other bars in the graphic or a bar that has an annotation when the other bars do not. The CTM extracts information from the caption (discussed later) and then passes the augmented XML representation to the message recognition module (MRM), which is described in the next section.

1.7 MRM: A Bayesian Inference System

We have developed a Bayesian inference system for recognizing the intended message of an information graphic. We hypothesize that this message can serve as the core content of an effective summary of the information graphic. This message inference methodology has several potential domain applications aside from accessibility for blind users, such as generating searchable summaries of information graphics for digital libraries, and tutoring systems aimed at improving the design of information graphics. Note that this chapter is centered on the issue of making bar charts accessible to blind users, and therefore this section is intended as an overview of our Bayesian inference system; further details on the network design and implementation can be found in [12].

1.7.1 *Communicative Signals*

We view information graphics that appear in popular media as a form of language with a communicative intention. Therefore, we have extended plan inference techniques that have successfully been used in recognizing the intention of natural language utterances (for example [24, 6]) to the novel domain of information graphics. A critical component of any plan inference system is the evidence, or communicative signals, that are identified and utilized by the system to infer the communicative intention of an agent. In this case, the MRM is attempting to infer the message that the graphic designer intended a viewer of the graphic to recognize. We have identified three categories of communicative signals that appear in simple bar charts.

Our first communicative signal is the salience of entities in the graphic. There are a number of strategies that can be employed by a graphic designer to make an entity or entities in a bar chart salient to the viewer. For example, a bar could be shaded or colored differently from the other bars, as is the case for the bar labelled CBS in Figure 1.5. Alternatively, the label of the bar could be made salient, as seen in the graphic in Figure 1.1, where the “Lancaster” label is shown in bold. Alternatively, a bar in a bar chart could be specially annotated. Mentioning a bar’s label in the caption of the graphic also makes the represented entity salient, since this draws attention to the bar. Salient entities are identified by the preprocessor and caption tagging module. The preprocessor analyzes the XML representation of the graphic and augments it to indicate entities that are salient due to graphic design decisions. — for example, the preprocessor will compare the colors of bars to determine if any bar is salient due to being the only bar with a unique color. The Caption Tagging Module uses a part-of-speech tagger to identify nouns in the caption, and then it augments the XML to indicate any bars whose labels match a noun in the caption.

A second communicative signal is the relative effort required for different perceptual and cognitive tasks. By perceptual tasks [17] we mean tasks that are performed by viewing the graphic, such as comparing the heights of two bars; by cognitive tasks, we mean tasks that require a mental computation such as interpolating between two labelled values on the dependent axis in order to determine the value represented by a bar whose top is not aligned with a labelled value. Here we are extending a hypothesis of the AutoBrief group [17]. The AutoBrief project was concerned with generating information graphics, and they hypothesized that a graphic designer chooses a design that best facilitates the perceptual and cognitive tasks that a viewer will need to perform on the graphic. We posit that if a graphic designer went to the effort of making a particular perceptual task easy to perform on a given graphic, then that perceptual task is likely to be part of the set of tasks that the viewer was intended to perform in deciphering the graphic’s message. This correlates with Larkin and Simon’s [19] observation that graphics that are informationally equivalent are not necessarily computationally equivalent — for example, if a set of bars are arranged in order of increasing height, then it will be much easier to identify the rank of an individual bar than if the bars were arranged in alphabetical order of their label, although the task of identifying the rank of a bar can be done on both graphics.

In order to rank the perceptual tasks that can be performed on an information graphic in terms of effort, we constructed a set of rules. Each rule is comprised of condition-computation pairs and can be used to estimate the effort of performing a particular perceptual task on a given information graphic. The rules contain conditions that capture design features of a graphic (such as whether the bars are sorted by height or whether the top of a particular bar aligns with a labelled tick mark on the axis) that impact the effort required to perform the task. The computations corresponding to the conditions capture the effort required for the task given that set of graphic features. The effort computations are based on work by cognitive psychologists, and our set of effort estimation rules have been validated through eye-tracking experiments [13].

A third communicative signal, and the only one external to the information graphic itself, is the presence of certain verbs and adjectives in a caption. In [11] we presented a corpus study showing that (1) captions are often very general or uninformative, and (2) even when captions convey something about the graphic’s intended message, the caption is often ill-formed or requires extensive analogical reasoning. Similar observations were also made by Corio and LaPalme [9], who studied captions with the objective of categorizing the kinds of information contained in captions in order to form rules for generating captions to accompany graphics. Rather than attempting to actually *understand* the caption, we utilize shallow processing of the caption in order to extract communicative signals. For example, the verb *lag* in the caption “*American Express’ total billings still lag*” suggests a message about an entity’s rank with respect to some measure. Similarly, we found that nouns derived from verbs, such as *rise* in the caption “*Cable on the Rise*”, and adjectives also suggest the general category of message. Using WordNet and a thesaurus, we identified verbs and adjectives that were similar in meaning and might signal one or more categories of message and organized them into verb classes. The Caption Tagging Module uses a part-of-speech tagger and a stemmer to analyze captions and extract nouns, adjectives, and the root form of verbs, adjectives, and nouns derived from verbs, and further augments the XML representation of the graphic to indicate the presence of one of our identified verb or adjective classes in the caption.

1.7.2 Reasoning about the Graphic’s Message

The communicative signals that we extract from an information graphic are then utilized by our system to reason about the intended message of the graphic within a Bayesian framework. For each new information graphic, we dynamically construct a Bayesian network. The top level of the network captures the various categories of messages that can be conveyed by a bar chart, such as conveying a change in trend (Change-Trend), conveying the rank of an entity in a bar chart (Get-Rank), comparing two entities (Relative-Difference), and so forth.

To identify the kinds of high-level messages conveyed by bar charts, we collected a corpus of bar charts from various publications, including business and news-

oriented magazines such as Newsweek, BusinessWeek, Time, Fortune and Money, as well as local and national newspapers. For each bar chart in the corpus, we asked two coders to identify 1) the intended message of the graphic using the provided list of possible message schemas, and 2) the instantiation of the parameters in the schema. The coders were also given the option of identifying new message schemas when appropriate. The high-level messages identified by our corpus study, and inferred by the SIGHT system, are shown in Figure 1.4.

Intention	Description
Get-Rank	Viewer to believe that $\langle param_1 \rangle$ is ranked $\langle rank \rangle$ among the elements in the graphic
Rank-All	Viewer to believe that the elements in the graph have an ordering $\langle element_1 \dots element_n \rangle$
Increasing-Trend	Viewer to believe that there is an increasing trend from $\langle param_1 \rangle$ to $\langle param_2 \rangle$
Decreasing-Trend	Viewer to believe that there is a decreasing trend from $\langle param_1 \rangle$ to $\langle param_2 \rangle$
Stable-Trend	Viewer to believe that there is a stable trend from $\langle param_1 \rangle$ to $\langle param_2 \rangle$
Change-Trend	Viewer to believe that there is a $\langle slope_1 \rangle$ trend from $\langle param_1 \rangle$ to $\langle param_2 \rangle$ and a significantly different $\langle slope_2 \rangle$ trend from $\langle param_2 \rangle$ to $\langle param_3 \rangle$
Contrast-Pt-Trend	Viewer to believe that there is a $\langle slope_1 \rangle$ trend from $\langle param_1 \rangle$ to $\langle param_2 \rangle$ and that the value of subsequent element $\langle param_3 \rangle$ contrasts with this trend
Relative-Difference	Viewer to believe that the value of element $\langle param_1 \rangle$ is $\langle comparison \rangle$ the value of element $\langle param_2 \rangle$, where $\langle comparison \rangle$ is greater-than, less-than, or equal-to.
Relative-Difference-Degree	Viewer to believe that the value of element $\langle param_1 \rangle$ is $\langle comparison \rangle$ the value of element $\langle param_2 \rangle$, where $\langle comparison \rangle$ is greater-than, less-than, or equal-to, and the $\langle degree \rangle$ of that difference is large, medium, or small
Maximum	Viewer to believe that $\langle param_1 \rangle$ has the largest value among the entities in the graphic
Minimum	Viewer to believe that $\langle param_1 \rangle$ has the smallest value among the entities in the graphic
Present-Data	Graphic simply presents data with no underlying message

Fig. 1.4 Categories of High-Level Intentions

Below each category of high-level message in the Bayesian network are nodes that capture the different possible instantiations of that message category. For example, if a graphic has five bars as in Figure 1.5, then the children of the Get-Rank node would be Get-Rank(BAR1,LABEL1), Get-Rank(BAR2,LABEL2), ..., Get-Rank(BAR5, LABEL5).

Our library of plan operators specifies how the communicative goals of a graphic designer for a bar chart can be achieved via the viewer performing certain perceptual and cognitive tasks. The operators break higher-level tasks into lower-level subgoals. Each subgoal in an operator is either a primitive with an associated effort rule or has an operator that decomposes it into a set of simpler subgoals. The operators determine the structure of our Bayesian network, in that subgoals in an operator become children of their goal node in the network. For example, consider the goal of the viewer getting the rank of a bar, given that the bar is salient, as in Figure 1.5. The operator for achieving this goal decomposes the goal into three subgoals: perceiving whether the bars are sorted in order of height, perceiving (ie., finding) the label associated with the bar, and perceiving the rank of that bar with respect to bar height. These subgoals would be part of the system’s inferred plan for getting the rank of CBS in terms of average ad price in Figure 1.5, as in `Get-Rank(4, CBS)`.

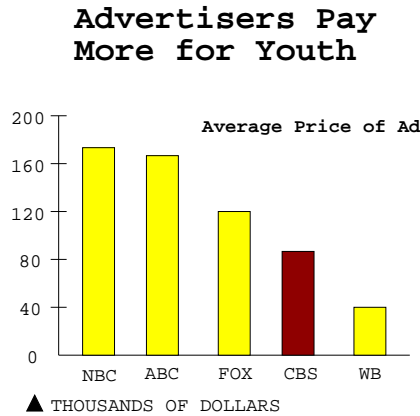


Fig. 1.5 Graphic with a Get-Rank Message⁵

To build the network for a given graphic, a node capturing the top-level message categories is initially entered into the network along with nodes capturing a set of low-level perceptual tasks. Ideally, this would include each possible instantiation of each low-level perceptual task; for example, the parameter `.bar` in the perceptual task `Perceive-Label(.bar)` could be instantiated with any of the bars that appear in a graphic. However, memory limitations restrict the size of the network and force us to include only instantiated perceptual tasks that are suggested by the graphic. The instantiations that produce perceptual tasks of lowest effort and any salient instantiations (see Section 1.7.1) are used to form the set of suggested low-level perceptual

⁵ Graphic from BusinessWeek, April 5, 1999. Note that in its original form, the graph was not a simple bar chart, because there was a secondary value (average age of values) also displayed on the bars, so it has been adapted to display only a single dependent value.

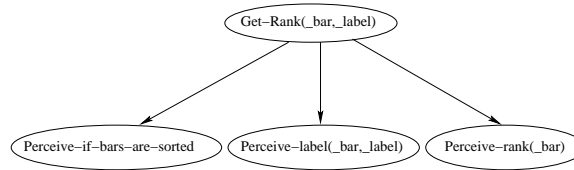


Fig. 1.6 A Piece of Network Structure

tasks that are initially entered into the network. Then chaining via the operators adds nodes until a link is established to the top-level; as new nodes are added, their sub-goals (as captured in the plan operators) are also added, so that the network is also expanded downwards. Once the network is constructed, evidence nodes are added as discussed in the next section.

As noted above, the communicative signals extracted from a graphic are represented in the Bayesian network as evidence nodes. As discussed in Section 1.7.1, this evidence includes the relative effort required for different perceptual tasks, the salience of entities in the graphic, and the evidence provided by verbs and adjectives in the caption that are members of one of our identified verb and adjective classes. Evidence nodes that pertain to the likelihood of a particular perceptual task being part of the graphic designer's intended plan for the viewer are attached to the perceptual task nodes in the network. For example, evidence nodes reflecting the amount of effort required for a perceptual task (categorized as low, medium, high, or impossible) and whether a parameter of a perceptual task is salient (via highlighting, annotating it, etc.) are attached to perceptual task nodes. Consider the graphic in Figure 1.5, the piece of network structure shown in Figure 1.6, and the evidence nodes that would be attached to the instantiated perceptual task node `Perceive-rank(BAR4)`. The effort evidence node would indicate that little effort is required for this task since the bars are sorted according to height. The highlighting evidence node would indicate that the instantiated parameter `BAR4` is highlighted in the graphic. The annotation and noun-in-caption evidence nodes would indicate respectively that no bars have special annotations and that none of the bar labels are part of the graphic's caption. In contrast, evidence nodes reflecting the presence of one of our verb or adjective classes in the caption provide evidence for a general category of message and thus are attached to the top level node in the network. For the graphic in Figure 1.5, this evidence node indicates the lack of verb or adjective evidence in the graphic's caption.

Associated with each child node in a Bayesian network is a conditional probability table that gives the conditional probability for each value of the child node given the value of the parent node. For example, one conditional probability table states that the probability that a particular bar (and only that bar) is annotated is 24.99% if the graphic designer intended for the viewer to perceive the rank of the bar, and the probability is 74.99% if that was not the designer's intention. The conditional probability tables for network nodes are learned from our corpus of graphics.

Once the network with its evidence nodes is built, the probabilities propagate through the network to hypothesize the intended message of the graphic. For the graphic in Figure 1.5, the system infers that the graphic is conveying the rank of CBS and produces the summary “*This bar chart titled ‘Advertisers pay more for youth’ shows that CBS has the second lowest rank in terms of the dollar value of average price of Ad compared with NBC, ABC, FOX, and WB.*”⁶

1.8 Generating the Summary

Once the intended message has been inferred by our Bayesian inference system, it is used as the core content of a textual summary of the graphic. One of the most challenging aspects of generating coherent natural language has been determining the full label for the measurement (or value) axis. In examining our corpus of bar charts, taken from a variety of magazines and newspapers, we have found that the measurement axis label might be very abbreviated and that full rendering of the label often requires extraction of words from text within the graphic (such as *AVERAGE LAPTOP PRICES* in Figure 1.3) or from the caption and/or second-tier descriptive text below the caption. We have constructed a set of heuristics for ranking the graphic’s components in terms of where to look for the measurement axis label and how to extract it from these textual pieces. Other heuristics augment the wording of the label.⁷ For example, one heuristic states that if the graphic’s text contains a single proper noun that does not match the label of a bar, then the measurement axis label should generally be preceded with that proper noun in possessive case. Consider, for example, the graphic in Figure 1.7. Here the measurement axis is labelled as *Percentage of unauthorized workers* and the unit of measurement is also captured by the % sign after the annotated values, but “*Percentage of unauthorized workers*” must be preceded with the proper noun “*United States*” in order to fully describe what is being measured. The natural language generation component of our system [10] uses a set of heuristics to generate a complete rendering of the measurement axis label, which is then used in templates to generate the appropriate wording for the particular category of inferred message. For example, the system hypothesizes that the graphic in Figure 1.7 is conveying a Rank-of-all message and generates the textual summary “*This bar chart titled ‘Workers without papers’ compares the entities Farming, Cleaning, Construction, and Food preparation with respect to United States’s percentage of unauthorized workers.*”

⁶ For further detail on the Bayesian network, see [12]

⁷ Note that the heuristics are impacting the wording of the summary, and not the inference of the intended message of the graphic; the message is inferred by the Bayesian network described in Section 1.7.

⁸ Graphic from USA Today, July 11, 2006 issue.

Workers without papers

Industries that require manual labor and little formal education draw heavily on illegal immigrants' labor. About 5% of workers in the United States are illegal immigrants.

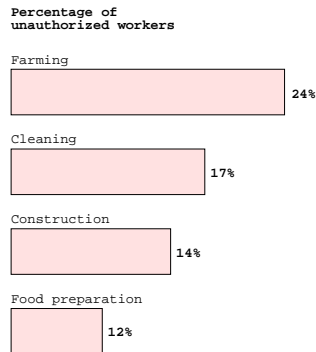


Fig. 1.7 Graphic with a Rank of All Message⁸

1.9 Example

In order to better illustrate how our system works, we will follow an example from the web page through to the generation of the textual summary of the bar charts contained on the web page. The web page that we are using for this example is based on an article from the April 5, 1999, issue of *BusinessWeek* magazine. The article is entitled, "A Rising House of Cards?" and covers approximately one column of a three-column page. The article contains a single bar chart and no other graphics. Our html version of the article (see Figure 1.8) contains the full text of the article and an image file containing a replication of the bar chart.⁹ Since the majority of graphics on the web do not have alt text to describe their content, we have not included any alt text in our html code.

Upon opening the web page in Internet Explorer (with our BHO installed and JAWS screen reading software running), the graphics on the page are scanned by our bar chart detection logic, and the bar chart is added to the tab order of the page and tagged with the alt text, "This image appears to be a bar chart. Press Control+Z to determine the summary." Note that the other graphics on the page, such as the *BusinessWeek* cover, are passed over by our application because they are not recognizable as bar charts. JAWS begins reading the article and eventually the user will

⁹ Minor modifications have been made to the appearance of the original bar chart in order to enable the image processing in VEM to successfully process the image: altering the font, eliminating gridlines, and removing text from inside the bars.

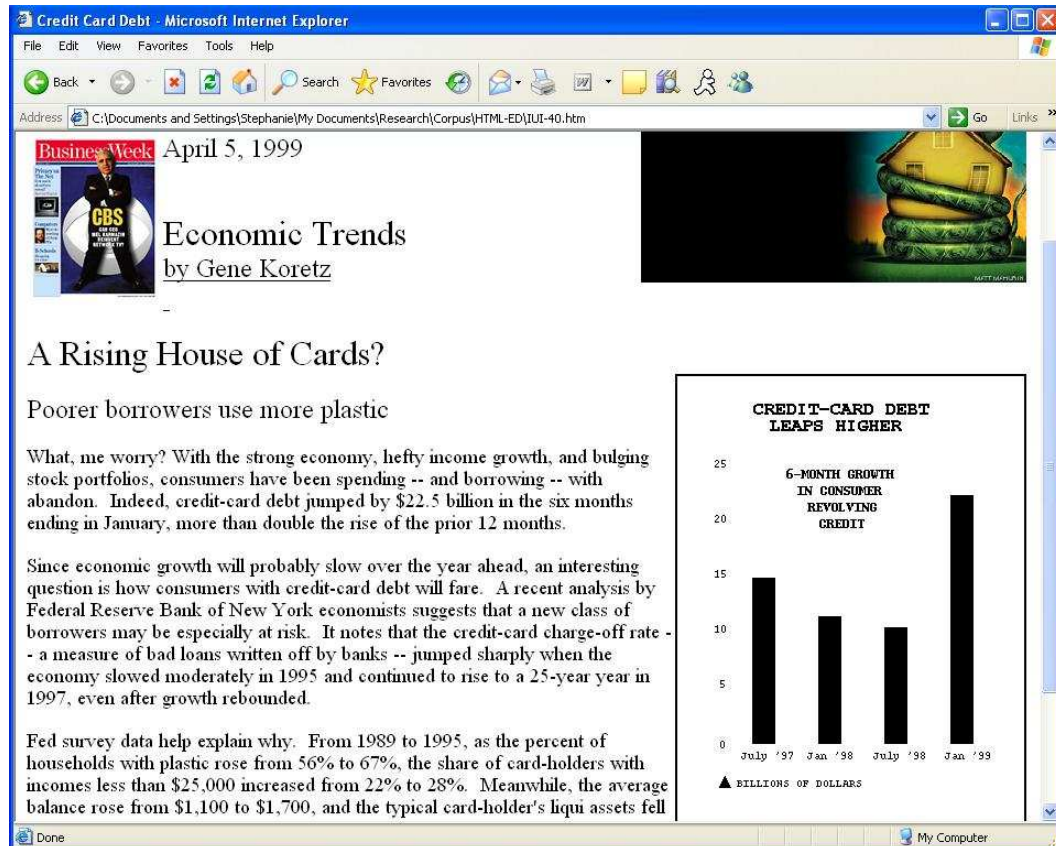


Fig. 1.8 Screenshot of Document in Web Browser

hear the message, “This image appears to be a bar chart.” Upon hearing this, the user can enter CONTROL+Z to request a textual summary of the graphic.

Within the SIGHT system, the BHO passes the graphic to the Visual Extraction Module (VEM). The XML representation produced by the Visual Extraction Module (VEM) for the graphic in Figure 1.9 is shown below; note that the measurements in the XML may not match the bar chart in Figure 1.9 since it has been resized for display purposes.

```
<?xml version="1.0" encoding="UTF-8"?>
<InformationGraphic>
<BarChart BarDirection="vertical">
  <Caption>
    <Content>
      CREDIT-CARD DEBT LEAPS HIGHER
```

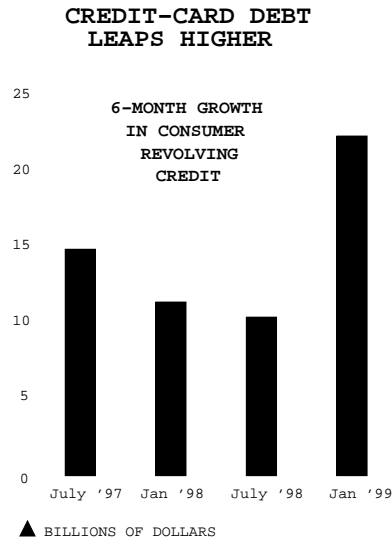


Fig. 1.9 Graphic with a Contrast Point with Trend Message¹⁰

```

</Content>
</Caption>
<TextInGraphic>
  6-MONTH GROWTH IN CONSUMER REVOLVING CREDIT
</TextInGraphic>
<MeasurementAxis Length="7.94">
  <Label>BILLIONS OF DOLLARS</Label>
  <Tickmark>
    <TickLabelled>true</TickLabelled>
    <TickValue>0.00</TickValue>
    <GridLine>>false</GridLine>
  </Tickmark>
  ... (xml for other vertical axis labels)
</MeasurementAxis>
<BarAxis Length="7.09">
</BarAxis>
<Bar>
  <Label>
    <Content>July '97</Content>
    <Color>0</Color>
    <Bold>>false</Bold>
  </Label>
  <Color>0</Color>
  <Height>5.33</Height>
  <AxisDistance>0.49</AxisDistance>
  <SightLine>>false</SightLine>
  <Value>14.70</Value>

```

```

</Bar>
<Bar>
  <Label>
    <Content>Jan '98</Content>
    <Color>0</Color>
    <Bold>>false</Bold>
  </Label>
  <Color>0</Color>
  <Height>4.09</Height>
  <AxisDistance>2.61</AxisDistance>
  <SightLine>>false</SightLine>
  <Value>11.20</Value>
</Bar>
... (xml for the other three bars)
</BarChart>
</InformationGraphic>

```

This XML is then passed to the Preprocessing and Caption Tagging Module (CTM). The preprocessor does not find any special annotations or highlighted bars, but it does determine that the bar labels form a timeline so that a trend is possible. The preprocessor analyzes possible divisions of the graphic where a trend could be recognized and augments the html with this information. The augmented xml is then passed to the CTM. The CTM does not find any nouns that match bar labels, nor any verbs from our list of helpful verbs.¹¹

The augmented XML is then passed to the Message Recognition Module (MRM). The MRM constructs the Bayesian network and hypothesizes that the intended message of the graphic is to contrast the increase in January of 1999 in 6-month growth in consumer revolving credit against the previously decreasing trend. This inferred message is assigned a probability of 75.6. If we had utilized the verb “leap” from the caption, this probability might have been higher; however, notice that the caption, “Credit-Card Debt Leaps Higher,” can be considered, at best, to partially convey the message of the graphic, in that it implies some sort of a sharp increase. However, it is unclear from the caption whether there might be a sharply rising trend or a single increasing data point. In either case, the caption fails to convey the previously decreasing trend from July of 1997 to July of 1998.

The SIGHT system now uses this inferred message and the XML representing the bar chart to construct the following natural language summary:

“This bar chart titled ‘Credit-card debt leaps higher’ shows that there is an increase in Jan ’99 in the dollar value of 6-month growth in consumer revolving credit in contrast with the decreasing trend from July ’97 to July ’98.”

The BHO creates a new window, which is a child window of Internet Explorer, containing this text (see Figure 1.10), and assigns that window the current focus. JAWS reads the textual summary, and the user closes the window using ALT+F4.

¹⁰ Graphic from BusinessWeek, April 5, 1999 issue.

¹¹ Note that although “leap” does appear to have the potential to be helpful in this case, it is not in our list of previously identified verbs.

The focus within Internet Explorer and JAWS returns to the bar chart, and JAWS continues to read the remaining content on the page.



Fig. 1.10 Screenshot of SIGHT output

1.10 Evaluation

The SIGHT system has been evaluated for the accuracy of the message inference process, as well as for the user experience. In order to evaluate the accuracy of the message inference, we utilized a corpus of 110 simple bar charts that had previously been annotated with their primary message by two human coders. We used leave-one-out cross validation, where each graphic is selected once as the test graphic, and the other 109 graphics are used to compute the conditional probability tables for the Bayesian network. We viewed the system as successful in recognizing the graphic's message if its top-rated hypothesis matched the message assigned by the coders and the system-assigned probability for the hypothesis exceeded 50%. The system's overall success rate, 79.1%, is the average of the results of all 110 experiments.

In assessing the success of the system, it is important to note that even in the cases where the system's inferred message does not match the preferred message identified by human coders, under no circumstances will the system supply incorrect facts to the user. That is, the system will not recognize a falling trend if one does not exist, or erroneously identify an entity as having the maximum or minimum value. Also, although 79.1% is quite a strong success rate for a system involving plan inference, most of the errors in our current evaluation are caused by sparseness of the data; for example, if we only have a single graphic using a verb like "stabilize" to indicate a stable trend, the leave-one-out validation will not have any evidence linking the verb class to the intention, and we may get an incorrect result when evaluating the

graphic. We strongly believe that as we expand our corpus, our success rate will improve.

However, even the most accurate system cannot be deemed a success if it fails to meet the needs of its users. Therefore, we regard user evaluations to be vital, particularly when dealing with users with disabilities. Participants in our user evaluations were given a guided introduction to the SIGHT system, and then explored various web pages containing bar charts, using the SIGHT system to access the textual summaries of the messages conveyed by the bar charts. The evaluation concluded with a taped interview regarding the participants' experience with the system. Two areas were assessed during the user evaluations: the ease of use of the browser interface, and the effectiveness of our overall methodology in enabling visually impaired users to access the content of bar charts.

Ten visually impaired users participated in our study. The participants had varying degrees of computer experience, though all were JAWS users. The participants also possessed varying degrees of vision, but none could view the graphics without substantial magnification (which was not used during the experiments). There was also a mix of congenitally and non-congenitally blind users.

Three of the evaluation questions were numerical in nature. These questions were:

- On a scale of 1–10 with 1 being not useful at all and 10 being extremely useful, how useful would software be that can provide the message and content of a graph?
- On a scale of 1–10 with 1 being extremely difficult and 10 being extremely easy, how easy was it to select the graph within the web page?
- In order to start our system, you had to press Control+Z. On a scale from 1–10 with 1 being extremely difficult and 10 being extremely easy, how easy was this to do?

Users' average rating for the usefulness of a system that can provide the message and content of a bar chart found on a web page was 8.2 (STDEV = 2.1). Users likewise found that our system was overall very easy to use, giving the system an average rating of 9.1 (STDEV = 1.6) for the ease of selecting the graphic and an average rating of 9.77 (STDEV = 0.4) for the ease of the keystroke. The responses from each of the users is shown in Table 1.

The response from the participants in our user evaluations was overwhelmingly positive. Overall, the participants were very excited about the possibility of being able to access the content of information graphics. One participant remarked, "...sometimes the information is what you're trying to get. If you go over that graphic and then you run through this whole article or document and you're not getting what you want, it might be in that little bit. You know, which has happened in other cases."

One of our open-ended questions was whether or not the participants would like to be able to "ask questions" or obtain further information from the graphics, and if so, what information they would like to be able to get. A striking, though in retrospect not surprising, difference emerged between the congenitally and non-congenitally blind users as to whether they could identify what other types of infor-

User	Impairment	Usefulness	Ease of Use	Ease of Keystroke
1	Congenital	6	8	No answer
2	Congenital	10	10	10
3	Congenital	10	10	10
4	Congenital	10	10	10
5	Legally	8	9	9
6	Congenital	5	10	10
7	Noncongenital	9	5	10
8	Noncongenital	9	9	9
9	Noncongenital	10	10	10
10	Congenital	5	10	10
Average		8.2	9.1	9.8
Stdev		2.1	1.6	0.4

Table 1.1 Quantitative User Evaluation Results

mation they might like to have. This inherently makes sense, since the congenitally blind users have never had the opportunity to view or use bar charts. As one participant noted, "... I don't know enough about graphs to answer that question. Having never been able to see." We posit that this distinction lends credence to the hypothesis that it would be difficult for congenitally blind users to form a mental map of a bar chart, based on an alternative description of its appearance, in order to extract the message of the graphic. It should also be noted that, although many congenitally blind participants felt that the system would be quite useful, the three participants who rated the system's potential usefulness below a 7 were congenitally blind and had never viewed a bar chart. We surmise that these lower ratings reflect their lack of familiarity with information graphics and the valuable information that they provide.

Originally, we intended to find a keystroke other than Control+Z to launch the system, since Control+Z is used as 'undo' in other applications. However, since 'undo' is not a valid action in web browsers, and the keystroke is a relatively simple one, the users overwhelmingly felt that the simplicity of the stroke outweighed any potential confusion.

Another very interesting result of the evaluations was that when asked "Did you have confidence that the output from our system helped you to understand what the graph was conveying?" all of the participants responded positively. While encouraging, this result also underscores the responsibility of a system to avoid misleading its users. It should be noted that even if our system incorrectly infers the graphic designer's intended message for a bar chart, the inferred message still reflects information that is present in the graphic.

Several participants expressed their frustration with the current lack of access to information graphics. One participant commented on current approaches to verbal descriptions of graphics by saying "...They'll be like, you know, heart attacks in red, smoking's in green, you know, and I'm like, you know, 'Okay. Thank you. 'Cause that helps'... unless, like, the question on the exam is like, 'What color was...' then

it's like, 'Alright! I've got this one...' ” The participants also expressed their overall approval of our methodology of providing the message conveyed by the graphic, with one participant saying “The summary was wonderful. [It's] kind of like a cliff note version.”

1.11 Future Work

Our system has been implemented and tested for simple bar charts. We eventually envision the system as providing a richer initial summary than it currently does (although the core content of the summary will still be the inferred intended message), and then allowing the user to obtain additional information through a keystroke-driven interface. We are using the results of our user evaluations, along with additional experiments currently being conducted, to identify 1) what additional information should be included in the initial summary and how that information can be identified for a given graphic, and 2) an appropriate design for the interface for obtaining additional information from the graphic. We are also working to make the image processing in VEM more robust, to improve the response time of the system, and to extend the system to other kinds of information graphics such as line graphs and pie charts, and to complex graphics, such as grouped bar charts.

1.12 Conclusion

Information graphics are an important part of many documents available on the world-wide web, yet they are largely inaccessible to visually impaired users. This is clearly unacceptable. As one of our participants stated, “I think that we deserve to have as much information as everyone else does.” This paper has presented a novel implemented interface that enables visually impaired users to gain access to the information provided by simple bar charts that appear on a web page. Our approach of presenting the message conveyed by the information graphic does not require specialized hardware, preparation work by a sighted individual, or for the user to construct a mental map of the graphic. Moreover, it places little cognitive load on the user and does not require that he or she construct a mental map of the graphic – a task that is very difficult for congenitally blind individuals. Our user evaluations indicate that our approach shows great promise for addressing the challenge of providing universal access to information graphics.

References

1. Alty, J.L., Rigas, D.: Communicating graphical information to blind users using music: The role of context. In: Proceedings of CHI-98, Human Factors in Computer Systems, pp. 574–581. ACM Press, Los Angeles (1998)
2. Bertin, J.: *Semiology of Graphics*. The University of Wisconsin Press, Madison, Wisconsin (1983)
3. Bigham, J.P., Kaminsky, R.S., Ladner, R.E.: WebInSight: Making web images accessible. In: Proceedings of the Eighth International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS), pp. 181–188 (2006)
4. Burns, R., Elzer, S., Carberry, S.: Estimating effort for trend messages in grouped bar charts. In: Proceedings of the Fifth International Conference on the Theory and Application of Diagrams (Diagrams '08) (2008)
5. Carberry, S., Elzer, S., Demir, S.: Information graphics: An untapped resource for digital libraries. In: Proceedings of SIGIR 2006. Seattle, WA (2006)
6. Charniak, E., Goldman, R.P.: A bayesian model of plan recognition. *Artificial Intelligence* **64**(1), 53–79 (1993)
7. Chester, D., Elzer, S.: Getting computers to see information graphics so users do not have to. In: Proceedings of the 15th International Symposium on Methodologies for Intelligent Systems (ISMIS), Lecture Notes in Artificial Intelligence 3488, pp. 660–668. Springer-Verlag (2005)
8. Cleveland, W.S.: *The Elements of Graphing Data*. Chapman and Hall, New York (1985)
9. Corio, M., Lapalme, G.: Generation of texts for information graphics. In: Proceedings of the 7th European Workshop on Natural Language Generation EWNLG'99, pp. 49–58 (1999)
10. Demir, S., Carberry, S., Elzer, S.: Effectively realizing the inferred message of an information graphic. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP), pp. 150–156 (2007)
11. Elzer, S., Carberry, S., Chester, D., Demir, S., Green, N., Zukerman, I., Trnka, K.: Exploring and exploiting the limited utility of captions in recognizing intention in information graphics. In: Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL), pp. 223–230 (2005)
12. Elzer, S., Carberry, S., Zukerman, I., Chester, D., Green, N., Demir, S.: A probabilistic framework for recognizing intention in information graphics. In: Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI), pp. 1042–1047 (2005)
13. Elzer, S., Green, N., Carberry, S., Hoffman, J.: A model of perceptual task effort for bar charts and its role in recognizing intention. *User Modeling and User-Adapted Interaction* **16**(1), 1–30 (2006)
14. Ferres, L., Verkhogliad, P., Lindgaard, G., Boucher, L., Chretien, A., Lachance, M.: Improving accessibility to statistical graphs: the inspectgraph system. In: Proceedings of the Ninth International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS). Tempe, AZ (2007)
15. Ina, S.: Computer graphics for the blind. *ACM SIGCAPH Computers and the Physically Handicapped* **55**, 16–23 (1996)
16. Kennel, A.R.: Audiograf: A diagram-reader for the blind. In: Second Annual ACM Conference on Assistive Technologies, pp. 51–56 (1996)
17. Kerpedjiev, S., Roth, S.: Mapping communicative goals into conceptual tasks to generate graphics in discourse. In: Proceedings of the International Conference on Intelligent User Interfaces, pp. 60–67 (2000)
18. Kurze, M.: Giving blind people access to graphics (example: Business graphics). In: Proceedings Software-Ergonomie '95 Workshop Nicht-visuelle graphische Benutzungsoberflächen. Darmstadt, Germany (1995)
19. Larkin, J., Simon, H.: Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science* **11**, 65–99 (1987)

20. Law, C., Jacko, J., Edwards, P.: Programmer-focused website accessibility evaluations. In: *Assets '05: Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, pp. 20–27. ACM Press, New York, NY, USA (2005). DOI <http://doi.acm.org/10.1145/1090785.1090792>
21. McGookin, D.K., Brewster, S.A.: Soundbar: exploiting multiple views in multimodal graph browsing. In: *NordiCHI '06: Proceedings of the 4th Nordic conference on Human-computer interaction*, pp. 145–154. ACM Press, New York, NY, USA (2006). DOI <http://doi.acm.org/10.1145/1182475.1182491>
22. Meijer, P.B.: An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering* **39**(2), 112–121 (1992)
23. Onestat: Microsoft's Internet Explorer global usage share is 85.85 percent. <http://www.onestat.com/>. Accessed on 29 January 2007.
24. Perrault, R., Allen, J.: A Plan-Based Analysis of Indirect Speech Acts. *American Journal of Computational Linguistics* **6**(3-4), 167–182 (1980)
25. Ramloll, R., Yu, W., Brewster, S., Riedel, B., Murton, M., Dimigen, G.: Constructing sonified haptic line graphs for the blind student: First steps. In: *Proceedings of ASSETS 2000*, pp. 17–25. Arlington, Virginia (2000)
26. Tufte, E.R.: *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT (1983)
27. W3C: WAI: Web accessibility initiative. URL: <http://www.w3.org/WAI/>. Accessed May 22, 2007
28. Wu, P., Carberry, S., Chester, D., Elzer, S.: Decision tree induction for identifying trends in line graphs. In: *Proceedings of the 17th International Symposium on Methodologies for Intelligent Systems (ISMIS-08)*, pp. 399–409 (2008)
29. Yu, W., Reid, D., Brewster, S.: Web-based multimodal graphs for visually impaired people. In: *Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology*, pp. 97–108 (2002)