Information Graphics:
An Untapped Resource for Digital Libraries

Sandra Carberry
Dept. of Computer Science
University of Delaware
Newark, DE 19716 USA
carberry@cis.udel.edu

Stephanie Elzer
Dept. of Computer Science
Millersville University
Millersville, PA USA
elzer@cs.millersville.edu

Seniz Demir
Dept. of Computer Science
University of Delaware
Newark, DE 19716 USA
demir@cis.udel.edu

ABSTRACT

Information graphics are non-pictorial graphics such as bar charts and line graphs that depict attributes of entities and relations among entities. Most information graphics appearing in popular media have a communicative goal or intended message; consequently, information graphics constitute a form of language. This paper argues that information graphics are a valuable knowledge resource that should be retrievable from a digital library and that such graphics should be taken into account when summarizing a multimodal document for subsequent indexing and retrieval. But to accomplish this, the information graphic must be *understood* and its message recognized. The paper presents our Bayesian system for recognizing the primary message of one kind of information graphic (simple bar charts) and discusses the potential role of an information graphic's message in indexing graphics and summarizing multimodal documents.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Content Analysis and Indexing

General Terms

Algorithms

Keywords

Summarization, graphics, multimedia, Bayesian reasoning

1. INTRODUCTION

Information graphics are non-pictorial graphics such as bar charts and line graphs that depict attributes of entities and relations among entities. Although much attention has been devoted to the summarization and categorization of text, and to effective methods for retrieving textual documents relevant to an individual's needs, relatively little attention has been given to information graphics that appear
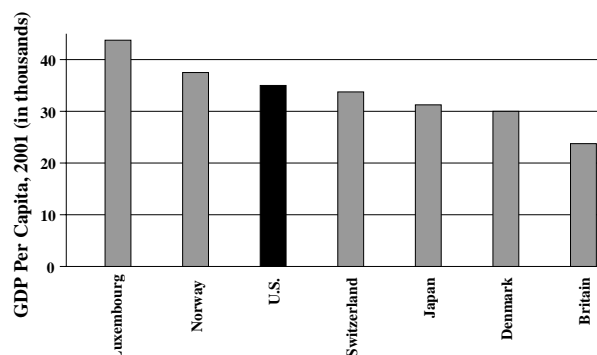
Figure 1: Graphic from U.S. News and World Report[2]

in documents. This paper addresses the inclusion of information graphics in digital libraries along with their role in the summarization of multimodal documents.

Section 2 presents a corpus study that explores how information graphics are used in multimodal documents. Section 3 argues (1) that information graphics are an important knowledge resource in their own right and that they should be retrievable from a digital library and (2) that information graphics should be taken into consideration in summarizing and indexing multimodal documents. Although some information graphics are only intended to display data, the majority of information graphics that appear in newspapers, magazines, and formal reports are intended to convey a message. For example, the information graphic in Figure 1 conveys the message that the U.S. ranked third (among the countries listed) in GDP (gross domestic product) per capita in 2001. Similarly, the graphic in Figure 2 conveys the message that there was a substantial increase in Delaware bankruptcy personal filings in 2001 compared with the preceding decreasing trend from 1998 to 2000. A graphic's primary or core message constitutes a brief summary of the graphic and captures its major contribution to

---

[2]In the original graphic, the bar for the United States was annotated. Here we have highlighted it in order to later show the XML that our system produces when bars are colored differently. We have also placed the dependent axis label alongside the dependent axis, instead of at the top of the graph, since our vision system is currently limited to standard placement of axis labels.
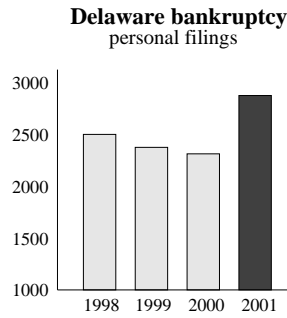
Delaware bankruptcy
personal filings

Figure 2: **Graphic from Wilmington News Journal**

Category	#
Category-1: Fully conveys message	22
Category-2: Mostly conveys message	17
Category-3: Conveys a little of message	26
Category-4: Does not convey message	35

Table 1: **Analysis of Text of Articles Containing Graphics**

the overall communicative goal of a multimodal document. Consequently, developing a methodology for recognizing this message is the first step to exploiting information graphics. Section 4 presents our implemented and evaluated Bayesian network for identifying the primary message conveyed by one kind of information graphic, simple bar charts. Section 5 discusses how the message recognized by our system can form the basis for summarizing an information graphic, for indexing and retrieving it from a digital library, and for constructing a richer summary of multimodal documents.

2. MULTIMODAL DOCUMENTS: A CORPUS ANALYSIS

Information graphics are an important component of many documents. In some cases, information graphics are standalone and constitute the entire document, as was the case for the graphic in Figure 2. However, in most cases, information graphics are part of a multimodal document consisting of both text and graphics. We conducted a corpus study 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. The articles varied in length: 27% were very short (a half-page or less), 20% were classified as short (one magazine length page), 22% were moderate in length (2 magazine length pages), and 31% were long (more than 2 magazine length pages). The graphics also varied in type: 33% were simple bar charts, 37% were simple line graphs, 10% were grouped bar charts, 16% were multiple line graphs (graphs consisting of multiple lines), and 4% were pie charts.

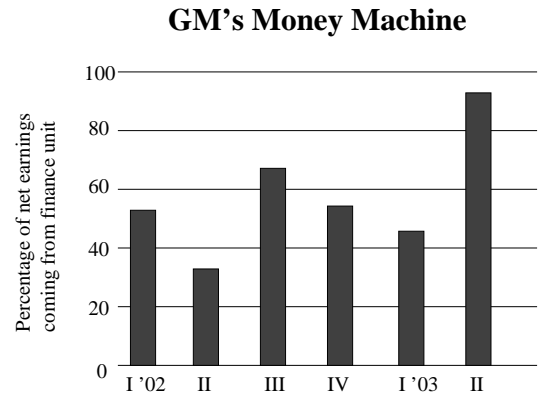We examined the text of each article and determined to what extent the text repeated the message conveyed by the

GM's Money Machine

Figure 3: **Graphic from Business Week**

information graphic. Table 1 displays the results. 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. An example is the graphic shown in Figure 3 which appeared in a *Business Week* article entitled "For GM, Mortgages are the Motor". The graphic's message is that there was a substantial increase in the percentage of GM's net earnings produced by its finance unit in the second quarter of 2003 in contrast with the preceding decreasing trend from the third quarter of 2002 to the first quarter of 2003. However, the pieces of text most closely related to the graphic were the following:

> "So where did the other $818 million in second-quarter profits come from? Try General Motors Acceptance Corp., GM's leading arm."

> "Until then, GM will struggle and remain dependent on its finance arm."

> "If not, GM would face lean times once GMAC is no longer minting money."

None of these text segments achieves the graphic's primary communicative goal. However, since the text segments talk about $818 million in second quarter profits coming from GM's finance unit and the finance unit making a lot of money (*minting money*), we judged the text to at least convey the high profitability of the finance unit and thus classified the text as conveying a little of the graphic's message.

Most surprising was the observation that in 35% of the instances in our analyzed corpus, the text failed to convey any of the graphic's message. An example is the grouped bar chart shown in Figure 4 which is taken from a *Newsweek* article entitled "Microsoft's Cultural Revolution"; this graphic conveys that the percentage of pirated software in China is much higher than in the world as a whole and that the decrease in pirated software in 2002 compared with 1994 was larger in the world than in China. Although the text is about Microsoft's efforts in China and the problem of pirated software, the closest that the text comes to capturing the graphic's message is the following statement:

> "Ninety percent of Microsoft products used in China are pirated."

**Percentage of
Software in Use
Which is Pirated**

Figure 4: Graphic from Newsweek
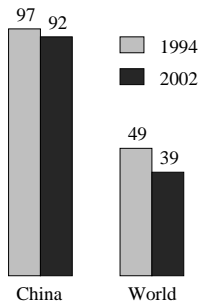
Median Income
In thousands of 2001 dollars

Figure 5: Graphic from Newsweek

But the text does not compare pirating in China with pirating in the world as a whole or compare the situation in 2002 with that in 1994. Thus we classified the text of this graphic as *failing to convey the graphic's message.*

Our further analysis of these multimodal documents has led us to conclude that graphics in multimodal documents generally have a communicative goal that, along with the communicative goals of the text segments, contributes to accomplishing the discourse purpose[11] of the overall article. For example, Figure 5 illustrates a graphic from a *Newsweek* article entitled "The Black Gender Gap"; the graphic conveys that the income of black women has risen dramatically over the last decade and has reached the level of white women. Although the text notes that the earnings of college-educated black women exceed both the median for all women and the median for all black working men, the text does not compare the earnings of all black women with those of all white women. Yet this comparison is more important (than those in the text) to achieving the overall communicative goal of this portion of the article — namely, convincing the reader that there has been a "monumental shifting of the sands" with regard to the achievements of black women. This example illustrates how authors distribute their communicative goals between the text and the graphics. We hypothesize that the communicative goals captured by information graphics are particularly central to the purpose of a document, since the graphic designer has chosen to draw attention to them via a graphic.

3. INFORMATION GRAPHICS AND DIGITAL LIBRARIES

Information graphics are themselves an important informational resource that should be stored in a digital library and be as accessible to humans as text documents. An individual might access an information graphic for the knowledge that can be gleaned from it, and for its use in planning, problem-solving, and decision-making. For example, a graphic such as the one in Figure 2 might be used by social service agencies in forecasting needed programs for the next year, or by legislators in arguing for or against bankruptcy legislation. Alternatively, graphics might be accessed for use in writing reports or proposals. They could also be used in educational settings for teaching students good analysis techniques.
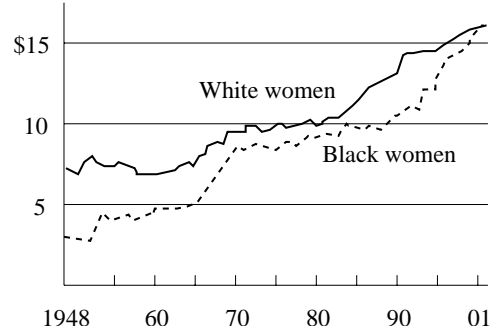
Moreover, as shown in Section 2, information graphics have communicative goals or intended messages that contribute to achieving the discourse purpose of a multimodal document but which are often not captured by the document's text. Thus the summarization of a multimodal document should take into account its information graphics.

We hypothesize that the core message of an information graphic (the primary overall message that the graphic conveys) can serve as the basis for an effective brief summary of the graphic. In a retrieval setting, this summary could be used to index the graphic and enable intelligent retrieval. In a question-answering setting, the summary might be used to directly answer a question or to determine whether the graphic should be analyzed in more detail as a possible source for answering the question. We further hypothesize that, since the core message conveyed by an information graphic captures its communicative goal, a graphic's summary based on this core message represents a good starting point for analyzing the contribution of the graphic to the overall discourse purpose of a document and for taking the graphic's contribution into account in constructing a rich summary of a multimodal document. The next section presents the methodology underlying our implemented and evaluated system for recognizing the core message conveyed by one kind of information graphic, a simple bar chart.

4. RECOGNIZING THE CONVEYED MESSAGE

We contend that a graphic's primary or core message constitutes a brief summary of the graphic and captures its major contribution to the overall communicative goal of a multimodal document. Therefore, our first step toward exploiting information graphics within digital libraries has been to develop a methodology for recognizing the message conveyed by a simple bar chart. Our message recognition system assumes as input an XML representation of the graphic that specifies its axes, the bars, their heights, color, labels, and any special annotations, the caption, etc. This is the responsibility of a Visual Extraction Module (VEM)[3]. This module currently handles only electronic images produced with a given set of fonts. 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 in a graphic is much more constrained than typical computer vision problems. The following displays the XML representation produced by the VEM for the graphic shown in Figure 1; the measurements in the XML may not match the bar chart in Figure 1 since it has been resized for display purposes.[3]

```xml
<?xml version="1.0" encoding="UTF-8"?>
<InformationGraphic>
<BarChart BarDirection="vertical">
 <Caption>
   <Content></Content>
 </Caption>
 <MeasurementAxis Length="5.96">
   <Label>GDP Per Capita' 2001 (in thousands)
   </Label>
   <Tickmark>
      <TickLabelled>true</TickLabelled>
      <TickValue>0.00</TickValue>
      <GridLine>true</GridLine>
   </Tickmark>
   ... (xml for other vertical axis labels)
 </MeasurementAxis>
<BarAxis Length="15.66"> </BarAxis>
<Bar>
   <Label>
      <Content>Luxembourg</Content>
      <Color>0</Color>
      <Bold>true</Bold>
   </Label>
   <Color>128</Color>
   <Height>5.68</Height>
   <AxisDistance>1.09</AxisDistance>
   <SightLine>false</SightLine>
   <Value>44.14</Value>
</Bar>
      ...(xml for bar labelled Norway)
 <Bar>
   <Label>
      <Content>U.S.</Content>
      <Color>0</Color>
      <Bold>true</Bold>
   </Label>
   <Color>64</Color>
   <Height>4.48</Height>
   <AxisDistance>5.54</AxisDistance>
   <SightLine>false</SightLine>
   <Value>34.88</Value>
 </Bar>
 ... (xml for the other four bars)
</BarChart>
</InformationGraphic>
```

4.1 A Bayesian Inference System

We view information graphics that appear in popular media as a form of language with a communicative intention.

---

[3]Although hand-coded XML was used to train and test the message recognition system since it was developed in parallel with the VEM, a variety of examples (such as the graphic in Figure 1) have been run through the complete system, with XML produced from the graphic by the VEM and sent to the message recognition system.

Language research has been based on a theory of speech acts[17, 18], where a speech act is the act of making an utterance with the intention of communicating. Speech act theory has posited that a speaker executes a speech act whose intended meaning the listener is expected to deduce. In addition, speech act theory posits that the listener deduces the utterance's intended meaning by reasoning about the communicative signals present in the utterance and the mutual beliefs of speaker and hearer[18, 10].

Our research draws on the AutoBrief project's work on generating information graphics[13, 14, 9]. The AutoBrief group proposed that speech act theory could be extended to the generation of graphical presentations. Given a desired communicative goal, AutoBrief employed a two-phase graphics generation process. First an algorithm mapped communicative goals into a set of *perceptual* and *cognitive* tasks that the graphic should support. By perceptual tasks we mean tasks that can be performed by simply viewing the graphic, such as determining which of two bars is taller in a bar chart; by cognitive tasks we mean tasks that require a mental computation, such as interpolating between labelled values on the dependent axis to compute the exact value of a point in a graphic. A fundamental hypothesis of the AutoBrief project was that graphic designers construct graphics that make important tasks (tasks that the viewer is intended to perform) as easy as possible. Thus the second step in AutoBrief's graph construction process used a constraint satisfaction algorithm to design a graphic that facilitated these important tasks as much as possible, subject to the constraints imposed by competing tasks.

We are inverting the process. Given an information graphic, we want to recognize its intended message or communicative goal by reasoning about the communicative signals present in the graphic. Thus while AutoBrief extended speech act theory to the generation of information graphics, our work extends speech act theory to the understanding and summarization of information graphics.

To recognize an information graphic's message, we make recourse to plan inference techniques that have been used in language understanding to infer the communicative goal of a sentence in a text or an utterance in a dialogue. Following the lead of Charniak and Goldman who used a probabilistic framework to model plan inference for language understanding[2], we have developed a Bayesian network to infer the message conveyed by one type of information graphic, simple bar charts.

The top-level of the Bayesian network represents the 12 categories of messages that we have identified for simple bar charts:

Get-Rank	Rank-of-All
Increasing-Trend	Stable-Trend
Decreasing-Trend	Change-Trend
Maximum	Minimum
Contrast-Point-with-Trend	
Relative-Difference-between-Entities	
Relative-Difference-with-Degree	
Other (Present Data)	

The next level of the network captures the possible instantiations of the general message categories for a given graphic. For example, if a graphic has three bars, then the children of the Get-Rank node might be Get-Rank(_label1, _bar1), Get-Rank(_label2, _bar2), and Get-Rank(_label3, _bar3). We use

```
Get-Rank(_label, _bar)
    1. Perceive-if-bars-are-sorted
    2. Perceive-bar(_label, _bar)
    3. Perceive-rank(_bar)
```
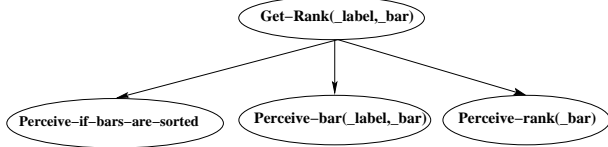
Figure 6: A Sample Operator

Figure 7: A Piece of Network Structure

operators to capture how communicative goals and perceptual tasks can be decomposed into a set of simpler subgoals or subtasks. Figure 6 displays the operator for the goal of Get-Rank(_label,_bar). It states that, to achieve the goal of the viewer getting the rank of the bar associated with a given label in a simple bar chart, the viewer must perceive whether the bars are sorted according to height, then perceive (ie., find) the bar associated with the specified label, and finally perceive the rank of that bar with respect to bar height. Subgoals in operators are either primitive perceptual tasks or they have associated operators that further decompose them. The operators determine the structure of our Bayesian network, in that subgoals in an operator become children of their goal node in the Bayesian network. For example, Figure 7 displays the piece of the Bayesian network produced by the Get-Rank operator.

The entire network is built dynamically for a new graphic. A node capturing the top-level message category is 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 *Get-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 4.2) are used to form the set of suggested low-level perceptual 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 subgoals (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.

4.2 Three Kinds of Communicative Signals

Bayesian networks need evidence for guiding the construction of a hypothesis. In natural language understanding, the observed evidence would include features of the utterance and the context in which it was made. For information graphics, the evidence consists of the communicative signals present in the graphic. We have identified three kinds of communicative signals that appear in simple bar charts:

the relative effort of perceptual tasks that the viewer might perform on the graphic, salient elements in the graphic, and signals from the verbs and adjectives in the caption.

We have adopted the AutoBrief hypothesis that the graphic designer constructs a graphic that makes intended tasks as easy as possible. Thus the relative difficulty of different perceptual tasks serves as a signal about which tasks the viewer was intended to perform in deciphering the graphic's intended message. This correlates with Larkin and Simon's observation[15] that graphics that are informationally equivalent are not necessarily computationally equivalent — that is, it might be possible to infer the same information from two different graphics, but it might be much easier to do so in one graphic than in the other. As a very simple example, consider comparing the height of two bars in a simple bar chart. If the bars are adjacent to one another and significantly different in height, then the task will be easy; however, if the two bars do not differ much in height, are not annotated with their values, and are separated by many intervening bars, then the task will be much more difficult. Similarly, if the bars in a bar chart are ordered by height, then perceiving the rank of a particular bar (see subgoal 3 in Figure 6) will be far easier than if the bars are unsorted. We constructed a set of rules for estimating the effort involved in performing different perceptual tasks in a simple bar chart. These rules have been validated by eyetracking experiments and are presented in [5].

The second kind of communicative signal is salience. An entity in a graphic can be made salient in a variety of ways. For example, its bar might be colored differently from other bars in the graphic, as is the bar for the U.S. in Figure 1, or the bar's label might be mentioned in the caption. To determine whether the labels on any bars appear as part of the caption, a caption processing module[4] extracts nouns from the caption found in the XML representation of the graphic using a part-of-speech tagger, matches the nouns against labels on the bars, and augments the XML representation to indicate those bars whose label matches a noun in the caption. By analyzing the augmented XML representation, the message recognition module can determine which entities have been made salient by virtue of their special color, special annotation, reference in a caption, etc.

The third kind of communicative signal is the presence in the caption of a verb or adjective that suggests a particular category of message.[4] For example, the verb "lag" might suggest a message about some entity being a minimum or about some entity falling behind some other entity in value. Using WordNet and a thesaurus, we identified classes of verbs and adjectives that are similar in meaning and which might suggest some general category of message. Our caption processing module uses a part-of-speech tagger and a stemmer to identify the presence of one of our identified verb or adjective classes in the caption, and the XML representation is augmented to reflect this.

The identified communicative signals must be entered into the Bayesian network as evidence that can influence the sys-

---

[4]In [4] we present 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. Thus we have chosen to perform a shallow analysis of captions that extracts communicative signals but does not attempt to *understand* the caption.

tem's hypothesis about the graphic's message. Nodes indicating the effort involved in performing a particular perceptual task and nodes capturing whether a parameter of a particular perceptual task is salient in the graphic are attached to the low level perceptual task nodes in the Bayesian network. Nodes reflecting the presence or absence of one of our identified verb and adjective classes are attached to the message category node that appears at the top of the Bayesian network, since verbs and adjectives serve to signal a general category of message.

4.3   Implementation and Evaluation

Associated with each child node in a Bayesian network is a conditional probability table that captures the probability of each value of a child node given the value of the parent node. In our network, the parent nodes are always goals or tasks, and thus the value of the parent node is always that it either is (or is not) part of what the viewer is intended to do in recognizing the graphic's message. Our conditional probability tables are computed from our corpus of graphics.

We implemented our Bayesian network for hypothesizing an information graphic's message using the Netica software tools[16] for Bayesian reasoning. The current system is limited to simple bar charts, but we believe that our methodology will be applicable to more complex graphics.

Using a corpus of 110 simple bar charts that had previously been annotated with their message, we evaluated our approach using leave-one-out cross validation in which 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. For each test graphic, the system was credited with success if its top-rated hypothesis matched the message assigned to the graphic by the human coders and the probability that the system assigned to the hypothesis exceeded 50%. The system's overall success rate was 79.1%. This can be compared with a baseline of the most common message category, *rising trend*, which occurred 23.6% of the time; however, our system's task was more difficult than just selecting the message category, since it also had to determine the instantiation of the parameters. For example, rather than just hypothesizing a Change-Trend message category, our system had to identify the two contrasting trends and the point at which the trend changed.

In order to determine whether the intentions being inferred by our system would meet the approval of users, we performed a second evaluation. This evaluation consisted of a survey in which we asked human subjects to examine a set of bar charts and rate a posited primary intention for each bar chart. Seventeen undergraduate students took part in the survey, which contained twenty-seven bar charts. For each bar chart, the participants were asked to answer a set of questions, including their level of agreement with the stated primary intended message of the graphic (strongly agree, agree, not sure, disagree, strongly disagree), and follow-up questions for cases where they did not agree with the stated message. For twenty of the twenty-seven bar charts, the statement matched the hypothesized intention of our system. For the remaining seven bar charts, we proposed messages that did not match the hypotheses of our system. When calculating the results of the survey, we assigned numeric values to the scale in the first question. An answer of "strongly agree" was counted as a four and "strongly disagree" a zero. For the twenty graphs where the proposed

The sound of sales
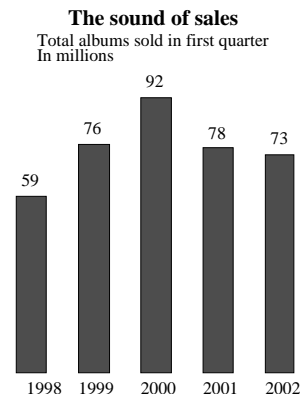Total albums sold in first quarter
In millions

Figure 8: A Variation of a Graphic from USA Today

message matched the output of our system, we expected the majority of participants to agree with the proposed message. This was, indeed, the case, given that the average agreement rate for the twenty graphs was 3.33 (a value between "agree" and "strongly agree" on our scale) with a standard deviation of 1.02 and a 95% confidence interval of .108. On the other hand, for the seven graphs where the proposed message did not match the output of our system, the average agreement rate was only 1.19 with a standard deviation of 1.46 and a 95% confidence interval of .261. The results of this survey demonstrate 1) that viewers of information graphics do tend to form consensus regarding the intended message of the graphic, and 2) that utilizing the intentions recognized by our system as the basis of a summary of a graphic should produce summaries which would be satisfactory to a majority of users.

5.   EXAMPLES AND RELEVANCE TO DIGITAL LIBRARIES

Our current research has focused on developing a methodology for recognizing the primary message conveyed by an information graphic. This work has several applications within digital libraries. The first is to provide a representation of the core message of an information graphic for indexing and retrieval of the graphic. Consider the graphic in Figure 8. A graphic stored using the caption *"The sound of sales"* or even the dependent axis label *"Total albums sold in first quarter in millions"* could have a variety of messages, such as conveying the distribution of record album sales among major recording studios or contrasting record album sales by rock artists with those by jazz singers. Our system hypothesizes that the graphic in Figure 8 is conveying a changing trend in record album sales, with sales increasing from 1998 to 2000 and then decreasing from 2000 to 2002. The logical representation of this message is

Change-trend(increasing, 1998, 2000, decreasing, 2002,
        total albums sold in first quarter in millions)

Indexing the graphic using this message would enable retrieval of the graphic if one wanted to know about album sales during the period 1998-2002, about trends in album sales, or if one wanted to know about trends that have changed between 1998 and 2002. Furthermore, the probability assigned to the graphic's message reflects the sys-
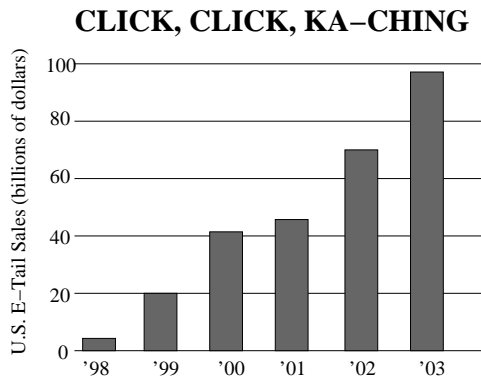
CLICK, CLICK, KA–CHING
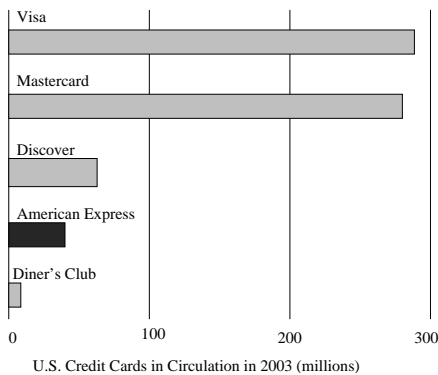
Figure 9: Graphic from Business Week

Figure 10: Graphic from Business Week

tem's confidence in its hypothesis and thus how clearly the graphic conveys the posited message. Thus the associated probability might be used as one criteria in ranking alternative graphics that are suggested for retrieval.

The author of a multimodal document ostensibly viewed certain communicative goals as important enough to warrant expending effort on designing graphics to convey them. This suggests that the messages conveyed by information graphics should be taken into account in summarizing the document. For example, consider the information graphic in Figure 9 which appeared in a *Business Week* article entitled "THE START OF A DOT-COMBACK?". Our system hypothesizes that the graphic's message is that there has been a rising trend in U.S. E-tail sales between '98 and '03. The graphic's message is not conveyed in the article, but the message represents a communicative goal which, together with the discourse purposes of the text segments, contributes to the overall discourse purpose of the article — namely that dot-com companies are now focusing on results and that prospects for infusions of venture capital are improving. Thus the graphic plays a central role in the article, and its conveyed message should be captured in the document's summary.

The third use of the messages produced by our system is as an indication of the focus or topic of a document. Consider the graphic shown in Figure 10. The highlighting of the bar associated with American Express is an important

communicative signal in this graphic; without the highlighting, the message inferred for the graphic by our system is that the graphic is conveying the relative rank of the different credit card companies in terms of U.S. credit cards in circulation in 2003. But with American Express highlighted, our system hypothesizes that the graphic's message is that American Express ranks fourth among the five credit card companies shown with respect to U.S. credit cards in circulation in 2003. Consequently, although the text of the article talks about a variety of credit card companies (for example, the text states that AmEx charges merchants higher fees than do Visa and MasterCard), the focus of the graphic is clearly on American Express and thus suggests that the article is about American Express. This is in fact the case. We hypothesize that the messages conveyed by information graphics in a multimodal document can be helpful in determining the focus of an article and thus useful in constructing a good summary.

Our future work includes not only extending our message recognition system for graph summarization to more complex graphics, such as line graphs and grouped bar charts, but also to investigating 1) the use of our inferred messages in indexing and retrieving information graphics in a digital library and 2) the utilization of the messages conveyed by information graphics in constructing good summaries of multimodal documents.

6. RELATED WORK

Within information retrieval research, the work most closely related to our project involves the indexing and retrieval of images. Bradshaw [1] notes that the work on image retrieval has progressed from systems that retrieve images based on low-level features such as color, texture and shape ([6, 21, 19, 12], among others), to systems which attempt to classify and reason about the semantics of the images being processed. These include systems that attempt to classify images according to attributes such as indoor/outdoor, city/landscape and man-made/artificial, and they often present these classifications as occurring along semantic axes (see [22], for example) or according to probabilistic labels (see [23, 1] for examples).

Srihari, Zhang and Rao have examined the possibility of combining text-based indexing techniques for the caption and any accompanying text with image-based techniques [20]. They argue that text-based methods alone are ineffective, and provide the example of a search for pictures of Clinton and Gore, which resulted in 941 images. After eliminating graphics and spurious images, they had 547 images remaining, but upon manual inspection, they discovered that only 76 of those images actually contained pictures of Clinton or Gore! They demonstrate, however, that when combined with image-based retrieval techniques, the collateral text can provide a rich source of evidence for improving the information retrieval process. Their work is similar to ours in that they are attempting to use multiple, disparate sources of evidence.

However, the work done in image retrieval is concerned with the semantics of images (what is physically represented in the image and the relationships between those objects, such as "Clinton at the White House" or "a bee on a flower"), whereas we are concerned with recognizing the communicative goal or discourse purpose of an information graphic — that is, the message that the graphic is intended to convey.

Very little research has been concerned with summarizing information graphics. Yu et. al.[24] used pattern recognition techniques to summarize interesting features of time series data generated by a gas turbine engine, but this was automatically generated data without a communicative intention. Futrelle and Nikolakis[8] developed a constraint grammar formalism for parsing vector-based visual displays and producing structured representations of the elements comprising the display. The goal of Futrelle's current research is to produce a summary graphic that captures the content of one or more graphics from a document[7]. However, the end result will itself be a graphic.

7. CONCLUSION

Information graphics are an important component of many documents, yet little consideration has been given to making them available in a digital library or to taking them into account when summarizing a multimodal document. This paper has presented the results of a corpus study which showed that the communicative goal or message of an information graphic is typically not repeated in the article's text. It has also presented our implemented and evaluated Bayesian network for recognizing the message conveyed by an information graphic. This message can form the core of a brief summary of the graphic for use in indexing the graphic in a digital library and for taking constituent information graphics into account when summarizing a multimodal document. To our knowledge, our research is the first to address the problem of enabling digital libraries to exploit the rich resource of information graphics.

8. ACKNOWLEDGEMENTS

9. REFERENCES

[1] B. Bradshaw. Semantic based image retrieval: a probabilistic approach. In *Proc. 8th ACM Int. Conf. on Multimedia*, pp. 167–176, 2000.

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

[3] D. Chester and S. Elzer. Getting computers to see information graphics so users do not have to. In *Proc. 15th Int. Symposium on Methodologies for Intelligent Systems*, pp. 660–668, 2005.

[4] S. Elzer, S. Carberry, D. Chester, S. Demir, N. Green, I. Zukerman, and K. Trnka. Exploring and exploiting the limited utility of captions in recognizing intention in information graphics. In *Proc. 43rd Meeting of Association for Computational Linguistics*, pp. 223–230, 2005.

[5] S. Elzer, N. Green, S. Carberry, and J. Hoffman. Incorporating perceptual task effort into the recognition of intention in information graphics. In *Proc. of Third Int. Conf. on Theory and Application of Diagrams*, pp. 255–270, 2004.

[6] M. Flickner, D. Petkovic, D. Steele, P. Yanker, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, and D. Lee. Query by image and video: The qbic system. *Computer*, 28(9):23–32, 1995.

[7] R. Futrelle. Summarization of diagrams in documents. In I. Mani and M. Maybury, editors, *Advances in Automated Text Summarization*. MIT Press, 1999.

[8] R. Futrelle and N. Nikolakis. Efficient analysis of complex diagrams using constraint-based parsing. In *Proc. of Third Int. Conf. on Document Analysis and Recognition*, 1995.

[9] N. Green, G. Carenini, S. Kerpedjiev, J. Mattis, J. Moore, and S. Roth. Atuobrief: An experimental system for the automatic generation of briefings in integrated text and graphics. *Int. Journal of Human-Computer Studies*, 61(1):32–70, 2004.

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

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

[12] A. Gupta and R. Jain. Visual information retrieval. *Communications of the ACM*, 40(5):71–79, 1997.

[13] S. Kerpedjiev, G. Carenini, N. Green, J. Moore, and S. Roth. Saying it in graphics: From intentions to visualizations. In *Proc. of IEEE Symposium on Information Visualization*, pages 97–101, 1998.

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

[15] J. Larkin and H. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11:65–99, 1987.

[16] N. S. C. Netica, 2005.

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

[18] J. R. Searle. Indirect Speech Acts. In P. Cole and J. Morgan, editors, *Syntax and Semantics: Speech Acts*, volume 3, pages 59–82. Academic Press, Inc., New York, New York, 1975.

[19] J. R. Smith and S. fu Chang. Querying by color regions using VisualSEEk content-based visual query system. In M. T. Maybury, editor, *Intelligent Multimedia Information Retrieval*, pages 23–41. AAAI Press/MIT Press, 1997.

[20] R. K. Srihari, Z. Zhang, and A. Rao. Intelligent indexing and semantic retrieval of multimodal documents. *Information Retrieval*, 2(2):1–37, 2000.

[21] M. J. Swain. Color indexing. *Int. Journal of Computer Vision*, 7(1):11–32, 1991.

[22] A. B. Torralba and A. Oliva. Semantic organization of scenes using discriminant structural templates. In *Proc. of Int. Conf. on Computer Vision*, 1999.

[23] A. Vailaya, M. Figueiredo, A. K. Jain, and H.-J. Zhang. Image classification for content-based indexing. In *Proc. of IEEE Conf. on Multimedia Computing and Systems*, pages 518–523, 1999.

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