Improving the Accessibility of Line Graphs in Multimodal Documents

Charles F. Greenbacker **Peng Wu** **Sandra Carberry** **Kathleen F. McCoy**
Stephanie Elzer* **David D. McDonald†** **Daniel Chester** **Seniz Demir‡**
Dept. of Computer & Information Sciences, University of Delaware, USA
[charlieg|pwu|carberry|mccoy|chester]@cis.udel.edu
*Dept. of Computer Science, Millersville University, USA elzer@cs.millersville.edu
†SIFT LLC., Boston, Massachusetts, USA dmcdonald@sift.info
‡TÜBİTAK BİLGEM, Gebze, Kocaeli, Turkey senizd@uekae.tubitak.gov.tr

Abstract

This paper describes our work on improving access to the content of multimodal documents containing line graphs in popular media for people with visual impairments. We provide an overview of our implemented system, including our method for recognizing and conveying the intended message of a line graph. The textual description of the graphic generated by our system is presented at the most relevant point in the document. We also describe ongoing work into obtaining additional propositions that elaborate on the intended message, and examine the potential benefits of analyzing the text and graphical content together in order to extend our system to produce summaries of entire multimodal documents.

1 Introduction

Individuals with visual impairments have difficulty accessing the information contained in multimodal documents. Although screen-reading software can render the text of the document as speech, the graphical content is largely inaccessible. Here we consider information graphics (e.g., bar charts, line graphs) often found in popular media sources such as *Time* magazine, *Businessweek*, and *USA Today*. These graphics are typically intended to convey a message that is an important part of the overall story, yet this message is generally not repeated in the article text (Carberry et al., 2006). People who are unable to see and assimilate the graphical material will be left with only partial information.

While some work has addressed the accessibility of scientific graphics through alternative means like touch or sound (see Section 7), such graphs are designed for an audience of experts trained to use them for data visualization. In contrast, graphs in popular media are constructed to make a point which should be obvious without complicated scientific reasoning. We are thus interested in generating a textual presentation of the content of graphs in popular media. Other research has focused on textual descriptions (e.g., Ferres et al. (2007)); however in that work the same information is included in the textual summary for each instance of a graph type (i.e., all summaries of line graphs contain the same sorts of information), and the summary does not attempt to present the overall intended message of the graph.

SIGHT (Demir et al., 2008; Elzer et al., 2011) is a natural language system whose overall goal is providing blind users with interactive access to multimodal documents from electronically-available popular media sources. To date, the SIGHT project has concentrated on simple bar charts. Its user interface is implemented as a browser helper object within Internet Explorer that works with the JAWS screen reader. When the system detects a bar chart in a document being read by the user, it prompts the user to use keystrokes to request a brief summary of the graphic capturing its primary contribution to the overall communicative goal of the document. The summary text can either be read to the user with JAWS or read by the user with a screen magnifier tool. The interface also enables the user to request further information about the graphic, if desired.

However, SIGHT is limited to bar charts only. In this work, we follow the methodology put forth by SIGHT, but investigate producing a summary of

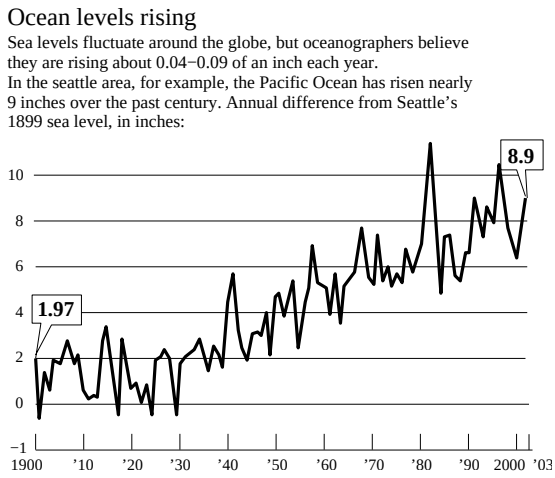

Figure 1: From "Worry flows from Arctic ice to tropical waters" in USA Today, May 31, 2006.

line graphs. Line graphs have different discourse goals and communicative signals than bar charts,[1] and thus require significantly different processing. In addition, our work addresses the issue of coherent placement of a graphic's summary when reading the text to the user and considers the summarization of entire documents — not just their graphics.

2 Message Recognition for Line Graphs

This section provides an overview of our implemented method for identifying the intended message of a line graph. In processing a line graph, a visual extraction module first analyzes the image file and produces an XML representation which fully specifies the graphic (including the beginning and ending points of each segment, any annotations on points, axis labels, the caption, etc.). To identify the intended message of a line graph consisting of many short, jagged segments, we must generalize it into a sequence of visually-distinguishable trends. This is performed by a graph segmentation module which uses a support vector machine and a variety of attributes (including statistical tests) to produce a model that transforms the graphic into a sequence of straight lines representing visually-distinguishable trends. For example, the line graph in Figure 1 is divided into a stable trend from 1900 to 1930 and a rising trend from 1930 to 2003. Similarly, the line graph in Figure 2 is divided into a rising trend from

[1]Bar charts present data as discrete bars and are often used to compare entities, while line graphs contain continuous data series and are designed to portray longer trend relationships.

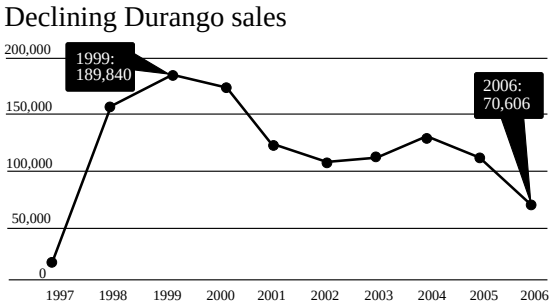

Figure 2: From "Chrysler: Plant had $800 million impact" in The (Wilmington) News Journal, Feb 15, 2007.

1997 to 1999 and a falling trend from 1999 to 2006.

In analyzing a corpus of around 100 line graphs collected from several popular media sources, we identified 10 intended message categories (including *rising-trend*, *change-trend*, *change-trend-return*, and *big-jump*, etc.), that seem to capture the kinds of high-level messages conveyed by line graphs. A suggestion generation module uses the sequence of trends identified in the line graph to construct all of its possible candidate messages in these message categories. For example, if a graph contains three trends, several candidate messages are constructed, including two change-trend messages (one for each adjacent pair of trends), a change-trend-return message if the first and third trends are of the same type (rising, falling, or stable), as well as a rising, falling, or stable trend message for each individual trend.

Next, various communicative signals are extracted from the graphic, including visual features (such as a point annotated with its value) that draw attention to a particular part of the line graph, and linguistic clues (such as the presence of certain words in the caption) that suggest a particular intended message category. Figure 2 contains several such signals, including two annotated points and the word *declining* in its caption. Next, a Bayesian network is built to estimate the probability of the candidate messages; the extracted communicative signals serve as evidence for or against each candidate message. For Figure 2, our system produces change-trend(1997, rise, 1999, fall, 2006) as the logical representation of the most probable intended message. Since the dependent axis is often not explicitly labeled, a series of heuristics are used to identify an appropriate referent, which we term the *measurement axis descriptor*. In Figure 2, the measurement axis descriptor is identified as *durango sales*. The

intended message and measurement axis descriptor are then passed to a realization component which uses FUF/SURGE (Elhadad and Robin, 1996) to generate the following initial description:

> *This graphic conveys a changing trend in durango sales, rising from 1997 to 1999 and then falling to 2006.*

3 Identifying a Relevant Paragraph

In presenting a multimodal document to a user via a screen reader, if the author does not specify a reading order in the accessibility preferences, it is not entirely clear where the description of the graphical content should be given. The text of scientific articles normally makes explicit references to any graphs contained in the document; in this case, it makes sense to insert the graphical description alongside the first such reference. However, popular media articles rarely contain explicit references to graphics. We hypothesize that describing the graphical content together with the most relevant portion of the article text will result in a more coherent presentation. Results of an experiment described in Section 3.3 suggest the paragraph which is geographically closest to the graphic is very often not relevant. Thus, our task becomes identifying the portion of the text that is most relevant to the graph.

We have developed a method for identifying the most relevant paragraph by measuring the similarity between the graphic's textual components and the content of each individual paragraph in the document. An information graphic's textual components may consist of a title, caption, and any additional descriptions it contains (e.g., the five lines of text in Figure 1 beneath the caption *Ocean levels rising*). An initial method (P-KL) based on KL divergence measures the similarity between a paragraph and the graphic's textual component; a second method (P-KLA) is an extension of the first that incorporates an augmented version of the textual component.

3.1 Method P-KL: KL Divergence

Kullback-Leibler (KL) divergence (Kullback, 1968) is widely used to measure the similarity between two language models. It can be expressed as:

$$D_{KL}(p||q) = \sum_{i \in V} p(i) log \frac{p(i)}{q(i)}$$

where i is the index of a word in vocabulary V, and p and q are two distributions of words. Liu et al. (Liu and Croft, 2002) applied KL divergence to text passages in order to improve the accuracy of document retrieval. For our task, p is a smoothed word distribution built from the line graph's textual component, and q is another smoothed word distribution built from a paragraph in the article text. Smoothing addresses the problem of zero occurrences of a word in the distributions. We rank the paragraphs by their KL divergence scores from lowest to highest, since lower scores indicate a higher similarity.

3.2 Method P-KLA: Using Augmented Text

In analyzing paragraphs relevant to the graphics, we realized that they included words that were germane to describing information graphics in general, but not related to the domains of individual graphs. This led us to build a set of "expansion words" that tend to appear in paragraphs relevant to information graphics. If we could identify domain-independent terms that were correlated with information graphics in general, these expansion words could then be added to the textual component of a graphic when measuring its similarity to a paragraph in the article text.

We constructed the expansion word set using an iterative process. The first step is to use P-KL to identify m pseudo-relevant paragraphs in the corresponding document for each graphic in the training set (the current implementation uses $m = 3$). This is similar to pseudo-relevance feedback used in IR (Zhai, 2008), except only a single query is used in the IR application, whereas we consider many pairs of graphics and documents to obtain an expansion set applicable to any subsequent information graphic. Given n graphics in the training set, we identify (up to) $m * n$ relevant paragraphs.

The second step is to extract a set of words related to information graphics from these $m * n$ paragraphs. We assume the collection of pseudo-relevant paragraphs was generated by two models, one producing words relevant to the information graphics and another producing words relevant to the topics of the individual documents. Let W_g represent the word frequency vector yielding words relevant to the graphics, W_a represent the word frequency vector yielding words relevant to the document topics, and W_p represent the word frequency vector of the

pseudo-relevant paragraphs. We compute W_p from the pseudo-relevant paragraphs themselves, and we estimate W_a using the word frequencies from the article text in the documents. Finally, we compute W_g by filtering-out the components of W_a from W_p. This process is related to the work by Widdows (2003) on orthogonal negation of vector spaces.

The task can be formulated as follows:

1. $W_p = \alpha W_a + \beta W_g$ where $\alpha > 0$ and $\beta > 0$, which means the word frequency vector for the pseudo-relevant paragraphs is a linear combination of the background (topic) word frequency vector and the graphic word vector.
2. $< W_a, W_g >= 0$ which means the background word vector is orthogonal to the graph description word vector, under the assumption that the graph description word vector is independent of the background word vector and that these two share minimal information.
3. W_g is assumed to be a unit vector, since we are only interested in the relative rank of the word frequencies, not their actual values.

Solving the above equations, we obtain:

$$\alpha = \frac{< W_p, W_a >}{< W_a, W_a >}$$

$$W_g = \text{normalized}\left(W_p - \frac{< W_p, W_a >}{< W_a, W_a >} \cdot W_a\right)$$

After computing W_g, we use WordNet to filter-out words having a predominant sense other than *verb* or *adjective*, under the assumption that nouns will be mainly relevant to the domains or topics of the graphs (and are thus "noise") whereas we want a general set of words (e.g., *"increasing"*) that are typically used when describing the data in any graph. As a rough estimate of whether a word is predominantly a verb or adjective, we determine whether there are more verb and adjective senses of the word in WordNet than there are noun senses.

Next, we rank the words in the filtered W_g according to frequency and select the k most frequent as our expansion word list (we used $k = 25$ in our experiments). The two steps (identifying $m*n$ pseudo-relevant paragraphs and then extracting a word list of size k to expand the graphics' textual components) are applied iteratively until convergence occurs or minimal changes are observed between iterations.

In addition, parameters of the intended message that represent points on the x-axis capture domain-specific content of the graphic's communicative goal. For example, the intended message of the line graph in Figure 1 conveys a changing trend from 1900 to 2003 with the change occurring in 1930. To help identify relevant paragraphs mentioning these years, we also add these parameters of the intended message to the augmented word list.

The result of this process is the final expansion word list used in method P-KLA. Because the textual component may be even shorter than the expansion word list, we do not add a word from the expansion word list to the textual component unless the paragraph being compared also contains this word.

3.3 Results of P-KL and P-KLA

334 training graphs with their accompanying articles were used to build the expansion word set. A separate set of 66 test graphs and articles was analyzed by two human annotators who identified the paragraphs in each document that were most relevant to its associated information graphic, ranking them in terms of relevance. On average, annotator 1 selected 2.00 paragraphs and annotator 2 selected 1.71 paragraphs. The annotators agreed on the top ranked paragraph for only 63.6% of the graphs. Considering the agreement by chance, we can calculate the kappa statistic as 0.594. This fact shows that the most relevant paragraph is not necessarily obvious and multiple plausible options may exist.

We applied both P-KL and P-KLA to the test set, with each method producing a list of the paragraphs ranked by relevance. Since our goal is to provide the summary of the graphic at a suitable point in the article text, two evaluation criteria are appropriate:

1. TOP: the method's success rate in selecting *the most relevant paragraph*, measured as how often it chooses the paragraph ranked highest by either of the annotators
2. COVERED: the method's success rate in selecting *a relevant paragraph*, measured as how often it chooses one of the relevant paragraphs identified by the annotators

Table 1 provides the success rates of both of our methods for the TOP and COVERED criteria, along with a simple baseline that selected the paragraph

geographically-closest to the graphic. These results show that both methods outperform the baseline, and that P-KLA further improves on P-KL. P-KLA selects the best paragraph in 60.6% of test cases, and selects a relevant paragraph in 71.2% of the cases. For both TOP and COVERED, P-KLA nearly doubles the baseline success rate. The improvement of P-KLA over P-KL suggests that our expansion set successfully adds salient words to the textual component. A one-sided Z-test for proportion based on binomial distribution is shown in Table 1 and indicates that the improvements of P-KL over the baseline and P-KLA over P-KL are statistically-significant at the 0.05 level across both criteria. The Z-test is calculated as:

$$\frac{p - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

where p_0 is the lower result and p is the improved result. The null hypothesis is $H_0 : p = p_0$ and the alternative hypothesis is $H_1 : p > p_0$.

3.4 Using relevant paragraph identification to improve the accessibility of line graphs

Our system improves on SIGHT by using method P-KLA to identify the paragraph that is most relevant to an information graphic. When this paragraph is encountered, the user is asked whether he or she would like to access the content of the graphic. For example, our system identifies the following paragraph as most relevant to Figure 2:

> *Doing so likely would require the company to bring in a new model. Sales of the Durango and other gas-guzzling SUVs have slumped in recent years as prices at the pump spiked.*

In contrast, the geographically-closest paragraph has little relevance to the graphic:

> *"We have three years to prove to them we need to stay open," said Sam Latham, president of the AFL-CIO in Delaware, who retired from Chrysler after 39 years.*

4 Identifying Additional Propositions

After the intended message has been identified, the system next looks to identify elaborative informational propositions that are salient in the graphic. These additional propositions expand on the initial description of the graph by filling-in details about the knowledge being conveyed (e.g., noteworthy points, properties of trends, visual features) in order to round-out a summary of the graphic.

We collected a corpus of 965 human-written summaries for 23 different line graphs to discover which propositions were deemed most salient under varied conditions.[2] Subjects received an initial description of the graph's intended message, and were asked to write additional sentences capturing the most important information conveyed by the graph. The propositions appearing in each summary were manually coded by an annotator to determine which were most prevalent. From this data, we developed rules to identify important propositions in new graphs. The rules assign weights to propositions indicating their importance, and the weights can be compared to decide which propositions to include in a summary.

Three types of rules were built. Type-1 (message category-only) rules were created when a plurality of summaries for all graphs having a given intended message contained the same proposition (e.g., *provide the final value for all rising-trend and falling-trend graphs*). Weights for type-1 rules were based on the frequency with which the proposition appeared in summaries for graphs in this category.

Type-2 (visual feature-only) rules were built when there was a correlation between a visual feature and the use of a proposition describing that feature, regardless of the graph's message category (e.g., *mention whether the graph is highly volatile*). Type-2 rule weights are a function of the covariance between the magnitude of the visual feature (e.g., degree of volatility) and the proportion of summaries mentioning this proposition for each graph.

For propositions associated with visual features linked to a particular message category (e.g., *describe the trend immediately following a big-jump or big-fall when it terminates prior to the end of the graph*), we constructed Type-3 (message category + visual feature) rules. Type-3 weights were calculated just like Type-2 weights, except the graphs were limited to the given category.

As an example of identifying additional proposi-

[2] This corpus is described in greater detail by Greenbacker et al. (2011) and is available at www.cis.udel.edu/~mccoy/corpora

	closest	P-KL	significance level over closest	P-KLA	significance level over P-KL
TOP	0.272	0.469	$(z = 3.5966, p < 0.01)$	0.606	$(z = 2.2303, p < 0.025)$
COVERED	0.378	0.606	$(z = 3.8200, p < 0.01)$	0.712	$(z = 1.7624, p < 0.05)$

Table 1: Success rates for baseline method ("closest"), P-KL, and P-KLA using the TOP and COVERED criteria.

tions, consider Figures 1 and 2. Both line graphs belong to the same intended message category: change-trend. However, the graph in Figure 1 is far more volatile than Figure 2, and thus it is likely that we would want to mention this proposition (i.e., "*the graph shows a high degree of volatility...*") in a summary of Figure 1. By finding the covariance between the visual feature (i.e., volatility) and the frequency with which a corresponding proposition was annotated in the corpus summaries, a Type-2 rule assigns a weight to this proposition based on the magnitude of the visual feature. Thus, the volatility proposition will be weighted strongly for Figure 1, and will likely be selected to appear in the initial summary, while the weight for Figure 2 will be very low.

5 Integrating Text and Graphics

Until now, our system has only produced summaries for the graphical content of multimodal documents. However, a user might prefer a summary of the entire document. Possible use cases include examining this summary to decide whether to invest the time required to read a lengthy article with a screen reader, or simply addressing the common problem of having too much material to review in too little time (i.e., *information overload*). We are developing a system extension that will allow users to request summaries of arbitrary length that cover both the text and graphical content of a multimodal document.

Graphics in popular media convey a message that is generally not repeated in the article text. For example, the March 3, 2003 issue of *Newsweek* contained an article entitled, "The Black Gender Gap," which described the professional achievements of black women. It included a line graph (Figure 3) showing that the historical gap in income equality between white women and black women had been closed, yet this important message appears nowhere in the article text. Other work in multimodal document summarization has relied on image captions and direct references to the graphic in the text (Bhatia et al., 2009); however, these textual elements do not necessarily capture the message conveyed by information graphics in popular media. Thus, the user may miss out on an essential component of the overall communicative goal of the document if the summary covers only material presented in the text.

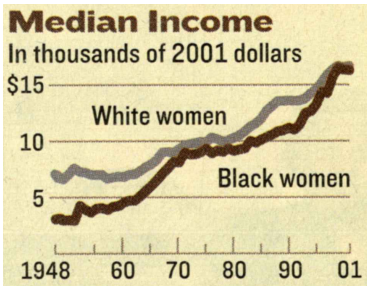

Figure 3: From "The Black Gender Gap" in Newsweek, Mar 3, 2003.

One approach to producing a summary of the entire multimodal document might be to "concatenate" a traditional extraction-based summary of the text (Kupiec et al., 1995; Witbrock and Mittal, 1999) with the description generated for the graphics by our existing system. The summary of the graphical content could be simply inserted wherever it is deemed most relevant in the text summary. However, such an approach would overlook the relationships and interactions between the text and graphical content. The information graphics may make certain concepts mentioned in the text more salient, and vice versa. Unless we consider the contributions of both the text and graphics together during the content selection phase, the most important information might not appear in the summary of the document.

Instead, we must produce a summary that *integrates* the content conveyed by the text and graphics. We contend that this integration must occur at the *semantic* level if it is to take into account the influence of the graphic's content on the salience of concepts in the text and vice versa. Our tack is to first build a single semantic model of the concepts expressed in both the article text and information graphics, and then use this model as the basis for generating an abstractive summary of the multimodal document.

Drawing from a model of the semantic content of the document, we select as many or as few concepts as we wish, at any level of detail, to produce summaries of arbitrary length. This will permit the user to request a quick overview in order to decide whether to read the original document, or a more comprehensive synopsis to obtain the most important content without having to read the entire article.

5.1 Semantic Modeling of Multimodal Documents

Content gathered from the article text by a semantic parser and from the information graphics by our graph understanding system is combined into a single semantic model based on typed, structured objects organized under a foundational ontology (McDonald, 2000a). For the semantic parsing of text, we use Sparser (McDonald, 1992), a bottom-up, phrase-structure-based chart parser, optimized for semantic grammars and partial parsing.[3] Using a built-in model of core English grammar plus domain-specific grammars, Sparser extracts information from the text and produces categorized objects as a semantic representation (McDonald, 2000b). The intended message and salient additional propositions identified by our system for the information graphics are decomposed and added to the model constructed by Sparser.[4]

Model entries contain slots for attributes in the concept category's ontology definition (fillable by other concepts or symbols), the original phrasings mentioning this concept in the text (represented as parameterized synchronous TAG derivation trees), and markers recording document structure (i.e., where in the text [including title, headings, etc.] or graphic the concept appeared). Figure 4 shows some of the information contained in a small portion of the semantic model built for an article entitled "Will Medtronic's Pulse Quicken?" from the May 29, 2006 edition of *Businessweek* magazine[5], which included a line graph. Nodes correspond to concepts and edges denote relationships between concepts; dashed lines indicate links to concepts not shown in this figure. Nodes are labelled with the name of the conceptual category they instantiate, and a number to distinguish between individuals. The middle of each box displays the attributes of the concept, while the bottom portion shows some of the original text phrasings. Angle brackets (<>) note references to other concepts, and hash marks (#) indicate a symbol that has not been instantiated as a concept.

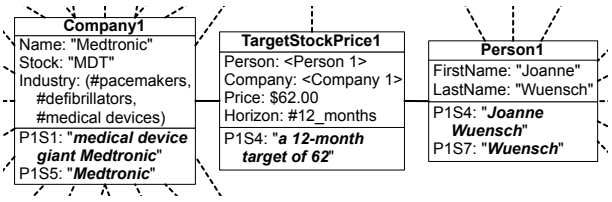

Figure 4: Detail of model for *Businessweek* article.

5.2 Rating Content in Semantic Models

The model is then rated to determine which items are most salient. The concepts conveying the most information and having the most connections to other important concepts in the model are the ones that should be chosen for the summary. The importance of each concept is rated according to a measure of *information density* (ID) involving several factors:[6]

Saturation Level Completeness of attributes in model entry: a concept's filled-in slots (f) vs. its total slots (s), and the importance of the concepts (c_i) filling those slots: $\frac{f}{s} * log(s) * \sum_{i=1}^{f} ID(c_i)$

Connectedness Number of connections (n) with other concepts (c_j), and the importance of these connected concepts: $\sum_{j=1}^{n} ID(c_j)$

Frequency Number of observed phrasings (e) realizing the concept in text of the current document

Prominence in Text Prominence based on document structure (W_D) and rhetorical devices (W_R)

Graph Salience Salience assessed by the graph understanding system (W_G) – only applies to concepts appearing in the graphics

[3] https://github.com/charlieg/Sparser

[4] Although the framework is general enough to accommodate any modality (e.g., images, video) given suitable semantic analysis tools, our prototype implementation focuses on bar charts and line graphs analyzed by SIGHT.

[5] http://www.businessweek.com/magazine/content/06_22/b3986120.htm

[6] The first three factors are similar to the dominant slot fillers, connectivity patterns, and frequency criteria described by Reimer and Hahn (1988).

Saturation corresponds to the completeness of the concept in the model. The more attribute slots that are filled, the more we know about a particular concept instance. However, this measure is highly sensitive to the degree of detail provided in the semantic grammar and ontology class definition (whether created by hand or automatically). A concept having two slots, both of which are filled-out, is not necessarily more important than a concept with only 12 of its 15 slots filled. The more important a concept category is in a given domain, the more detailed its ontology class definition will likely be. Thus, we can assume that a concept definition having a dozen or more slots is, broadly speaking, more important in the domain than a less well-defined concept having only one or two slots. This insight is the basis of a normalization factor ($log(s)$) used in ID.

Saturation differs somewhat from repetition in that it attempts to measure the amount of information associated with a concept, rather than simply the number of times a concept is mentioned in the text. For example, a news article about a proposed law might mention "Washington" several times, but the fact that the debate took place in Washington, D.C. is unlikely to be an important part of the article. However, the key provisions of the bill, which may individually be mentioned only once, are likely more important as a greater amount of detail is provided concerning them. Simple repetition is not *necessarily* indicative of the importance of a concept, but if a large amount of information is provided for a given concept, it is safe to assume the concept is important in the context of that document.

Document structure (W_D) is another important clue in determining which elements of a text are important enough to include in a summary (Marcu, 1997). If a concept is featured prominently in the title, or appears in the first or final paragraphs, it is likely more important than a concept buried in the middle of the document. Importance is also affected by certain rhetorical devices (W_R) which serve to highlight particular concepts. Being used in an idiom, or compared to another concept by means of juxtaposition suggests that a given concept may hold special significance. Finally, the weights assigned by our graph understanding system for the additional propositions identified in the graphics are incorporated into the ID of the concepts involved as W_G.

5.3 Selecting Content for a Summary

To select concepts for inclusion in the summary, the model will then be passed to a discourse-aware graph-based content selection framework (Demir et al., 2010), which selects concepts one at a time and iteratively re-weights the remaining items so as to include related concepts and avoid redundancy. This algorithm incorporates PageRank (Page et al., 1999), but with several modifications. In addition to centrality assessment based on relationships between concepts, it includes apriori importance nodes enabling us to incorporate concept completeness, number of expressions, document structure, and rhetorical devices. More importantly from a summary generation perspective, the algorithm iteratively picks concepts one at a time, and re-ranks the remaining entries by increasing the weight of related items and discounting redundant ones. This allows us to select concepts that complement each other while simultaneously avoiding redundancy.

6 Generating an Abstractive Summary of a Multimodal Document

Figure 4 shows the two most important concepts (Company1 & Person1) selected from the Medtronic article in Section 5.1. Following McDonald and Greenbacker (2010), we use the phrasings observed by the parser as the "raw material" for expressing these selected concepts. Reusing the original phrasings reduces the reliance on built-in or "canned" constructions, and allows the summary to reflect the style of the original text. The derivation trees stored in the model to realize a particular concept may use different syntactic constituents (e.g., noun phrases, verb phrases). Multiple trees are often available for each concept, and we must select particular trees that fit together to form a complete sentence.

The semantic model also contains concepts representing propositions extracted from the graphics, as well as relationships connecting these graphical concepts with those derived from the text, and there are no existing phrasings in the original document that can be reused to convey this graphical content. However, the set of proposition types that can be extracted from the graphics is finite. To ensure that we have realizations for every concept in our model, we create TAG derivation trees for each type of graphi-

cal proposition. As long as realizations are supplied for every proposition that can be decomposed in the model, our system will never be stuck with a concept without the means to express it.

The set of expressions is augmented by many built-in realizations for common semantic relationships (e.g., "is-a," "has-a"), as well as expressions inherited from other conceptual categories in the hierarchy. If the observed expressions are retained as the system analyzes multiple documents over time, making these realizations available for later use by concepts in the same category, the variety of utterances we can generate is increased greatly.

By using synchronous TAG trees, we know that the syntactic realizations of two semantically-related concepts will fit together syntactically (via substitution or adjunction). However, the concepts selected for the summary of the Medtronic article (Company1 & Person1), are not directly connected in the model. To produce a single summary sentence for these two concepts, we must find a way of expressing them together with the available phrasings. This can be accomplished by using an intermediary concept that connects both of the selected items in the semantic model, in order to "bridge the gap" between them. In this example, a reasonable option would be TargetStockPrice1, one of the many concepts linking Company1 and Person1. Combining original phrasings from all three concepts (via substitution and adjunction operations on the underlying TAG trees), along with a "built-in" realization inherited by the TargetStockPrice category (a subtype of Expectation), yields this surface form:

> *Wuensch expects a 12-month target of 62 for medical device giant Medtronic.*

7 Related Work

Research into providing alternative access to graphics has taken both verbal and non-verbal approaches. Kurze (1995) presented a verbal description of the properties (e.g., diagram style, number of data sets, range and labels of axes) of business graphics. Ferres et al. (2007) produced short descriptions of the information in graphs using template-driven generation based on the graph type. The SIGHT project (Demir et al., 2008; Elzer et al., 2011) generated summaries of the high-level message content conveyed by simple bar charts. Other modalities, like sound (Meijer, 1992; Alty and Rigas, 1998; Choi and Walker, 2010) and touch (Ina, 1996; Krufka et al., 2007), have been used to impart graphics via a substitute medium. Yu et al. (2002) and Abu Doush et al. (2010) combined haptic and aural feedback, enabling users to navigate and explore a chart.

8 Discussion

This paper presented our system for providing access to the full content of multimodal documents with line graphs in popular media. Such graphics generally have a high-level communicative goal which should constitute the core of a graphic's summary. Rather than providing this summary at the point where the graphic is first encountered, our system identifies the most relevant paragraph in the article and relays the graphic's summary at this point, thus increasing the presentation's coherence. System extensions currently in development will provide a more integrative and accessible way for visually-impaired readers to experience multimodal documents. By producing abstractive summaries of the entire document, we reduce the amount of time and effort required to assimiliate the information conveyed by such documents in popular media.

Several tasks remain as future work. The intended message descriptions generated by our system need to be evaluated by both sighted and non-sighted human subjects for clarity and accuracy. We intend to test our hypothesis that graphics ought to be described alongside the most relevant part of the text by performing an experiment designed to determine the presentation order preferred by people who are blind. The rules developed to identify elaborative propositions also must be validated by a corpus or user study. Finally, once the system is fully implemented, the abstractive summaries generated for entire multimodal documents will need to be evaluated by both sighted and sight-impaired judges.

Acknowledgments

This work was supported in part by the by the National Institute on Disability and Rehabilitation Research under grant H133G080047 and by the National Science Foundation under grant IIS-0534948.

References

Iyad Abu Doush, Enrico Pontelli, Tran Cao Son, Dominic Simon, and Ou Ma. 2010. Multimodal presentation of two-dimensional charts: An investigation using Open Office XML and Microsoft Excel. *ACM Transactions on Accessible Computing (TACCESS)*, 3:8:1–8:50, November.

James L. Alty and Dimitrios I. Rigas. 1998. Communicating graphical information to blind users using music: the role of context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '98, pages 574–581, Los Angeles, April. ACM.

Sumit Bhatia, Shibamouli Lahiri, and Prasenjit Mitra. 2009. Generating synopses for document-element search. In *Proceeding of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, pages 2003–2006, Hong Kong, November. ACM.

Sandra Carberry, Stephanie Elzer, and Seniz Demir. 2006. Information graphics: an untapped resource for digital libraries. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '06, pages 581–588, Seattle, August. ACM.

Stephen H. Choi and Bruce N. Walker. 2010. Digitizer auditory graph: making graphs accessible to the visually impaired. In *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, CHI '10, pages 3445–3450, Atlanta, April. ACM.

Seniz Demir, Sandra Carberry, and Kathleen F. McCoy. 2008. Generating textual summaries of bar charts. In *Proceedings of the 5th International Natural Language Generation Conference*, INLG 2008, pages 7–15, Salt Fork, Ohio, June. ACL.

Seniz Demir, Sandra Carberry, and Kathleen F. McCoy. 2010. A discourse-aware graph-based content-selection framework. In *Proceedings of the 6th International Natural Language Generation Conference*, INLG 2010, pages 17–26, Trim, Ireland, July. ACL.

Michael Elhadad and Jacques Robin. 1996. An overview of SURGE: a re-usable comprehensive syntactic realization component. In *Proceedings of the 8th International Natural Language Generation Workshop (Posters and Demonstrations)*, Sussex, UK, June. ACL.

Stephanie Elzer, Sandra Carberry, and Ingrid Zukerman. 2011. The automated understanding of simple bar charts. *Artificial Intelligence*, 175:526–555, February.

Leo Ferres, Petro Verkhogliad, Gitte Lindgaard, Louis Boucher, Antoine Chretien, and Martin Lachance. 2007. Improving accessibility to statistical graphs: the iGraph-Lite system. In *Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '07, pages 67–74, Tempe, October. ACM.

Charles F. Greenbacker, Sandra Carberry, and Kathleen F. McCoy. 2011. A corpus of human-written summaries of line graphs. In *Proceedings of the EMNLP 2011 Workshop on Language Generation and Evaluation*, UCNLG+Eval, Edinburgh, July. ACL. (to appear).

Satoshi Ina. 1996. Computer graphics for the blind. *SIGCAPH Newsletter on Computers and the Physically Handicapped*, pages 16–23, June. Issue 55.

Stephen E. Krufka, Kenneth E. Barner, and Tuncer Can Aysal. 2007. Visual to tactile conversion of vector graphics. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 15(2):310–321, June.

Solomon Kullback. 1968. *Information Theory and Statistics*. Dover, revised 2nd edition.

Julian Kupiec, Jan Pedersen, and Francine Chen. 1995. A trainable document summarizer. In *Proceedings of the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '95, pages 68–73, Seattle, July. ACM.

Martin Kurze. 1995. Giving blind people access to graphics (example: Business graphics). In *Proceedings of the Software-Ergonomie '95 Workshop on Nicht-visuelle graphische Benutzungsoberflächen (Non-visual Graphical User Interfaces)*, Darmstadt, Germany, February.

Xiaoyong Liu and W. Bruce Croft. 2002. Passage retrieval based on language models. In *Proceedings of the eleventh international conference on Information and knowledge management*, CIKM '02, pages 375–382.

Daniel C. Marcu. 1997. *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts*. Ph.D. thesis, University of Toronto, December.

David D. McDonald and Charles F. Greenbacker. 2010. 'If you've heard it, you can say it' - towards an account of expressibility. In *Proceedings of the 6th International Natural Language Generation Conference*, INLG 2010, pages 185–190, Trim, Ireland, July. ACL.

David D. McDonald. 1992. An efficient chart-based algorithm for partial-parsing of unrestricted texts. In *Proceedings of the 3rd Conference on Applied Natural Language Processing*, pages 193–200, Trento, March. ACL.

David D. McDonald. 2000a. Issues in the representation of real texts: the design of KRISP. In Lucja M. Iwańska and Stuart C. Shapiro, editors, *Natural Language Processing and Knowledge Representation*, pages 77–110. MIT Press, Cambridge, MA.

David D. McDonald. 2000b. Partially saturated referents as a source of complexity in semantic interpretation. In *Proceedings of the NAACL-ANLP 2000 Workshop on Syntactic and Semantic Complexity in Natural*

Language Processing Systems, pages 51–58, Seattle, April. ACL.

Peter B.L. Meijer. 1992. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, 39(2):112–121, February.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November. Previous number: SIDL-WP-1999-0120.

Ulrich Reimer and Udo Hahn. 1988. Text condensation as knowledge base abstraction. In *Proceedings of the 4th Conference on Artificial Intelligence Applications*, CAIA '88, pages 338–344, San Diego, March. IEEE.

Dominic Widdows. 2003. Orthogonal negation in vector spaces for modelling word-meanings and document retrieval. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - Volume 1*, ACL '03, pages 136–143, Stroudsburg, PA, USA. Association for Computational Linguistics.

Michael J. Witbrock and Vibhu O. Mittal. 1999. Ultra-summarization: a statistical approach to generating highly condensed non-extractive summaries. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '99, pages 315–316, Berkeley, August. ACM.

Wai Yu, Douglas Reid, and Stephen Brewster. 2002. Web-based multimodal graphs for visually impaired people. In *Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology*, CWUAAT '02, pages 97–108, Cambridge, March.

Chengxiang Zhai. 2008. *Statistical Language Models for Information Retrieval*. Morgan and Claypool Publishers, December.