Knowledge Extraction from Diagram and Text for Media Integration

Yuichi NAKAMURA, Miwa TAKAHASHI, Masayuki ONDA, Yuichi OHTA
Institute of Information Sciences and Electronics, University of Tsukuba
Tsukuba, Ibaraki, 305, JAPAN (yuichi@is.tsukuba.ac.jp)

1 Introduction

For the effective use of multimedia, it is essential to integrate different kinds of media in their semantics, and to organize a new information medium by combining them. So far the correlation between the multiple media has been analyzed by hand and the links among them are manually generated and utilized for this purpose. However, an automatic or semi-automatic analysis and integration of multiple media is a quite important task we have to tackle with, since enormous amount of document data have been accumulated for a long time and are still increasing.

We previously proposed a framework to integrate diagram and text by finding the mutual relationship between them[5]. It showed that simple semantic interpretation of diagram elements can be realized by using the correspondences between diagram and text. The semantics of each element in a digram, for instance, is clarified by the words in a text attached to the diagram, also the essence of the text can be extracted by referring the the diagram.

In this paper, the semantic reasoning is augmented by using the knowledge of the typical diagrammatic patterns. This is based on the idea that important information is often carried by a typical combination of figure elements.

2 Basic Idea for Media Integration

Let us consider a simple example shown in Fig.1: the text describes the topic accurately; the diagram shows the flow of "yen check" with a simple diagrammatic structure, and is easy to understand at a glance. Such a combination of the two media is often seen in written documents and electronic documents. The aim of this research is the information extraction from such media and the construction of an integrated medium.

The target data of this research is a set of diagrams and texts taken from a textbook or an encyclopedic dictionary. They are usually well organized, and the argument claimed by the author is clear. The diagrams are well described by texts, and vice versa.

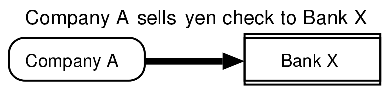


Figure 1: Example of a diagram

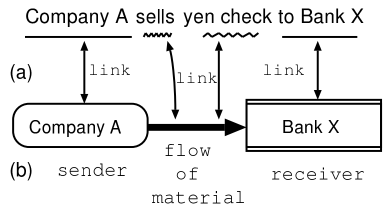


Figure 2: Integration of Diagram and Text

Semantics at Element Level

A diagram element usually expresses one notion, and relations between them, ex. a connecting line or an arrow, expresses their relation. Similarly, other attributes, that are shape, size, color, etc., sometimes have special meanings which are essential for the topic. The usage of a figure element, however, heavily depends on situations. An arrow, for example, can be regarded as a flow of material, while another arrow can be regarded as the movement (trajectory) of objects.

To cope with this problem, we previously proposed the method to semantic interpret the elements in a diagram[5]. Our basic idea of the integration is to generate the following descriptions:

link: Link shows the correspondence between an element of diagram and an element of natural language text. An example is shown in Fig.2(a).

semantic interpretation: By using semantics of the word linked to a figure elements, the semantic interpretation, *i.e.* a category of notions, is attached to each element. An example is shown in Fig.2(b).

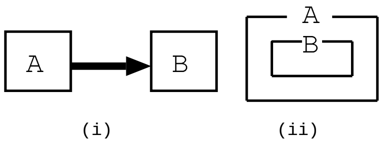


Figure 3: Basic structures

Semantics at Compound Structure Level

In the example in Fig.2(b), it is often difficult to determine the appropriate role of a figure element by using only the single element and the single word linked to it.

For this purpose, typical combination pattern of figure elements in a diagram, the compound structure, can be quite important means for the analysis. In other words, the structure of a diagram is usually based on a set of empirical rules and the fact offers us the important keys to understand the meaning of the diagram. For example, a diagram expressing '*process order*' is mainly composed of two sets of important elements; a set of figure elements which express processes and a set of elements which express orders. If the underlying rule to compose the diagram is correctly obtained, the outline of the topic can be estimated. This is a quite useful key to the disambiguation of the detailed meaning of each element as well as a key to understanding the discourse structure of the text.

For this purpose, two typical patterns shown in Fig.3 are examined in detail as basic structures. Tens of example diagrams including these patterns are manually analyzed, and are classified into several categories of semantics: *flow* (*input*, *output*, *input&output*), *effect*, *transformation*, *spatial structure*, and *others*. We focused on the first five semantics in our research. The rest two are left for future works.

For each basic structure, the role, *i.e.* detailed meaning, of an element is shown in Fig.4. For example, if the basic structure (i) expresses the flow of *input&output*, each element, *e.g. A or B*, has both semantics as a sender and a receiver. The arrow may express the flow of material sent/received, or the process which causes the flow. Also important relationships such as causality, subsumption, or hierarchy may be inferred in various cases.

There also exists considerable degree of ambiguities in the classification of real diagrams. They can be disambiguated by using the correlation with texts similarly to the reasoning at element level.

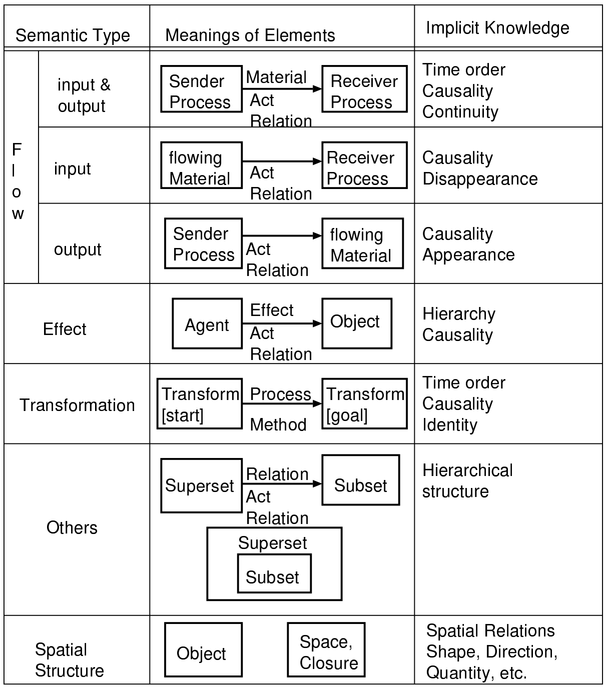


Figure 4: Detailed semantics for basic structure

3 Semantic Interpretation

We propose a new method to utilize typical compound structures in a diagram to integrate the text and diagram media. First, the semantic interpretation of the basic structures is performed, then the results are combined to obtain the semantics for larger structures.

3.1 Analysis for Basic Structure

There are two kinds of information which are the keys to detect the semantics of a basic structure: the lexical knowledge of a word linked to the elements in the structure and the knowledge about the case frame related to the structure. Since the sentence including the words linked to the elements in a diagram usually states about the attributes or relations of the elements, the verbs and case frames are the most important information to detect them. For example, the basic structure (i) in Fig.3 is often accompanied by a sentence whose verb is "send", "receive", and so on.

Lexical Knowledge of Verbs

We have manually analyzed the verbs in the texts attached to diagrams, and have recorded each correspondence between a verb and the semantics of basic structure in BGH. An example of this classification is shown in Table 1, in

Table 1: Example of verbs corresponding to the semantics

Semantic Type	Example of Verbs
Input&Output	send, flow, give
Input	receive, use, buy
Output	send, generate, sell
Effect	touch, hit, help
Transformation	change, transform

Table 2: Examples of the case frames defined in IPAL

Agent(動作主), Cause(起因), Object[beneficiary](相手), Object[for attitude](対象 [態度の対象]), Object[effected](対象 [受影]), Object[changing](対象 [変化]), Object[disappear](対象 [消滅]),

which typical verbs frequently used with the basic structures and example sentences are shown. This classification was performed at the third level of BGH hierarchy[1]. The number of entries at this level is about one thousand, which can be classified with tractable efforts.

Given the classification at the third level, the classification for each verb is easily obtained: the node at the third level which is the ancestor node of the verb is automatically searched; then the semantics assigned to the node are inherited to the verb.

Case Frame

The case frame is more useful than the lexical knowledge of a verb. For about one thousand Japanese verbs, fortunately, there is IPAL dictionary[4] which precisely describes case structures. We utilize the case frame information when the verb of a case frame is included in the dictionary. Examples of other cases defined and used in IPAL are shown in Table 2.

We manually analyzed the consistency between the case structure and the semantics of basic structure, and collected typical examples. Several examples are shown in Table 3. In this example, the correspondence between the semantics for the basic structure and typical case patterns are shown. By accumulating these patterns, and by comparing them with an input diagram and a text, the semantics of the input can be detected.

[1]The words in BGH are stored in a tree structure with six levels. The lowest level is words, the second lowest is synonyms, and the third level is a small group of synonyms

Table 3: Example of Correspondences between Semantics and Case Frames

Semantic Type & Related Cases
(Input&Output) Sender = Agent, Place[spatial start], Receiver = Place[spatial goal], Object[beneficiary] flowing material = Object[effected], Object[others]
flowing process = verb
(Transformation) Transform[start] = Agent, Object[non-spatial start] Transform[goal] = Place[non-spatial goal]
transform process = verb

“verb” in the above table shows that the verb of a case frame indicates the name of process, etc..

Procedure

First, we assumed that the links between elements across two media are obtained in detail. For simple diagrams and texts, links are well obtained automatically by the method previously proposed by us. For complicated cases, some manual correction might be necessary.

The interpretation process is as follows. First, the semantics structure from the diagrammatic structures, those from verbs, and those from case frames are obtained. Then the intersection of the semantics is regarded as the most plausible semantics for the basic structure. However, all candidates are left for the next step, *i.e.* interpretation for a group of these structures.

3.2 Analysis of Global Structure

Similar elements in a diagram have similar meanings. This is an important customary rule to compose a diagram. From this point of view, the semantics of two or more similar basic structures are disambiguated by using the candidates for each basic structure.

For this purpose, similar basic structures are grouped: similarity measure between basic structures is defined by the attributes of their elements; similarity among two or more basic structures is counted only if they share one or more elements. The semantics common throughout a group is selected as the most plausible semantics for the group:

1. Collect the candidates of semantics from every basic structure in a group as the votes for each candidate.
2. The semantics which get the maximum number of votes is considered as the most plausible semantics for the group.

If two or more groups share some elements, they are treated independently to each other [2].

4 Experiments

We have processed ten examples by the method described above. The input data are taken from textbooks for high school or junior high school in Japan. The input data of the diagrams are prepared by using *tgif*, a widely used drawing tool. They are manually redrawn with this drawing tool so as to be similar to the original data[3]. The input texts are taken from textbook and coded manually.

Then, we reorganized the input diagram and text manually: a complicated figure element, for example, a realistic shape of human, is replaced by a simple closure; a complex sentence or a compound sentence is manually divided into simple sentences; demonstrative pronouns are replaced by their referents.

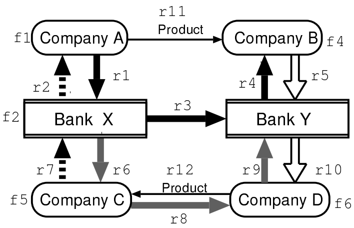
One of the input diagrams is shown in Fig.5 The linking result for the upper left part of the diagram is already shown in Fig.1(b). In interpretation process, $(f1, f2, r1)$ and a sentence S1 is analyzed[4]. The verb "sell" and its case frame pattern is used for the interpretation. As a result, *flow (input&output)* is obtained as the semantics of the basic structure. Then the interpretation of the larger structures is performed. Partial results are shown in Fig.6. For a group of $(f1, f2, f3, f4, r1, r3, r4)$, the semantics "*flow*" was obtained through the interpretation by the voting from every basic structure. By using this result, the flowing material is inferred for each arrow of $(r1, r2, r3)$, and the result is (yen check(円手形), check(手形), check(手形)).

As shown in the above example, the proposed method works well for relatively simple and well organized diagrams and texts. Generally, if the correspondence between the two media, *i.e.* links, are well obtained, the semantic interpretation is not difficult. Usually, the flowing materials or other items which are not explicitly mentioned in a diagram are correctly detected by the integration process. However, misinterpretation often occurs when the semantics for each basic structure often hold multiple meaning, *i.e.* multiple sentences mention multiple aspects of the diagram. It is often the case that complicated and indirect reference of elements cause an error. For example, "A company orders a product to another company" often means "The company receives a product from the other company". But the latter is often omitted in the text. These are left for future works.

[2]It will be reasonable that an element is a sender in a group and an affected object affected in another group

[3]Although this input process is important media processing for diagram understanding, most of diagrams newly drawn at present or in the future will be stored once in such a format as mentioned above. So, we skip the stage of diagram recognition

[4]The analyzes are performed in Japanese, then we translated the input texts and the results into English for the convenience of readers.



S1: Japanese exporting Company A sells the yen check to Bank X.

S2: Company A receives cash from Bank X.

S3: Bank X send the check to Bank Y.

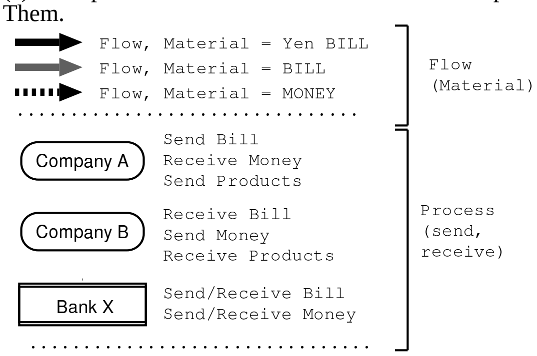
S4: American importing Company B receives the check from Bank Y.

(The following sentences are omitted. because of lack of space.)

Figure 5: Input diagram

(f1, f2, r1):	Input&Output, Sending Process = f1, Receiving Process = f2, Flowing Material = Yen Check
(f2, f1, r2):	Input&Output, Sending Process = f2, Receiving Process = f1, Flowing Material = Cash (YEN)

(a) Example of Obtained Basic Structures and Interpretation of Them.



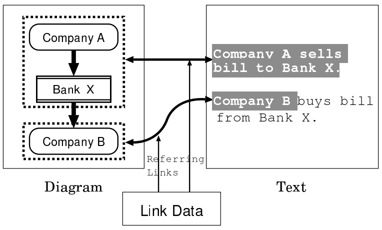
(b) Semantics for Elements based on Basic Structures.

Figure 6: Result of analysis

5 Utilization of Integrated Media

Given the result of media integration, we have developed a prototype system which can generate various kinds of explanations. This is not only an explanation system but also a demonstration of the utilities of the obtained data.

When a user specify an element to request its explana-



When an element or a set of elements in a diagram or a text, corresponding part of the other medium is highlighted. The upper arrow shows the correspondence between a basic structure and a case frame, *i.e.* a sentence. Therefore, when either of the sentence or the basic structure is specified, the other is highlighted.

Figure 7: Example of highlighting

tion, the following explanation generation will take place in our system.

(a) Emphasizing or highlighting of identical or relating parts in both media is given.
   - A part in one medium is highlighted which corresponds to a specified elements in the other medium.
   - A sentence in the text is read out, when the corresponding part of the diagram is specified.

   An example is shown in Fig.7.

(b) A new sentence is generated for the specified part of a medium. An example is shown in Fig.8.
   - An explanation sentence of a specified diagram element is generated.
   - A summary sentence of a specified sentence in a text is generated.

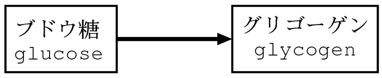
   For this purpose, we prepared several template sentences, one of which is shown in Fig.8. 3. When a part in a medium is specified, an appropriate template sentence is selected, then each slot of the template is filled by a word which has the semantics corresponding to the slot.

Details are omitted for lack of space.

6 Conclusion

In this paper, a new idea of media integration for diagrams and texts was proposed. First the formalism of media integration and its utilization was presented. Then the several types of semantics for diagrammatic structures are discussed and categorized. By using this categories, semantic interpretation of a diagram and a text can be performed so as to generate integrated media. In the media, each typical structure in a diagram is attached semantics, which can be used for further reasoning of the media, and for generating explanations as a function of the integrated medium. A prototype system was constructed to prove the usefulness of the knowledge obtained by integrating the two media.

1. Specified part.



2. Semantics for the basic structure:
   Transformation
3. Selected template for explanation:
   (Transform[before]) changes into
   (Transform[after]).
4. Corresponding words are searched using semantics and links:
   (Transform[before]) = “glucose”
   (Transform[after]) = “glycogen”
5. Replacing by found words:
   Glucose changes into glycogen.

Figure 8: Example of sentence generation

References

[1] Y. Arens, L. Miller, and N. Sondheimer. *Proc. ACM SIGCHI Workshop on Architectures for Interfaces*, pages 93–107, 1988.

[2] S. Feiner and K. McKeown. *Proc. AAAI ’90*, pages 442–449, 1990.

[3] R. Futrelle. *Proc. 10th ICPR*, pages 403–408, 1990.

[4] Japan Information-technology Promotion Agency. IPA, 1987.

[5] Y. Nakamura, R. Furukawa, and M. Nagao. *Proc. Int. Conference on Document Analysis and Recognition*, pages 614–618, 1993.

[6] A. Rizk, N. Streitz, and J. Andre, editors. *”Hypertext: concepts, systems and applications”*. The Cambridge Press, 1990.

[7] S. Roth, J. Mattis, and X. Mesnard. *Proc. ACM SIGCHI Workshop on Arhitetures for Interfaces*, pages 109–128, 1988.