Diagram Understanding Utilizing Natural Language Text

† Yuichi NAKAMURA ‡ Ryo FURUKAWA †† Makoto NAGAO

† Inst. of Information Sciences and Electronics
University of Tsukuba
Tsukuba-City, 305, JAPAN

‡ Advanced Inst. of Science and Technology
8916-5, Takayama
Ikoma, 555, JAPAN

†† Dept. of Electrical Eng. II
Kyoto University
Sakyo, 606, Kyoto, JAPAN

Abstract

Diagram understanding and its cooperative use with other media are important subjects in both pattern understanding and communication. However, it is quite difficult to understand diagrams without supplementary explanation by other media. For this purpose, we propose a new framework for semantic understanding of a diagram by utilizing textual information. In this framework, the elements in a diagram are tightly linked to words or other structure in a natural language text, and the semantic structure of a diagram is interpreted clearly by using natural language information. The obtained result contributes to the construction of hyper-text and multi-media.

1 Introduction

Generally, diagram and text are used complementary as effective means of communication in many situations. Although a text can convey precise information with one dimensional sequence of letters, textual data are sometimes quite complicated and redundant including a lot of miscellaneous information. On the other hand, a diagram can express an idea very clearly in a two dimensional structure. People often use diagrams[1]to describe their concepts to their audience in their presentation. However, it is difficult to understand the semantic structure of a diagram without textual information. Consequently, we have to investigate the mechanism of cooperative usage of both information sources. As one important topic of this field, we have investigated the diagram understanding by using natural language text analysis.

There are a few research topics related to this: researches to recognize the semantic structures of a diagram as the extension of diagram analysis (ex. [Lak87, Fut90, PSSW89]); researches to summarize textual data and generate as diagrams (ex. [Tab91, ET91]); researches on multimedia for user interface or presentation (ex. [AMS88, NS88, RMM88, FM90]).

These researches, however, analyze diagrams or generate diagrams by using knowledge about diagrams quite separately from natural language information. There are stronger demands for developing a framework to integrate textual information and diagram information for multimedia database and knowledge base.

For this purpose, we propose a new framework for the integration of diagram information and textual information for semantic understanding of a diagram. In this paper, we show the basic idea of the integration across media, the methods for elements linking across media, and semantic interpretation of elements.

The experiments for the diagrams and texts taken from the Encyclopedic Dictionary of Computer Science (hereafter abbreviated as EDCS. Iwanami Publ. 1990) showed the effectiveness of our framework.

2 Diagram Understanding

2.1 Understanding across Two Media

The aim of this research is to clarify the way of integrating diagram information and natural language information.

First we assume that the information in each medium is organized in a network after some process in which elements are extracted and organized by certain relationships. In other words, elements (ex. x_i) in each medium are extracted and relations (ex. $relation(\{x_1, x_2\}, reltype)$) between them are extracted. Then, we consider the integration of media as the combination of two important processes, linking and semantic interpretation.

[1]For the simplicity of notation, we use the term “diagrams” for figures and other style of diagrams.

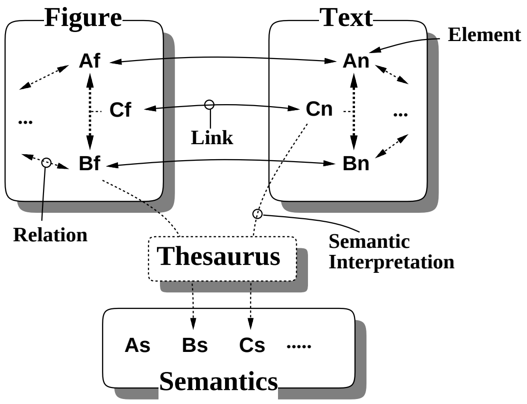


Figure 1: Integration of a Diagram and Text

linking: An element of a diagram and the corresponding element of natural language text (word, etc.) have a link (across media).

semantic interpretation: To each element, an appropriate semantic interpretation is attached.

We can see an example in Fig.1. Initially, each element in the two media (diagram and text) is not related to each other at all. By analyzing each medium, the elements are connected by relations within each medium. By linking corresponding elements (for example, Af and An) between the two media, the data of diagram and text are integrated. At the same time, the structure is semantically interpreted by using thesaurus information. Assume, for example, that Bf has semantic interpretation Bs, and that there is a link which shows the equality of Bf and Bn. In this case, the two elements from different sources can share the same semantic interpretation.

Let us describe the two important notions briefly in the followings.

2.2 Linking between Two Media

This operation can be considered as a process to generate links denoted as follows.

$$link(d_i, n_j, linktype) \qquad (1)$$

where $linktype$ is a category of a link (mentioned below). d_i is a diagram element or a relation, and n_j is a natural language element or a relation.

There can be many kinds ($linktype$) of correspondences between elements in two media. In this research, however, we use only three kinds of links as shown below, since it is difficult to detect delicate relationships at the current stage.

$identical$**:** A link of this type shows equality of elements in a diagram and a text.

Table 1: Example of Semantic Categories

transmission: transmit, send, inform, etc.
transformation: transform, deform, etc.
operation: repeat, do, work, etc.
creation: create, make, generate, etc.
appearance/extinction: appear, disappear, etc.
observation/record: observe, record, analyze, etc.

$example$**:** A link of this type shows that a figure element is one example of a natural language element, or vice versa.

$description$**:** A link of this type shows that a natural language element is a description for a figure element (but no natural language element which is identical to the figure element can be found).

2.3 Semantic Interpretation

The semantic interpretation is considered as a set of mapping (si) from an element (or a relation) to a semantic category. We prepared 21 categories (several are shown in Table 1) to roughly meet variety of notions expressed in diagrams in encyclopedic dictionary or other texts. This categorization is made by human by using thesaurus information[2].

With these semantic categories, let us denote semantic interpretation across media as follows:

$$si(d_i, n_j, sc_k) \qquad (2)$$

The above formula shows that a diagram element/relation d_i has a correspondence to a natural language element/relation n_j. The semantic interpretation that these two share is sc_k, that is one of the semantic categories.

3 Diagram Information

In this research, the following structure of a diagram, that we call *physical structure*, is handled.

physical element: letters (letter, word, sentence), line (straight line, dotted line, curve, arrow), geometric elements (circle, ellipse, rectangle, polygon), and enclosure.

physical relation: *spatial relationship* (inclusion, overlapping, touch, stab, adjacent, parallel, collinear/alignment (horizontal, vertical, oblique)), *relationship of attributes* (shape, inner region (color, texture), and boundary (line width, dotted, line color))

[2] Since this categorization is adjusted for the analysis of EDCS, further tuning may be required for the diagrams in other fields

Table 2: Physical Elements which can be interpreted as logical relations

line: a line (on the condition that it has adjacent to some other elements), an arrow. This element can be interpreted as relation of *connected by a line (arrow)*.

enclosure: an enclosure in which more than one element are located. This element can be interpreted as relation of *enclosed in the same enclosure*

However, we need more abstract structure of a diagram in which notions and relationships between notions are extracted and related to each other. For this purpose, we classified figure elements into three groups, *logical elements*, *logical relations*, and *labels*.

logical element: An element which corresponds to an atom of notion.

logical relation: An element or a relationship which corresponds to relationships between "logical elements"

label: The words attached to logical elements or relations for their explanation.

We call this structure the *logical structure* of a diagram. To get logical structure, we need the conversion from the physical structure, since they are not directly corresponding. Currently, our criterion for the conversion is simple:

> Every physical element/relation can be translated into any of logical element, relation and label, if it has possibility.

Therefore, the following conversion is performed.

- Special kinds of physical elements shown in Table 2 can be interpreted as both logical elements and logical relation.
- A word (letter) can be interpreted both as an element and as a label.
- Other physical elements are interpreted as logical elements.
- Physical relations are interpreted as logical relations.

For example, an arrow, which is a physical element of a diagram, may be classified not only as a *logical element* but as a *logical relation* because it simply expresses some relationship between physical elements.

In addition to that, each word (sequence of letters in a diagram) is attached to the nearest logical element as a label. The pair is considered as an element of a diagram. The label is used in checking relationships in natural language expression, while the logical element is used in checking relationships between logical elements in a diagram.

4 Textual Information

We consider a text as a collection of words tightly related to each other. In this sense, the natural language elements in the formulae (1) and (2) are the words in a text and the relations between words. We handle this structure in three forms:

word: Nouns and verbs in any part of a text. Words are basis of linking and semantic interpretation.

case frame: The meaning of each word or a relationship between words can be detected by checking up to which slot of a case frame the word corresponds. The followings are the pieces of information held by case frames.

(a) A relation (of other types) between words.
ex. A verb(wd_l) B
$\Rightarrow relation(\{wd_i, wd_j, ...\}, wd_l)$,
where wd_l is a word specifying the relation.

(b) A direct mapping from diagram contents to a word or semantics.
ex. A (for instance, "upper left box") is B.
$\Rightarrow link(d_i, wd_j, identical)$

(c) A direct semantic description of a word.
ex. A means B $\Rightarrow si(wd_i, sc_j)$

(d) Equivalence of the semantic categories of words.

(e) Equivalence of words in a context.

coordinate structure: A coordinate structure shows that the coordinate words have the same meaning or they are replaceable.
ex. A, B and C are ....
$\Rightarrow relation(\{wd_j, wd_k\}, coordinate)$

These are obtained through natural language processing: morphological analysis [Nag92]; dependency analysis [KN92]; finally case frame analysis by pattern matching. We have to skip details for lack of space.

5 Integration Process

As mentioned in Section 2, the integration process is mainly composed of linking process and semantic interpretation process. Since the intermediate result obtained these two process is not reliable, two more process are performed to obtain better results. We show these four processes in the followings.

5.1 Linking Process

During the linking process, links mentioned in Section 2 are generated. The conditions to generate links are the similarities of words in letter expression and case frame information obtained by text analysis shown in the previous section.

similarity of words: The similarity in letter expression between the label (la) and a word (wd). For example, "subway" and "way" is similar in letter expression. Exact match generates an $identical$ link. Partial match with high score generates an $identical$ link, while partial match with low score generates an $example$ link.

direct pointing: The direct pointing of a physical element by words, for instance, "the upper right box". This generates an $identical$ link.

description: If a case frame includes words which correspond to elements in either medium, $description$ links are generated.

5.2 Semantic Mapping by Thesaurus

To find the category of a word and to find the similarity between words, we use a thesaurus organized in a network. Although there are many thesauruses we can read on computer, there are only a few in which words are hierarchically organized. 'Bunrui Goi Hyou' (abbreviated as BGH)[Lan64] is one of these thesaurus in which "verb"s and "noun"s are hierarchically organized according to the semantics. BGH has a six layer abstraction hierarchy, and more than 60,000 words are assigned to its leaves. A fragmentary portion of the structure of BGH is shown in Fig.2.

By checking the location of a word in BGH, we can estimate rough meaning of a word. In other words, we can regard BGH as a mapping function (Map) from a word (wd) to a semantic category (sc).

$$Map(wd_i) = sc_j \;\;\Rightarrow\;\; si(-, wd_i, sc_j) \qquad (3)$$

where "$-$" means something that is not specified.

In this research, each pre-defined semantic category shown in Table 1 has a certain scope in BGH. The semantic interpretation of every word in inside the scope can be regarded as this category.

It is often the case that a word is classified into more than one category, since the boundaries of categories can overlap. For example, the word "way" can be used as either "place(road)" or "means". In this case, the semantic interpretation of the word is a set of categories.

5.3 Propagation and Refinement

The interpretation obtained by the previous step is not reliable enough. One reason is the multiple interpretation mentioned in the previous section, another is the sparseness of the links. Not many elements in both media can be semantically interpreted, since diagram elements which have

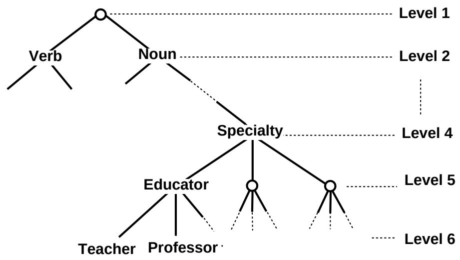


Figure 2: Brief overview of BGH

links to words are sparse in usual case. We need further operations for the rest. For this purpose, two operations are performed, that are propagation of semantic interpretation and refinement after that.

For the propagation process, we picked up stable structures which appear frequently in many situations. Fig.3 shows three of them which include *identical* links. (a) is the simplest one that appears frequently. (b) is a more general one in which elements are connected by relations and links.

Using these structures, the propagation process attach semantic interpretations to the elements which have not been attached yet. The basic idea is to complete the stable structure that is mentioned above by giving an appropriate semantic interpretation. Assume that a diagram element d_1 (which has an interpretation sc_1) and natural language element n_1 have an *identical* links between them. Then n_1 has strong possibility of having the same semantic interpretation sc_1. Therefore, this operation gives sc_1 to d_1.

This operation can be formally denoted as the following heuristic rules.

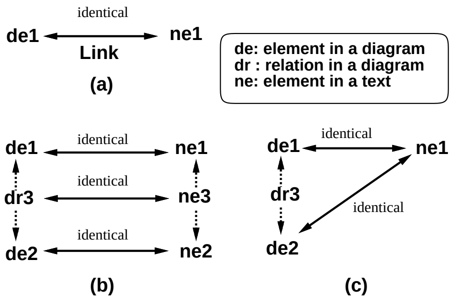
$$si(d_1, -, sc_1) \wedge link(d_1, n_1, identical) \;\Rightarrow\; si(d_1, n_1, sc_1)$$

We have ten more rules for propagation according to variety of structures which include variety of links. Using these rules, the propagation process repeats the following operation:

> For each element that has no semantic interpretation, attach appropriate interpretations by creating the stable structures.

After all possible propagations are performed, the refinement process takes place. In this process, multiple interpretations which are mutually contradictory are reduced to the most possible candidate(s) by checking the constraints to other elements tightly related.

The operation we took is very simple. For each element that has more than two semantic categories, the intersection to the categories of elements which are related by strong relations are checked. Finally, we get the unified result.



(a) $link(de_1, ne_1, identical), si(de_1, ne_1, sc_1)$
(b) $link(de_1, ne_1, identical), si(de_1, ne_1, sc_1)$
$link(de_2, ne_2, identical), si(de_2, ne_2, sc_2)$
$link(dr_3, ne_3, identical), si(dr_3, ne_3, sc_3)$
(c) $link(de_1, ne_1, identical), link(de_2, ne_1, identical)$
$dr_3 = relation(de_1, de_2, coordinate)$,
$si(de_1, ne_1, sc_1), si(de_2, ne_1, sc_1)$

Figure 3: Basic Structure for Semantic Propagation

6 Experimental Result

Targets

We applied this system to about ten topics which contain relatively simple diagrams. Topics are selected according to the following criteria:

- Diagrams which contain physical elements in Section 3 are chosen.
- Topics in which a text does not describe the diagram enough are chosen.

Flow of Operations

The flow of whole operations is as follows:

1. The input diagrams are selected from EDCS. They are encoded as a set of physical structures mentioned in Section 3 by human hand.
2. Independent analysis of a diagram and text.
3. Linking and Semantic Interpretation.
4. Interpretation propagation and Refinement

Results

Through experiments, about half of the elements in diagrams were correctly interpreted. Rest half of elements are left un-interpreted or mis-interpreted. This is mainly because of the lack of description by text and the mismatch of the semantic categories (pre-defined in Table 1) to the semantic structure of a diagram. Although obtained links between the two media are rather sparse, most of them are correctly generated.

The linking result before refinement process is partially shown in Fig.4. The translations of Japanese words are presented in Table 3. The result of semantic interpretation is shown in Fig.5. In this example, most figure elements in the diagram are given semantic interpretations which include correct semantic categories, although several are given multiple semantic categories (The "方式設計 (system architecture design)", for example, has three interpretations). A few of them are interpreted incorrectly because the wrong interpretations are propagated through wrong *identical* links.

Table 3: Japanese Text and English Translation

設計工程 (ハード ウェアの) design process
所望のハード ウェアを実現するために, 一定の設計手法 (design methodology) に従って設計の段階を進めていく作業の流れの総称. この作業に引きつづき, 製造工程 (manufacturing process), 試験工程 (test process) を経てハード ウェアが完成する. 設計工程は図に示すように, 方式設計, 機能設計, 論理設計, 回路設計, 実装設計 (主として LSI の場合), 試験設計などが含まれる. 設計で使用する概念モデルは, ...

Design process (Hardware)
The flow of operations in which designs of several steps are proceeded to attain objective hardware according to design methodology. Manufacturing process and test process follow this process to obtain the hardware we want. The process is roughly composed of ....

Translation of important terms in the diagram
方式設計: system architecture design, 機能設計: functional design, 回路設計: circuit design, 論理設計: logic design 試験設計: test design, 実装設計: physical design, ライブラリーデータベース: library database, レイアウト設計: layout design, 製造: manufacturing, 試験: test, 完成: completion, 製造工程: manufacturing process, 試験工程: test process, 試験データ: test data

7 Conclusion

We proposed a new framework for the integration of diagram information and textual information. The following structure for both media are proposed: physical structure and logical structure for diagrams; words, case frames, and coordinate structure in natural language text. Two important integration processes are proposed: linking elements across media; semantic mapping with thesaurus. The experimental system worked out fairly well for relatively simple diagrams in EDCS.

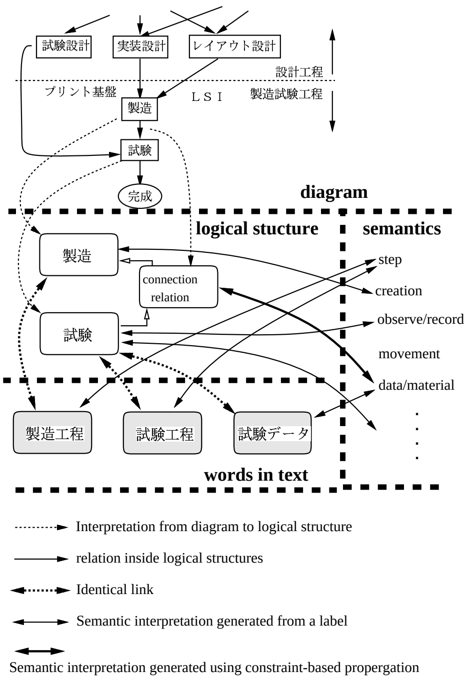


Figure 4: Linking Result

References

[AMS88] Y. Arens, L. Miller, and N. Sondheimer. Presentation planning using an integrated knowledge base. *Proc. ACM SIGCHI Workshop on Arhitetures for Interfaces*, pp. 93–107, 1988.

[ET91] T. Endo and K. Takaoka. Drill text understanding with integrated natural language and picture processing (In Japanese). *Symposium on Integration in Natural Language Processing*, pp. 64–71, 1991.

[FM90] S. Feiner and K. McKeown. Coordinating text and graphics in explanation generation. *Proc. AAAI '90*, pp. 442–449, 1990.

[Fut90] R. Futrelle. Strategies for diagram understanding: Generalized equivalence, spatial/object pyramids and animate vision. *Proc. 10th ICPR*, pp. 403–408, 1990.

[KN92] S. Kurohashi and M. Nagao. Dynamic programming method for analyzing conjunctive structures in japanese. *Proc. of the 14th International Conference on Computational Linguistics*, pp. 170–176, 1992.

[Lak87] F. Lakin. *VISUAL LANGUAGES*, chapter Spatial Parsing for Visual Languages, pp. 35–85. Plenum Press, 1987.

[Lan64] National Institute Japanese Language. *Bunrui Goi Hyou (in Japanese)*. Shu-ei Shuppan, 1964.

[Nag92] Nagao Laboratory, Dept. of Electrical Eng. Kyoto University. *JUMAN Manual (In Japanese)*, 1992.

[NS88] J. Neal and S. Shapiro. Inelligent multi-media interface technology. *Proc. ACM SIGCHI Workshop on Arhitetures for Interfaces*, pp. 69–91, 1988.

[Oka87] N. Okada. Towards a unified understanding of natural language and picture patterns. *Language and Artificial Intelligence*, 1987.

[PSSW89] R. Plant, S. Scrievner, A. Schappo, and A. Woodcock. Usage and generality of knowlede in the interpretation of diagrams. *Knowledge-Based Systems*, Vol. 2, No. 2, pp. 99–108, 1989.

[RMM88] S. Roth, J. Mattis, and X. Mesnard. Graphics and natural language as components of automatic explanation. *Proc. ACM SIGCHI Workshop on Arhitetures for Interfaces*, pp. 109–128, 1988.

[Tab91] A. Tabuchi. Text structure visualization (In Japanese). *Symposium on Integration in Natural Language Processing*, pp. 56–63, 1991.

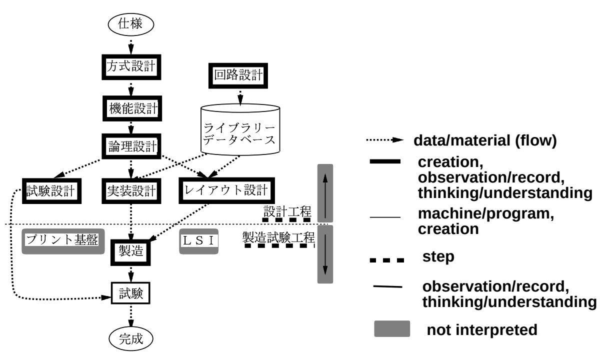


Figure 5: Final Interpretation Result