Computers ind. Engng Vol. 13, Nos 1-4, pp.304-308, 1987
Printed in Great Britain.

0360-8352/87 $3.00+0.00

A Graphic Identification Procedure For An Expert Authoring System

RAMASWAMY RAMESH
Department of Management Science, School of Management
SUNY at Buffalo, NY 14260
and
CHEICKNA SYLLA
Department of Engineering Management, College of Engineering
Drexel University, Philadelphia, PA 19104

Abstract

To reduce the cost of producing text-graphic materials, computer routines are used to generate illustrations and to automate page compostions. The present paper describes an on going contribution to this effort.

Text-graphic material involved are from a narrative technical procedure description for the training of lower skilled mechanics of the U.S. Navy to operate and maintain complex technical equipment (e.g., an aircraft maintenance procedure). The procedure description text is written by the equipment designer and contains: (1) A description of the equipment, (2) a description of the procedures to be followed in using any part of that equipment.The graphic identification project is a module of the Expert Authoring System, which constructs a sementic network from the equipment and the procedure descriptions and then: (1) identify the relationships between the equipment and its components, between components and components, (2) identify the components needed for each task, (3) identify the results of each task, and (4) identify the relationships between the task steps.

The paper describes the preliminary phases of the development of a graphic identification procedure from natural language processing of a technical text written in english. These phases involve the design of a semantic network from a given text, and the design of the support data base.

Introduction

The Naval Technical information Presentation Program (NTIPP) is an effort by the Navy to improve the efficiency of producing and maintaining the technical information required to support equipment. One part of this effort has been to develop computer aided authoring programs to lower the cost of preparing technical information.

One type of computer aided authoring program developed for NTIPP is used to prepare two closely related types of material: Job Performance Aids (JPA's) and Paper Procedure Training Aids (PPTA's). This type of computer authoring program contains automatic page layout routines. The first two generations of this type of authoring program are PLA and PowerPLA. Materials created with these programs are being successfully used to support the operations and training on three types of Navy equipment (aircrafts and submarines). Materials authored with these procedures, are instruction and repair procedures performed by entry level personnel.

JPAs and PPTAs are made with text-graphic pages which use illustrations to describe the steps in a procedure and text to clarify the meaning of the illustrations. Figure 1 shows an example of a text-graphic page for performing simple calibration procedures on an oscilloscope. Also shown are the needed learning guidelines included.. This type of text-graphic page presents information on the components of that portion of the system under consideration. The component names, locations, and functions are presented using:

- -- an overview picture--of the equipment under consideration--in form of diagram or photograph,
- -- close-up view pictures of the components--also in the form of diagrams or photographs,
- -- label boxes--containing the task steps or information of the procedure to be carried out,
- -- the pointers (darts)-- connecting the close-up view pictures to general location on the overview and connecting label boxes to the close-up view pictures.

FORMAT MODEL
RECALLING FACTS ABOUT EQUIPMENT

A general format for use in designing training materials which present steps of a procedure to be performed from memory.

Recalling Facts About Equipment Format Model- Page 1

Use this page format to present each step in a procedure.
The purpose of this page format is to present:
- a word description of the step—emphasize human action.
- a visual display of the step—emphasize human action.
- the purpose of the step.
- the location of actions on equipment.
- the system response to actions taken.
- notes—additional needed information.

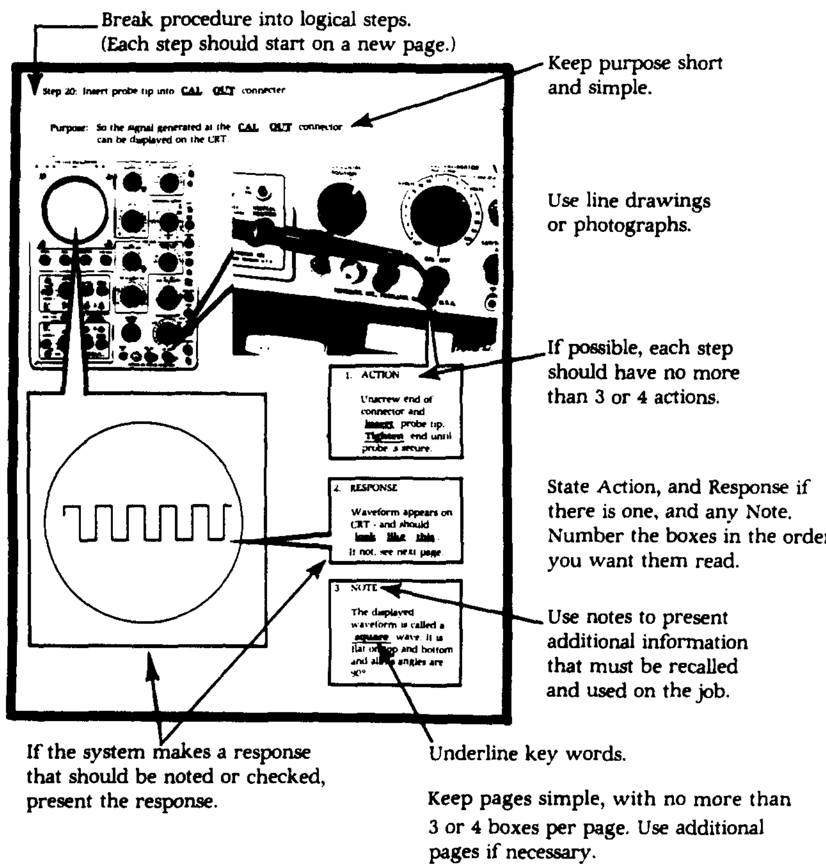

Figure 1. An Example of a typical page layout showing pictures, labels and pointers.

The format also includes a word description of each step, with emphasis on human action; a visual display of the step, again with emphasis on the action; the purpose of the step; the location of actions on equipment; the system response to actions taken; and notes, such as additional needed information.

When this format is also to be used for training materials, the text-graphic pages are organized to focus the student's attention on key words, and the same pages are reproduced with the key words dropped to provide students with exercises to recall name, location, and function of each component. In addition, one or more examples of a special page format called paper mock-up is added at the end showing the complete equipment to provide a finger-tracing exercise to aid students in recalling the complete sequence of steps (Polino and Braby, 1980).

The early versions of the text-graphic page layout system (Babu and Sylla, 1981, Terrell, 1982) were page-oriented systems in which the writer chose the information to be displayed on a page. The writer input, this information to a file in the computer, then the computer generarated a layout and printed a camera-ready copy, except that pictures and darts still had to be inserted by hand. One serious problem encountered with these versions is that the writer must monitor the computer while the pages are being formatted. If the information designated for a page does not fit on the page, the writer had to remove information from the page. This overflow information then had to be added by hand to the next page, changing the content of this and perhaps subsequent pages.

The current software system for these Job Performance and Procedure Training Aids based on work by Sylla (1983) have made it possible for the computer to format an entire document without human interference and with unprecedented speed and savings (Braby, 1984). However, the required data preparation is intensive and slow, because in transforming the narrative documents and the subsequent course related documents into the procedure documents, the author still must to:

1. decide about all of the procedure steps,
2. decide how and where to illustrate from the text,
3. compose the text of labels,
4. decide the picture-label relationships
5. decide how and where to derive other supportive course materials, e.g., the instructor guide, the excercises, and the test questions and answers,...., etc.

Therefore, to assist the author to deal with the above problems we must develop intelligent software capable of:

- Accepting technical language text as inputs;
- Determining what "facts" are linguistically presupposed by the input text in the domain of interest;
- Expressing the facts in formal structures connected according to their logical interdependence;
- Recording the facts in a data base.

The organization of the data base should permit an automatic, reliable, and speedy retrieval of various kinds of information by the computer, from an input supplied in English, and to supply the required inputs directly usable by computer to produce the PPA's and JPA's and related course documentation without a human intervention. The present research is a contribution to this effort.

Clearly, with active advance research in AI, and the appreciable results already obtained in many projects (including the philosophy of language, computational semantic, and information sciences) it is possible to develop the required models and theories of algorithm and to design an efficient software with the above objective results. EXAS is such an expert authoring system software. It is being designed with the objective to remove the above labor intensive authoring tasks, and make it possible to prepare all the instructional and text-graphic materials at one computer terminal from a simple English text, without human interference. This paper describes the research effort for the development of EXAS and the results obtained to date. Subsequent discussion covers the graphic identification procedure which is the main pillar of EXAS.

EXAS Capabilities and How it Works

Initially, the author will submit a product description text and a procedure description text of his course to EXAS which will then proceed in the following chronological main steps:

1. EXAS will analyze, simplify and improve the text. This phase of EXAS is handled by a module called CRES (for Computer Readability Editing System, see Kincaid et al.,1981, Ramesh et al.,1985). CRES performs syntactical analysis, spelling corrections and other lexical analysis, and finally a readability analysis (which is mainly statistical). During this phase, CRES works with a common words dictionary. At the end of this phase, the uncommon technical words not found in the dictionary will be saved for a later review during the feedback sessions with the author.

2. If the text is for procedure training aids (PTA), EXAS will provide the following elements to the author upon request:

    a. A breakdown of the text in logical sequence of task steps (i.e., action, response, note, warning, caution,.. etc.). That is, EXAS will be able to decide the steps in the procedure and to compose the text of the labels and also identify the key words.

    b. The logical illustration sequence of the steps obtained above will also be made available. That is, EXAS will decide the picture label relationships.

3. Upon request, EXAS will generate and printout the page format layouts of the entire procedure in a text-graphic document while satisfying all the page design requirements (as done in Power PLA). A graphical terminal display of the text-graphic pages will be made available.

Step 2, described above, is centered on the graphical identification system from the two input texts to recognize the major equipment(s) in use and its components, and to capture the various relationships of interest. These relationships include the relationships between the components themselves, between the components and tasks, and between the task steps. The derivation of these relationships requires a linguistic treatment of the English texts to represent the internal knowledge, and later convert that knowledge to build the support data base for the text-graphic pages. The remaining section provides a brief description of the preliminary developments of the graphic identification procedure system which consists of a knowledge acquisition and representation phase.

The Graphic Identification Procedure System

The overall structure of the graphic identification procedure system (GIPS) is seen in Figure 2. Initially, a procedure description text or a product description text is analyzed and processed one paragraph at a time. When a paragraph is used as a processing unit, it is segmented (i.e., parsed) into key sentences through the application of linguistic, heuristic and template matching rules. Next, the sentences derived above are processed one at a time with an extensive parser to derive the needed logical elements in order to construct a semantic network representation of the knowledge included in both of the input texts. The end result of the parsing phase is two types of tree representations for the procedure description (task steps , represented as a linked list) and for the the product description (product tree).

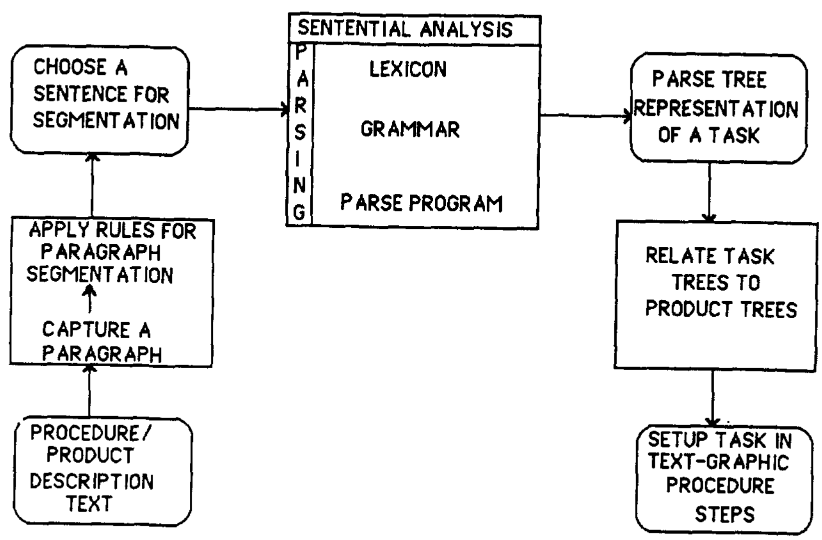

FIGURE 2. OVERALL STRUCTURE OF GIPS

The final step of (GIPS) is a synthesis phase during which the interrelationships between the resulting trees are analysed and the logical elements are assembled to derive a data base to serve as input to generate the text-graphic pages.

During the above synthesis phase the entire set of tasks which constitute one complete procedure would be linked together based on the same sequence of the sentences in the input text. This would initially form a task network. When a product name is encountered in this task network, a tree (or a network) consisting of that product and its components would be formed with links defined by the task steps (namely the verb used in the task sentence i.e., active or passive verbs). When another component is encountered, another link would be formed between that product and its component according to the task structure. Therefore at the end of this phase, all links between the product and its components would inherit the task structures that exist between the components themselve in both the procedure and product descriptions.

During all the phases of GIPS, the sentences are processed through text parsing. In order to parse the sentence, it is necessary to use a grammar that describes the structure of the strings that will be encoutered in the text. Several types of grammars (e.g., context-free, syntactic, semantic, case, etc.) are being analyzed and evaluated for possible applications. The design of the parser for GIPS has proved to be the most complex problem in both theory and implementation.

Currently, we are concerned with finding the specifications of the grammar to be employed. We expect these considerations to run into many of the general questions of computer science and AI concerning process control of input text and manipulation of the knowledge in the natural language.

References

1. Babu, A.J.G. and Sylla, C. (March 1981). COMALAY: Automatic Layout Of Naval Training Course Manual. (T.R. 81-B1); SUNYAB: Department of Industrial Engineering.
2. Braby R., Kincaid P.J., and Scott, G.P. (1984) "Computer Aided Authoring of Procedural Taining Documents", in the 6th Proceedings of Inter/Service/Industry Training Equipment Conference, pp 277-282.
3. Kincaid, J.P., Aagard, J.A., O'Hara, J.W. and Cottrell, (1981) "Computer Readability Editing System", IEE Trans. Prof. Comm. Vol. PC-24, No. 1 pp 38-4.
4. Polino, A.M., and Braby, R. (1980) Learning of Procedures in Navy Technical Training. An Evaluation of Stategies and Formats (T.R. 84), Orlando, Florida, 32813.
5. Ramesh, R., Babu, A.J.G., Kincaid, J.P. (1985) "Efficient Algorithms For Variable-Depth Trie Index Optimization: An Expert Systems Applications" A Working Paper No. 654, SUNYAB: School of Management.
6. Terrel, W.R. (1982) Computer Automated PAge Layout (PLA) For Text-Graphic Materials. User's Guide (TAEG T.R. 137), Orlando, Florida, 32813.
7. Sylla, C. (January 1983) Computer aided Layout of Job Performance and Procedure Training Aids. Unpublished Ph.D. dissertation. SUNYAB: Department of Industrial Engineering.