Acquiring Diagnostic Assembly Knowledge from Documents
For the Domain of Assembly of Aircraft Structures

Madhusudanan N.[1], Gurumoorthy B.[2] and Amaresh Chakrabarti[2]

[1]*Virtual Reality Lab, Centre for Product Design and Manufacturing, Indian Institute of Science, Bangalore, India*

[2]*Centre for Product Design and Manufacturing, Indian Institute of Science, Bangalore, India*

1 INTRODUCTION

To give a brief background on the proposed research being discussed in this paper, the domain of research is that of knowledge acquisition from documents. The implementation is planned for the domain of manual assembly of aircraft structures. This research aims to develop a knowledge acquisition system for a diagnostic system to detect potential issues in assembly during the planning stages itself. This is to avoid such issues being faced whilst during actual assembly, and to avoid iterations in assembly planning that may result. The research challenge is to acquire the required knowledge from documents and other texts. The procedure to acquire knowledge from text would be tested with texts pertaining to assembly diagnostics and the acquired knowledge is validated by applying it to example assemblies.

2 STAGE OF THE RESEARCH

The research being reported here is in the initial stages of developing a means of segmenting out relevant sections of a document. The motivation for tackling the research problem, as well as some background of related work has already been investigated. There has also been some progress made towards modeling of assembly situations so as to define them for purpose of application of acquired knowledge.

3 OUTLINE OF OBJECTIVES

As mentioned in Section 1 the goal of this research is to acquire diagnostic knowledge for assembly, from documents. Towards this, the detailed objectives of the research are,

- Identify documentary sources of knowledge about assembly in general and assembly issues in particular. Examples of such documents could be incident reports, process description sheets, standards, etc.
- Segregate sections that are potentially related to the assembly domain and issues related to assembly. The documents may be only about assembly or from some other domain, but containing some sections pertaining to assembly.
- Extract necessary knowledge pieces as required for diagnosing assemblies. Knowledge that is available in documents cannot be directly used to infer issues in assembly. They have to be further processed to structure the knowledge for making inferences.
- Translate acquired knowledge into a usable system such as a knowledge based system. This representation of knowledge would influence how the knowledge gets applied.
- Provide for suitable means of applying knowledge to assemblies. The challenge here is to define what an assembly situation means, and how to represent it for a knowledge-based system to use it.

4 RESEARCH PROBLEM

4.1 Knowledge Acquisition

The proposed research described in this paper attempts to address a long-standing problem in the field of building knowledge based systems, namely that of *knowledge acquisition*. It has long been a bottleneck in the construction of knowledge based systems, since it involves translating knowledge possessed by human experts into machine-understandable form. Such automation has been targeted previously too, with a variety of tools, methodologies and methods. Previously, there has been an effort made to acquire knowledge by means of a dialogue with experts (Madhusudanan and Chakrabarti, 2011a). This was successful in acquiring and using diagnostic knowledge. However,

N. M., B. G. and Chakrabarti A. (2013).
Acquiring Diagnostic Assembly Knowledge from Documents - For the Domain of Assembly of Aircraft Structures.
In *Doctoral Consortium*, pages 37-41

acquisition of knowledge in this manner is still difficult, due to constraints in availability and expressiveness of experts.

4.2 Documents

In organizations, documents represent the combined knowledge of one or more sources. They are usually prepared with considerable care, and are subject to reviews, and revisions. Thus they represent a refined and concise source of knowledge that has been compiled and can be considered authoritative sources. There are many different types of documents, based on the role they play, such as documentation of standard processes, best practices, instructional documents, incident reports, etc. Hence the knowledge to be acquired from a document may be influenced by which category the document belongs to. For example, diagnostic knowledge may be found in an incident report, whereas the domain knowledge needed to understand definitions and terms for the diagnostic knowledge may be found in a standards document.

In summary, documents represents the experts' knowledge in a machine-processible and comprehensive form for knowledge acquisition.

4.3 Research Challenge

Although documents serve as useful sources of knowledge, if the process requires a human to comprehend documents and then build a knowledge base, we go back to the original problem of automating knowledge acquisition. The challenge in this research is to develop a computer based tool that can *understand* the content of documents in order to acquire knowledge connected with a certain domain topic. Here, by *understanding* we mean to recognize the content and meaning of documents with a specific purpose. The purpose, in our research is to look for diagnostic knowledge (as well as other knowledge that is not directly diagnostic in nature but supplements the diagnosis, such as definitions).

The understanding may not be comparable to how a human being completely understands a given text. Human beings perform this task in a casual manner, with many factors such as experience, language skills, context and visual media to help in the process. To replicate the same factors for a machine-based language understanding system may not be practically feasible. However one could narrow their focus based on the specific knowledge that is being searched for. In this paper, the specific knowledge is diagnostic knowledge for aircraft assembly.

4.4 Research Questions

To structure the research problem discussed above the following are the detailed research questions that the research would like to address:

- What knowledge is required to diagnose issues in assembly ?
- What type of documents contain some, or all of this knowledge ?
- How can only related portions of the documents be segmented ?
- How can the required knowledge be extracted from the segments ?
- How should the extracted knowledge be used in knowledge based systems ?
- How can assemblies be represented for applying the acquired knowledge ?

5 STATE OF THE ART

As far as the state of the art in the domain, literature reports various approaches that have been used for the research challenge reported above. To summarize the readings, we discuss below the various groups of literature as per the specific issues they address.

One portion of literature looks at machine learning and such (mathematical) methods on text, with the goal of mining patterns out of strings. Garcia and Bueno (Baena-Garcıa and Morales-Bueno, 2012) discuss string pattern mining for interesting patterns in strings, using a classifier from machine learning. A measure is defined for interestingness and used to enhance the classifier performance. Also in this group are methods like multi criteria fuzzy decision making (Aydin et al., 2012) and clustering using equivalence groups for effective search(Zhang et al., 2012). Another research work for fuzzy classification of gene expressions looks for patterns in strings to classify(Khashei et al., 2012), however the semantics of texts do not seem to be involved. Hence this group of literature mostly uses mathematical methods to process string data to satisfy specific purposes.

The next group of literature is about knowledge discovery in databases (Fayyad et al., 1996; Frawley et al., 1992). This is mostly about mining patterns from databases, which is different from what is considered knowledge for our purposes. For example, things like association rules, dependancy modeling, regression and data summarization could be mined from databases(Dehuri and Cho, 2010). The idea in this group of literature is to seek out patterns in data, and convert this into knowledge.

Other notable literature include WordNet based semantic interpretation of texts using an ontology(Gomez and Segami, 2007) and context aware applications using knowledge based systems(Sánchez-Pi et al., 2012). In (Gomez and Segami, 2007) domain knowledge is constructed from texts, using a reference ontology, namely WordNet. Inferences are acquired from the meaning of English sentences. Measures have been developed to define semantic distances, as shown in (Patwardhan and Pedersen, 2006). A large collection of research aims at using existing knowledge bases such as Wikipedia for entity-topic linking(Han and Sun, 2012). This collection aims at identifying the most relevant entity to which a reference in a text can be linked to. For example, given a sentence in a document such as ”The assembly witnessed a lot of discussions“ the target is to identify the correct wiki entry for the word *assembly* (as in *legislative assembly* as opposed to *product assembly*). An attempt to acquire knowledge from a constrained syntax text of limited forms is reported in (Iwanska et al., 2000).

Regarding the modeling of assemblies for knowledge application, many previous models for characterizing situations have been used. They include Petri nets, Situation Hidden Markov Models (HMM), probabilistic frame based representation language (for modeling possibilities and types of cyber-attacks), discrete event simulations, and an integrated, object oriented definition of assembly (Madhusudanan and Chakrabarti, 2011b). Jun et. al (Jun et al., 2005) consider parts, sequence and tasks but not liaisons or process details - the latter are actually concerned with documents like assembly process sheets.

As discussed above, many of the machine learning based techniques are quite capable of classifying relevant portions of text. However the semantics of the content are largely ignored. Also, such approaches demand considerable amounts of training data. For our purposes, to acquire both the diagnostic, as well as the related domain knowledge, it is important to understand the meaning and context of documented information. Hence we chose to focus on the semantics of text, rather than process large amounts of data.

6 METHODOLOGY

6.1 Segregation of Relevant Document Portions

Referring back to the original goal of knowledge acquisition from texts, the field of *natural language understanding* provides methods for decoding the structure of sentences and in turn, their meaning. With the use of such methods, it becomes possible to separate out relevant sections from a large collection of documents and process the information further. However, even seemingly rudimentary tasks as identifying the portions related to assembly, from a larger document, are a challenge for automation. This is because when a human reads the document, a large number of factors are at work such as prior knowledge, current context of reading, prior history of how the document was obtained (what he searched for). The same factors may not be emulated when a machine tries to understand the document. As discussed in the literature (Section 5), it could involve approaches such as learning a classifier, or using a large knowledge base like Wikipedia to determine the most suitable context for a current sentence. However in the domain of assembly, resources such as a large, annotated knowledge base cannot be expected to be exist. Hence alternative sources of domain knowledge are required. Ontologies are domain-specific abstract classifications of objects in a domain. They have been extensively used in product information systems to model products and processes. Hence an ontology in the assembly domain that can cover the necessary portions of the domain that we are interested in, could serve as an abstract domain model. The details that are desirable currently include (Dawari et al., 2011),

- *Product Information* - this includes information such as the geometry of the parts, mating constraints for assembly, material information and other such details. These are typically found in the CAD files of assemblies.
- *Process Information* - this is about the assembly processes such as riveting, welding, etc. and the sequence in which the processes are carried out. Process details usually contain a description of what pre-processes and post-processes have to be performed, along with detailed task-wise steps.
- *Assembly Environment* - This concerns the conditions around the place of assembly - factors that are not classified under any other category in this list - such as the temperature of the surroundings, tool rack placement etc.
- *Tools used* - this is about the assembly set-up and tools, such as riveting gun. They may be modeled along with the part information if necessary.
- *Human Operator* - since aircraft assembly contains human involvement to a large extent, the details about operator constraints would prove useful.

Once such an ontology is available, it may be used as a reference to evaluate whether the *meaning / context* of a current portion of text is related to assembly or not.

6.2 Extraction of Necessary Knowledge

After segregating the relevant portions from a text, diagnostic knowledge must be extracted from these portions. For this, enough information that can be used as knowledge for diagnosis must be acquired. The extent of information necessary for performing diagnosis is a research question in itself, for which we already have some basis (Madhusudanan and Chakrabarti, 2011a).

When enough information is available for constructing diagnostic knowledge, a knowledge base must then be constructed from it. Translating the acquired knowledge to a knowledge base must also be done carefully, since it influences how the knowledge would be used. The choice of different types of knowledge based systems, such as rule-based systems, frames, semantic nets, or logic systems, plays a crucial role here, subject to practical constraints such as implementation.

6.3 Applying the Acquired Knowledge

Once the knowledge base is constructed, the knowledge has to be applied on assemblies to diagnose issues. This is where another research question is asked - *how does one represent assembly to the knowledge base ?* A good clue can be found in Section 6.1, where the various information related to assembly were presented. A model of assembly that can cover these aspects would be ideal to represent assembly situations for applying knowledge. The progress that has been made in this respect is presented in the Appendix.

7 EXPECTED OUTCOMES

The expected outcomes of this research are some of the following:

- A method of identifying context of text using ontologies and similar structures as reference
- A means of extracting assembly diagnostic knowledge from documents
- A knowledge base to diagnose assemblies using acquired knowledge
- A method of modeling assembly situations that covers various practical facets of assembly both as a product and a process

REFERENCES

Aydin, S., Kahraman, C., and Kaya, İ. (2012). A new fuzzy multicriteria decision making approach: An application for european quality award assessment. *Knowledge-Based Systems*, 32:37–46.

Baena-Garcıa, M. and Morales-Bueno, R. (2012). Mining interestingness measures for string pattern mining. *Knowledge-Based Systems*, 25(1):45–50.

Dawari, A., B, S., Venkatayogi, C., Chakrabarti, A., Sen, D., Gurumoorthy, B., and Appelman, H. (2011). Developing a virtual environment for aiding assessment and improvement of assemblability of aerospace structures. In *Research into Design Supporting Sustainable Product Development*.

Dehuri, S. and Cho, S. (2010). Theoretical foundations of knowledge mining and intelligent agent. *Knowledge Mining Using Intelligent Agents*, 6:1.

Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37.

Frawley, W. J., Piatetsky-Shapiro, G., and Matheus, C. J. (1992). Knowledge discovery in databases: An overview. *AI magazine*, 13(3):57.

Gomez, F. and Segami, C. (2007). Semantic interpretation and knowledge extraction. *Knowledge-Based Systems*, 20(1):51–60.

Han, X. and Sun, L. (2012). An entity-topic model for entity linking. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 105–115. Association for Computational Linguistics.

Iwanska, L., Mata, N., and Kruger, K. (2000). Fully automatic acquisition of taxonomic knowledge from large corpora of texts: Limited-syntax knowledge representation system based on natural language. In *In LM Iwanksa and SC Shapiro, editors, Natural Language Processing and Knowledge Processing*. Citeseer.

Jun, Y., Liu, J., Ning, R., and Zhang, Y. (2005). Assembly process modeling for virtual assembly process planning. *International Journal of Computer Integrated Manufacturing*, 18(6):442–451.

Khashei, M., Zeinal Hamadani, A., and Bijari, M. (2012). A fuzzy intelligent approach to the classification problem in gene expression data analysis. *Knowledge-Based Systems*, 27:465–474.

Madhusudanan, N. and Chakrabarti, A. (2011a). An interactive questioning based method to acquire knowledge for knowledge-based systems. In *International conference on trends in product life cycle, modeling, simulation and synthesis- PLMSS*.

Madhusudanan, N. and Chakrabarti, A. (2011b). A model for visualizing mechanical assembly situations. *Research into Design-Supporting Sustainable Product Development (ICoRD'11)*, pages 238–246.

Madhusudanan, N. and Chakrabarti, A. (2013). Implementation and initial validation of a knowledge acquisition system for mechanical assembly. In *CIRP Design 2012*, pages 267–277. Springer.

Patwardhan, S. and Pedersen, T. (2006). Using wordnet-based context vectors to estimate the semantic relatedness of concepts. In *Proceedings of the EACL 2006 Workshop Making Sense of Sense-Bringing Computational Linguistics and Psycholinguistics Together*, volume 1501, pages 1–8.

Sánchez-Pi, N., Carbó, J., and Molina, J. M. (2012). A knowledge-based system approach for a context-aware system. *Knowledge-based Systems*, 27:1–17.

Zhang, J., Wei, Q., and Chen, G. (2012). An efficient incremental method for generating equivalence groups of search results in information retrieval and queries. *Knowledge-Based Systems*, 32:91–100.

APPENDIX

Assembly Situation Model is a means of modeling assembly situations within the constraints of available assembly product and process information at the planning stage. At the most elementary level it represents an assembly step as a transition from a set of unconstrained parts (or subassemblies) to that of the assembled set of parts. This model can be extended to larger assemblies in the same manner as an assembly tree. An example is shown in the figure below. This model enables to represent the assembly process, parts and subassemblies at various stages in time, in different configurations. The model contains both product and process information, wherein the product information (parts and subassemblies) can be obtained from CAD assembly information, and the process information can be obtained from process sheets, and an assembly process model. This has been already been independently implemented and tested in a knowledge acquisition tool (Madhusudanan and Chakrabarti, 2013).

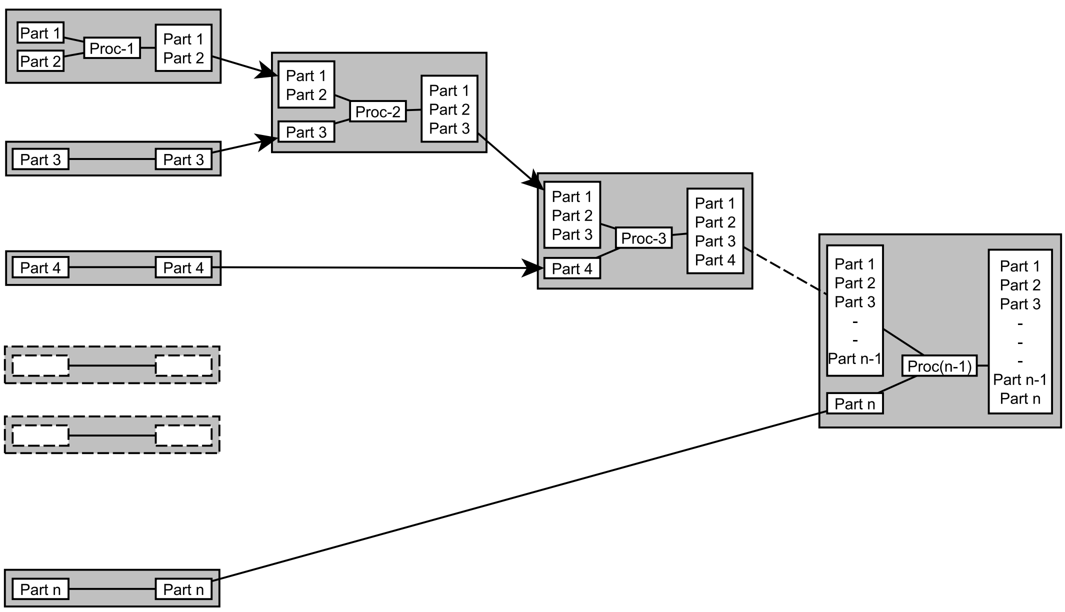

Figure 1: Assembly situation model for representing a non-trivial assembly process.