KNOWLEDGE ACQUISITION AND CONSTRUCTION

Transfer of Knowledge

• Knowledge acquisition is the process of extracting knowledge from whatever source including document, manuals, case studies, etc.

• Knowledge elicitation is a type of the knowledge acquisition where the only knowledge source is the domain expert. Techniques:
  – Interviews (unstructured to structured)
  – Protocol analysis (on-line, off-line)
  – Concept sorting
Difficulties in Knowledge Elicitation

- Technical nature of specialist fields that hinders knowledge elicitation by non-specialist knowledge engineers.
- Experts tend to think less in terms of general principles and more in terms of typical objects and commonly occurring events.
- Difficulties in searching for a good notation for expressing domain knowledge and a good framework for fitting it all together.
Stages of Knowledge Acquisition

- Identify problem characteristics
- Find concepts to represent knowledge
- Design structure to organize knowledge
- Formulate rules to embody knowledge
- Validate rules that organize knowledge

Reformulations
  - Redesign
  - Refinements

Revisions
  - Requirements
  - Concepts
  - Structure
  - Rules

Ch. 10 Knowledge Acquisition & Construction 3
# Knowledge Acquisition Document

**Elicitation Report**

<table>
<thead>
<tr>
<th>Date:</th>
<th>Knowledge Engineer:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Session#:</td>
<td>Topic:</td>
</tr>
<tr>
<td>Location:</td>
<td>Source:</td>
</tr>
<tr>
<td>Start time :</td>
<td>End time :</td>
</tr>
</tbody>
</table>

Type: 
- [ ] Interview
- [ ] Protocol analysis
- [ ] Concept sorting
- [ ] Other....................

Session Goals:

Session Summary:

Rules derived :
Document Analysis Guidelines

• Look at the document structure for how it have been organized
• Analyze the contents to extract the major linguistic categories.
• Map these categories as follows:
  – Nouns ==> objects and concepts
  – Verbs ==> relations
  – Modifiers ==> properties and values
  – Connectives ==> rules and links
Expert System Development Life Cycle

- Knowledge Engineering Methodology
- Software Engineering Methodology

The success of an expert system is affected by three main factors: Theoretical bases, practical implementation, workflow organization. This presentation ties these factors together and presents a complete methodology for the management of expert systems development.
Expert systems development goes through a number of stages that encapsulate knowledge engineering and software engineering activities. The adopted spiral model for expert systems development demonstrates the interaction between activities belonging to software and knowledge engineering paradigms.

According to this model, the development methodology consists of two main components: Knowledge Engineering, and Software Engineering. These two components are interacting with each other. In other words, they are not sequential in nature. Some phases of the software engineering methodology may be applied before the completion of the knowledge engineering part and vice versa.

As illustrated in the above figure, the adapted methodology includes three main activities, that are directed in iterations, to produce successive versions of the expert system, starting from research prototype and ending by the production version. These activities are:

- Knowledge acquisition,
- Knowledge analysis & modeling, and
- Knowledge verification.
Knowledge Engineering Methodology

• Knowledge Acquisition.
• Knowledge Modeling.
• Knowledge Verification.

**Knowledge engineering** is term used to describe the overall process of developing an expert system. The task of building an expert system involves: information gathering, domain familiarization, analysis, design, and implementation efforts.

• **Knowledge acquisition** is considered the bottleneck of the expert system building process. One of the major difficulties at this stage is to explicitly identify and capture knowledge relevant to the intended application.

• **Knowledge modeling.** A model, represents the problem solving steps, is constructed or selected. The developed models help in defining the set of domain models to be acquired from the domain expert, hence decrement unfruitful knowledge elicitation efforts, and direct the process in an organized manner. The model is used to construct the design of the target expert system.

• **Knowledge verification** is the stage whereby we make quality assurance of the acquired knowledge. Actually there are two points of interest: review procedure, and multiple expert conflict resolving procedure.
Knowledge Acquisition

• What is meant by knowledge Acquisition?
• Whom are the key personnel in knowledge engineering?

Knowledge acquisition is the most important and problematic aspects in developing expert systems. It alternately has been tagged knowledge extraction, knowledge elicitation, and knowledge acquisition. It refers to the transfer and transform of problem solving expertise from a knowledge source (e.g., human experts, books, etc.).

The process of knowledge acquisition involves a variety of personnel (e.g., knowledge engineer, domain expert, programmers).

• Knowledge engineer: is the individual responsible for structuring and/or constructing an expert system. He/she assumes the task similar to those carried out by the system analysts. Those tasks include: Analyzing information, determine program structure, working with experts to obtain knowledge, and performing design function.

• Domain Experts: is an individual selected for expertise in a given field and for his/her ability to communicate that knowledge.
Knowledge Types

- Declarative Knowledge
- Procedure knowledge
- Meta-Knowledge

**Declarative knowledge** represents surface level of information that experts can verbalize. The primary distinction between procedural and declarative knowledge focuses on the ability to verbalize or express the knowledge. It is useful in the initial stages of knowledge acquisition. But is of less value in later stages.

**Procedure knowledge** includes the skills an individual knows how to perform. The procedure for carrying out these skills are deeply embedded and linked sequentially. That is completing one step in the procedure may serve as the mental trigger to complete the next step. Consequently, these steps may be so highly complied that are difficult for the expert to identify or discuss.

**Meta-knowledge** can be described as conscious awareness of what and how we use what we know. It concerns knowledge about how to use the knowledge that we have. In another saying it is knowledge used to help domain experts to retrieve their knowledge.
Knowledge Acquisition types and Techniques

- Knowledge acquisition versus Knowledge elicitation.
- Document analysis guidelines
- Knowledge elicitation techniques
  - Interviews (Unstructured to Structured)
  - Protocol Analysis (On-line, off-line)
  - Concept Sorting

Knowledge Acquisition is the process of eliciting knowledge from whatever source including documents, manuals case studies etc.

Knowledge elicitation is a type of the knowledge acquisition where the only knowledge source is the domain expert. Therefore, several techniques are used for this purpose, e.g., interviews, protocol analysis, and concept sorting.

Document and text analysis
- look at the document structure for how it have been organized
- Analyze the contents to extract the major linguistic categories.
- Map these categories as follows:
  - nouns ----> objects and concepts
  - verbs ----> relations
  - modifiers ----> properties and values
  - connectives ----> rules and links
Interviews

• Need the willing co-operation of the expert
• Selecting an expert
• Preparation

• This technique, as with all techniques need the willing co-operation of the expert. There are potential obstacles to this. For instances:
  • Status differences
  • Age differences
  • Differences of interest
• Selecting an expert : if possible he/she should:
  • have recent practical experience
  • be communicative and articulate
  • be easy to work with
  • have management support to commit time to the project.
• Preparation: This activity includes:
  • Identifying exact function of the proposed ES
  • Identifying the end-users
  • Studying the domain background
  • Arranging Pre-KA meetings
Types Of interviews

• Unstructured interview
  – Not planned sessions
  – used in the first knowledge acquisition stages
  – For identifying and understanding the problem

• Structured interview
  – Planned sessions
  – time scheduled
  – Probes specified
  – KE in control

• **Unstructured interview**: Given the general knowledge acquisition session goal, the expert functions as a lecturer. The KE asks questions to clarify understanding and take notes in an outline format. This interview method is often used in the initial knowledge acquisition stages. Problems include: lack of focus, dependence upon the domain expert’s ability to “teach”, and the possibility of not using expert’s time efficiently.

• **Structured interview**: The KE outlines specific goals and questions for the knowledge acquisition session. The expert is provided with session goals and sample lines of questioning. This type of interview is a mainstay of knowledge acquisition. It is used in all phases of the process to clarify or extend information received via other techniques.
Protocol Analysis

- Asking expert to report on or demonstrate their decision making process for specific problem.
- Behavioral may include self-report
- On-line or Off-line

Protocol analysis is, in fact, a set of techniques that allows KE to determine a domain expert’s *train of thought* while he/she completes a task or reaches a conclusion.

Protocol analysis requires that KE use one of several techniques to analyze the protocols. For example, in one type of protocol analysis session the domain expert may be asked not only to solve a problem, but also to *think aloud* while doing so which is called then *On-line* protocol analysis type. On the other hand, *Off-line* type of protocol analysis means that verbal protocol will be gathered after solving the problem given to him/her.
Concept Sorting

- Set of cards
- Expert sorts into piles along salient dimensions
- If expert dries use triadic presentation

Concept sorting is a psychological technique that is useful in tapping organization knowledge.

To apply this technique, the KE follows the following steps:

1- First, the KE consults a textbook, training manual, or in-house domain expert to identify the major top-level concepts represented in the domain.
2- Place each concept on a note card
3- Next, the KE asks the domain expert to begin sorting these cards placing them in groups according to those that belong together.
4- As the domain expert sorts the cards the KE uses questioning techniques to determine why they are placed together.
5- Repeat steps 3, and 4 until the expert dry up.
6- If the expert dries up, The KE applies the triadic presentation, in which any three cards are taken, and then by asking the domain expert for giving information about the relation between any two of them such that this relation is not hold for the third one.
Model Driven Knowledge Acquisition

• Knowledge modeling
• Model instantiation
• Model validation

Improving the process of knowledge acquisition has been important motivation in developing second generation ES. It is thus not surprising it has been a very active field of research during the last decade. In the first generation ES, knowledge acquisition was seen as a problem of transferring knowledge: extracting the expert's knowledge and translating it into the implementation language constructs (rules). Knowledge acquisition is now considered to be a modeling task, composed at least of three distinct phases: building a model of KBS (modeling), filling the model with domain knowledge (instantiation), and finally validating the developed KB.
Knowledge-level models are useful for knowledge acquisition since they often fill the gap between expert's discourse and implementation. Modeling, however, is rarely done from scratch, on the contrary, new models are often created by adapting and refining generic, or at least previously developed, models. Generally, there have been three different approaches for model construction namely: data-driven modeling, select-and-modify, and Compositional modeling from library elements.
Data-driven Modeling

- Knowledge model is built from scratch
- Widely used in first generation ES
- Time consuming

Data-driven modeling: This denotes the approach, widely used in the first generation ES, whereby a knowledge model is built up from scratch on the bases of expert data stemming from interviews, protocol analysis, task observation, and other techniques. In its extreme form, no predefined models from libraries or literature are employed, but model construction is fully driven by data acquired from domain experts.
Select-and-modify approach: This refers to the approach, whereby a predefined and rather complete model of expertise is selected from a library, and subsequently adapted to the application needs.

The above figure shows that, an initial problem solving model is selected from library of complete generic problem solving models. This selection is done based on the available task features. The selected model is then validated with the domain experts by tracing its behavior against the desired behavior, and recording any discrepancies. These discrepancies form the bases of any further modifications. The process of validation and modification loop until we have an empty discrimination list, in this case the model is used to guide the knowledge acquisition process.
Compositional Modeling

- Second generation ES
- Library of generic components is available
- More powerful

**Compositional modeling from library elements**: This indicates the approach whereby an expertise model is constructed in an incremental fashion from available generic components, as in CommonKADS library. The grain size of the used library elements is typically lower than in the select-and-modify approach, whereas variety and flexibility are larger.
Model Instantiation

• Knowledge acquisition is started for filling the model obtained from the first phase
• The model indicates what knowledge should be acquired
• Automatic knowledge acquisition tools may be designed to be used for particular model

The second phase of the knowledge acquisition process is driven by the model built in the first phase. This model indicates what knowledge should be acquired and clearly defines the role each piece of knowledge will play in the problem solving process. Using the models to guide the knowledge acquisition process has been proven to be very powerful technique. Many of the tools that support the knowledge acquisition process can be classified as dedicated to a particular model (e.g., SALT, MOLE).
Model Validation

- The validation of a knowledge base involves validating:
  - The model itself
  - The acquired knowledge

The validation of a knowledge base involves validating both the model itself and the acquired knowledge. Both can benefit from a knowledge-level approach. Knowledge-level models act as a functional specification of the system that can be discussed with the experts. They also provide a precise, unambiguous meaning to the acquired knowledge.
Diagnostic Problem Solving Example

- **Example scope**: Given four different disorders that affect Tomato Crop, and we need to identify which disorder exists.
- **Input**: Set of observations.
- **Output**: One or more identified disorder

For simplicity, this example assumes that there are only four disorders affect the Tomato plant. These disorders are: root-rot, drought, white fly, and Alternaria leaf.spot.

We need to develop the appropriate reasoning model that can explain the reasons for the given abnormal plant observations.
Expertise Model Construction

• Expertise model consists of:
  – Task knowledge
  – Inference knowledge
  – Domain knowledge

The knowledge model, also called expertise model, consists of three category of knowledge:

• **Domain knowledge**: as the factual knowledge about the application domain.

• **Inference knowledge**: as knowledge about how the domain knowledge can be applied in the reasoning process, thus serving as a kind of interface between the task and domain knowledge.

• **Task knowledge**: as the knowledge about the control of the reasoning process such that a solution can be found in an effective and efficient way.
Several diagnostic strategy could be applied for diagnosis. The strategy shown above called Hierarchical Classification (HC).

The above figure shows the relation between task, and inference knowledge in this diagnostic strategy. The task knowledge indicates that there are two main sub-tasks namely: generate hypotheses, and test hypotheses. The former sub-task’s goal is to generate a set of hypotheses based on some initial observations. The latter sub-task’s goal is to test, and hence, omit those disorders which are not consistent with the rest of observations. The task knowledge also indicates the control over these subtasks. This control shows that these two sub-tasks are in loop until we could reach a conclusion.

Each sub-task is mapped to its corresponding inference. For instance generate hypotheses is mapped to establish hypotheses, and test hypotheses is mapped to refine hypotheses. Each inference has its input and output, and operates on a specific domain model. For instance establish hypotheses takes the initial observation (complaints) and produces a set of hypotheses. These hypotheses is taken as input to the refine hypotheses inference step and omit those hypotheses which are not consistent with the rest of observations. Both inferences uses domain knowledge called disorder hierarchy model.
Hierarchical Classification (HC) problem solving method requires that the domain knowledge should be organized in a tree-like structure. Each node in this tree represents a cluster of disorders that share common observations (symptoms) which is called specialist.

The specialist at top of the hierarchy represent the most general hypotheses, with more and more specific sub-hypotheses distributed in layers beneath. The control regime is a top-down establish and refine mechanisms, in which each specialist, when invoked, attempts to determine if the evidence of the current case supports the diagnostic hypothesis the specialist represents (i.e. it attempts to establish itself) and then if it establishes, it will call on its sub-specialists to refine the hypothesis (i.e. it attempts to refine itself). By pruning the hypothesis space at high levels of the generality, establish-refine cuts through some of the computational complexity inherent in the diagnostic problem.