3 Knowledge Acquisition and Validation

Lecture Overview

1. Sources of Knowledge
2. Difficulties in knowledge acquisition
3. Methods of knowledge acquisition
4. Knowledge base – verification and validation

Terminology

- **Knowledge** is a collection of specialized **facts**, **procedures** and **judgment rules**

- **Domain:** some area of interest e.g. banking, food industry, photocopiers, car manufacturing

- **Domain Expert (DE):** A DE is someone who has a prolonged or intense experience through practice and education in a particular field

- **Knowledge Engineer (KE):** KE translate information elicited from domain experts into terms which cannot be easily communicated by the highly technical domain expert. KE interpret and organize knowledge in how to make expert systems

Knowledge Engineering (KE)

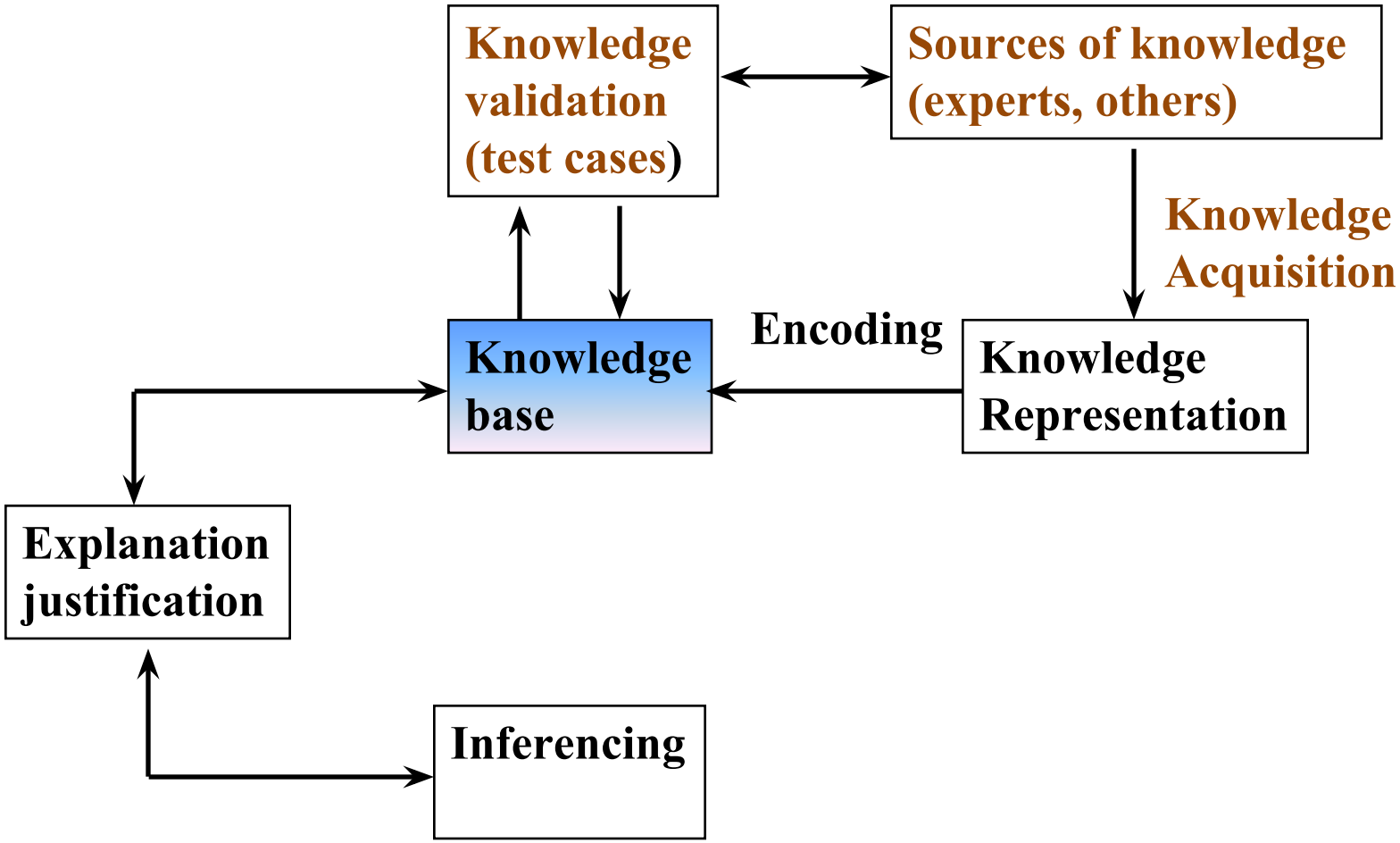
- KE is the art of bringing the principles and tools of AI research to bear on **difficult application problems** requiring experts' knowledge for their solutions
- KE involves technical issues such as acquiring, representing and using knowledge appropriately to construct and explain lines-of-reasoning
- KE is the art of building complex computer programs that **represent** and **reason** with knowledge of the world

  – (Feigenbaum and McCorduck [1983])

Knowledge Engineering Process Activities

1. **Knowledge Acquisition**
2. **Knowledge Validation**
3. Knowledge Representation (topic 4)
4. Inferencing (topic 8)
5. Explanation and Justification

The Knowledge Engineering Process



Knowledge Acquisition definitions

1. Knowledge acquisition is the process of extracting, structuring and organizing knowledge from one or more sources, usually **human experts**, so it can be used in software such as an ES. This is often the major obstacle in the development of expert systems and other A.I systems.

2. Knowledge acquisition is the extraction of knowledge from *sources of expertise* and its transfer to the knowledge base and sometimes to the inference engine.

Knowledge Acquisition

- Knowledge acquisition is readily **acknowledged to be the most difficult aspect of developing an expert system.**
- Obviously, the **selection of an appropriate expert** and the **correct use of an appropriate knowledge acquisition technique** can have a significant impact on the ultimate success of the expert system
- **The success of knowledge based systems lies in the quality and extent of the knowledge available to the system.**
- Extracting knowledge from a human is often called knowledge **elicitation**

Knowledge Sources

- Documented (**books, manuals**, reports etc.)
- Undocumented (in people's minds) – **From people, from machines**
- Knowledge Acquisition **from Databases**
- Knowledge Acquisition **Via the Internet**

Knowledge Levels

- **Shallow knowledge :** A representation of only surface level information that can be used to deal with very specific situations

- **Deep knowledge :** A representation of information about the internal and causal structure of a system that considers the interactions among the system's components

Knowledge Levels (2)

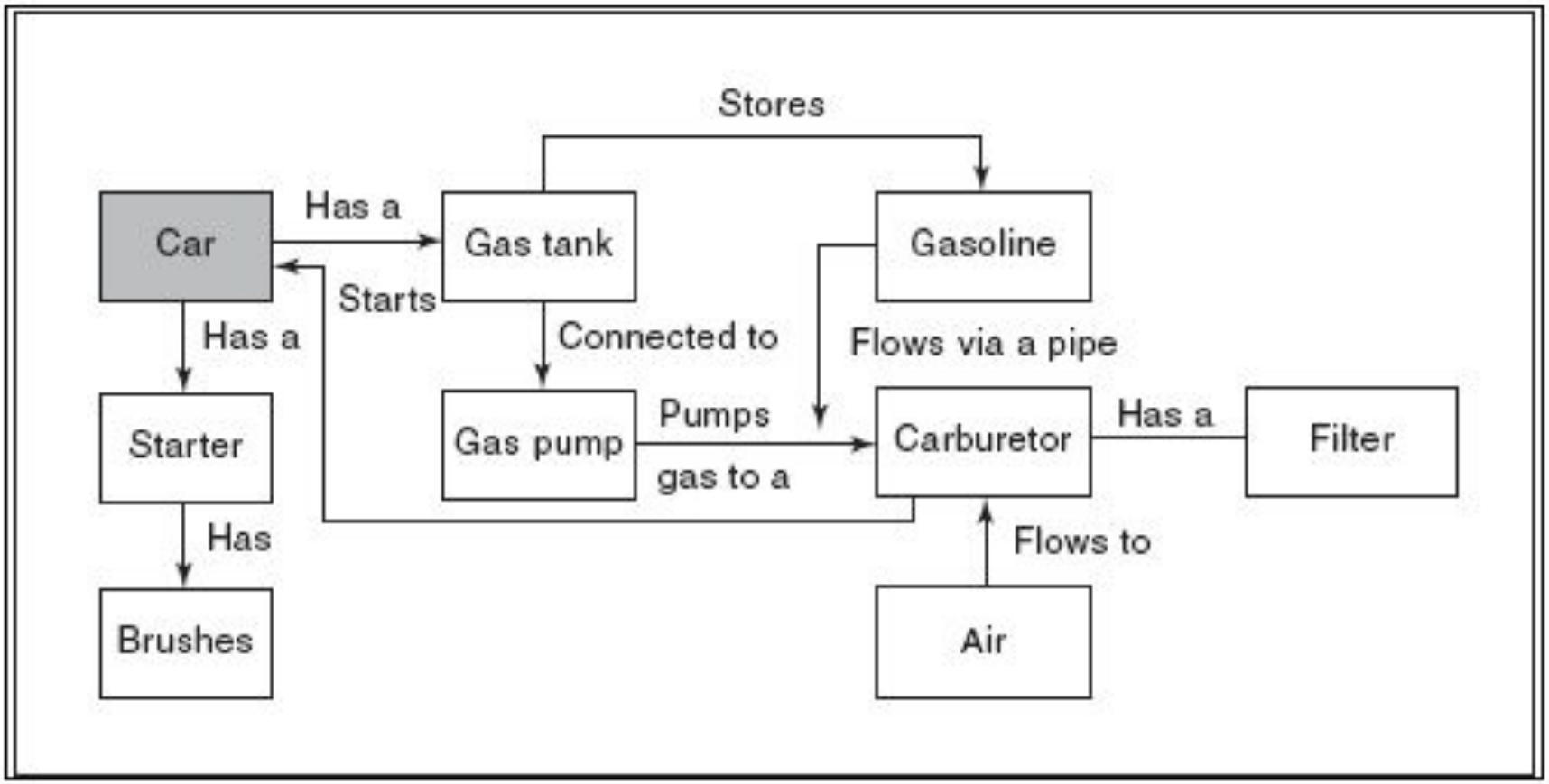


FIGURE W18.3 Schematic Representation of Deep Knowledge: Automobile Gas System

Types of Knowledge

1. **Declarative Knowledge -** tells us the facts
   - It is the descriptive representation of knowledge
   - It relates to a specific object. It Includes information about the meaning, roles, environment, resources, activities, associations and outcomes of the object
   - It is often shallow knowledge, expressed in factual statements, and is important in the initial phase of knowledge acquisition

Types of Knowledge (2)

2. **Procedural Knowledge** - tells us what to do
   - This is knowledge about **procedures employed** or **involved** in the **problem solving phase**
   - It considers the **manner in which things work under different sets of circumstances**
   - It includes **step-by-step sequences** and **how-to types of instructions**
   - It may also **include explanations** and may also tell how to use declarative knowledge and how to make inferences

Types of Knowledge (3)

3. **Meta-knowledge**
   - This is Knowledge about Knowledge.
   - In ES, It refers to **knowledge about the operation of knowledge-based systems** and its **reasoning capabilities**

Tacit vs. Explicit

- Explicit knowledge refers to knowledge that is transmittable in formal, systematic language. This is knowledge of concepts, relations, routine procedures, facts , heuristic; and classificatory knowledge

- Tacit (aka implicit) knowledge is deeply rooted in actions, experience, and involvement in a specific context. It consists of cognitive element (mental models) and technical element (know-how and skills applicable to specific work).

Knowledge Acquisition Paradox

- The more competent a Domain Expert (DE) becomes, the less able they are to describe the knowledge they use to solve problems

3.2 Difficulties in Knowledge Acquisition

Problems in Transferring Knowledge

- Problems in
  1. Expressing Knowledge
  2. Transferring Knowledge
  3. Structuring Knowledge
  4. Transfer to a Machine
  5. Number of Participants

1. Difficulty in Expressing the knowledge

- A human expert performs a two-step process to solve a problem. 1)Input information about the external world into the brain. 2)The expert then uses an **inductive, deductive**, or other problem-solving approach on the collected information, and finally output a **recommendation** on how to solve the problem.
- Because this process is usually internal, the Knowledge Engineer (KE), must ask the expert to be introspective about his decision-making process and the inner experiences that are involved in it. It is very difficult for an expert to express this process, especially when these experiences are made up of sensations, thoughts, sense memories, and feelings. The expert is often unaware of the detailed process used to arrive at a conclusion. Therefore, the rules used by the expert to solve real-life problems may actually be different than those stated in a knowledge acquisition interview.

2. Difficulty in Transferring Knowledge

- **Transferring information is difficult**
  - Transferring mechanisms—**written words, voice, pictures, music**…—no one of them is perfect.
  - There are also problems in transferring any knowledge, even simple messages.
  - Transferring knowledge in ES is even more difficult.

3. Difficulty in Structuring Knowledge

- In ES it is necessary to elicit not only knowledge but also its **structure**. We must represent the knowledge in a structured way. (e.g., as rules).
- We can thus lose a significant amount of knowledge when structuring implicit knowledge

4. Difficulties in transferring Knowledge to a machine

- Machines works at a more **basic level**, but the **human expert seldom operates at a basic level.**
- Knowledge is transferred to a machine in a very organized and particular manner. Humans on the other hand, do not always remember all the intermediate steps used by their brains in transferring or processing knowledge.
- Thus, there is a mismatch between computers and experts.

5. Number of participants

- In a normal transfer of knowledge process, there are usually two participants. In an ES, however there can be four participants (plus a computer): domain expert, **knowledge engineer**, **system designer**, **user**. Some times more participants (programmers and vendors).
- These all **have different backgrounds**, **use different terminology**, **possess different skills** and **knowledge**.
- The **domain expert** may thus **know very little** about **computers**, in addition the **knowledge engineer** may **know very little about** the **problem area** and that can pose as a huge challenge.

Other Reasons

1. A domain expert may **lack time (**i.e. be unavailable) or **uncooperative**
2. **Testing and refining** knowledge is complicated
3. Poorly defined methods for knowledge elicitation
4. System builders may collect knowledge **from one source**, but the **relevant knowledge** may be scattered across several sources
5. May collect documented knowledge rather than use experts

Other Reasons (2)

6. The knowledge collected may be incomplete
7. Difficult to recognize **specific knowledge** when mixed with irrelevant data
8. Experts may change their behaviour when observed and/or interviewed
9. **Problematic interpersonal communication** between the knowledge engineer and the expert

Overcoming the Difficulties

- Use of Knowledge acquisition tools with ways to decrease the representation mismatch between the human expert and the program ("learning by being told")
- Simplified rule syntax
- Natural language processor to translate knowledge to a specific representation
- Impacted by the role of the three major participants: **Knowledge engineer**, **domain expert** and **end user**

Overcoming the Difficulties (2)

- The domain experts should take a very active role in the creation of knowledge base.
- The knowledge engineer should act more like a teacher of knowledge structuring, a tool designer, and a catalyst at the interface between the expert and end users.
- If the above 2 are done, this could minimize problems such as inter-human conflicts, knowledge engineering filtering, and end-user acceptance of the system and reduce knowledge maintenance problems.
- Think of the participants as playing one or more roles, acting as knowledge sources, agents, targets for knowledge acquisition processes.

Overcoming the Difficulties (3)

- Critical is the **ability** and **personality** of the knowledge engineer.
  - The KE, must develop a positive relationship with the expert
  - The KE must create the right impression
- Critical are also **computer-aided** knowledge acquisition tools

Required Knowledge Engineer Skills

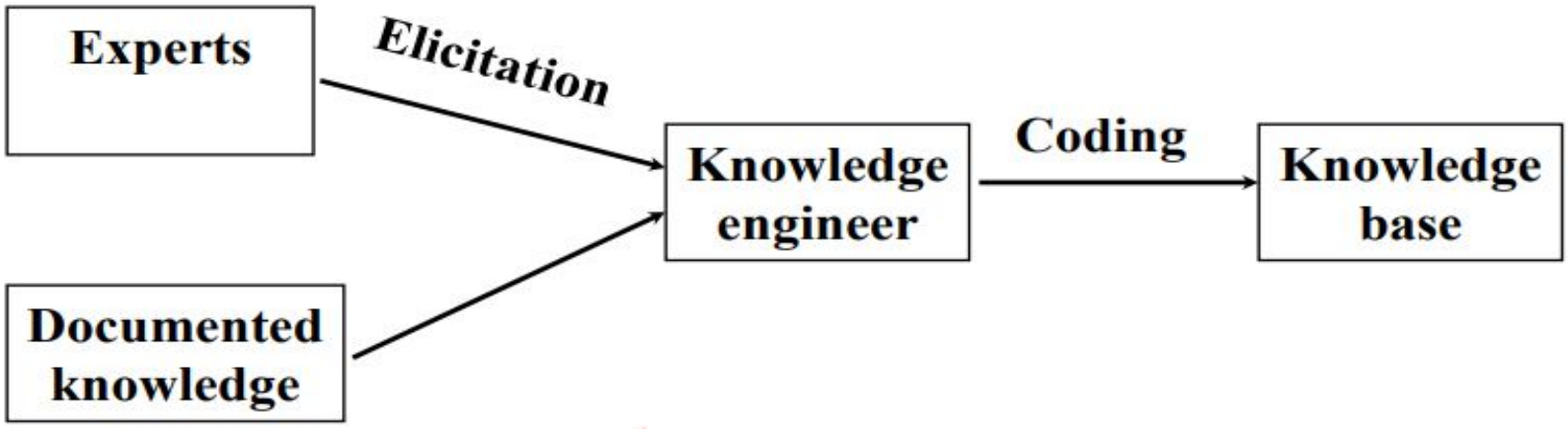
- Computer skills
- Tolerance and ambivalence
- Effective communication abilities
- Broad educational background
- Advanced, socially sophisticated verbal skills
- Fast-learning capabilities (of different domains)
- Must understand organizations and individuals
- Wide experience in knowledge engineering
- Intelligence
- Empathy and patience
- Persistence
- Logical thinking
- **Versatility** and inventiveness
- Self-confidence

Knowledge Acquisition Methods: An Overview

1. Manual
2. Semiautomatic
3. Automatic (Computer Aided)

1. Manual Methods

- A human-intensive method for knowledge acquisition, such as interviews and observations, used to elicit knowledge from experts
- *Here Domain Experts* and *Knowledge Engineers* are actively involved



1 Interviews

- The most common method of knowledge acquisition is the face-to-face interview. Through conversation, experts are asked to verbalize how they solve problems
- Information is collected most often with the aid of a tape recorder, and subsequently transcribed, analysed, and coded

Interviews (2)

- AI researchers have found that the interview is one of the most important tools for facilitating the transfer of human knowledge
- Most knowledge engineering sessions begin with an informal interview to get acquainted with the expert and to gain a basic understanding of the basic structure of the problem domain
- Once this has been accomplished, a more structured knowledge acquisition technique is employed

Interviews (3)

- The greatest advantage of using the interviewing technique is that it is a natural process which is easily understood by both the knowledge engineer and the expert .
- However, interviewing is often more than just simply sitting down and talking with an expert. The interview relies on the expert's ability to articulate the information used to work through a task. UNFORTUNATELY, experts often have a difficult time verbalizing how they go about solving problems
- As a result, the interview is not always a reliable way to obtain complete, objective, or well-organized descriptions of complex cognitive processes.

2. Questionnaires

- Questionnaires can be a very effective and efficient method of accessing an expert's knowledge. Questionnaires are useful for acquiring **explicit knowledge**.
- They can be used to determine objects from the knowledge domain, relationships among those objects, and uncertainties about those relationships

Questionnaire (2)

- The major **advantages** of using a questionnaire are that it is less time consuming for the knowledge engineer than the interview, it is an efficient method for gathering information, and the expert can answer the questions at the expert's leisure.
- The major **disadvantage** of a questionnaire is that it is unable to pursue unanticipated information

3. Task Observation.

- Task observation involves observing experts work at real problems to determine how they make decisions .
- Using the task observation method, the observer will watch the **behaviour** and **activities** of the experts as they typically proceed through a problem, and then the observer can ask questions.
- Recording the experts' performance can be accomplished by simply taking notes, using a tape recorder, or even videotaping the process.

Task Observation (2)

- Task observation is a straightforward approach to knowledge acquisition and, like the interview, is easily understood by both the expert and the knowledge engineer.
- Task observations allows the KE to acquires tacit knowledge through practical domain experience with the expert
- Observations of actual performance can reveal the inference strategy and can also correct misleading or incomplete verbal descriptions of the problem solving process

Task Observation (2)

- Task observation is not suited to achieving all knowledge acquisition goals, and there are **LIMITATIONS** to this method.1 Access to the people and places to be observed is foremost among the problems .
- 2 The problem may also take a considerable amount of time to solve
- 3 Experts may act unnaturally when they are aware that they are being watched
- Additionally, all activities during the observation period are being observed. As a consequence, 4 large quantities of data may be collected from which little actual problem solving knowledge may be useful

4. Protocol Analysis.

- Similar to task observation, protocol analysis involves having an expert perform actual or simulated problem solving scenarios
- Unlike task observation, however, the expert is asked to provide a running verbal commentary on his thought processes as he solves the problem
- Protocol analysis, presents the expert with a series of realistic problems to solve aloud, probing for the rationale behind the reasoning steps (solve the problem verbally)

Protocol Analysis (2)

- One of the knowledge engineer's primary roles is to keep the expert talking, but not to ask specific questions. A detailed analysis of the subsequent transcriptions provides the facts and rules to be used in the knowledge base. Protocol analysis can be used to collect both implicit and explicit expert knowledge

Protocol Analysis (3)

- The main advantage protocol analysis has over the interview and task observation is that information collected is directly related to the problem solving process, and the knowledge engineer is not required to infer the steps involved in solving the problem.
- Protocol analysis is limited to processes that lend themselves to verbalization. Protocol analysis requires a time consuming dissection of the transcripts to produce a usable model of the expert's knowledge

Protocol Analysis (4)

- Other weaknesses of protocol analysis are that the protocols may fail to tap the full range of an expert's knowledge, and the very act of verbalizing the problem solving process may affect the actual way an expert solves a particular problem
- Widely used in psychology, Ex: Dermatology-Psoriasis, an Expert System to diagnose Psoriasis asks questions like how long rash lasts? Where is the rash? etc,.

5. Interruption Analysis.

- Interruption analysis is another method for accessing explicit forms of expert knowledge.
- Like task observation, the expert is observed solving relevant tasks. The expert is not asked to provide a verbal commentary on his decision making process, ***but whenever the expert does something that the knowledge engineer does not comprehend,*** the knowledge engineer interrupts the expert and asks exactly what the expert did

Interruption Analysis (2)

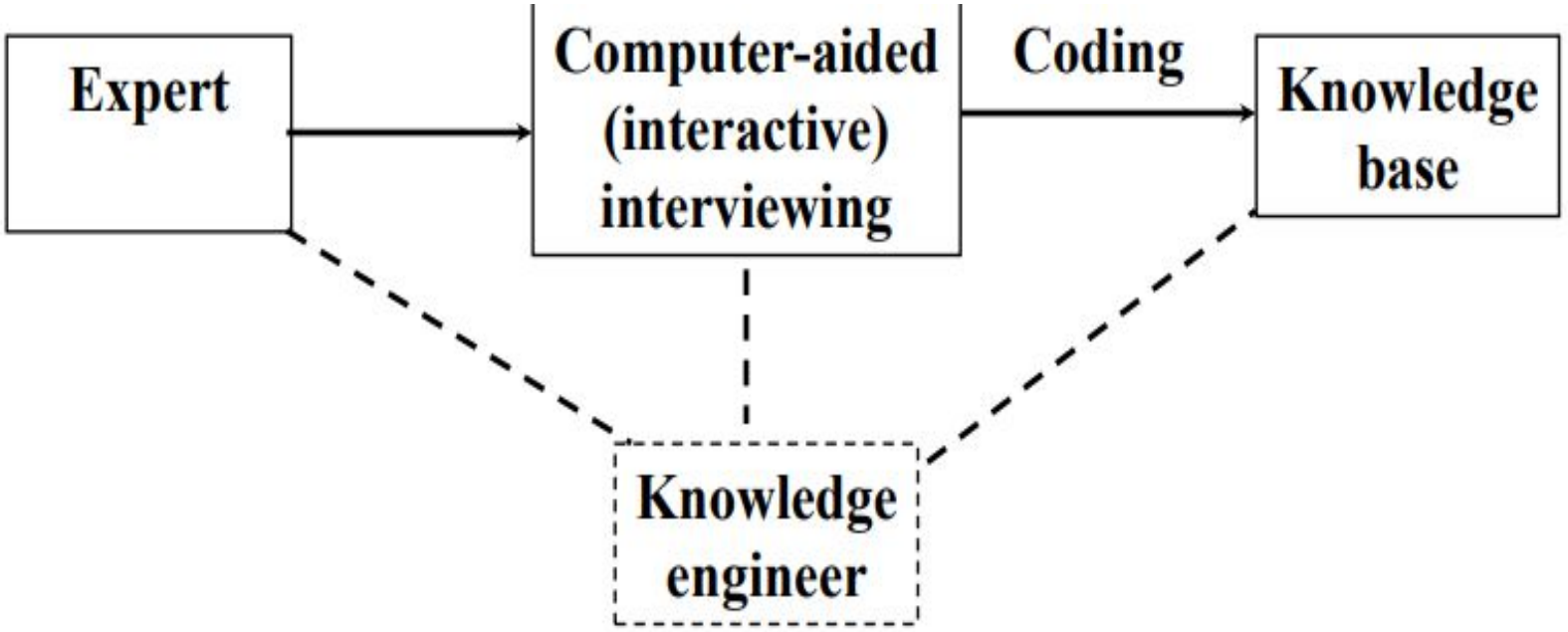
- The main advantage of interruption analysis is that the knowledge engineer is able to capture an expert's knowledge **"at the moment the focus of attention" is greatest**. This advantage comes at the expense of interrupting the thought process and risking not being able to restart it.
- This technique provides the best results when used after a prototype system has been developed, and the system's performance is being compared to that of an expert

Other Manual Methods

- Case analysis
- Critical incident analysis
- Discussions with the users
- Commentaries
- Conceptual graphs and models
- Brainstorming
- Prototyping
- Multidimensional scaling
- Johnson's hierarchical clustering
- Performance review

2. Semi-Automatic Methods

- A knowledge acquisition method that uses computer-based tools to support knowledge engineers in order to facilitate the process



Semi-Automatic Methods (2)

- **Knowledge Engineer Support**
  - Knowledge Acquisition Aids
  - Special Languages
  - Editors and Interfaces
  - Explanation Facility
  - Revision of the Knowledge Base
  - Pictorial Knowledge Acquisition (PIKA)

Semi-Automatic Methods (3)

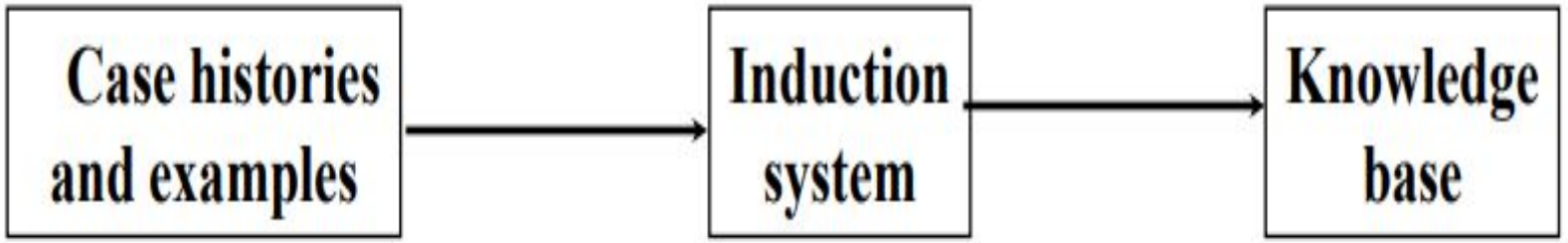
- Integrated Knowledge Acquisition Aids
  - PROTÉGÉ-II
  - KSM
  - ACQUIRE
  - KADS (Knowledge Acquisition and Documentation System)
- Front-end Tools
  - Knowledge Analysis Tool (KAT)
  - NEXTRA (in Nexpert Object)

3. Automatic Methods

- Manual and semiautomatic elicitation methods: slow and expensive
- Other Deficiencies
  - Frequently weak correlation between verbal reports and mental behavior
  - Sometimes experts cannot describe their decision making process
  - System quality depends too much on the quality of the expert and the knowledge engineer
  - The expert does not understand ES technology
  - The knowledge engineer may not understand the business problem
  - Can be difficult to validate acquired knowledge

Automatic Methods

- An automatic knowledge acquisition method that involves using computer software to automatically discover knowledge from a set of data
- Here expert's and/or the knowledge engineer's roles are minimized (or eliminated)



Automatic Methods

- Automatic knowledge acquisition is the process of using computers to extract knowledge from data. This is sometimes called machine learning or *knowledge discovery*
- Two reasons for the use of automated knowledge acquisition: 1 Good knowledge engineers are highly paid and difficult to find 2 Domain experts are usually busy and sometimes uncooperative

Computer-aided Knowledge Acquisition, or Automated Knowledge Acquisition Objectives

1. Increase the productivity of knowledge engineering
2. Reduce the required knowledge engineer's skill level
3. Eliminate (mostly) the need for an expert
4. Eliminate (mostly) the need for a knowledge engineer
5. Increase the quality of the acquired knowledge

Automatic Methods (2)

- **Rule Induction**
  - Knowledge Discovery and Data Mining
  - Include Methods for Reading Documents and Inducing Knowledge (Rules)
  - Here the DE provides some examples similar to data mining, then **Statistical/Mathematical techniques** such as Multivariate Regression or Artificial Neural Net (ANN) are applied to solve the problem

Automatic Methods (4)

- **Case-based Reasoning**
  - For Building ES by Accessing Problem-solving Experiences for Inferring Solutions for Solving Future Problems
  - Here cases and resolutions Constitute a Knowledge Base
  - This method works by making use of previous cases: Example: Help Desk » Printer not functioning » Refer to previous case from “n” weeks ago

Automatic Methods (5)

- **Model-based Reasoning**
  - Mainly applicable to design of an engineering application
  - Give me specifications of some hardware
  - Used often in NASA
  - Build a model using DE knowledge

Acquiring Knowledge from Multiple Experts

- Major purposes of using multiple experts:
  - To better understand the knowledge domain
  - To improve knowledge-base validity, consistency, completeness, accuracy, and relevancy
  - To provide better productivity
  - To identify incorrect results more easily
  - To address broader domains
  - To be able to handle more complex problems and combine the strengths of different reasoning approaches

Knowledge Verification and Validation

- Knowledge acquired from experts needs to be evaluated for quality. Quality control includes evaluation, validation and verification.

1. The main objective of ***evaluation*** is to assess an ES's overall value. Evaluation analyses whether the system would be **usable, efficient** and **cost-effective**

Knowledge Verification and Validation (2)

2. ***Validation*** is the part of evaluation that deals with the performance of the system as compared to the expert. It tries to answers the question, **was the "right" system built** (acceptable level of accuracy?)

3. ***Verification*** *answers the question,* **was the system built right?** i.e. is it correctly implemented according to the user requirements or specifications

To Validate an ES

- Test
  1. The extent to which the system and the expert decisions agree
  2. The inputs and processes used by an expert compared to the machine
  3. The difference between expert and novice decisions