Cognitive Building Blocks of a Database Expert System

Dale Lee Harris
AT&T Network System
Network Software Center
2600 Warrenville Rd
Lisle, IL 60532

Abstract
This paper addresses the practical application of cognitive psychology. It first gives a model for how human thought abstraction works. It then shows how that model was used as the framework for the design of the ATHENA Expert System. The system is intended to assist engineers in quality assurance of database design.

Building Blocks in Cognitive Science
To develop a software system that performs the same job a human does, it is best to first understand how the human does the job. The greater the insight into how the human expert performs, the better one can expect the software system to be. The best expert system can be expected to be one that is constructed from the building blocks of cognitive science.

Cognitive psychology is the theoretical foundation of expert systems and gives a basic grasp of how the human system works. Of particular importance is an understanding of the levels of human information or memory.

Permanent memory has been categorized by time (life experience) and by concepts (ideas). "As this fact has become clearer to psychologists, a distinction has been drawn between two types of permanent memory, episodic memory and semantic memory (Tuiling,1972)." Episodic memory refers to the memory a person has for temporally dated events and for relations between these events. "Semantic memory, on the other hand, refers to the organized knowledge a person has about words, symbols, formulas, concepts, and rules." An object is an entity which can be described in terms of its features. The interaction of objects in a time frame is an event.

The time order of existence of objects and occurrence of events is the life experiences about objects. Grouping objects together by their features or by the way they interact with other objects is development of concepts about objects.

Understanding of ideas can be viewed as being either conceptual association of objects or event ordering. "The ACT model proposes a basic distinction between two kinds of knowledge which humans have. One kind, called declarative knowledge, refers to facts, concepts, and beliefs we have about our environment.... The second kind of knowledge, called procedural knowledge, refers to that which we know how to do." A concept can be thought of as the grouping of objects into categories. A procedure is an ordered occurrence of events.

One can view information in the human system to be on several levels of abstraction. The three lowest levels may be referred to as: data, knowledge, and rules. Data is stored information which describes an object. For example "Jack is 6 feet tall" is data. Knowledge consists of associations between objects. It can be time dependent or time independent. The knowledge "Jack is taller than Jill" is an object association not related to time and is a concept. The knowledge "Jill is a human female" is also a concept, since it relates the object "Jill" to the object class "human female". The knowledge "Jack tumbled down the hill" is a time related interaction between objects and is an event. An event can be viewed as a relationship between objects in a given time frame. Rules are relationships between events or between concepts. Rules can be procedures to follow under given conditions to be met, that is, an event chain. Rules can also be formulas that tell how to derive new knowledge, concepts, from existing knowledge.
Objects are the lowest level of information abstraction. Associating objects together, either by a shared feature or a shared event interaction, can be viewed as the next higher level of information abstraction. Concept derivation and procedures can be viewed as the next higher level of mental abstraction. The three lowest levels of memory can be viewed as:

- rules The time order of events or the logical implication of causality.
- knowledge The association of objects or events together by common properties.
- data The detailed description of objects or events, stored mostly in temporary memory.

These levels of information can be viewed to form the first three levels of information abstraction. In ascending order the levels are:

1. Feature descriptions of objects and events.
2. Associations between objects, events, and features.
3. Logical implication or time order of events.

Knowledge flows from descriptions, to associations, and finally to implications. An example of the first three levels of knowledge abstraction, can be given using arithmetic objects.

1. \((x = 6), (y = 7)\)
2. \((y > x)\)
3. if \((z > y)\) then \((z > x)\)

In the human mental process, information at lower levels is processed to create information at higher levels of abstraction. There are certain procedures or rules that govern the upward flow of knowledge toward higher levels of abstraction. In the above example it is the rules of mathematics which control the upward flow or abstraction of information. From the fact that \((x=6)\) and \((y=7)\) and the definition of numeric order, one can derive the association that \((y > x)\). From the association that \((y > x)\) and the transitivity of numeric ordering, one can derive the implication that \("(z>y)\) implies \((z>x)\)". Thus in the discipline or knowledge domain of mathematics, the laws of arithmetic govern the derivation of new knowledge from existing knowledge and the flow of knowledge from lower to higher levels of abstraction.

Knowledge Flow in Database Design

To provide AI support to the human process of quality checking the formal specification of a database design, it is necessary to first understand the human mental process used. The term cognitive ergonomics means design of a system to reflect and adapt to the mental process and methodology that humans use in the task. A software system can only mesh well with and supplement the human system for problem solving if it is built on an understanding of how the human system works. A system well designed to assist human experts is one that uses the principles of cognitive psychology as applicable to the task being performed. Perhaps many expert system designs should be reviewed by experts in cognitive psychology before proceeding to implementation. At least some research or background knowledge of cognitive psychology should be used to develop the design of the expert system.

The ATHENA expert system is used to support evaluation of the population rules. Human evaluation in general uses upward flow of knowledge. The knowledge abstraction process for database design is controlled by expert rules for database design. The basic components of the upward flow of knowledge are shown below.

1. Definitions and value ranges.
2. Cognitive associations and schema information.
3. Data dependencies.

The schema information for a database relation includes the list of attributes for the relation, the list of key attributes, and the underlying domain of each attribute. Some domains consist of a list of enumerated values, each of which serve to flag a given condition. For example, the attribute "porttype" is "TYPE_LINE" or "TYPE_TRUNK" to indicate the type of port that is being used. The names of the relations, attributes, domains, and enumerated values can be viewed as labels. Associated with each of these labels is an English description of the item and what it is used for. Human experts use descriptive information and certain expert rules to determine the cognitive associations between database labels. For example, "a line is a type of port" is an understanding of the cognitive association between those two entities. It is the combining of schema type information and cognitive information which allows the mental derivation of data dependencies or checks. There are involved expert rules which control the upward flow of knowledge, but the process of upward flow in the context of database design can be stated in a few simple formulas:
Expert rules are used to combine schema information and cognitive associations between database labels to derive data dependencies. Below appears a rule that is used together with the cognitive information to derive the listed check. The derived check is shown in a database Population Rule Language named "PRL". Note that "RL" as a prefix to a name means it is the name of a relation. The names of attributes are "fully qualified" by preceding the name of the attribute with the name of the relation and a dot as a connector to the attribute name. Thus "RLa.a" is a relation name and "RLa.a" is an attribute name.

**Rule:**

If an attribute of one relation is a type of a key attribute of another relation,

Then the existence of a tuple in the first relation implies the existence of a tuple in the second with equivalence of the two attributes as the condition of the linkage of the two tuples.

**Schema:**
RLa.a is an attribute in RLa.
RLb.b is the key of RLb.

**Cognitive:**
RLa.a is a type of RLb.b.

**Derived Check:**
for every tuple in RLa
begin
must find exactly 1 tuple in RLb
where (RLb.b equals RLa.a);
end

The goal is a cooperative effort between human experts and AI software. Human experts do what can best be done by humans, determining label relationships and validating software derived checks. Software does what it can do best: automation of tedious and repetitive aspects of the job which are too time consuming for humans to do.

**Expert System use of Primitive Learning**

The ATHENA expert system seeks to evolve in the future to be able to gain cognitive information from descriptions of database labels. To do that, it must apply some methods of machine learning. The following classification is proposed by Michalski, et. al. [4]:

1. **Rote learning and direct implanting of new knowledge.**
2. **Learning from instruction.** "input language to an internally usable representation"
3. **Learning by analogy.** "transforming and augmenting existing knowledge ...." 
4. **Learning from examples.** "Given a set of examples and counterexamples of a concept, ...."
5. **Learning from observation and discovery.** "inductive learning that includes ... classification criteria to form taxonomic hierarchies, and similar tasks without benefit of an external teacher."

Learning in the ATHENA expert system will be done as follows:

1. **Rote - Programming.**
2. **Instruction - Database design to knowledge base encoding.**
3. **Analogy - Checks to Cognitive to Checks.**
4. **Example - Conceptual Clustering.**

The learning by example will, in fact, be done mostly by human effort with some software support. It will allow human experts to do the more complex task of natural language processing and interpretation and encode conclusions in a formal language for knowledge about the switch information. Software will then use the background knowledge about the switch information to derive cognitive associations between database labels.

The form of machine learning called *conceptual clustering* is described by Ryszard S. Michalski and Robert E. Stepp as follows:

"Given:

- A set of objects (physical or abstract).
- A set of attributes to be used to characterize the objects,
- A body of background knowledge, ....

Find:

- A hierarchy of object classes, ...."[6]

In the ATHENA expert system the theory can be applied by observing the following parts.
1. Objects - database labels.
2. Attributes - descriptive phrases.
4. Learned Rules - label hierarchy.

Database labels are the names of relations, attributes, domains, and enumerations. Association of phrases with labels can be done with new cognitive statements. Background knowledge is the relationships between phrases and can be encoded in new cognitive statements. The derived information is the hierarchy of database labels, which is represented in current cognitive statements.

The plan is for cooperative effort between human experts and AI support software. The first step is for human experts to derive and encode background knowledge and descriptions of database labels. For example an English comment like the following would be read and interpreted by a human expert.

```/*
For a packet switching DSL line,
*  service class is ISDN.
*/```

The human expert would then encode the following information.

"packet switching" implies "ISDN",
RLrc.cninf defined_as "packet switching",
FC.ISDN defined_as "ISDN",

From the above information the ATHENA expert system will be able to derive the following association between database labels.

RLrc.cninf restricts RLrc.lcc.servclass to { FC.ISDN },

From the above cognitive information, other cognitive information, and schema information, the following check can be derived.

```
for every tuple in RLrc.cninf begin
  must find exactly 1 tuple in RLrc.lcc
  where (RLrc.lcc.lcc equals RLrc.cninf.lcc
   and RLrc.lcc.rax equals RLrc.cninf.rax)
  when found
  begin
    RLrc.lcc.servclass must_be FC.ISDN;
  end
end```

Of course the accuracy of the information derived will depend on the exactness of the encoding of the knowledge about the switch information. A more general implication is given as follows.

"packet switching" implies "ISDN",
RLrc.cninf defined_as "packet switching",

But perhaps that is not exactly right, in which case erroneous checks may be derived. It may be more accurate to encode the following more restrictive implication.

"packet switching DSL line" implies "ISDN",
RLrc.cninf defined_as
"packet switching DSL line",

Building an expert system requires the proper tools. This means PROLOG, LISP, C code software, or some combination of software tools in a programming environment. It also requires skilled programmers. But the process of building also requires a blueprint for the design. The blueprint for building an expert system comes from an evaluation of Cognitive Science and how it applies to the system planned. The Computer Science industry will no doubt continue doing research into development of better tools for building expert systems. However, the evolution of the sophistication of expert systems also requires further research into the human mental system and how the latest theories of Cognitive Science may be applied to the building of expert systems. The ATHENA Expert System for database data quality assurance is being designed with that in mind.

References

[2] Ibid. (page 83)
[3] Ibid. (pages 92 through 94)