A Categorization of KR&R Methods for Requirement Analysis of a Query Answering Knowledge Base

Vinay K. CHAUDHRI[a], Bert BREDEWEG[b], Richard FIKES[c], Sheila MCILRAITH[d], and Michael P. WELLMAN[e]

[a] *Artificial Intelligence Center, SRI International, Menlo Park, CA, USA 94025*

[b] *University of Amsterdam, Science Park 107, 1098 XG Amsterdam, The Netherlands*

[c] *Computer Science Department, Stanford University, Stanford, CA, USA 94305*

[d] *Department of Computer Science, University of Toronto, Toronto, Canada M5S 1A4*

[e] *Computer Science & Engineering, University of Michigan, Ann Arbor, MI 48109*

Abstract. Our long-term goal is to build a query answering system that can answer questions on a wide variety of topics and explain the answers. In such a situation, a designer faces the challenge of how to specify the KR&R requirements that are needed to answer questions. In this paper, we introduce a categorization of KR&R methods, and apply it to specifying the requirements for answering questions in six different domains: Physics, Chemistry, Biology, Environmental Science, Microeconomics, and U.S. Government & Politics. Drawing from the corpus of about 500 questions that we analyzed, we consider an example question in each domain and show the analytical process that we used to derive the requirements in terms of the KR&R categorization. We analyze the effectiveness of the current KR&R categorization, and identify directions for future work suggesting how this categorization can be further evolved by community participation.

Keywords. Domain Analysis, Knowledge Engineering, Query Answering, E-Science

Introduction

We examine the task of identifying the Knowledge Representation and Reasoning (KR&R) requirements for a query answering knowledge base (KB) designed to answer questions on a wide variety of topics. One possible approach for doing such an analysis is to assume a specific KR language and a system, and attempt to reduce each query answering task to the capabilities of that particular system. One could go to the extreme of actually implementing each question, and be very concrete in terms of the KR&R needs. The use of a specific KR language or system, however, is not desirable if one wants the results to be general across systems. Undertaking actual implementations can be far too expensive for the initial design and planning stage of a project. Motivated by these considerations, we focus in this paper on an analysis methodology to derive KR&R requirements from a set of questions to be answered by the KB. Our methodology is based on a categorization of KR&R methods. As a corpus of questions, we consider about 500 questions from Advanced Placement (AP)

Exams in the domains of Physics, Chemistry, Biology, Microeconomics, Environmental Science, and U.S. Government & Politics. We use an example question from each domain to show the analysis process for deriving the requirements, and then study the effectiveness of the KR&R categorization. Our analysis methodology and the use of KR&R categorization is a novel contribution, as a comparable analysis on a similar scale has never before been attempted.

This work is part of a larger effort called Project Halo that is staged, long-range research by Vulcan Inc. to develop a "Digital Aristotle" — a reasoning system capable of answering novel questions and solving advanced problems in a broad range of scientific disciplines and related human affairs (see http://www.projecthalo.com). We have connected this long-term goal to the grand challenge of representing a college-level textbook, and then answering questions at the back of that book [1, 2]. The AP exam is a college-level test administered in the United States (see http://www.collegeboard.com/student/testing/ap/about.html for more details). We chose the AP test as an evaluation criterion because it is a widely accepted standard for testing whether a person understands the content of a given subject.

1. A Categorization of KR&R Methods

Our approach for developing the categorization of KR&R methods was to review the proceedings of the recent international KR&R conferences for the topics of current research (http://www.kr.org) and standard textbooks on the subject [3, 4]. We used these resources to develop an initial inventory of KR&R categories. That inventory was refined by using it in the analysis of the six AP domains.

The KR&R categorization is divided into three broad subcategories: representation capabilities, built-in theories, and reasoning capabilities. The representation capabilities refer to the representation features of the KR language. The built-in theories refer to the capabilities that require special-purpose knowledge and reasoning methods. The reasoning capabilities refer to general-purpose inference or query answering methods supported by the system.

For this paper, we simply enumerate below the different KR&R categories. A more complete document with the definitions of each category is available online (http://www.ai.sri.com/halo/public/fois2010/).

1 Representation Capabilities
- 1.1 Atomic Propositions
  - 1.1.1 Binary Relationships
  - 1.1.2 N-Ary Relationships
- 1.2 Boolean Combinations of Atomic Propositions
- 1.3 Quantification
  - 1.3.1 Universal and Existential Quantification over Objects
  - 1.3.2 Universal and Existential Quantification over Predicates
- 1.4 Qualifying the Truth Value of Propositions
  - 1.4.1 Propositions qualified by "necessary" and "possible"
  - 1.4.2 Propositions qualified by "obligatory", "permitted", and "forbidden"
  - 1.4.3 Propositions qualified by "was", "will be", and "always"
- 1.5 Structured Objects

3.8 Reasoning with Uncertain Knowledge
  3.8.1 Reasoning with Probabilities
  3.8.2 Reasoning with Approximate Values
3.9 Defeasible Reasoning
  3.9.1 Default Reasoning
    3.9.1.1 Prioritized Default Reasoning
    3.9.1.2 Default Inheritance
  3.9.2 Abduction
  3.9.3 Making Assumptions
3.10 Reasoning about Processes
  3.10.1 Prediction
  3.10.2 Comparative Projection
  3.10.3 Planning
  3.10.4 Scenario Explanation
  3.10.5 Verification
  3.10.6 Process Description Analysis
3.11 Reasoning with Aggregates
3.12 Reasoning with Teleological Knowledge
3.13 Compare and Contrast
3.14 Description Comparison
  3.14.1 Determining Closeness of Match
3.15 Determination of Applicability
3.16 Detection and Resolution of Discrepancies
3.17 Generation of Reasonable Values
3.18 Identification of Essential Information
3.19 Explanation Generation
3.20 Reasoning about Preferences
3.21 Qualitative Reasoning
3.22 Equality Reasoning
3.23 Case-Based Reasoning
  3.23.1 Analogical Reasoning
3.24 Reasoning with Diagrams
  3.24.1 Reading Curves in a Diagram
  3.24.2 Recognizing Objects
  3.24.3 Determining Correspondence to an Equation
  3.24.4 Plotting an Equation
  3.24.5 Extracting Physical Models
  3.24.6 Matching Curves
3.25 Causal Reasoning
3.26 Reasoning with Examples

2. Requirement Analysis for Query Answering

We used the KR&R categorization in identifying the requirements for query answering for each of the six domains. We worked with one full-length AP Exam in each of these domains. The official descriptions of these exams and the course descriptions are available from the website of the College Board (http://www.collegeboard.com/student/testing/ap/subjects.html). Each practice exam was drawn from the *Princeton Review* guide (http://www.princetonreview.com/), which is commonly available test preparation material. A scanned copy of each sample exam, detailed answers to all questions of each sample exam, and a copy of our analysis of each complete example exam are available online (http://www.ai.sri.com/halo/public/fois2010/). Even a cursory review of these exams will reveal that the breadth of the topics covered (from Cell structure to laws of supply and demand) is enormous. The total number of questions on each exam that was analyzed is shown in Table 1. For our discussion, we will cover one question per domain.

	Multiple Choice	Free Response
Physics	70	6
Chemistry	75	8
Biology	100	4
Environmental Science	100	4
Microeconomics	60	3
U.S. Govt & Politics	60	4

Table 1: Number of questions analyzed in each domain

Since our goal was to derive KR&R requirements, the analyst should understand enough about the domain as well as KR&R to do an adequate job. In an initial iteration of the analysis that was limited to Physics, Chemistry and Biology, we tried to divide the work between a domain expert and a KR&R expert. Since that division was not very effective, for the current analysis we identified analysts with KR&R expertise who also knew enough about each domain to do the analysis.

Our analysis factors out the natural language processing of a question and focuses on the KR&R requirements for answering the question. We assumed that each question would be represented using predicates and functions that capture the logical content of the question, but not different ways to phrase the question using natural language. A separate analysis analyzing the natural language issues in understanding a question was undertaken and is available elsewhere [5]. Answering some questions can also require the use of open-ended knowledge about the world. Analyzing and specifying such world knowledge was also kept outside the scope, as we wanted to focus on knowledge that is explicitly written down and specified in a textbook.

2.1. Physics

In most Physics questions, a key challenge is to take a problem description and identify a suitable model to solve that problem – for example, choosing whether to use Newton's laws or equations of motion. This challenge is captured by the reasoning requirement labeled as determination of applicability (KR&R category 3.15).

Every question requires representing and reasoning with structured objects – in some cases, it is just an object and its properties, and in other cases it also includes a class-subclass hierarchy. This requirement is captured by KR category Structured Objects (KR&R category 1.5). Similarly, the use of units and measures (KR&R category 2.1.1) is required very frequently.

Equations are also used all across the domain and in at least three different ways: to make qualitative inferences (KR&R category 2.4.2.2), in numerical calculations (KR&R category 2.4.2.4), and for symbolic manipulation (KR&R category 2.4.2.3). As a specific example, consider question number 58 from the sample AP Physics test (The correct answer choice is shown in italics):

> How much force is required to lift a 50-newton object with an acceleration of 10 m/s^2?
>
> (A) 10 N
> (B) 50 N
> *(C) 100 N*
> (D) 150 N
> (E) 200 N

This question involves numeric solution of equations (KR&R category 2.4.2.4) and knowledge of units and measures (KR&R category 2.1.1). We also have to make a default assumption that the lifting action is vertical. This question requires establishing a free body diagram that can be thought of as a domain-specific micro-theory, which is a method used quite frequently in answering Physics questions.

2.2. Chemistry

Just like Physics, in most Chemistry questions a key challenge is to take a problem description and identify a suitable model to solve that problem. In many Chemistry questions, such model selection can be reduced to classification reasoning of the sort seen in description logics (KR&R category 3.5). Most questions also need to represent and reason with structured objects (KR&R category 1.5) and logic program rules (KR&R category 1.7). Many questions require the use of units and measures (KR&R category 2.1.1). As a specific example, consider the following question from the sample AP test:

> A 100 ml sample of 0.10 molar NaOH solution was added to 100 ml of 0.10 molar H3C6H5O7. After equilibrium was established, which of the ions listed below was present in the greatest concentration:
>
> *(A) $H2C6H5O7^-$*
> (B) $HC6H5O72^-$
> (C) $C6H5O73^-$
> (D) OH^-
> (E) H^+

Representing this question requires the use of structured objects (KR&R category 1.5), units and measures (KR&R category 2.1.1), and a representation for chemical formulas that are specialized forms of structured objects. Answering this question requires two problem-solving steps. (1) Determine how much of each solvent will be left after the neutralization reaction is over. In this case, the two solvents will completely neutralize each other. (2) Determine the remaining ions from the equation representation, which is usually a logic programming style rule (KR&R category 1.7). One can determine from the equation that the highest remaining concentration will be H2C6H5O7-. To represent the problem-solving steps, we need a representation of the problem-solving knowledge referred to as *computational knowledge* in our categorization (KR&R category 1.18).

2.3. Biology

Process descriptions are used heavily in Biology; nearly half the questions refer to processes in one way or other (KR&R category 1.11). The representations needed are almost always discrete and nonnumeric (KR&R category 1.11.1). Questions frequently ask about steps in a process, outcomes of a process (including products), roles of various entities in the process, and so on. Locations of processes and movement of entities in processes are also frequently considered.

Most process-oriented questions in Biology are answerable by inspection of process structure and characteristics or by relatively straightforward inference about them. Wherever the inspection requires process-specific special-purpose procedures, we refer to this reasoning as Process Description Analysis (KR&R category 3.10.6). Process-specific representational elements such as inputs, outputs, preconditions, and effects are needed for many of these questions. Most Biology questions require some form of semantic matching to match up question formulation with the appropriate content or problem-solving model in the KB (KR&R category 3.14). As a specific example, consider question number 6 from the sample AP Biology test:

Which of the following is NOT a characteristic of bacteria?

(A) Circular double-stranded DNA
(B) Membrane-bound cellular organelles
(C) Plasma membrane consisting of lipids and proteins
(D) Ribosomes that synthesize polypeptides

In this question, we are asked which of the given characteristics does not apply to Bacteria. This question can be answered using a structured object representation of Bacteria. Since this question queries for negated information, either such information should be directly represented in the KB, or the system should provide a way to infer this information from a representation.

2.4. Environmental Science

The domain of Environmental Science requires rich representations of relevant qualitative processes (KR&R category 1.15, 1.11 & 3.10). Finding the appropriate set is typically achieved by classifying (KR&R category 3.5) a specific situation to match generic descriptions of structured objects (KR&R category 1.5). The notion of classification should be taken from a broad perspective, including all kinds of properties of concepts and manipulations. This reasoning is often a prerequisite for doing the 'real' analysis in terms of (*i*) establishing causal dependencies between different quantities resulting from different processes (concerning a single system state), (*ii*) determining behavior paths by finding transition from one state to another, and (*iii*) comparative analysis (KR&R category 3.13) to compare two or more possible behaviors (e.g., worse/better, more/less impact). There is some need for concepts of law and legal reasoning to represent issues that intend to regulate the behavior of humans in relation to the physical world that surrounds them. As a concrete example, consider question number 32 from a sample exam on Environmental Science:

> Which of the following organisms is the first to be adversely affected by thermal pollution in a stream?
>
> (A) Trees along the bank
> *(B) Insect larvae in the water*
> (C) Large fish migrating up stream
> (D) Birds drinking the water
> (E) Bacteria in the water

In a coarse representation for answering this question, one would define a class hierarchy for organisms, life stages, and so on (KR&R category 1.5). Young organisms would be situated as indicator species that are vulnerable to different kinds of pollution, including thermal. For this question, "Insect larvae in the water" would be the only case with a young life stage and hence the only indicator species being more vulnerable for (thermal) pollution. However, such a representation is brittle because it strongly depends on the correct wording of the organisms in the class hierarchy, and it does not capture much explanation of the phenomenon and therefore would limit the capabilities of the reasoner to explain the answer. A better solution is to have a generic class hierarchy of organisms, objects and substances, means to describe the structural organization of a subset of these in a particular context, qualitative descriptions of elementary processes capturing the possible causes of change (e.g., growth of species, heat exchange), and exogenous factors (such as polluters) [6]. This knowledge must be qualified with assumptions or perspectives that govern the inclusion (or exclusion) of certain details. In the case of solving the example question, organisms would first be positioned with respect to the water (e.g., in, above, next to) as well as how they use the water. Instantiating the applicable processes to the situation delivers the causal dependencies between the quantities involved (e.g., indicating a causal influence from the environment temperature on the growth rate of the populations). From the species actually affected by the polluter, comparative analysis reveals which organism is most vulnerable for thermal pollution. Some default reasoning may be required – e.g. to assume alternative water resources for ``Birds drinking the water'', and thereby exclude certain answer options as less plausible.

2.5. Microeconomics

Many questions in the Microeconomics domain call for reasoning about mathematical functions, and in particular about qualitative properties of these functions. The domain also requires the ability to move back and forth between analytic and geometric descriptions of the functions and relations – that is, between properties of functions and properties of their depiction on a graph. This is required for interpreting questions (many include graphs), understanding text (e.g., shifting a curve to the right corresponds to a positive monotone transform of the function), and producing graphs for the free-form questions. As a specific example, consider the following question from the sample AP test:

What are the effects on supply and demand curves for Frisbees if a new procedure reduces the cost of making a Frisbee®?

	Demand	***Supply***
(A)	Shifts right	Shifts right
(B)	No change	Shifts left
(C)	*No change*	*Shifts right*
(D)	Shifts left	Shifts right
(E)	Shifts right	Shifts left

This question posits a potentially causal factor, and asks separately for effects on supply and demand, two fundamental concepts in microeconomic analysis. A new procedure that reduces the cost of making a product (here, "Frisbees") would make a firm willing to supply more of the product, at any given price. This is the definition of a supply curve shift to the right. The question does not mention any factor that would affect the value of Frisbees to consumers, or influence any other known ingredient of demand. Thus, by default we would assume that there is no shift at all in the demand curve.

Each answer choice is a conjunction of effects on both the supply and demand curves. The reasoner would separately analyze what happens to supply and demand, and select the answer choice that contains the correct conjunction of these changes. The knowledge to perform this reasoning would be represented using structured objects (KR&R category 1.5), and basic axioms of supply and demand represented as condition-action causal rules (KR&R category 1.8).

We could also express this reasoning directly in terms of supply and demand as mathematical functions, specifying the quantity supplied or demanded at various prices. Expressed as curves on a graph, a shift in supply, for example, can be represented as the relation between two functions, the supply curve before and after. A shift to the right means that quantity supplied *after* is greater than *before* at every price, as noted above.

2.6. U.S. Government & Politics

Building an automated system that can be successful on an AP test for this domain seems to involve satisfying several requirements. First, since this domain uses a broad vocabulary, the system developers need to closely match the domain knowledge they build into the system with the domain knowledge that is likely to be asked about on an AP test (KR&R category 3.14). Second, the system needs to be capable of using heuristic methods to assess the likelihood that a suggested answer to a question is the correct answer, since in many cases the system may not have a sufficient understanding of either the question or the suggested answers to determine the correct answer with certainty. When such assessments determine that one or more of the suggested answers to a question are highly unlikely to be correct, it is advisable for the system to guess the correct answer by selecting one of the remaining suggested answers.

The reasoning required for this domain is primarily theorem proving (KR&R category 3.1) and description comparison that determines closeness of match (KR&R category 3.14.1). As a specific example, consider the following question from the sample AP test:

Which of the following correctly states the relationship between the federal and state judiciaries?

(A) Federal courts are higher courts than state courts and may overturn state decisions on any grounds.
(B) The two are entirely autonomous, and neither ever hears cases that originate in the other.
(C) The two are generally autonomous, although federal courts may rule on the constitutionality of state court decisions.
(D) State courts are trial courts; federal courts are appeal courts.
(E) State courts try all cases except those that involve conflicts between two states, which are tried in federal courts.

As an illustration, we consider here the detailed analysis of the answer option (A). There is no hierarchical relationship between state courts and federal courts. If a KB explicitly states that no hierarchical relationship exists, then one could trivially prove that the first clause of (A) is false. Otherwise, one could only fail to prove that the first clause of (A) is true, thus requiring negation as failure. Proving that the second clause of (A) is either true or false requires knowing whether or not there can be appeals of state court decisions that are not on constitutional grounds. If one knows that there can be, then the second clause of (A) can be proved to be false. The KB would need either an example of an appeal of a state court decision that is not on constitutional grounds (thus, reasoning from examples) or need an axiom stating that such appeals exist.

3. Evaluation of the KR&R Categorization

In reporting on two forms of evaluation of the KR&R categorization, we first analyze its usage in the context of the six domain analyses that we have presented in this paper. Second, we report on the implementations of query answering that were informed by the KR&R categorization.

3.1. Usage of KR&R Categorization in Domain Analyses

We used the KR&R categorization in analyzing the six domains. We analyzed each question in an exam, and stated the requirements using a KR&R category from our categorization. At the end of their analysis, we counted how frequently each category was used, as shown in Table 2. Some categories, such as structured objects (KR&R category 1.5), rules (KR&R category 1.7), and explanation generation (KR&R category 3.19), are so pervasive that almost every question requires them. Therefore, they are not shown in the table. The number in each column denotes the number of questions in a full-length exam that required that KR&R category for answering it. We can draw several general conclusions from this raw data.

First, only a fraction of the KR&R categories from the full categorization were used. That observation helps suggest a specific list of features that a system for answering AP questions for these six domains will need to support. It also helps focus attention on specific kinds of KR&R methods that should be studied more precisely and further detailed. The breadth of the KR&R categories used also suggests that any system capable of answering questions in these six domains will need to cover several different kinds of representation and reasoning methods.

Second, the data also help us see differences across the domains. Most strikingly, it is apparent that both Biology and Environmental Science make heavy use of process knowledge. Physics and Chemistry rely heavily on reasoning with equations. Computational knowledge is needed quite significantly in Chemistry. Such insights are useful in understanding the requirements for these domains and would not have been possible without the use of the KR&R categorization introduced in this paper.

Cat No	KR&R Category	P	C	B	E	G	M
1.10	Diagrams	30	7	14			8
1.11	Processes	3	4	44	72	1	
1.12.2	Vague knowledge						2
1.15	Qualitative modeling	3	20	6	72		
1.18	Computational knowledge		19	8			
2,1	Continuous dimension						6
2.3	Spatial relations and directions	4	1				
2.4	Mathematics						15
2.4.2.1	Simultaneous equations	1	6				
2.4.2.2	Qualitative reasoning with equations	18	5	1			6
2.4.2.3	Symbolic reasoning with equations	19	18	1			
2.4.2.4	Numerical solution of equations	24	4	2			
3.1	Theorem proving					9	
3.2	Query answering					17	
3.7	Reasoning by elimination	8	5	6			
3.9.2	Abduction			7			
3.9.3	Making assumptions	10	2				
3.10.3	Planning	2					
3.10.6	Process description analysis					1	
3.13	Compare and contrast			9	4	3	1
3.14	Description comparison					26	
3.21	Qualitative reasoning	2	19		30		15
3.25	Causal reasoning				30		

Table 2: Usage of KR&R Categorization. (P – Physics, C – Chemistry, B – Biology, E – Environmental Science, M – Microeconomics, G – U.S. Government & Politics)

Third, and most important, is the question of how uniformly and effectively the analysts were able to use this categorization? It is apparent that the usage of the categorization in the domains of Microeconomics, U.S. Government & Politics, and Environmental Science is sparser than in the domains of Physics, Chemistry, and Biology. We attribute this to the analysis approach in those domains. For example, it is quite obvious that some questions in each domain will require query answering and theorem proving, but those requirements were used only by the analyst in U.S.

Government & Politics. Another source for this difference could be that sometimes there is more than one way to answer a question, depending on what one assumes about what knowledge is represented in the KB versus what knowledge is derived. This limitation suggests a direction for further work to refine the methodology so that analysts can use these categories uniformly and consistently. Finally, the categories are overlapping. For example, for environmental science, wherever there was need for process representation (KR&R category 1.11), we needed to also use qualitative modeling (KR&R category 1.15), and wherever we needed causal reasoning (KR&R category 3.25), we also needed qualitative reasoning (KR&R category 3.21).

3.2. Implementations Informed by the KR&R Categorization

Informed by the results of the domain analysis, we have undertaken two different implementations of query answering. The first implementation is a system called AURA that is a mature prototype and has undergone significant testing and evaluation for the three science domains of Physics, Chemistry, and Biology [5, 7]. The second implementation is only preliminary, covering a small number of questions from each of the six domains, and is based on a representation language called SILK [8].

The current scope of AURA is to enable domain experts to construct KBs from 50 pages of a science textbook for each of Physics, Chemistry, and Biology in a way that another user can pose questions similar to those in an AP exam and get answers and explanations. In the near future, we will use AURA to represent a full textbook in one of the domains.

The overall concept of operation for AURA is as follows: a *knowledge formulation engineer* (KFE) with at least a graduate degree in the discipline of interest undergoes 20 hours of training to enter knowledge into AURA. A different person, called a *question formulation engineer* (QFE), with a high-school-level education undergoes 4 hours of training and asks questions of the system. Knowledge entry is inherently a skill-intensive task and therefore requires more advanced training in the subject as well as in using the system. A QFE is a potential user of the system, and the training requirement is low. KFEs do knowledge entry through a graphical user interface, and QFEs pose questions by converting them from original English into controlled English.

The current AURA system supports the KR&R categories of structured objects (KR&R category 1.5), rules (KR&R category 1.7), processes (KR&R category 1.11), equations (KR&R categories 2.4.2.1 and 2.4.2.4), process description analysis (3.10.1), query answering (KR&R category 3.2), compare and contrast (KR&R category 3.14), description comparison (KR&R category 3.14), and spatial relations and directions (KR&R category 2.3).   Since the KR&R categories implemented in AURA are only a subset of all the required categories listed in Table 1, we do not expect the KBs authored by KFEs using AURA to be able to answer all the questions on a full-length AP exam.

In a recent evaluation, AURA was used by domain experts at SRI to construct KBs in the three domains from a 50-page syllabus, and these KBs were tested on a suite of questions known to the development team ahead of time. These KBs were able to correctly answer 79% of the questions in Physics, 73% of the questions in Chemistry, and 71% of the questions in Biology. The AURA system was also independently evaluated by having newly trained users construct KBs that were tested on novel questions. AURA was able to correctly answer 49% of the questions in

Biology, 42% in Physics, and 22% in Chemistry. Since the questions are posed through controlled English, QFEs usually introduce some simplifications, and therefore these results should not be viewed as an indication of the system's ability to answer questions posed in original English. These results, nonetheless, do demonstrate that the requirements analysis done using our described methodology did enable the construction of a system with a high degree of performance.

We recently started a small-scale project to prototype query answering using SILK with the goal to distribute some examples using a traditional logic programming framework. The implemented SILK examples are available online (see http://www.ai.sri.com/halo/public/fois2010/). These examples provide a concrete realization of the analysis work. For example, the SILK implementation of a Physics question shows how the logic programming rules can be used in combination with different approaches for stating defaults. While the domain analysis may suggest broad KR&R requirements, when it comes to actually producing an implementation; many much more specific decisions need to be made.

4. Related Work

Several researchers have investigated the problem of representing textbook knowledge and answering questions about that knowledge, for example, [9] [10] and [11]. No one has previously attempted to build a system that would function across the range of domains considered here. The six domains are different in many ways ranging from purely descriptive to highly mathematical. The query answering tasks range from lookup, to conventional rule-based or taxonomic reasoning, to qualitative reasoning and problem solving. In spite of these divergent requirements, our challenge has been to synthesize KR&R features that can be incorporated into a functional system. Such a task has never before been undertaken.

Work has been done on using competency questions to specify the requirements of a KB [12]. The problem we consider here is very different because of the sheer scope and diversity of the questions, and the need to assess the requirements much before reducing the task to a set of axioms.

The work most closely related to our KR&R Categorization is on information metrology, reasoning, and representation languages (IMRRL) done at NIST [13]. The NIST approach is similar to our KR&R categorization in that it studies the representation and the reasoning aspects separately. The NIST approach, however, is closely tied to very specific languages and logics – for example, Common Logic, description logic, RDF/S (Resource Description Framework Schema). The NIST approach also lacks the level of breadth and the coverage considered in our KR&R categorization, and it was done outside the context of a specific application need. We do believe, however, that there is considerable synergy between our categorization and IMRRL, and it will be promising to unify the two in our future work.

5. Summary and Conclusions

We considered the problem of specifying requirements for constructing a KB for answering questions in six diverse domains: Physics, Chemistry, Biology, Microeconomics, Environmental Science, and U.S. Government & Politics. The work is part of a larger effort to build a system to answer questions on a wide variety of topics and explain the answers. No established methodology helps a designer determine the features needed in such a system. The problem has been especially challenging because for answering some questions in these domains, the required methods are unresolved and topics of research. For requirements analysis, it is not realistic to expect a detailed design and implementation for each question. Therefore, we need a methodology that is coarse, yet serves as an informative analysis tool. Our analysis using the KR&R categorization presented here is a novel contribution because of its effectiveness in giving some quantitative measure of the KR&R needs across multiple domains, and by providing useful input in the design and implementation of AURA. The approach has several weaknesses, of course, as its usage by different analysts is not always consistent, the category definitions are sometimes overlapping, and more detailed guidelines and definitions need to be developed. The overall task, however, is too large for one group to undertake, and we hope to engage others in the community in further developing the categorization and the analysis methodology.

6. Acknowledgments

This work was funded by Vulcan Inc. under its Halo Advanced Research Program. We thank Bruce Porter and Ken Barker for conceptualizing the earlier drafts of the domain analysis framework. We thank Shirin Sohrabi, Hojjat Ghaderi, and David Martin for their contributions to the SILK implementations and domain analysis. We thank Benjamin Grosof for his inputs to the drafts of this work.

References

[1] Feigenbaum, E., Some Challenges and Grand Challenges for Computational Intelligence. *Journal of the Association of Computational Machinery,* 2003. 50(1): p. 32-40.

[2] Reddy, R., Three Open Problems in AI. *Journal of the Association of Computational Machinery,* 2003. 50(1).

[3] Brachman, R.J. and H.J. Levesque, *Knowledge Representation and Reasoning.* 2003, San Francisco: Morgan Kaufmann.

[4] Russell, S. and P. Norvig, *Artificial Intelligence: A Modern Approach.* 2003: Prentice Hall.

[5] Clark, P., et al. Capturing and Answering Questions Posed to a Knowledge-Based System. *International Conference on Knowledge Capture Systems (KCAP).* 2007. Whistler, Canada.

[6] Bredeweg, B. and P. Salles, eds. *Mediating Conceptual Knowledge Using Qualitative Reasoning. Handbook of Ecological Modeling and Informatics,* ed. S.V. Jorgensen, T.-S. Chon, and F.A. Recknagel. 2009, Wit Press: Southampton, UK. 351-398.

[7] Chaudhri, V.K., et al., Enabling Experts to Build Knowledge Bases from Science Textbooks, *International Conference on Knowledge Capture Systems (KCAP).* 2007: Whistler, Canada.

[8] Wan, H., et al. Logic Programming with Defaults and Argumentation Theories. *25th International Conference on Logic Programming.* 2009. Pasadena, CA.

[9] Bobrow, D.G. A Question-Answering System for High School Algebra and Word Problems. *Proceedings of the Fall Joint Computer Conference.* 1964. San Francisco, CA.

[10] Klenk, M. and K. Forbus, Analogical Model Formulation for Transfer Learning in AP Physics. *Artificial Intelligence,* 2009. 173(18): p. 1615-1638.

[11] Mukherjee, A. and U. Garain, A Review of Methods for Automatic Understanding of Natural Language Mathematical Problems. . *Artificial Intelligence Review,* 2008. 29: p. 93-122.

[12] Gruninger, M. and M.S. Fox, The Role of Competency Questions in Enterprise Engineering, *Proceedings of the IFIP WG 5.7 Workshop on Benchmarking – Theory and Practice. 1994.*

[13] Bock, C., et al., *Evaluating Reasoning Systems: Information Metrology, Reasoning, and Representation Languages.* 2006, National Institute of Standards and Technology.